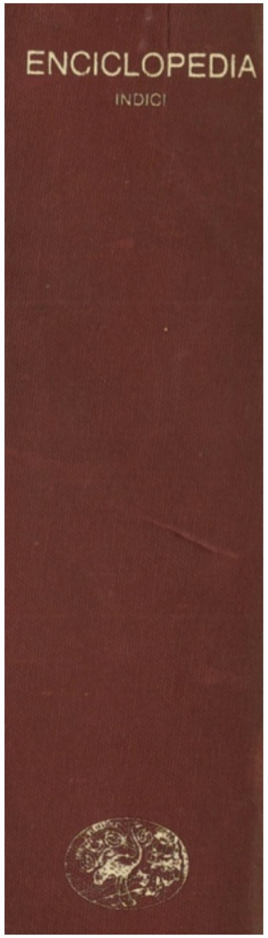


ENCICLOPEDIA EINAUDI [1982]

PROBABILITÁ

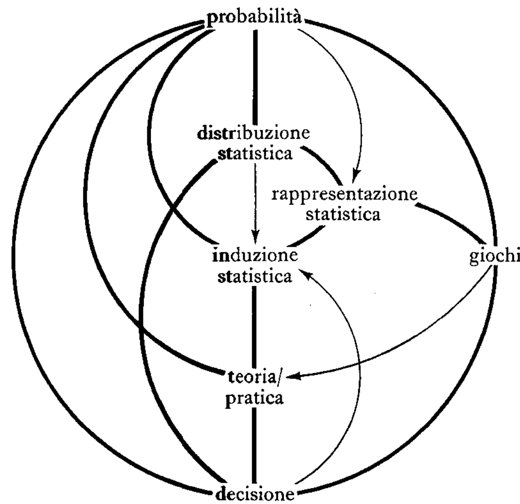
	anthropos	argomentazione	arti	calcolo	capitale	cervello	civiltà	classi	cognizione	combinatoria	comunicazione	concetto	conoscenza	corpo	creatività	cultura mat.
decisione	2	4	·	·	7	·	4	5	·	3	5	5	3	·	2	2
distribuzione statistica	·	·	·	3	4	·	·	·	2	4	·	·	·	·	·	·
giochi	4	·	2	·	4	·	3	·	2	4	4	·	4	2	·	·
induzione statistica	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
probabilità	2	3	2	3	5	2	·	·	4	4	5	2	4	·	3	2
rappresentazione statistica	·	·	·	4	·	·	·	·	·	·	·	·	·	·	·	·
teoria/pratica	5	1	7	5	6	3	·	5	4	4	5	4	7	6	7	4

	memoria	metodo	modo di produz.	mythos/logos	natura	normale/anormale	orale/scritto	organico/inorgan.	organismo	parentela	politica	probabilità	produzione/distrib.	regione	religione	rito
decisione	·	4	2	4	·	2	·	·	·	·	4	7	4	5	·	·
distribuzione statistica	·	3	·	·	·	·	·	3	·	·	·	8	·	2	·	·
giochi	·	4	·	·	·	·	2	·	·	·	·	8	·	·	·	3
induzione statistica	·	·	·	·	·	·	·	2	·	·	·	6	·	·	·	·
probabilità	·	4	·	·	2	3	·	5	2	4	·	9	·	4	·	·
rappresentazione statistica	·	·	·	·	·	·	·	·	·	·	·	6	·	·	·	·
teoria/pratica	2	5	4	4	6	6	6	4	4	5	6	6	6	7	7	5

	probabilità	decisione	teoria/pratica	distribuzione statistica	giochi	induzione statistica	rappresentazione statistica
probabilità	·	7	5	6	4	7	5
decisione	5	·	6	3	3	5	·
teoria/pratica	4	5	·	1	·	4	3
distribuzione statistica	4	4	2	·	3	7	5
giochi	5	6	6	1	·	1	4
induzione statistica	3	·	4	·	2	·	5
rappresentazione statistica	·	·	2	2	2	5	·

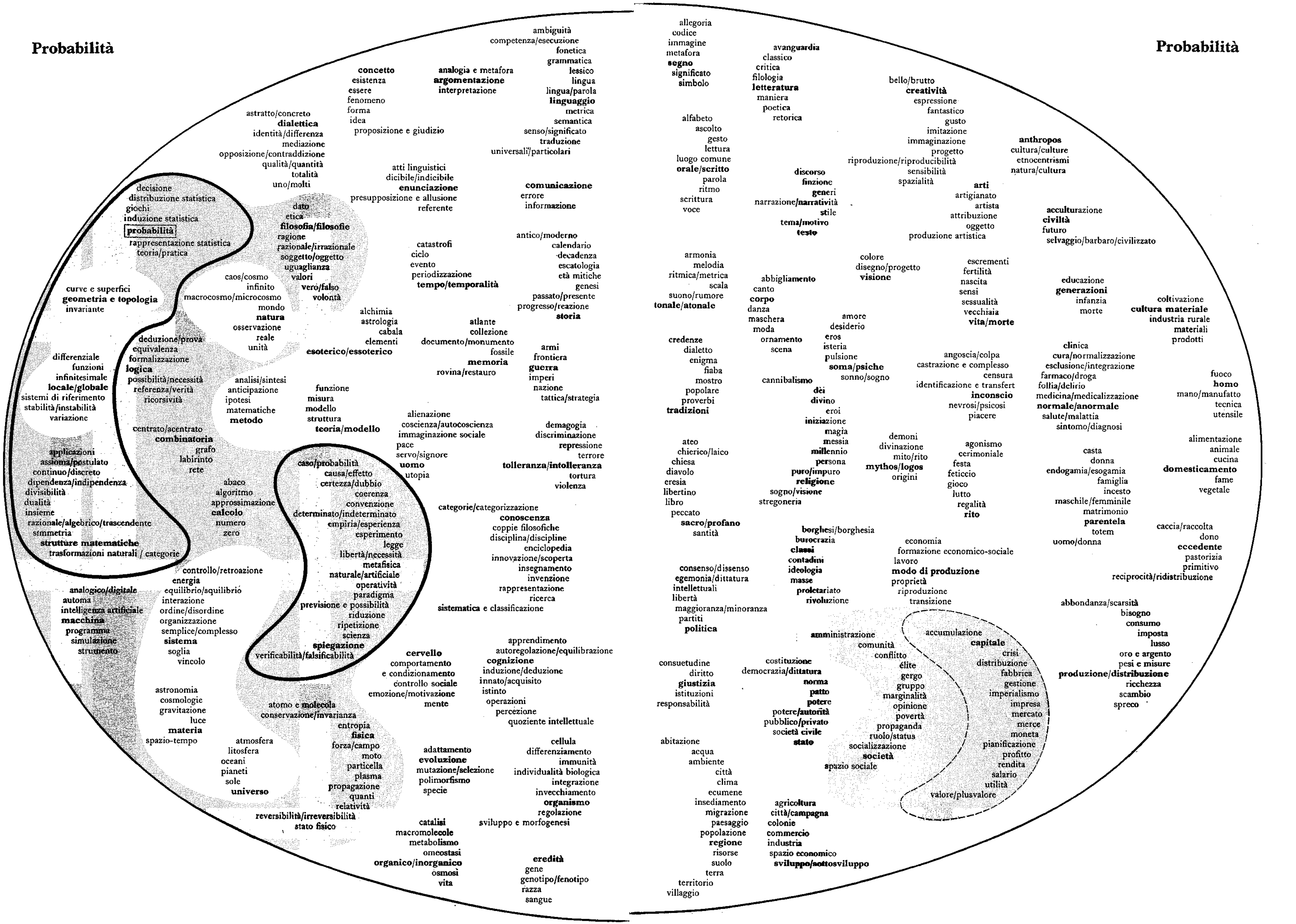
	dialettica	domesticam.	eccedente	enunciazione	eredità	**esoterico/essoter.**	evoluzione	filosofia/filosofie	fisica	generazioni	geometria e topol.	giustizia	guerra	homo	inconscio	letteratura	linguaggio	locale/globale	logica	macchina	materia
decisione	2	·	·	2	·	·	·	5	4	·	3	4	5	·	2	·	3	·	6	5	·
distribuzione statistica	·	·	·	·	3	·	4	2	8	·	·	·	·	·	·	·	2	·	3	5	6
giochi	3	·	·	·	·	·	2	5	·	·	·	·	6	·	·	·	·	·	3	6	2
induzione statistica	·	·	·	·	·	·	·	·	3	·	·	·	·	·	·	·	·	·	3	·	·
probabilità	4	·	·	·	·	·	5	3	8	·	·	3	3	·	·	·	3	4	5	5	5
rappresentazione statistica	3	·	·	·	·	·	·	·	6	·	·	·	·	·	·	·	·	4	·	·	·
teoria/pratica	6	2	5	4	2	4	3	6	5	3	3	5	4	3	5	5	6	5	5	7	5

	sacro/profano	segno	sistema	società	soma/psiche	**spiegazione**	stato	storia	strutture matem.	sviluppo/sottosvil.	tempo/temporalità	teoria/modello	testo	tolleranza/intoll.	tonale/atonale	tradizioni	universo	uomo	visione	vita/morte
decisione	2	5	4	6	·	7	5	2	·	3	2	2	·	·	·	·	2	4	·	·
distribuzione statistica	·	·	6	3	·	7	·	·	6	·	3	·	·	·	·	·	5	·	·	·
giochi	·	4	2	4	2	3	4	·	·	3	·	4	·	·	·	3	·	2	·	2
induzione statistica	·	·	·	·	·	2	·	·	3	·	·	·	·	·	·	·	·	·	·	·
probabilità	2	4	6	5	·	7	·	3	7	3	4	3	2	·	·	·	4	·	·	·
rappresentazione statistica	·	·	·	2	·	·	·	·	6	·	·	·	·	·	·	·	·	·	·	·
teoria/pratica	4	5	5	7	6	8	6	6	6	2	6	7	4	5	5	5	3	6	6	6



Probabilità

Probabilità



Probabilità

Decisione, Distribuzione statistica, Giochi, Induzione statistica, Probabilità, Rappresentazione statistica, Teoria/pratica

1. *Davanti a un'urna.*

Immaginate – o meglio immagina *tu*, dato che ciascuno dovrà prendere la *propria* ♦decisione♦ – di trovarti davanti a un'urna con tutte le tue idiosincrasie. La persona cui l'urna appartiene – diciamo il signor Caso – ti spiega che essa è stata estratta (naturalmente a caso!) da una «superurna» contenente cento urne di due tipi. Quelle del primo tipo, dette H-urne, contengono 4 palle rosse e 6 nere e quelle del secondo tipo, dette K-urne, 9 rosse e 1 nera. A questo punto, il signor Caso ti regala un biglietto in cui s'impegna a pagare 100 000 lire al possessore del biglietto se l'urna estratta è una H-urna e, dopo averti fatto dare uno sguardo alla superurna, ti chiede di specificare il prezzo minimo a cui tu saresti disposto a cedere ad altri il biglietto stesso. Lo sguardo non ti ha consentito di contare *quante* siano le urne dei due tipi, ma solo di farti una vaga idea del loro numero. Questa è tutta l'informazione che inizialmente hai dell'urna che ti sta davanti. Pensaci bene allora: qual è il prezzo *minimo* a cui sei disposto a cedere il biglietto? In altri termini: qual è la cifra x tale che tu sei *indifferente* tra il ricevere con certezza x lire e il riceverne 100 000 se l'urna estratta è una H-urna?

Supponiamo che tu ponga, dopo una riflessione sufficientemente lunga, $x = 20\,000$; questo significa che tu sei disposto a cederlo per prezzi maggiori, ma non per prezzi minori di 20 000. In tal caso si dice che la ♦probabilità♦ *per te* dell'evento 'l'urna estratta è una H-urna', è pari a 2/10 (20 000/100 000). Piú in generale, sia E qualche evento incerto e sia ℓ_E la lotteria che dà k lire se E si realizza e nulla se $\bar{E}$ non si realizza. Se ℓ_E è per te indifferente a x lire, allora x/k viene detta la probabilità personale, cioè per te, di E. La probabilità personale di E rappresenta dunque il grado di credenza che tu hai rispetto al realizzarsi di E.

Tuttavia, per giustificare questa «definizione» si debbono fare almeno due cose. In primo luogo mostrare che la nozione cosí definita è una probabilità nel senso della teoria matematica, mostrare cioè che soddisfa gli assiomi:

1) La probabilità di ogni evento è compresa tra 0 e 1.
2) La probabilità dell'evento certo è pari a 1.
3) La probabilità che almeno uno di due eventi incompatibili si realizzi è uguale alla somma delle loro probabilità.

In secondo luogo mostrare che il valore di x/k non dipende dalla scelta della lotteria. Sfortunatamente, questo è falso per la maggior parte delle persone: x/k dipende in generale proprio da questa scelta. Di nuovo, pensaci bene: se il signor Caso moltiplicasse per 100 il valore del premio, saresti tu disposto a moltiplicare per 100 anche il valore di x? L'avversione al rischio spinge molte persone a dare

una risposta negativa. Questo significa che per ottenere una definizione generale si dovrà trovare un modo per «scontare» l'avversione al rischio. È a questo scopo che deve essere introdotta la nozione di utilità e la definizione precedente riformulata nei suoi termini e non in termini di valori monetari. È chiaro che se la tua funzione di utilità è *lineare* nei valori monetari, le due formulazioni coincidono.

Tuttavia, la nozione moderna di utilità assiomatizzata nel 1944 da Neumann e Morgenstern in *Theory of Games and Economic Behavior* presuppone una qualche nozione di probabilità. Sembrerebbe cosí di essere intrappolati in un circolo vizioso: non si può definire la probabilità senza l'utilità, né l'utilità senza la probabilità. Questo problema, oltre a quello di mostrare che le probabilità personali sono probabilità nel senso della teoria matematica, è stato brillantemente risolto da Savage in *The Foundations of Statistics* (1954), assiomatizzando *simultaneamente* probabilità e utilità nei termini di una sottostante relazione di preferenza. Da allora, le assiomatizzazioni simultanee sono proliferate mettendo in luce sempre meglio le ipotesi da fare perché abbia senso parlare di probabilità personali e di utilità nel contesto di decisioni in condizioni d'incertezza. Tali ipotesi definiscono in effetti una nozione di *coerenza* per il sistema di preferenze di un dato individuo, richiedendo ad esempio che *se* egli preferisce la lotteria ℓ alla lotteria ℓ', e la lotteria ℓ' alla lotteria ℓ'', *allora* egli preferisce la lotteria ℓ alla lotteria ℓ''. Una lotteria ℓ risulterà specificata una volta che si siano specificati n eventi $E_1, \ldots, E_n$, esclusivi ed esaustivi, e n premi, $c_1, \ldots, c_n$, dove c_i è il premio assegnato se E_i si realizza, per $i = 1, \ldots, n$. Si scriverà cosí $\ell = (c_1E_1, \ldots, c_nE_n)$. Il risultato fondamentale di Savage è allora che se il tuo sistema di preferenze è coerente, nel senso definito dalle ipotesi cui abbiamo accennato, allora esiste un'*unica* funzione di probabilità $\mathbf{P}$ e una funzione di utilità U, unica a meno di trasformazioni lineari crescenti, tali che tu preferisci la lotteria $\ell = (c_1E_1, \ldots, c_nE_n)$ alla lotteria $\ell' = (c'_1E_1, \ldots, c'_nE_n)$ se e solo se

$$(1) \qquad \sum_{i=1}^{n} \mathbf{P}(E_i)\, U(c_i) \geq \sum_{i=1}^{n} \mathbf{P}(E_i)\, U(c'_i),$$

dove le sommatorie in questione vengono dette rispettivamente l'utilità prevista delle lotterie ℓ e ℓ', o, in una terminologia piú classica, le loro speranze morali.

Risulta cosí giustificato dall'ipotesi della coerenza il principio fondamentale della razionalità, e cioè il principio di massimizzazione dell'utilità prevista: se tu sei coerente, prenderai le tue decisioni, che tu lo sappia o meno, in modo da massimizzare l'utilità prevista. «Conseguentemente», affermava Laplace nel suo *Essai philosophique sur les probabilités* (1819) «dovremmo sempre, nella condotta della vita, scegliere in modo che il prodotto del beneficio sperato per la sua probabilità sia almeno uguale all'analogo prodotto relativo alla perdita»; e concludeva: «Questa regola conduce a risultati in armonia con le indicazioni del senso comune, che può in tal modo essere valutato con esattezza».

Naturalmente, la funzione $\mathbf{P}$ che compare nella (1) è una funzione di probabilità nel senso della teoria matematica e rappresenta inoltre probabilità personali.

2. *La sfida di Hume.*

La probabilità, nel senso della teoria matematica, può dunque essere interpretata come una misura dei gradi di credenza di una persona coerente. (Per chi non gradisse il nesso tra probabilità e decisioni implicato dalla nozione di coerenza di Savage, vi è un'altra nozione di coerenza, dovuta a Cox, che dà essenzialmente la stessa conclusione senza riferimenti a processi decisionali, bensí soltanto a processi inferenziali).

Questo risultato, come tale, non avrebbe però molto interesse se non fosse per le sue conseguenze sul problema dell'♦induzione statistica♦. Di fatto, vi sono altre interpretazioni – formalmente non meno legittime – della nozione matematica di probabilità. Ma questa – in termini di gradi di credenza di una persona coerente – è l'unica che consenta di applicare l'*intero* formalismo matematico al mondo reale, e di avviare cosí a soluzione il problema dell'induzione statistica.

Torniamo alla nostra urna, supponendo che tu sia coerente e che la tua funzione di utilità sia lineare nei valori monetari. In queste circostanze, tu hai codificato la tua informazione iniziale sotto forma di una distribuzione di ♦probabilità♦ ai due eventi H (una H-urna è stata estratta), e K (una K-urna è stata estratta). Siano $\mathbf{P}(H)$ e $\mathbf{P}(K)$ le probabilità assegnate rispettivamente a H e a K. Si avrà naturalmente che $\mathbf{P}(H)+\mathbf{P}(K)=1$ e che l'utilità prevista del biglietto che ti ha regalato il signor Caso è pari a $\mathbf{P}(H)\times(100\,000)$. $\mathbf{P}(H)$ e $\mathbf{P}(K)$ vengono dette probabilità iniziali o antecedenti di H e K. Ora, tuttavia, essendo riuscito a determinare la tua opinione iniziale circa H e K, il signor Caso passa ad un'altra offerta. Ti chiede di dirgli se l'urna è di tipo H o K e quindi:

a) se dici H ed è H, vinci 100 000
b) se dici H ed è K, perdi 5 000
c) se dici K ed è K, vinci 40 000
d) se dici K ed è H, perdi 20 000.

Inoltre, per un pagamento di 7500 egli ti consente di estrarre 5 palle dall'urna con reimbussolamento (rimettendo cioè la palla estratta nell'urna prima dell'estrazione successiva). Ricorda che 1) se l'urna è una H-urna, contiene 4 palle rosse e 6 nere, e che 2) se l'urna è una K-urna, contiene 9 palle rosse e 1 nera. È ragionevole accettare l'offerta? È ragionevole cioè raccogliere *nuove* informazioni intorno all'urna al prezzo stabilito? Certo la tua informazione *iniziale* non è molta: tu sai solo che è stata estratta da una superurna della cui composizione hai un'idea molto vaga che ti sei fatto dando uno sguardo veloce al suo interno. Ma la nuova informazione costa 7500 lire. L'ultima offerta del signor Caso equivale per te alla possibilità di svolgere un esperimento i cui esiti possibili saranno denotati con «$E(r,n)$». $E(r,n)$ è naturalmente un campione di 5 palle di cui r rosse e n nere, con $r+n=5$. L'ipotesi della coerenza implica allora che $\mathbf{P}(E(r,n))$ sia definito per $0\leq r,n\leq 5$. Abbreviato «$E(r,n)$» con «E», sia $\mathbf{P}(H|E)$ il tuo grado di credenza in H, avendo a tua disposizione, oltre all'informazione iniziale, l'informazione che è stato estratto il campione E. Di nuovo, l'ipotesi della coerenza implica non solo che $\mathbf{P}(H|E)$ è una probabilità nel senso della teoria matematica, ma inoltre che

$$\mathbf{P}(H|E)=\frac{\mathbf{P}(H \text{ e } E)}{\mathbf{P}(E)},$$

purché naturalmente $\mathbf{P}(E)\neq 0$. Allora, semplici trasformazioni algebriche implicano immediatamente che

$$\mathbf{P}(H|E)=\frac{\mathbf{P}(H)\times\mathbf{P}(E|H)}{\mathbf{P}(E)} \tag{2}$$

$$\mathbf{P}(E)=(\mathbf{P}(H)\times\mathbf{P}(E|H))+(\mathbf{P}(K)\times\mathbf{P}(E|K)). \tag{3}$$

$\mathbf{P}(H|E)$ viene detta probabilità finale o susseguente di H relativamente a E oppure ancora probabilità di H condizionata a E, anche se è forse meglio evitare quest'ultimo termine dato che, in un senso importante, anche le probabilità iniziali sono condizionate, condizionate cioè al tuo stato d'informazione iniziale.

La semplice formula (2) è il famoso teorema di Bayes. Esso asserisce che la probabilità finale di un evento è proporzionale alla sua probabilità iniziale moltiplicata per il fattore $\mathbf{P}(E|H)$, detto verosimiglianza per E di H. Esso determina perciò il modo in cui la tua opinione iniziale circa un evento «incognito» (rappresentato qui da H) dovrebbe modificarsi alla luce della conoscenza di certi altri eventi (rappresentati qui da E) e di conseguenza la forma piú generale di ragionamento induttivo. Ragionamenti di questa forma risultano giustificati dall'ipotesi di coerenza, che offre cosí la miglior risposta alla sfida di Hume. Tuttavia, la sfida di Hume era rivolta alla giustificazione di una forma piú specifica di ragionamento induttivo: quella per cui si tende a valutare la probabilità di un evento in accordo con la frequenza osservata di eventi «analoghi». Fortunatamente, l'ipotesi di coerenza non giustifica in generale questa forma di ragionamento; di fatto, essa non è valida, come ha mostrato il paradosso di Goodman. Tuttavia, grazie al cosiddetto teorema di rappresentazione di De Finetti, è noto sotto quale condizione addizionale essa è valida. Si tratta della condizione di scambiabilità. È proprio la sua assunzione che – come ha mostrato in dettaglio per primo Lindley – consente una soluzione bayesiana della maggior parte dei problemi d'inferenza statistica – stime puntuali, zonali, test di significatività, ecc. – tradizionalmente risolti – nell'ambito della statistica frequentista (che identifica – per definizione – probabilità e frequenza) con una pletora di metodi ad hoc. Non solo: essa consente anche una soluzione bayesiana di molti problemi che sfuggono interamente a tali metodi.

Un tipico esempio di problema di ♦induzione statistica♦ è quello che tu devi risolvere per dare una risposta al signor Caso. Vediamo. Il tuo problema è determinare $\mathbf{P}(H|E)$ e $\mathbf{P}(K|E)$. Per il teorema di Bayes, questo implica due cose:

1) la determinazione di un modello dei dati, qui espressi da $E(r,n)$, sotto forma di una funzione di verosimiglianza, $\mathbf{P}(E(r,n))|X)$, dove $X=H$ oppure $X=K$;
2) la determinazione di una distribuzione iniziale a H e a K, $\mathbf{P}(H)$ e $\mathbf{P}(K)$.

Qui, l'ipotesi di coerenza ci abbandona: essa afferma solo che se tu sei coerente, allora sono determinati sia un modello dei dati sia una distribuzione iniziale. Ma quali sono i piú appropriati alle circostanze del caso?

Cominciamo da 1). Pensiamo ai campioni $E(r,n)$ come se fossero ottenuti, invece che facendo cinque successive estrazioni da X con reimbussolamento, facendo cinque estrazioni simultanee da cinque «repliche» di X. In questo caso, lo spazio di tutti i possibili campioni di cinque elementi è il prodotto cartesiano $X_1 \times \ldots \times X_5$. L'ipotesi piú plausibile (ma indipendente da quella di coerenza) è allora che la probabilità di $E(r,n)$ sia uguale alla frequenza relativa di tutti i campioni, contenenti una palla per ciascuna urna, che consistono di r palle rosse e n palle nere. Questo implica la scelta della seguente ♦distribuzione statistica♦, detta binomiale, come modello dei dati:

$$\mathbf{P}(E(r,n)|X) = \binom{5}{r} p^r (1-p)^{5-r},$$

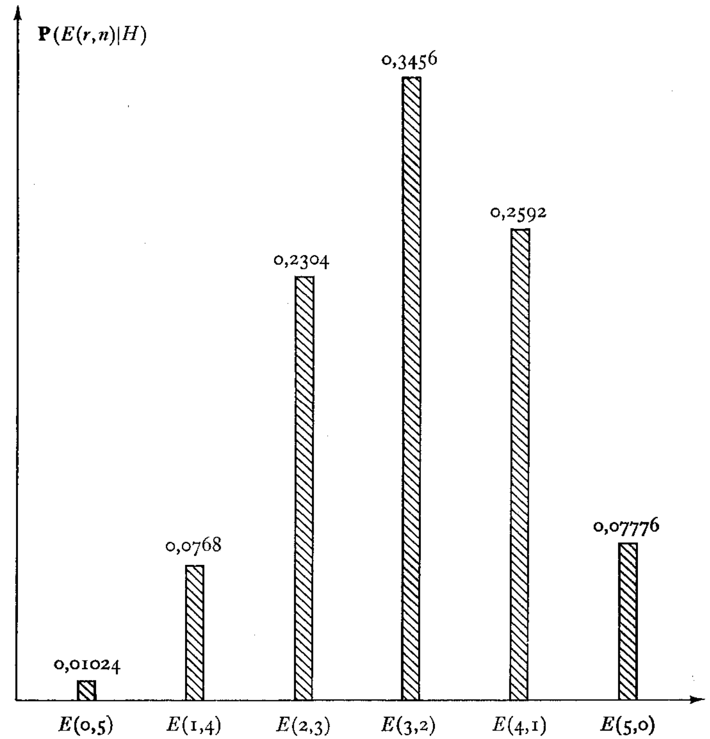


Figura 1.

Distribuzione statistica dei campioni con r palle rosse per $r=1, \ldots, 5$, e $X=H$ (i valori sono approssimati).

dove $p=4/10$ nel caso di $X=H$ e $9/10$ nel caso di $X=K$. Il modo migliore per comprenderne il significato è di darne una ♦rappresentazione statistica♦. Ponendo in ascissa il numero di palle rosse (r) e in ordinata la frequenza relativa dei campioni corrispondenti, si ottiene la rappresentazione statistica della figura 1 per $X=H$. Per $X=K$ si ottengono d'altra parte i seguenti valori approssimati:

$\mathbf{P}(E(0,5)\,|\,K) = 0{,}00001$
$\mathbf{P}(E(1,4)\,|\,K) = 0{,}00045$
$\mathbf{P}(E(2,3)\,|\,K) = 0{,}0081$
$\mathbf{P}(E(3,2)\,|\,K) = 0{,}0729$
$\mathbf{P}(E(4,1)\,|\,K) = 0{,}32805$
$\mathbf{P}(E(5,0)\,|\,K) = 0{,}59049$.

Quanto a 2), il problema è quello di trovare la distribuzione iniziale che meglio codifichi la tua informazione iniziale. Qui, la regola piú plausibile è, come ha mostrato Jaynes, quella della massimizzazione dell'entropia. Questo significa massimizzare l'uniformità della distribuzione iniziale rispettando i vincoli posti dall'informazione iniziale. Cosí, se il tuo sguardo all'urna ti ha convinto che la proporzione di H-urne è maggiore o uguale al 10 per cento ma minore o uguale al 30 per cento, dovresti considerare equiprobabili tutti i valori compresi in questo intervallo ed assegnare probabilità trascurabile a quelli esterni. Ponendo il vincolo $0{,}1 \leq \mathbf{P}(H) \leq 0{,}3$ questa regola implica porre $\mathbf{P}(H)=0{,}2$ e $\mathbf{P}(K)=0{,}8$.

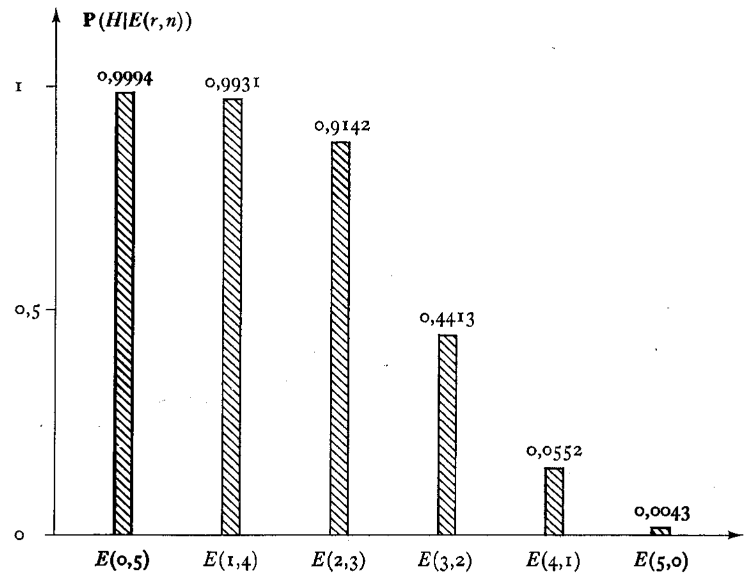


Figura 2.

Distribuzione finale per H come funzione di r.

A questo punto il teorema di Bayes dà per H la distribuzione finale rappresentata nella figura 2. Per K si avrà naturalmente $\mathbf{P}(K|E(n,r)) = 1 - \mathbf{P}(H|E(n,r))$.

A questo punto sono disponibili tutti i dati per calcolare il valore previsto dell'informazione convogliata dall'esperimento in questione. Per rendere concettualmente piú chiara la situazione, supporremo che l'esperimento consista di una sola estrazione. Denotiamo con «$V(e)$» tale informazione, dove e è l'esperimento «semplificato». $V(e)$ sarà evidentemente pari alla differenza tra l'utilità prevista della nuova offerta del signor Caso *con* l'esperimento e quella della stessa *senza*. Quest'ultima sarà pari al massimo tra

a) $\mathbf{P}(H)\times(100\,000)-\mathbf{P}(K)(\ 5\,000)$
b) $\mathbf{P}(K)\times(\ 40\,000)-\mathbf{P}(H)(20\,000)$,

e cioè a 28 000 lire, in base al principio di massimizzazione dell'utilità prevista. Nel primo caso (*con* l'esperimento) si hanno due possibilità: R e N. Se si realizza R, si ottiene una utilità prevista pari al massimo tra

a') $\mathbf{P}(H|R)(100\,000)-\mathbf{P}(K|R)(\ 5\,000)$
b') $\mathbf{P}(H|R)(\ 40\,000)-\mathbf{P}(K|R)(20\,000)$.

Se si realizza N, si ottiene una utilità prevista pari al massimo tra

a'') $\mathbf{P}(H|N)(100\,000)-\mathbf{P}(K|N)(\ 5\,000)$
b'') $\mathbf{P}(H|N)(\ 40\,000)-\mathbf{P}(K|N)(20\,000)$.

Dunque, l'utilità prevista dell'offerta *con* l'esperimento sarà pari a

$$\mathbf{P}(R)\times(\text{massimo tra } (a') \text{ e } (b')) + \mathbf{P}(N)\times(\text{massimo tra } (a'') \text{ e } (b'')).$$

Applicando la (2) e la (3), semplici calcoli consentono di stabilire che tale utilità è pari a 35 200 lire. Cosí, il valore previsto dell'informazione è pari a 7200 lire (35 200 − 28 000). Essendo il suo prezzo 7500 lire, chiaramente devi respingere l'opportunità di sperimentare. Inoltre, devi dire K e non H, dato che l'utilità prevista della prima decisione (28 000) è maggiore dell'utilità prevista della seconda (16 000).

In tal modo, applicando passo passo la teoria bayesiana della razionalità, si è giunti a una soluzione semplice e evidente dell'intero problema, del tutto rappresentativa, nonostante il suo carattere artificiale, di una vasta gamma di problemi d'induzione statistica.

3. *Il genio maligno.*

Fin qui è stata affrontata la questione delle decisioni individuali «non-competitive» e si è visto che il comportamento razionale è quello che massimizza l'utilità prevista. Il signor Caso era una semplice finzione retorica: non rappresentava un individuo che cercasse di trarre qualche vantaggio personale dalla situazione descritta, un avversario con interessi parzialmente o totalmente opposti ai nostri, bensí una Natura indifferente rispetto alle conseguenze dei vari eventi

possibili. Consideriamo ora la questione seguente, sollevata nel 1713 da Raymond de Montmort in una lettera diretta a Nicolas Bernoulli: «Un padre intende fare al figlio un regalo. Lo chiama e gli dice: metterò nella mia mano destra un numero di gettoni pari o dispari, a mia scelta; fatto questo:

a) se tu dici pari e il numero è pari, ti regalerò 4 scudi;
b) se tu dici dispari e il numero è pari, non ti regalerò alcuno scudo;
c) se tu dici dispari e il numero è dispari, ti regalerò 1 scudo;
d) se tu dici pari e il numero è dispari, non ti regalerò alcuno scudo».

Naturalmente, se il padre non fosse interessato alle conseguenze, il problema per il figlio non differirebbe da quello affrontato in precedenza: egli dovrebbe procedere ad assegnare una probabilità ai due eventi possibili e quindi prendere la decisione (dire pari o dire dispari) che massimizza la sua utilità prevista. Ma Montmort prosegue cosí: «Il problema è: 1) quale regola bisogna prescrivere al padre perché economizzi al massimo il suo denaro; 2) quale regola bisogna prescrivere al figlio perché massimizzi il suo guadagno; 3) qual è il valore del regalo che il padre fa al figlio, supponendo che ciascuno dei due terrà la linea di condotta che gli è piú vantaggiosa». Qui naturalmente, il punto 1) è cruciale: si assume infatti non solo che il padre non sia indifferente alle conseguenze ma inoltre che farà del suo meglio per minimizzare il suo esborso. Anzi, 1) e 2) equivalgono ad assumere che gl'interessi di padre e figlio sono diametralmente opposti. Il figlio ha a che fare non con una Natura indifferente ma con un genio maligno! E cosí il padre. Decisioni di tipo competitivo come questa sembrano sfuggire al principio di massimizzazione dell'utilità prevista (essenzialmente già disponibile al tempo di Montmort) tanto che Montmort, rassegnato, concludeva: «Sarà assolutamente impossibile prescrivere una regola per questo gioco, se i giocatori sono entrambi ugualmente intelligenti e perspicaci». Solo recentemente, John Harsanyi e Richard Selten hanno esteso la teoria bayesiana della ✦decisione✦ non-competitiva in modo tale che essa implichi una soluzione anche per decisioni competitive. Nel caso particolare considerato, essa implica la stessa soluzione già proposta da Neumann e Morgenstern in *Theory of Games and Economic Behavior* (1944) per i cosiddetti ✦giochi✦ a due persone a somma zero, di cui quello costruito da Montmort è un esempio. Vediamola. Rappresentiamo in primo luogo il gioco con la seguente matrice:

		Padre	
		Pari	Dispari
Figlio	Pari	+4; −4	0; 0
	Dispari	0; 0	+1; −1

dove, ad esempio, la prima casella in alto a sinistra significa che la coppia di scelte strategiche (pari; pari) dà al figlio un pagamento uguale a $+4$ e al padre a -4. (Il gioco viene detto a somma zero proprio perché la somma dei pagamenti per ogni coppia di scelte strategiche è uguale a zero). Ora, se l'insieme di scelte strategiche disponibili a padre e figlio si riduce all'insieme {pari, dispari}, la conclusione rassegnata di Montmort è corretta. Diciamo che una strategia s del primo giocatore è la *miglior risposta* alla strategia s' del secondo giocatore se e solo se, data s', s è la strategia che massimizza il pagamento al primo giocatore. Cosí, per il figlio, la miglior risposta a «pari» è «pari», e la miglior risposta a «dispari» è «dispari». Diciamo inoltre che una coppia di strategie, (s, s'), rispettivamente del primo e secondo giocatore, è una coppia d'equilibrio, o un punto d'equilibrio, se e solo se ciascuna di esse è la miglior risposta all'altra. Ora, non è difficile controllare che nel gioco di Montmort, se le scelte strategiche sono limitate all'insieme {pari, dispari}, *non esistono coppie di strategie d'equilibrio.* Ma – e questa fu l'intuizione di Neumann – perché mai padre e figlio dovrebbero limitare le loro scelte all'insieme in questione? perché non dovrebbero prendere in considerazione l'insieme di strategie cosí definite: p (pari); $(1-p)$ (dispari), per ogni valore di p tale che $0 \leq p \leq 1$, affidando cosí la propria scelta a un meccanismo casuale che con probabilità p dà la scelta «pari» e con probabilità $(1-p)$ dà la scelta «dispari»? Tali strategie vengono dette casualizzate oppure miste, in opposizione a quelle precedentemente considerate dette pure. Naturalmente, le strategie pure sono solo un caso particolare di quelle miste; cosí la strategia pura, «pari», è equivalente alla strategia mista: 1 (pari); 0 (dispari). L'interpretazione intuitiva delle strategie miste non è affatto chiara; molto chiara è invece la loro portata matematica. Il teorema di Neumann afferma infatti che in ogni gioco a due persone a somma zero esiste un punto di equilibrio, (s, s'), dove s e s' sono strategie miste. Questo significa: in questa classe di giochi comportarsi razionalmente significa affidare la propria scelta al caso! Nel nostro esempio, non è difficile controllare che la coppia di strategie miste: $s=(1/5)$ (pari); $(4/5)$ (dispari), e $s'=(1/5)$ (pari); $(4/5)$ (dispari) costituisce un punto di equilibrio. Inoltre, la risposta al terzo problema di Montmort è che il valore del regalo che il padre fa al figlio è pari a $4/5$ di scudo.

La teoria bayesiana di Harsanyi e Selten estende il risultato di Neumann a tutti i cosiddetti giochi non-cooperativi, ai giochi cioè in cui i giocatori non possono contare sul rispetto di accordi strategici eventualmente stipulati nel corso del gioco: per ciascuno di essi, esiste un punto di equilibrio, e dunque una nozione di comportamento razionale.

4. *Lo spettatore simpatetico e imparziale.*

Decisioni individuali in situazioni non-competitive e competitive non esauriscono ancora lo spazio delle decisioni. In alcuni casi, gli individui non agiscono semplicemente al fine di massimizzare la *propria* funzione di utilità (o come se questo fosse il loro obiettivo) ma tenendo conto anche degli interessi di altri individui, prendono cioè la loro ✦decisione✦ riconoscendo come parti interessate anche altri individui. Qui, non è affatto ovvio quale delle varie funzioni di utilità debba essere massimizzata. Una via d'uscita consiste nel riconoscere al gruppo d'individui in questione una funzione d'utilità di gruppo non necessariamente identica ad alcuna delle funzioni di utilità degli individui che costituiscono il gruppo. Tale funzione naturalmente esisterà se il sistema di preferenze del gruppo è coerente, cosí che il gruppo si comporterà esternamente come un unico individuo bayesiano massimizzando tale funzione. Tuttavia, queste nozioni di «funzione d'utilità di gruppo» e «sistema di preferenze di gruppo» non hanno alcun contenuto operazionale fino a che non venga specificata una procedura per determinarle. Potremmo allora dire: la funzione di gruppo è quella che sarebbe costruita da uno spettatore simpatetico e imparziale che desse ugual peso agl'interessi di ciascuno degli individui coinvolti.

Siano $a_1, \ldots, a_n$ tali individui, $U_1, \ldots, U_n$ le loro rispettive funzioni di utilità individuali e A e B le decisioni possibili. In tal caso, uno spettatore *simpatetico* si metterà nei panni di ciascuno degli n individui in modo da scegliere una stessa scala e origine per le n funzioni (si ricordi che le funzioni di utilità sono uniche a meno della scelta dell'origine e dell'unità di misura!) e uno spettatore *imparziale* assumerà che vi è la stessa probabilità, pari a $1/n$, di trovarsi nei panni di ciascuno degli n individui, e cioè di avere il *suo* sistema di preferenze. Cosí, il problema per uno spettatore simpatetico e imparziale assumerà la forma:

	$\frac{1}{n}$		$\frac{1}{n}$
A	$U'_1(A)$	...	$U'_n(A)$
B	$U'_1(B)$	...	$U'_n(B)$

dove $U'_1, \ldots, U'_n$ sono connesse a $U_1, \ldots, U_n$ da opportune trasformazioni lineari crescenti. Sarà quindi un problema di decisione individuale in condizioni di incertezza. Il principio di massimizzazione dell'utilità prevista implica allora che la funzione di utilità di gruppo, e cioè dello spettatore simpatetico e imparziale, W_0, abbia la forma seguente:

$$W_0 = \frac{1}{n}\sum_{i=1}^{n} U'_i.$$

Una decisione di gruppo razionale è perciò quella che massimizza l'utilità media del gruppo, come ha sostenuto l'intera tradizione utilitarista a partire da Francis Hutchenson con il famoso slogan secondo cui nel confrontare le qualità morali delle azioni al fine di scegliere tra le varie azioni proposte, oppure di scoprire quale di esse è moralmente migliore, siamo portati dalla nostra percezione morale della virtú a giudicare che è migliore quell'azione che produce la maggior felicità per il maggior numero (*An Inquiry into the Original of our Ideas of Beauty and Virtue*, 1725).

5. *Teoria e pratica.*

Cosí, la teoria bayesiana della razionalità è, almeno in linea di principio, in grado di dare una risposta circa ogni questione pratica: decisioni individuali noncompetitive, competitive, decisioni di gruppo... È quindi il punto d'arrivo del progetto razionalista di superare la spaccatura fra i termini della coppia ♦teoria/pratica♦ e di assoggettare la pratica a criteri di razionalità. «I razionalisti», afferma Paul Feyerabend nella *Scienza in una società libera* (*Erkenntnis für freie Menschen*, 1980), «vogliono che ci si comporti sempre in modo razionale, ossia che si prendano decisioni secondo regole e criteri che essi e i loro amici considerano importanti e fondamentali. L'esempio della scienza indica che un tale comportamento non conduce ad alcun risultato: il mondo fisico è troppo complesso per poter essere dominato e compreso con l'ausilio di metodi "razionali". Ma il mondo sociale, il mondo del pensiero e del sentimento umano, della fantasia umana, il mondo della filosofia, della poesia, delle scienze, il mondo della convivenza politica è ancora piú complicato. Ci si deve forse attendere che i razionalisti abbiano successo in questo mondo, dopo aver fallito nel mondo fisico?»

In questa affermazione, Feyerabend è solo l'eco piú recente della tradizione che Kant brillantemente descrisse con il motto popolare «Questo va bene in teoria, ma non in pratica», tradizione rappresentata in modo appena piú sofisticato da Vico nel *De nostri temporis studiorum ratione* (1708), dove afferma che non operano saggiamente coloro che negli usi pratici della prudenza civile s'avvalgono dei medesimi criteri di giudizio adoperati nella scienza.

Ora, c'è un'unica interpretazione dell'affermazione di Feyerabend che la renda intelligibile, e cioè che la teoria bayesiana della razionalità ha un ambito di *applicazioni pratiche*, di applicazioni a problemi concreti della vita di tutti i giorni, estremamente ristretto. In questa interpretazione, anzi, essa è vera. Eppure, le applicazioni pratiche riuscite, di cui l'esempio piú noto è quello dell'industria estrattiva, sembrano corroborare l'aspettativa opposta a quella di Feyerabend, e cioè che il programma di ricerca bayesiano, se sufficientemente sviluppato, consentirà, non solo ai razionalisti, ma a tutti gli uomini, di avere «successo in questo mondo», o meglio, di avere piú successo di quanto non ne avrebbero seguendo i princípî dei programmi alternativi attualmente disponibili. [M. M. e S. M.].

Carnap, R., e Jeffrey, R.
1971 (a cura di) *Studies in Inductive Logic and Probability*, vol. I, University of California Press, Berkeley Cal.

Costantini, D.
1977 *Introduzione alla probabilità*, Boringhieri, Torino.

Costantini, D., e Geymonat, L.
1982 *Filosofia della probabilità*, Feltrinelli, Milano.

Cox, R. T.
1961 *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore.

De Finetti, B.
1970 *Teoria delle probabilità. Sintesi introduttiva con appendice critica*, Einaudi, Torino.

Feller, W.
1950-66 *An Introduction to Probability Theory and its Applications*, 2 voll., Wiley, New York.

Good, I. J.
1965 *The Estimation of Probabilities*, Mit Press, Cambridge Mass.

Harsanyi, J. C.
1975 *The tracing procedure: a Bayesian approach to defining a solution for n-person non-cooperative games*, in «International Journal of Game Theory», IV, pp. 61-94.
1976 *Essays on Ethics, Social Behavior, and Scientific Explanation*, Reidel, Dordrecht-Boston.
1977 *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press, New York.
1982 *Papers in Game Theory*, Reidel, Dordrecht.

Jaynes, E. T.
[1958] *Probability Theory in Science and Engineering. Colloquium Lecture in Pure and Applied Science*, Socony Mobil Oil Field Research Laboratory, Dallas 1959.
1968 *Prior probabilities*, in «IEEE Transactions on System Science and Cybernetics», SSC-4, pp. 227-41; ora in V. M. Rao Tummala e R. C. Henshaw (a cura di), *Concepts and Applications of Modern Decision Models*, Michigan State University Press, East Lansing Mich. 1976.
[1973] *Confidence Intervals vs Bayesian Intervals*, in W. L. Harper e C. A. Hooker (a cura di), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, II. *Foundations and Philosophy of Statistical Inference. Proceedings of a Colloquium Held at the University of London, 10-13 May, 1973*, Reidel, Dordrecht 1976, pp. 175-213 e 229-57.
1980 *The Intuitive Inadequacy of Classical Statistics*, in D. Costantini (a cura di), *Atti del convegno internazionale sui fondamenti della probabilità e della statistica, Luino, 1980*, ciclostilato.

Jeffrey, R.
1965 *The Logic of Decision*, McGraw-Hill, New York.
1980 (a cura di) *Studies in Inductive Logic and Probability*, vol. II, University of California Press, Berkeley Cal.

Krantz, D. H., e altri
1971 *Foundations of Measurement*, Academic Press, New York.

Lindley, D. V.
1965 *Introduction to Probability and Statistics from a Bayesian Point of View*, Cambridge University Press, London.
1971 *Making Decisions*, Wiley, New York.

Luce, R. D., e Raiffa, H.
1957 *Games and Decisions. Introduction and Critical Survey*, Wiley, New York 1967[7].

Neumann, J. von, e Morgenstern, O.
1947 *Theory of Games and Economic Behavior*, Princeton University Press, Princeton N.J. 1947[2].

Raiffa, H.
1968 *Decision Analysis. Introductory Lectures on Choice under Uncertainty*, Addison-Wesley, Reading Mass.

Raiffa, H., e Schlaifer, R.
1961 *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, Boston.

Rosenkrantz, R. D.
1977 *Inference, Method and Decision: toward a Bayesian Philosophy of Science*, Reidel, Dordrecht.

Savage, L. J.
1954 *The Foundations of Statistics*, Wiley, New York; ed. Dover, New York 1972.

Schlaifer, R.
1959 *Probability and Statistics for Business Decisions; an Introduction to Managerial Economics under Uncertainty*, McGraw-Hill, New York.

Suppes, P.
1981 *La logique du probable*, Flammarion, Paris.

Tribus, M.
1962 *The Use of the Maximum Entropy Estimate in the Estimation of Reliability*, in E. Machol e P. E. Gray (a cura di), *Recent Developments in Information and Decision Processes*, Macmillan, New York, pp. 102-40.

Zellner, A.
1971 *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Decisione

1. *Introduzione.*

1.1. Certezza, incertezza, giochi.

Tutto ciò che avviene al mondo, dai fatti piú insignificanti ai piú salienti, è prodotto, od almeno condizionato, dall'effetto congiunto di innumerevoli piccole o grandi decisioni di ciascuno di noi: decisioni spesso prese per abitudine, quasi senza riflettere, con piú o meno sensatezza acquisita, e altre volte meditate, con finalità volute, con azioni adeguate.

Grande è pertanto l'importanza delle decisioni e degli studi intesi a mettere in luce, nei suoi vari aspetti, la problematica cui dànno luogo. Menzioniamo in particolare gli aspetti attinenti alla psicologia (intesi soprattutto a sperimentare la coerenza e razionalità di vari soggetti), all'economia (dove teorie normative sono sorte, specie in vista di problemi tipo ricerca operativa), e alla politica (ove una migliore elaborazione dell'informazione sarebbe di valido aiuto per migliori provvedimenti); su tali tre aspetti si possono vedere (nell'ordine) Edwards [1968], Marschak [1968], Robinson [1968].

Nel presente articolo il problema delle decisioni viene considerato essenzialmente dal punto di vista normativo, chiedendosi cioè quale sia il criterio migliore di decisione per raggiungere quanto piú possibile i risultati desiderati.

Il problema si presenta, schematizzando, in tre distinte condizioni, via via piú complesse, che andranno introdotte e studiate separatamente, con successive estensioni dell'impostazione.

Il primo caso, il piú semplice, è quello di decisioni in situazione di certezza: ad ogni scelta di una tra le alternative disponibili corrisponde un risultato certo, e tutto si riduce pertanto al confronto di preferibilità tra i risultati possibili.

Il caso di decisioni in situazione di incertezza – dove, cioè, il risultato dipende non solo dalla decisione presa ma anche da circostanze non note (dal «caso», come spesso si dice) – richiede che il precedente confronto di preferibilità sia non solo qualitativo ma quantitativo, e che vengano valutate le probabilità di dette circostanze sconosciute.

Il caso, infine, di decisioni in situazione di incertezza competitiva (o di «teoria dei giochi»), dove cioè esiste un competitore (o piú) che può influire con le sue scelte sul risultato di ogni nostra scelta, è il piú complesso. Al riguardo ci limiteremo a cenni sul caso piú semplice e significativo (giochi tra due persone, a «somma nulla»: cioè dove quel che uno perde è guadagnato dal competitore, senza apporti o prelievi da altra fonte). I casi piú complessi presentano una grande varietà di situazioni e di problemi; un cenno anche sommario su tale campo riuscirebbe inevitabilmente troppo lungo pur senza riuscire a integrare in modo significativo l'idea già fornita.

1.2. Chiarimento antimetafisico.

Si tratta di un chiarimento puramente terminologico che forse si potrebbe saltare ma che è invece necessario sottolineare fin d'ora per evitare fraintendimenti. Non si tratta di argomentazioni contro la metafisica, che qui non c'entra, ma contro il rischio di interpretazioni «metafisiche» delle precedenti considerazioni in cui si parla di certezza e di incertezza, e, come conseguenza, di interpretazioni metafisiche della nozione di probabilità che quanto prima introdurremo.

Di per sé, indipendentemente dalle conoscenze di questo o quell'individuo, un evento (ossia l'affermazione, la proposizione, la frase, o una formula che lo esprime) è o vero o falso (anche se si riferisce, ad esempio, a un fatto futuro imprevedibile o a un fatto storico di cui non si hanno notizie). Questa distinzione, in vero e falso, ha carattere oggettivo.

Per un dato individuo, nel suo presente stato d'informazione (per esempio «io»; spesso si preferisce dire «tu» per far immedesimare nella parte del soggetto), un evento è o certo, o impossibile, o possibile, a seconda che, in base a quanto «sa», lo ritiene certamente vero, o certamente falso, o non è certo di nessuna delle due alternative. Normalmente accadrà (e in genere si supporrà che cosí sia) che il giudizio si basi su informazioni esatte, cosicché il certo sia vero e l'impossibile sia falso (ma, come è ben noto, non sempre è cosí). In quanto tali distinzioni (incerto, o possibile, o impossibile) si riferiscono al soggetto che si considera, e al suo momentaneo stato d'informazione, esse hanno carattere soggettivo.

In queste precisazioni non c'è nulla di metafisico.

Sarebbe invece metafisico chiedersi se un evento futuro (di questo o quel tipo) sia fin d'ora «predestinato» a risultare vero (o falso), sia in base a leggi deterministiche, o ad opera del «destino», o in base ad altre locuzioni del genere, oppure, al contrario, «casuale». Peggio ancora sarebbe farsi un'idea di pseudo-determinismo basato sulla probabilità e sulle «leggi del caso»; anche se certe considerazioni probabilistiche assegnano un'alta probabilità a certe previsioni (ed è naturale attendersi che si verifichino), bisogna ben distinguere la previsione da una predizione (che è un'affermazione pura e semplice, categorica).

Queste considerazioni sono anticipazioni, di per sé premature e quindi di approssimativa comprensione, ma tuttavia opportune per mettere tempestivamente in guardia contro preconcetti e distorsioni che tendono a intrufolarsi nei modi piú subdoli nel seno delle interpretazioni delle considerazioni probabilistiche.

2. *Un groviglio di problemi.*

2.1. Da dove cominciare?

Cominciare dal caso piú banale è forse cosa troppo banale, ma è probabilmente il solo modo per giungere al panorama piú largo introducendo via via nuovi aspetti senza trovarsi subito in un groviglio.

E quali sono le circostanze che definiscono questo caso fortunato, fortunato per il privilegio di risultare «il piú banale»? Forse è piú appropriato invertire la domanda: quali sono le circostanze che complicano i problemi, e la cui assenza costituisce il privilegio caratterizzante il caso piú banale?

Ogni decisione è una scelta tra piú alternative (due, parecchie, molte, infinite) fatta (da chi?) in base a delle preferenze (quali?)

Da chi? Il caso piú semplice è ovviamente quello di un'unica persona responsabile, che, per usare la denominazione ormai invalsa, chiameremo *decision maker*. Se la decisione deve venir presa di concerto fra piú persone sorgono evidentemente molti problemi, sostanziali e procedurali. (Ne vedremo tosto, nel § 2.2, un tipico esempio).

E in base a quali preferenze? Al riguardo vi sono sempre molte difficoltà (anche per una scelta puramente personale), a seconda della minore o maggiore numerosità delle alternative (poche, parecchie, un numero grande ma finito, oppure una infinità: discreta, o continua, o funzionale, ...) Ma, piú che il numero, influisce sulla conseguente complessità del problema il fatto che le alternative siano disparate (ad esempio, la scelta di una persona tra un gruppo di aspiranti dalle caratteristiche molto diverse per pregi e difetti: scelta «multi-attributi»), oppure omogenee (ad esempio la scelta del numero n di azioni che un capitalista pensa di sottoscrivere). In quest'ultimo caso si presenterà naturalmente (in modo piú o meno semplice) una funzione di preferenza $f(n)$ che indichi ragionevolmente le motivazioni per contenere o ampliare l'investimento, e si tratterà di scegliere un n in prossimità del massimo (e abbastanza «tondo»).

Forse è d'obbligo chiedere scusa per quest'ultima osservazione tra parentesi, che – come qualcuno certo dirà o penserà – «svilisce la Teoria». Ma la sua motivazione sta proprio nell'opportunità di raccomandare – cogliendo tale occasione – di non mitizzare la teoria (il che, se ci sono persone di buon senso, significa anche ridicolizzarla). Non bisogna accettarne le conclusioni come apodittiche nella loro esattezza sia pure illusoria, mentre esigenze altrettanto irrilevanti come quella della comodità di quantità «tonde» possono ben competere ed avere giustamente il sopravvento.

È fin troppo facile, ma altrettanto rovinoso, comportarsi con noncuranza per le piccole cose, magari ripetendo con vanteria che «de minimis non curat praetor». Ma è, in gran parte, proprio per colpa di tale diffusa stortura se il complesso di tutte le cose va sempre peggio proprio quando piú sarebbe facile raggiungere il meglio, correggendo e capovolgendo la mentalità dominante.

È stranissimo infatti come lo scrupolo e il dispregio per l'esattezza avvengano proprio a rovescio rispetto a ciò che sarebbe appropriato e opportuno. Conteggi per importi grandi o piccoli vengono spesso presentati con l'esattezza fittizia alla lira di antica memoria, mentre trascurata – e dai piú ignorata – è l'esigenza di attenersi esattamente a norme di unificazione (Uni, spesso originariamente Din) che – per riferirsi al caso piú banale ma che riguarda tutti – eviterebbero la scomoda mescolanza di formati difformi nella corrispondenza con l'adozione generale del formato mm 210×297 (lettera), o metà, mm 148×210 (memorandum), tutti derivati (come gli altri della serie A) dal foglio di 1 m^2

(1,19 × 0,84) per successivi dimezzamenti (che mantengono le proporzioni, essendo le misure in progressione geometrica di ragione $\sqrt{2}$). Analoga progressione, utile per dimensionare una serie di oggetti (per esempio pentole) di uguale forma, è quella dei «numeri normali» (di ragione $\sqrt[10]{10}$, con opportuni arrotondamenti), e che meriterebbe di entrare nell'uso corrente in molti piú campi: ad esempio, per un aumento di prezzi o tariffe, basterebbe slittare la scala di uno o piú gradini (e idem per sconti). Ogni gradino corrisponderebbe a uno scatto del 7,2 per cento circa $(=\sqrt[10]{2}-1)$.

Con queste considerazioni non si vuol contestare che si debba dare maggiore importanza alle cose piú importanti, ma è certo piú grande – causa la lamentata incuria delle cose piccole – il danno che proviene dal non darne abbastanza alle cose minori e minime che, complessivamente, hanno non minore influenza su tutto ciò che ci circonda e condiziona.

Tutto ha il suo giusto valore e i suoi giusti limiti. Anche nel caso dei problemi di cui trattiamo (come in ogni altro campo, e nel caso delle religioni e di ogni altra ideologia) sono atteggiamenti sciocchi ed esiziali tanto l'insensibilità quanto il bigottismo.

E chiudiamo queste premesse generiche richiamando l'attenzione su un altro aspetto, diverso, in certo senso collaterale, ma assai importante per gli effetti che ne derivano. Se ogni individuo (o i piú), nel fissare le sue preferenze e i propri obiettivi e nel decidere le proprie azioni, trascura le ripercussioni negative che possono derivarne ad altri (disturbandoli, creando pericoli, inquinamenti, dissapori), gli eventuali vantaggi che ciascuno con sforzo cerca di assicurare a se stesso saranno inevitabilmente annullati dalla mancata serenità ed armonia dell'ambiente totale in cui si svolge, bene o male e in questo caso male, l'inevitabile convivenza di una comunità civile (ma – in genere – non troppo).

2.2. Decisioni collettive e paradossi.

Sono ben note le difficoltà che si presentano nell'interpretare risultati di elezioni o di inchieste intese a rivelare le preferenze della popolazione (degli elettori, di un campione di cittadini, di un gruppo di competenti, ecc.) riguardo a problemi e correnti di idee di vario tipo. Le cifre parlano abbastanza, a volte in modo perentorio, ma ciò che piú importa, il perché che c'è sotto, la molla che ha fatto scattare un rafforzamento o un'inversione di tendenza, rimangono opinabili.

Ma, a parte queste difficoltà in certo senso collaterali, è bene ricordare come il problema di ricavare, da opinioni o preferenze individuali, una opinione o preferenza «collettiva», su cui basare una eventuale decisione collettiva, comporti autentici paradossi matematici, noti da lungo tempo (Condorcet, Dodgson, ecc.), e recentemente ristudiati e generalizzati da Arrow [1951]; parecchi scritti ha dedicato all'argomento anche Black [1948-49; 1958].

Il paradosso piú semplice e noto è il seguente: facendo indicare a dei votanti l'ordine di preferenza in cui collocano alcuni candidati, può risultare una conclusione contraddittoria, quale: A è preferito a B; B è preferito a C; C è preferito ad A. Basta infatti supporre che tre votanti diano le tre graduatorie seguenti: primo-secondo-terzo, rispettivamente A-B-C, B-C-A, C-A-B; risulta che due (il 1° e il 3°) preferiscono (antepongono) A a B; due (il 1° e il 2°) preferiscono B a C; due (il 2° e il 3°) preferiscono C ad A.

Né si pensi che il paradosso si possa verificare solo eccezionalmente, in questo e qualche altro caso artificioso con pochi individui. Esso appare sempre (nel nostro esempio), qualunque sia il numero di votanti, quando uno dei due gruppi di terne (sinistrorso: ABC, BCA, CAB; destrorso: CBA, BAC, ACB) prevalga sull'altro in modo abbastanza netto e uniforme.

La presenza di complicazioni e paradossi del genere esemplificato conferma il pessimismo circa la possibilità di definire procedure di decisione collettiva che si impongano come «razionali»; pessimismo avvalorato dall'esperienza di discussioni in cui, in ogni occasione, i sostenitori di opposte decisioni insistono per applicare – portando plausibili motivazioni per sostenere che è obiettivamente la «migliore» – quella procedura che caso per caso reputano piú favorevole al conseguimento del risultato desiderato dalla loro parte.

Purtroppo, i dubbi vanno anche oltre: Arrow ha dimostrato che (anche a prescindere dal modo di decidere) neppure esistono soluzioni soddisfacenti ad un insieme di requisiti di cui una situazione sembra dovrebbe godere per esser giudicata «ottimale» riguardo alle esigenze della collettività.

E allora? La conclusione, alquanto sconfortante ma realistica, sta forse nel ritornare – naturalmente, interpretandolo con spirito piú egualitario – all'antico precetto di «distribuire equamente il malcontento» secondo il rapporto tra il calcolo economico dei singoli e della collettività, come affermato e discusso dagli «Italian Writers on Public Finance» (come De Viti De Marco e Luigi Einaudi), citati da Duncan Black a riconoscimento dell'influsso avuto dal loro pensiero sulle sue riflessioni.

Primo ed essenziale tra i *Requisiti per un sistema economico accettabile in relazione alle esigenze della collettività* [De Finetti 1973], anche nel senso delle precedenti citazioni, è (a parere dello scrivente) quello di costituire un «optimum paretiano»; di tali situazioni ne esistono però infinite e di per sé possono comportare disuguaglianze non solo piccole ma anche grandi quanto si vuole. La condizione egualitaria – almeno nel senso di ridurre le disuguaglianze ad un livello tollerabile – è il secondo requisito, che sembrerebbe assurdo non aggiungere, anche se sarebbero di parere opposto quanti sembrano disposti a difendere qualunque ingiustizia e mostruosità giustificandola con le circostanze storiche e le concezioni giuridiche in cui è sorta, come se situazioni e concezioni attuali e piú progredite non consentissero – ed anzi imponessero – il superamento di non piú sopportabili retaggi di barbarie.

La teoria delle decisioni andrebbe applicata soprattutto alla ricerca di un optimum per la collettività, e solo poi, in via subordinata, all'analoga ricerca a livello settoriale o regionale o addirittura aziendale o familiare o individuale. Si dovrebbe anzi pensare, avanti a tutto, alla preservazione della vita nella biosfera, e quindi all'uomo col compito della sua regolazione, se saprà raccogliere il messaggio di rari chiaroveggenti come Peccei, Huxley, Salk, messaggio che sembra prefigurato nell'ammonimento di Dante: «Fatti non foste a viver come bruti, | ma per seguir virtute e conoscenza» [*Inf.*, XXVI, 119-20], o ancora prima, in forma sublime quanto semplice, nel *Cantico delle creature* di san Francesco. Soltanto in questa prospettiva può esserci speranza per il futuro.

2.3. Decisioni individuali e collettive: coerenza.

Primo requisito per ogni decisione è la coerenza (nel senso che verrà precisato). Prima ancora di tale precisazione conviene però sottolineare come tale esigenza, intrinseca, necessaria e sufficiente perché una decisione sia formalmente «accettabile», valga allo stesso modo per decisioni individuali (ove può mancare solo per svista) e per decisioni collettive (dove occorre invece particolare cura per conseguirla).

Un problema di decisione s'incontra ogni qual volta ci si offre la possibilità, o ci si trova nella necessità, di scegliere tra diverse alternative possibili. (Tra le due formulazioni non c'è sostanzialmente nessuna diversità; dicendo «possibilità» si dice solo che oltre alle alternative esplicitamente formulate c'è anche quella di non far nulla, o nulla oltre al già prestabilito).

La decisione dipende da un confronto tra i benefici che ci si possono attendere da ogni alternativa (o da un insieme di alternative eventualmente compatibili); in termini mercantili o manageriali si parlerebbe di costi e ricavi, perdite e profitti, e si può anche dire che, metodologicamente, si tratta della stessa cosa, benché l'essenza cambi molto dovendosi intendere nel confronto anche, e spesso in misura preminente, tutti gli altri elementi non monetizzabili; oltre i guadagni: il piacere, la soddisfazione, il divertimento, gli apprezzamenti, le acquisizioni culturali, gli svaghi; oltre i costi: i sacrifici, le ansie, i pericoli, le controversie, gli intoppi, i piccoli e grandi passi falsi e insuccessi.

La teoria delle decisioni insegna pertanto, grosso modo, a tradurre tutti questi elementi (considerati e detti, spesso, «imponderabili» o, all'inglese, «intangibili») in termini monetari affinché abbiano il peso che intendiamo loro attribuire agli effetti di un confronto meno piatto e unilaterale di quello che si limita agli aspetti finanziari. (Il quale, beninteso, conserva tutto il suo valore agli effetti suoi propri).

Dato che in molti casi (in genere, i piú interessanti) molti elementi sono aleatori (possono presentarsi o meno, oppure presentarsi in misura piú o meno grande), interviene la probabilità (che qualcosa avvenga o no, o che l'intensità ne sia piú o meno grande). In tale caso si considereranno non i valori monetari effettivi bensí la loro previsione (o speranza matematica, o valor medio in senso probabilistico); se ne vedrà la definizione quando introdurremo la nozione di probabilità che ne è alla base (§§ 4.1 sgg.).

Parallelamente all'introduzione della probabilità, si presenta appropriato un altro perfezionamento: quello di sostituire al valore in senso monetario degli importi da mettere in conto (guadagni e perdite) il valore in termini di utilità (in cui si tiene conto del fatto che l'utilità marginale di ogni guadagno – o, in senso negativo, perdita – decresce al crescere della ricchezza posseduta).

Tutte queste valutazioni e previsioni dipendono dallo stato di informazione di chi ne giudica; per migliorarle conviene in genere acquisire maggiore informazione su aspetti rilevanti al riguardo. Potrà trattarsi di consultare dati statistici, di eseguire sperimentazioni statistiche (ad esempio collaudi), ma anche di sentire il parere di esperti e le loro analisi e previsioni, di confronti tra la situazione attuale e le precedenti nel campo che lo riguarda, e via dicendo. Teoricamente, ogni acquisizione d'informazione ha un certo costo (se non monetario, in tempo, ecc.) ed un certo valore (per chi deve prendere delle decisioni): il maggior beneficio derivante da una decisione presa con migliore cognizione di causa.

Senza entrare in ulteriori dettagli, possiamo dire che il problema delle decisioni (piú specificamente: delle decisioni in condizioni di incertezza) è in tal modo delineato. Si tratterà di sviluppare l'impostazione concretamente, precisando e chiarendo i concetti qui accennati in forma sintetica e preliminare.

Quanto detto vale per decisioni in generale, senza distinzione fra decisioni individuali e collettive. Va però notato che ciò implica che tutte le valutazioni siano fatte in modo *coerente*, secondo criteri concordati e applicati in modo uniforme da tutti coloro che partecipano alla decisione, sia per le probabilità e sia per le utilità. Un complesso di decisioni parziali, ciascuna coerente in sé ma relativamente a criteri diversi di individui diversi, non è nel suo insieme una valutazione coerente. Per esprimerci con una formuletta – che dovrebbe risultare chiara come sintesi di quanto detto – per una decisione collettiva ci si deve *basare su una «media di opinioni»*, e NON *ricorrere a una «media di decisioni»*.

Una riprova della bontà di questo precetto è data, per contrasto, dal mosaico di decisioni parziali scoordinate e piú o meno incompatibili che scaturisce da altrettante decisioni collegiali indipendenti. E purtroppo questo metodo sembra costituire l'accomodamento piú consueto in simili situazioni: basti pensare all'incoerenza interna che può prodursi in una legge qualora nelle votazioni sui singoli articoli ed emendamenti abbiano prevalso a volta a volta tendenze e maggioranze diverse.

Una *conclusione provvisoria*, che dovrebbe esser già apparsa sufficientemente ovvia in base alle preliminari considerazioni finora svolte, e che comunque *supporremo acquisita e accettata fin d'ora*, consiste nel dire che:

Unico criterio corretto di scelta in una decisione in condizioni di incertezza è quello basato sulla massimizzazione della utilità sperata (approssimativamente sostituibile, per decisioni con valori in gioco limitati, con la massimizzazione del guadagno sperato).

Il senso è sufficientemente chiaro, intuitivo, anche se le appropriate precisazioni teorico-critiche troveranno il loro posto naturale soltanto nei §§ 4.1 sgg., dopo introdotti i concetti fondamentali sulla probabilità.

2.4. Preconcetti e distorsioni.

Un difetto inevitabile (ma bisogna cercare almeno di limitarlo!) consiste nel non saper immaginare tutte le conseguenze possibili, vicine e lontane, di ogni atto (per forza: sono infinite!)

Non ricordo quale personaggio aveva battezzato «Postulato di mia moglie» la seguente osservazione, o ammonimento, che essa gli rammentava spesso (e, naturalmente, aveva sempre ragione): «In qualunque faccenda è sempre facile entrarci ma assai piú difficile uscirne». E ciò è appunto conseguenza del non poter prevedere tutti gli imprevisti possibili, che quasi sempre complicano le cose anziché semplificarle.

Altro difetto assai diffuso è la tendenza ad applicare «regole» piú o meno tradizionali per determinati tipi di questioni, regole che inevitabilmente sono alquanto «rozze», posto che vengano indicate come ricette fisse, senza controindicazioni o adattamenti. In particolare, anche ogni procedura «esatta» per una data decisione in una data situazione in un dato istante diventa distorta se la si applica in altra occasione senza le modifiche richieste per sopravvenute variazioni in ogni tipo di aspetti: nei rischi, nei costumi, nei prezzi, in altri elementi rilevanti.

Piú genericamente ancora possono influire, dando una piccola ma decisiva spinta finale in un senso o nell'altro, i piú generici tra i preconcetti generici: quelli espressi in massime che intendono indurre, indiscriminatamente, a innovare o no, a rischiare o no; ad esempio: «Chi non risica non rosica», «Chi lascia la via vecchia per la nuova male si ritrova».

Un diverso tipo di preconcetti è quello che fa ritenere opportuno, a taluni, di prendere la decisione «subito», «d'istinto», quasi «per ispirazione», e ad altri di prenderla con calma, lasciando «maturare» da sé, nel loro intimo, una convinzione in un senso o nell'altro.

Cosa dire? Non c'è dubbio che, in varia misura (a volte modesta, a volte sbalorditiva), molte persone (e, al rispettivo livello, anche molti animali) riescono a rispondere in modo indovinato a stimoli e problemi «come se» avessero elaborato correttamente (o lentamente, o, talvolta, pressoché istantaneamente) una moltitudine di dati in relazione all'opportunità di scegliere l'azione piú appropriata. (Si pensi alla prontezza di riflessi e all'automatismo di reazioni grazie a cui degli automobilisti – e in analoghe occasioni anche animali – riescono a evitare in extremis di trovarsi coinvolti in improvvisi incidenti di cui altrimenti sarebbero rimasti vittime).

Alla domanda «Cosa dire?» sarebbe preferibile rispondesse uno psicologo. Ritengo comunque che sarebbe eccessivo confidare in modo troppo miracolistico in queste facoltà misteriose, ma che sarebbe peggio, e dannoso, il consiglio di diffidarne ripudiandole o scambiando l'una con l'altra (con calma anziché subito, o viceversa, contrastando il proprio istinto). Tuttavia, anche per chi abbia motivo di confidare in siffatte doti istintive, è ben ragionevole pensare che l'aggiunta di una conoscenza precisa dei termini logici, matematici, psicologici, della questione e della teoria che vi si basa, se intesa a innestarsi corroborando dette facoltà e non a soppiantarle, costituirebbe pur sempre un notevole arricchimento. Tutto, in fondo, piú o meno, andrebbe a fondersi con la parte istintiva.

Tale acquisizione serve poi, comunque, anche a formare quell'intima comprensione della natura dei problemi che può aiutare a intravvedere la soluzione e, meglio ancora, il *perché* della soluzione. Può aiutare i presunti «esperti d'istinto» a comprendere e controllare, migliorandoli, i loro ragionamenti inconsci, e gli altri a intravvedere qualcosa nella medesima direzione, come sempre avviene per chi apprenda e approfondisca un ragionamento logico, non come mera tecnica ma come forma di pensiero e di visione.

2.5. Il ripudio dell'incertezza.

Oltre ai preconcetti di carattere generico di cui si è detto, ne esistono altri di carattere piú tecnico, strettamente connesso alla tematica del problema: del problema delle decisioni in condizioni d'incertezza.

Vi sono molte persone, anche tra quelle che si devono occupare e si occupano di problemi in cui interviene l'incertezza, che dimostrano chiaramente, sia nel modo di pensare che nel modo di agire, di «non poter soffrire l'incertezza». È stato detto, del resto, da un celebre psicologo, Cohen [1960], che «uncertainty is not easy to bear»: l'incertezza non è facile da sopportare. E ciò si può collegare (forse... ma direi «certamente») a un'altra acuta osservazione del medesimo autore: fin dalla scuola i ragazzi vengono «istruiti a credere che esistano soltanto *cut and dried questions*» (questioni predisposte e risecchite) alle quali si debba sempre e soltanto rispondere con un'unica altrettanto *cut and dried answer* (una risposta predisposta e risecchita).

Ciò è tanto piú deplorevole dato che già Bacone aveva criticato l'antieducativa frettolosità nel soffocare il dubbio e far accettare una certezza senza lasciar tempo ad una adeguata maturazione del giudizio. In tal modo anche le verità vere vengono inculcate come pregiudizi.

E, peggio ancora, addirittura paradossale e inesplicabile è il fatto che la forma piú estrema di tale atteggiamento alligni proprio nel campo di coloro cui, come tecnici o dirigenti, spetta di prendere delle decisioni, o, come esperti o statistici, spetta di proporre metodi e dare consigli.

In tali ambienti, infatti, prevale tuttora (benché la situazione sembra stia sensibilmente migliorando) lo sforzo di trattare dell'incertezza fingendo di ignorare l'incertezza o limitandosi a tenerne conto mediante qualche ripiego empirico.

La soluzione piú semplice (e semplicista) consiste nell'ignorare l'incertezza riferendosi all'ipotesi di una situazione *certa* intermedia fra quelle possibili e piú o meno probabili. Una siffatta riduzione del problema non può peraltro risultare valida sotto tutti gli aspetti: per ogni aspetto si richiederebbe la considerazione di un tipo appropriato di media, e per certi aspetti ciò sarebbe comunque inadeguato. (Non si potrebbero certo sostituire ragionevolmente i rischi, come ad esempio quello dell'incendio di un intero stabilimento, con probabilità 1 per mille per anno, con l'ipotesi della distruzione certa di 1/1000 di esso ogni anno).

Peggio ancora se, anziché riferirsi a una qualche ipotesi «media» (che, nonostante le precedenti critiche, o, meglio, messe in guardia contro interpretazioni acritiche, ha una certa validità orientativa), si considerasse come «certo» il caso «piú probabile» (nozione priva di senso: occorrerebbe riferirsi, per darglielo, ad una determinata, e sempre arbitraria, suddivisione in «sottocasi»), o addirittura (secondo una curiosa tesi di Shackle) a due casi su cui focalizzare l'attenzione: la piú favorevole e la piú sfavorevole delle situazioni ipotetiche «ragionevolmente prevedibili». (Sarebbe come dire «il piú grande dei nani e il piú piccolo dei giganti»).

Occorre invece tener presente l'intera gamma delle situazioni possibili e vagliarne le rispettive probabilità, come vedremo in seguito (§ 4), discutendo del «come si dovrebbe» decidere. Ma allo stesso scopo giova certamente anche ciò che stiamo per dire, segnalando aspetti piú o meno accettabili e rispettivamente piú o meno condannabili riguardo al «come si decide», secondo i metodi standard piú abituali.

2.6. *Adhockeries* per «accettare» o «rifiutare».

È stato Irving Good ad inventare la felice e appropriata denominazione di *adhockeries* (che in italiano è stata tradotta 'adhoccaggini') per indicare i metodi «ad hoc»: le regole empiriche di decisione (o, secondo i casi, di stima) che, per rispondere a certi scrupoli fuori luogo e per inavvertenza di scrupoli doverosi, devono sottostare a dei «tabú» che impediscono l'impostazione naturale. Se ne trovano a bizzeffe negli *statistical cookbooks* (e anche questa denominazione – consueta tra gli stessi angloamericani – è appropriata, perché si tratta di ricette per dare una risposta arbitraria ad un problema volutamente mal posto).

Di *adhockeries* ce n'è per ogni problema e di tipi svariati, ma lo schema di «ragionamento» piú frequente consiste nel «rigettare» una «ipotesi» se un esperimento ha dato un risultato che, ammettendo quell'ipotesi come *vera*, sarebbe risultato estremamente improbabile.

Abbiasi un'urna contenente palline bianche e nere, e la «ipotesi» sia che i due colori siano in ugual numero. L'esperimento consista nel fare un certo numero (per esempio 10) di estrazioni (per fissare le idee, con reimbussolamento; è anche il caso piú semplice), per decidere, in base al suo esito, se accettare l'ipotesi o rigettarla.

I risultati possibili dell'esperimento sono undici: il numero di estrazioni di una pallina bianca può essere 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10; che cosa possiamo concluderne?

Dal punto di vista puramente logico, soltanto questo: nei casi da 1 a 9, che tra le palline ce n'è almeno una bianca e una nera; nei casi 0 e 10 invece è anche possibile che le palline siano tutte nere, o, rispettivamente, tutte bianche.

Dal punto di vista adhoccagginesco, si pseudoragionerebbe a ruota libera dicendo che nei casi estremi: sempre bianco, o sempre nero (e magari anche nei casi vicini con solo una o due eccezioni), l'ipotesi di uguale numerosità va rigettata. E perché? Chiaro: perché se essa fosse vera il risultato ottenuto sarebbe estremamente improbabile (probabilità $(1/2)^{10}=1/1024$, meno di 1 su 1000), cioè «praticamente impossibile»... cioè impossibile... quindi... come volevasi dimostrare.

Disgraziatamente, qualunque successione di risultati si fosse osservata, con qualunque numero e disposizione di bianco e nero (ad esempio BBNBNNNNBB), la sua probabilità (in base alla «ipotesi» da testare) sarebbe stata sempre la stessa, 1/1024, e la risposta pure!

Povero oggettivista: è K.O.; eppure in fondo la sua convinzione è fondata. Possiamo aiutarlo mostrando che nel suo ragionamento e nei suoi calcoli manca un ingrediente essenziale, e, colmando tale lacuna, tutto va a posto (purché non rigetti l'ingrediente essenziale giudicandolo osceno: la probabilità iniziale, soggettiva, delle «ipotesi»).

Egli dovrebbe dirci la sua opinione iniziale (spesso si dice «a priori», ma ciò avrebbe un forviante sapore metafisico); ad esempio, se le palline sono 10, dovrebbe dirci la probabilità che attribuisce al fatto che le bianche siano 0, o 1, o 2, ecc., fino a 10. Allora sí il precedente esperimento diventa informativo, in quanto ogni estrazione costituisce un'informazione che altera via via le probabilità iniziali, rafforzando quelle delle composizioni piú vicine alla frequenza osservata. Non esiste un salto tra «ignoranza» e «certezza», bensí un progressivo adeguamento della valutazione probabilistica all'effetto dell'esperienza combinato coll'opinione iniziale. Cresce sempre la probabilità di avvicinarsi alla quasi certezza del valore vero; ma piú di questa «grande probabilità» la statistica non ci può mai dare.

2.7. Per le decisioni, Sí; per le ipotesi, No.

Il fulcro della differenza fra i due modi di concepire la situazione appare chiaro se pensiamo che il problema precedente sia connesso a un problema di decisione: a seconda che venga «accettata» una delle ipotesi, sia la H_s, prenderemo una corrispondente decisione, D_s; ad esempio potrebbe esser stato pattuito che la compravendita della partita avrà luogo con una riduzione di prezzo pari alla percentuale di pezzi difettosi trovati nel campione.

In tale ipotesi, se ispezionando 100 pezzi se ne sono trovati 12 difettosi (cioè il 12 per cento), la partita in questione, di (poniamo) 1000 pezzi verrà pagata col 12 per cento di sconto, ossia, in altra forma, si pagheranno non tutti i 1000 pezzi ma solo 880, convenendo in via presuntiva concordata di fare *come se* quelli difettosi fossero in tutto 120 (12 per cento di 1000). Come teoria delle decisioni tutto va bene: decidiamo (basta ci sia il comune accordo fra compratore e venditore) di comportarci cosí, di prendere questa decisione: dato l'esito del collaudo si doveva «accettare» questa alternativa e «rifiutare» le altre.

Ma gli oggettivisti non distinguono le due fasi; non dando senso a valutazioni di probabilità (riguardo al vero numero di pezzi difettosi) essi non dicono solo che è stata accettata la vendita a quel prezzo, ma anche che è stata accettata... l'ipotesi corrispondente, cioè che i pezzi difettosi fossero proprio 120 (o magari, poniamo, tra 110 e 130).

Sembra ovvio che tale identificazione di due cose totalmente diverse (anche se collegate tramite un contratto di per sé estraneo alla problematica statistica) è del tutto priva di senso. Non solo perché di un fatto non si può decidere se è vero o falso finché non ne abbiamo certezza (e prima sarà per noi soltanto piú o meno probabile), ma perché ovviamente, in base ai 12 pezzi difettosi su 100 trovati nel campione, possiamo concludere logicamente soltanto che sui 1000 della partita in vendita quelli difettosi non sono meno di 12 né piú di 912 (nulla potendosi escludere riguardo ai 900 non ispezionati).

Se uno si divertisse proprio, o ne facesse un puntiglio, a dire che ha deciso di fare *come se* i pezzi difettosi fossero 120, o magari a dire che è stato constatato che il «numero convenzionale di pezzi difettosi» è 120 (in quanto 120=12 per cento di 1000), poco male. Se uno vuole automistificarsi, buon pro gli faccia. Possiamo anche, con un po' di malafede o compatimento, fingere di assecondarlo nel suo vaneggiamento. Non certo, però, lasciarsi contagiare dai suoi arzigogoli.

Per concludere, in sintesi: 'accettare' e 'rifiutare' (sulla base di esperimenti statistici), sono termini che si addicono alla Decisione (cioè alle varie alternative di una decisione); non invece alle Ipotesi, per le quali un'informazione modifica soltanto la probabilità (soggettiva), salvo il caso limite che essa trapassi in certezza o impossibilità. (Vedi nell'esempio qui sopra: non meno di 12 né piú di 912).

2.8. Comportamento induttivo e ragionamento induttivo.

Dato che il confronto tra la validità dell'una o dell'altra fra le due concezioni sul significato e ruolo della probabilità e statistica per le decisioni nell'incertezza costituisce il punto cruciale per orientarsi sull'argomento, sarà opportuno completare un po' piú la descrizione – sia pur sempre preliminare – di alcuni aspetti sia concettuali che tecnici.

Lo scrupolo, o pregiudizio, che offusca e limita la visuale della concezione «oggettivista» della probabilità, e dei suoi fautori, consiste nel restringere l'uso del concetto e del termine 'probabilità' al caso di eventi che siano «prove ripetibili di un medesimo fenomeno».

Anzi, nella loro terminologia, tutto è assai confuso: si parla di «prove di un medesimo evento» (intendendo evidentemente 'evento' nel senso generale per cui avevamo invece usato 'fenomeno'), si dice che la probabilità di un evento è la frequenza con cui si presenta in numerose «prove», ma pare abbia senso anche per le singole prove dato che si precisa che esse devono essere... «ugualmente probabili» e magari «indipendenti». Non pretendo certo di chiarire questo guazzabuglio, né posso ammettere che ciò sia possibile.

Spiegando i termini della questione come appaiono dal punto di vista soggettivistico, ogni artificiosa confusione e superfetazione dilegua; il ragionamento induttivo è una conseguenza immediata del teorema di Bayes, e il comportamento induttivo non è né altro può essere se non quello di seguire il ragionamento induttivo.

Sviluppiamo un semplice esempio a scopo illustrativo, e cioè mostrando come funziona ma senza spiegare il perché (che si vedrà a suo luogo: § 4.5).

Sappiamo che un'urna contiene palline rosse e nere, con percentuale di rosse o il 25, o il 50, o il 75 per cento. Inizialmente potremo attribuire alle tre ipotesi delle probabilità suggerite da motivi qualunque, in particolare giudicarle equiprobabili (ciascuna, 1/3); nell'esempio che svilupperemo saranno rispettivamente del 30, 50, 20 per cento, ossia staranno nel rapporto 3:5:2. Poi possiamo fare delle estrazioni (con reimbussolamento, cosí che la composizione dell'urna rimanga sempre la stessa). Qual è l'effetto di ogni estrazione sulle probabilità che attribuiamo alle tre ipotesi? La regola (applicando il teorema di Bayes) è molto semplice, specie in questo caso: ad ogni colpo, le probabilità dei tre casi vanno moltiplicate rispettivamente per 1, 2, 3 se esce pallina rossa e per 3, 2, 1 se nera (e poi «normalizzando», cioè dividendo per la somma in modo che il totale torni 1). Oppure si può fare a meno di dividere, contentandosi di sapere ad esempio che stanno nel rapporto 3 : 10 : 6 anziché conoscere i valori in percentuale dati dalla divisione per $3+10+6=19$, e cioè 15,19 per cento, 52,63 per cento, 31,58 per cento.

Si veda sulla tabella 1 il proseguimento del processo di adeguamento delle opinioni ai risultati delle estrazioni, quale si ha tenendo conto degli inevitabili e indispensabili fattori soggettivi (probabilità iniziali) che invece gli oggettivisti vogliono nascondere o negare.

Al contrario, le adhoccaggini – che vogliono tener conto delle sole informazioni «oggettive» sostituendo quelle soggettive col nulla – non consentono alcun «ragionamento» induttivo. E, ad onor del vero, gli stessi fautori dell'impostazione oggettivistica non pretendono che il loro modo di procedere costituisca un «ragionamento» induttivo, bensí soltanto un «comportamento» induttivo: *inductive behaviour* anziché *inductive reasoning*.

Ed è, infatti, esclusivamente, il ragionamento bayesiano che costituisce un *ragionamento*, partendo necessariamente, per colmare la lacuna, da un giudizio soggettivo, e cosí infrangendo un sacro e venerato tabú.

2.9. Impossibile? (col «quasi»?)

La radice di tanti equivoci (forse di tutti?) risiede nel non distinguere, o nel non distinguere abbastanza nettamente, tra «impossibilità» (assoluta, logica) e probabilità molto piccola (spesso chiamata, in modo improprio e atto ad ingenerare equivoci, «impossibilità pratica»).

È il piú puerile dei sotterfugi tentati per trasformare delle previsioni corrette prossime alla certezza in prestigiose predizioni ammantate di certezza assoluta benché fasulla.

Purtroppo anche scienziati famosi si sono intrappolati in arzigogoli del genere; basti rammentare le elucubrazioni di Borel, che tuttavia temperò l'asserto parlando non di «impossibile» in senso assoluto, bensí di «praticamente impossibile» rispettivamente «alla scala umana, terrestre, cosmica, e universale» per un evento di probabilità inferiore a 10^{-6}, 10^{-15}, 10^{-50} e 10^{-1000}; anche ciò può ingenerare confusione, ma di per sé (a parte l'uso forviante del termine 'impossibilità') si tratta solo di far vedere quanto tali probabilità siano piccole (ma non nulle).

Tabella 1.
Andamento della valutazione di probabilità in base all'informazione data dal risultato di successive osservazioni.
Nel presente esempio consideriamo 10 successive estrazioni da un'urna di composizione sconosciuta, sapendo però che essa è stata «scelta a caso» fra 10 urne che contengono palline rosse (R) e nere (N) in proporzioni diverse:

- 2 urne hanno la composizione A: 25%R+75%N; ad esempio 1 rossa e 3 nere;
- 3 urne hanno la composizione B: 50%R+50%N; ad esempio 2 rosse e 2 nere;
- 5 urne hanno la composizione C: 75%R+25%N; ad esempio 3 rosse e 1 nera.

Dopo ogni estrazione le probabilità che l'urna sia del tipo A o B o C si alterano proporzionalmente ad 1 : 2 : 3 (o viceversa: 3 : 2 : 1) perché l'estrazione di pallina rossa (rispettivamente nera) favorisce nel rapporto 1 : 2 : 3 (rispettivamente 3 : 2 : 1) le ipotesi secondo cui il colore estratto figura con il 25 o 50 o 75%.
La tabella spiega elementarmente – con riferimento all'esempio sopra riportato – il meccanismo del modo corretto (bayesiano) di «imparare dall'esperienza».

		Risultati		*Rapporto tra le probabilità*				*Probabilità di estrazione*	
Successive estrazioni		*singoli*	*complessivi*	A	B	C	*totale*	R	N
I.	probabilità			2	3	5	10	42,5	57,5
	risultato	R	(1, 0)	×3	×2	×1			
II.	probabilità			6	6	5	17	51,4	48,6
	risultato	R	(2, 0)	×3	×2	×1			
III.	probabilità			18	12	5	35	59,3	40,7
	risultato	N	(2, 1)	×1	×2	×3			
IV.	probabilità			18	24	15	57	35,5	64,5
	risultato	R	(3, 1)	×3	×2	×1			
V.	probabilità			54	48	15	117	41,7	58,3
	risultato	N	(3, 2)	×1	×2	×3			
VI.	probabilità			54	96	45	195	48,8	51,2
	risultato	R	(4, 2)	×3	×2	×1			
VII.	probabilità			162	192	45	399	37,0	63,0
	risultato	R	(5, 2)	×3	×2	×1			
VIII.	probabilità			486	384	45	915	35,5	64,5
	risultato	R	(6, 2)	×3	×2	×1			
IX.	probabilità			1458	768	45	2271	33,5	66,5
	risultato	N	(6, 3)	×1	×2	×3			
X.	probabilità			1458	1536	135	3129	37,4	62,6
	risultato	R	(7, 3)	×3	×2	×1			
Probabilità di una estrazione ulteriore				4374	3072	135	7581	35,0	65,0

Va notato, anzi, che neppure le probabilità 0 ed 1 significano impossibilità o certezza: la probabilità di colpire esattamente un qualunque dato punto del bersaglio è nulla, ma se ciò si traducesse nel dire che significa «impossibile» ne conseguirebbe che è impossibile colpire il bersaglio (in uno qualunque dei suoi punti, non importa quale).

Un tipo di discussioni che si ripete con scarse variazioni da secoli fa a tutt'oggi è quello di cui riportiamo un esempio. Estraendo dei segni alfabetici «a caso» si è ottenuta la successione C-O-N-S-T-A-N-T-I-N-O-P-L-E (cosí in un testo francese dell'epoca in cui questo era il nome, in francese, di Istanbul). «Doveva esserci un trucco», si disse, perché la probabilità di ottenere proprio questa successione era piccolissima: $(24)^{-14}$, 1/24 per ogni lettera da scegliersi tra 24, e le lettere sono 14. Ma l'osservare che la probabilità era piccola non significa niente: qualunque altra successione di 14 lettere è altrettanto improbabile, e quindi suscettibile, se uscisse, di uguale sospetto di trucco. Il sospetto può sorgere solo se c'è, e preesisteva, un motivo di pensare che qualcuno avesse interesse, per scherzo o per altro, a far apparire quel certo nome (o qualcosa di curioso, ad esempio 14 lettere uguali, o tutte vocali, o in ordine alfabetico, e via dicendo).

Purtroppo, tra le «storture oggettivistiche», figura anche questa: come risultato di un esperimento, anziché aggiornare la valutazione di probabilità della circostanza che interessa in conformità ad esso, la si «accetta come vera» oppure la si «respinge» come «impossibile». In tal modo l'informazione, spesso assai ricca, fornita dall'esperimento, anziché venire sfruttata razionalmente (come nella statistica bayesiana), viene sperperata, e la decisione perde ogni elasticità.

A parte il campo delle decisioni statistiche o di natura economica, ecc., sarebbe opportuno ispirare riluttanza nel dire che qualcosa è certo o impossibile (sia pur nel senso sottinteso di «quasi-»); l'esperienza mostra (ad esempio nei pronostici calcistici) che c'è una tendenza diffusa all'esagerazione, nel senso di dare valutazioni troppo piccole alle probabilità piccole e troppo grandi alle probabilità grandi (cioè come subendo un'attrazione dei poli estremi, del «certo» e dell'«impossibile»).

2.10. La versione soggettivista.

Parlare della versione soggettiva è piú facile, sia perché (se è lecito un bisticcio) per un soggettivista essa è oggettivamente piú semplice e naturale (oltre che oggettivamente esatta), e sia perché è assai difficile parlare di una concezione opposta alla propria esprimendo debitamente il proprio dissenso e mostrandone i «perché», e tuttavia evitare di far apparire inutile o ingiustificabile l'apporto e l'atteggiamento di portatori di idee opposte, in un campo che da secoli – praticamente dai suoi inizi – è tormentato dallo scontro di concezioni metafisiche, di equivoci logico-terminologici, di ambiguità di scelte di fronte ai molteplici campi di applicazione e le contrastanti mentalità ed esigenze loro proprie.

Si può solo aggiungere un'osservazione banale, ma forse potenzialmente risolutiva (benché sia utopistico pensare a un tale miracolo). La teoria soggettiva permette a chiunque di valutare le probabilità come crede, purché siano rispettate le regole che tutti accettano. Perciò tutte le impostazioni oggettiviste, se accettano tali regole sia nel valutare le probabilità che nel vagliare le decisioni correttamente in base ad esse (e non secondo *adhockeries*), conducono chiunque le segue correttamente ad una valutazione accettabile come sua propria valutazione soggettiva. Unica differenza è che ciascuno (persona o teoria) deve rinunziare a sostenere che le sue valutazioni sono le sole esatte, e gli altri sbagliano. Tutte le opinioni coerenti sono logicamente ammissibili senza distinzione; ciascuno potrà portare argomenti per convincere altri a modificare le opinioni che gli sembrano stravaganti, ma non avrà diritto di dire che sono *sbagliate* perché ciò non ha senso (salvo il caso d'incompatibilità, come dare due probabilità di somma diversa da 1 al fatto che un certo evento sia vero o falso). *Ciascuno a suo modo* (purché sia coerente).

Naturalmente, in tema di decisioni, sono soggettive anche le preferenze (salvo i casi ovvi, come fra guadagni in moneta, o in altro bene omogeneo, ove si presume tutti preferiscano il piú, o quando esiste facoltà di scelta, ove il poter scegliere tra A o B è preferibile sia ad A che a B senza possibilità di scelta). Ma anche qui le conclusioni astratte che in astratto appaiono logicamente ovvie possono dar luogo a paradossi: la facoltà di scegliere tra A e B, se tanto l'uno che l'altro dei due oggetti o premi è assai desiderabile per l'individuo in questione, può trasformare la maggior fortuna in sfortuna obbligando a una scelta penosa, per cui il dispiacere di dover rinunziare di sua volontà ad uno dei due premi offusca il piacere che gli avrebbe dato quel premio che sceglie se non ci fosse stato il travaglio della decisione.

Naturalmente, è giusto che in una teoria si ammetta come norma (anzi come assioma, o postulato) ciò che corrisponde a criteri di logica astratta ineccepibili, trascurando eccezioni in certo senso «patologiche», spiegabili tuttavia in termini psicologici soggettivi. Ma penso sia sempre opportuno rendersi conto di ciò che viene in tal modo soffocato, e rammentarsene di quando in quando.

Neppure nei piú aridi deserti della logica formale è scusabile il rimanervisi immersi, totalmente sordi a quanto di contraddittorio, di irrazionale, di pirandelliano, è inevitabilmente presente nella fantasia e nella psicologia umane.

2.11. Giochi: cenni preliminari.

Ci rimane da considerare il caso dell'incertezza competitiva, menzionato come ultimo nei cenni informativi del § 1.1. L'incertezza «competitiva» è quella concernente guadagni e perdite dipendenti dal risultato di giochi in cui i competitori, o giocatori, scelgono fra le «mosse» ammissibili.

Il caso piú fondamentale e semplice – quello piú abitualmente studiato e cui saranno limitati i presenti cenni – considera i giochi tra due persone a somma nulla, nei quali cioè il risultato economico è «a somma nulla» nel senso che la vincita dell'uno è la perdita dell'altro (senza alcun apporto o prelievo da parte di terzi).

Un'ulteriore radicale semplificazione è quella di limitarsi al caso in cui tutto si riduca a un'unica mossa, o decisione, da prendersi simultaneamente (e senza conoscere quella dell'altro) da ciascuno dei due competitori. Ma non si tratta di un caso troppo banale per riuscire utile: risulta invece che proprio ad esso ci si può concettualmente ricondurre (nonostante ovvie complicazioni) per trovare il bandolo della intricata matassa.

Introduciamo, per cominciare, la tabella dei pagamenti (*pay-off table*), nel piú semplice caso in cui ciascuno dei due giocatori ha la scelta fra due sole alternative: SÍ o NO, oppure 1 o 2, oppure Pari o Dispari (da fare senza conoscere quella dell'altro: per esempio scoprendo simultaneamente la risposta, o mostrando simultaneamente la mano con aperto un certo numero di dita (da 0 a 5) pari o dispari). Si hanno due risultati possibili: risposte uguali o diverse, ovvero, equivalentemente, somma pari o dispari; e sia stabilito, per fissare le idee, che il primo vince (+1) nel caso di somma pari (ovvero di concordanza) e perde (−1; vince +1 l'avversario) nel caso di somma dispari (cioè di discordanza) (cfr. tab. 2).

Per ulteriore chiarimento, e per illustrare fatti che non possono aver luogo nel precedente caso troppo semplice, riporteremo tosto (§ 2.12) un'analoga tabella 6 × 6 (in cui, cioè, ciascuno dei competitori ha sei scelte: per analogia, ad esempio, aprire 1, 2, ..., 5 dita oppure – sesto caso – nessuna, col. 6).

Ma conviene prima completare alquanto lo sguardo preliminare, un po' panoramico, per dare una certa visuale anche a chi non avesse interesse ad addentrarsi in aspetti piú tecnici, e in qualche caso invece – forse – per suscitare tale interesse.

Si tratta, principalmente, di riprendere quanto già accennato sopra: che, cioè, il caso ben piú complesso di giochi che comportano piú mosse successive (per solito: alternatamente spettanti al primo e al secondo giocatore) si può ricondurre al caso piú semplice in cui tutto si condensa in una «tabella (o matrice) dei pagamenti» (come quella 2 × 2 già vista o quella 6 × 6 annunziata; in genere, però, di dimensione assai grande). L'aumento della dimensione rende poco pratica o addirittura proibitiva un'effettiva utilizzazione della tabella, ma, indipendentemente da ciò, la visione concettuale e molte conclusioni qualitative riguardanti il caso semplice di un'unica mossa si estendono automaticamente al caso di «strategie» consistenti in predisposizioni di «mosse» da attuare in cia-

Tabella 2.

Vince (+1) il I giocatore se entrambi scelgono pari o entrambi dispari; vince il II (il I perde −1) se le due scelte sono discordi (un pari e un dispari).

		II: P	II: D
I	P	+1	−1
I	D	−1	+1

scuna delle possibili «situazioni» susseguenti a non importa quante precedenti decisioni di entrambi i competitori.

Per avere un'immagine concreta, pensiamo a un gioco consistente (come quelli della dama o degli scacchi) nel muovere dei pezzi sulla solita scacchiera (8×8 caselle), ma piú semplice. Non è il caso (né sarei in grado) di inventare un esempio di gioco né troppo banale né troppo complicato; tanto per fissare le idee supponiamo che ciascuno abbia 8 pedine (rispettivamente bianche e nere) disposte inizialmente su lati opposti (e non mangiabili né trasformabili in dame od altro), e che ad ogni mossa (alternativamente) uno debba spostare una delle proprie pedine in direzione diagonale a sua scelta: avanti-destra, avanti-sinistra, indietro-sinistra, indietro-destra, fino alla prima casella libera (saltando le pedine intermedie; se esse vanno fino all'estremo della scacchiera, quella scelta è non ammessa).

Il gioco prosegue finché uno riesce a raggiungere posizioni (o ad ottenere configurazioni) indicate come scopo, e vince. Non so (né, ripeto, m'interessa) se tale schema possa dar luogo a un gioco sensato e interessante; conta solo, qui, il fatto che si presta ad esemplificazioni schematizzate in modo piú comodo che se si dovesse distinguere pedine e dame, mangiare pezzi e quindi poterne avere piú o meno, e via dicendo.

Qui le situazioni possibili sono tutte e sole le disposizioni di 8 pedine bianche ed 8 nere sulle 64 caselle (in numero di $64! / (8!)^2 \cdot 48!$: è un numero di circa 240 cifre!)

Ogni mossa trasforma la situazione esistente in quella ottenuta eseguendola: seguire una «strategia», per un giocatore, significa aver adottato un atlante contenente le mappe di tutte le praticamente innumerevoli situazioni con indicato su ciascuna il pezzo da muovere e la direzione (di quante caselle debba essere lo spostamento è già fissato dalla regola di fermarsi alla prima casella libera, e, se non ce ne fosse, tale mossa non potrebbe figurare in nessuna strategia).

Questo esempio, coi suoi vistosi difetti ma proprio grazie ad essi, chiarisce quanto preannunziato: che, cioè, concettualmente, tutto rimane altrettanto semplice che negli esempi banali, ma, dato il vertiginoso aumento delle combinazioni, nessun riflesso praticamente utile ne deriva per effettive applicazioni o calcoli.

A prescindere dalla numerosità, la conclusione è però sempre la medesima che vale nell'esempio della nostra tabella 6×6 indicando solo i tre risultati *guadagno, pareggio* o *perdita* (coi segni $+$, 0, $-$: come nella tabella 3) anziché coi valori numerici (e analogamente potremmo in entrambi i casi riferirci ai valori numerici): la conclusione, in termini pratici, è che, se due giocatori si accingono a iniziare un gioco applicando rispettivamente due prestabilite strategie, A e B, potrebbero risparmiarsi la fatica (ma rinunziando al divertimento) consultando, se esistesse, l'immensa tabella in cui troverebbero immediatamente, guardando all'incrocio della riga e colonna corrispondenti ad A e a B, l'esito della partita che avrebbero giocato.

2.12. Sulla decisione nel caso dei giochi.

Nel precedente § 2.11 ci siamo limitati alla descrizione dei giochi tra due persone a somma nulla, nel caso piú semplice di giochi decidibili con un'unica decisione da parte di ciascuno dei due competitori, e all'indicazione del modo in cui casi piú complessi possono ricondursi al medesimo schema (sia pure, naturalmente, accrescendo in misura enorme, molto spesso spropositata, le dimensioni).

Sappiamo cosí che si può sempre ragionare sul caso di un'unica mossa, ottenendo conclusioni valide in generale; e per enunciarle e discutere sarà sufficiente riferirsi a un caso di dimensioni piccole (ma non troppo): abbiamo scelto il caso di una tabella 6×6 (e precisamente ci riferiremo a quella presentata, con note esplicative, come tabella 3).

Cosa ci suggerirebbero dunque le considerazioni di teoria delle decisioni già svolte?

In certo senso la risposta è pronta, e non ci sarebbe nulla da aggiungere: identificandoci col I individuo, abbiamo da scegliere una tra le sei colonne (C_1, ..., C_6) ciascuna delle quali indica sei valori (di guadagni o perdite a seconda del segno $+$ o $-$); quale sarà l'effettivo guadagno o perdita dipende dalla scelta della riga, che spetta all'altro (e che avrà fatta, o dovrà fare, senza sapere quale sia la colonna prescelta). E quali probabilità attribuire alla scelta delle varie righe da parte dell'altro? (il quale evidentemente si sforzerà di attribuire analogamente delle probabilità alle mie scelte). Si tratta di un doppio problema doppiamente psicologico, in quanto ciascuno è obbligato a rimuginare il dantesco «Cred' ïo ch'ei credette ch'io credesse» [*Inf.*, XIII, 25] pensando: «Quali valutazioni mi conviene fare pensando a quelle che farà l'altro cercando di immaginare le mie?»

Non mi consta, e non credo, possano esistere regole al riguardo: se esistesse una «regola ottima», e si sapesse che uno la segue, ci sarebbe la massima facilità di batterlo, perché l'elemento decisivo è l'imprevedibilità.

La teoria dei giochi, anziché guardare il problema in questi che sono i suoi veri termini, preferisce anche qui rifugiarsi in *adhockeries*, con qualche maggiore giustificazione, considerato il circolo vizioso in cui di fatto ci si trova intrappolati.

Cioè, mi scuso, non sono *adhockeries*, nel senso di regole grossolane che intendono sostituire quelle esatte nel rispondere a un dato problema; si tratta di abbandonare il problema insolubile (di trovare la risposta esatta od ottima) e ripiegare su questioni diverse, come ad esempio: «In quale modo posso evitare (con certezza) risultati troppo sfavorevoli?»

Rispondere a questo problema è elementare: con riferimento alla tabella 3, se I sceglie la colonna C_3, dove il minimo è -2, è certo di non poter ottenere, qualunque sia la scelta dell'avversario, un risultato peggiore di -2 (il massimo tra i minimi delle colonne); e cosí se II sceglie la riga R_4 (il minimo tra i massimi delle righe) è certo che il risultato per lui non sarà peggiore di -3, cioè $+3$

risultato per I. Se, in particolare, entrambi seguono questa strategia «prudente», il risultato sarà necessariamente compreso tra detti limiti; infatti, il risultato della casella all'incrocio della riga e colonna prescelte non può se non essere intermedio fra il massimo dei minimi e il minimo dei massimi, a meno che non coincida con uno di essi, o con entrambi: caso in cui il criterio indicato suggerisce

I

II	C_1	C_2	C_3	C_4	C_5	C_6
R_1	−7	+8	−1	+1	+4	−6
R_2	+1	−3	−2	+4	−1	+9
R_3	−4	+1	+7	−5	0	+4
R_4	−2	−8	+2	0	+3	−1
R_5	+6	+5	−1	−3	+5	−8
R_6	−5	0	+6	+7	−8	+1

Colonne

Righe	1	2	3	4	5	6
1	−	+	−	+	+	−
2	+	−	−	+	−	+
3	−	+	+	−	0	+
4	−	−	+	0	+	−
5	+	+	−	−	+	−
6	−	0	+	+	−	+

una soluzione univoca: la soluzione *minimax*. (Si usa anche dire minimax e maximin per «minimo dei massimi» e «massimo dei minimi»: «minimum maximorum» e «maximum minimorum»; piú spesso si parla di minimax solo riferendosi al caso di coincidenza).

Nella teoria dei giochi si considerano però, oltre alle strategie consistenti nella scelta di una riga o di una colonna (o strategie pure) direttamente, anche strategie consistenti nella scelta casuale (per sorteggio, o simili) tra due o piú strategie (anche tutte). Tali strategie si dicono strategie miste (o strategie randomizzate); in tal caso anche il risultato rimane aleatorio: se ad esempio I e II, anziché C_3 ed R_4 soltanto, avessero deciso di giocare C_2 o C_3 con probabilità 70 e 30 per cento, rispettivamente R_4 ed R_5 con probabilità 40 e 60 per cento, il risultato sarebbe stato −8 con probabilità 28 per cento (70% × 40%), +2 con probabilità 12 per cento, +5 con probabilità 42 per cento, −1 con probabilità 18 per cento.

Verrà spontanea la domanda: perché un sorteggio? Se esso porta a scegliere «a caso», tramite il sorteggio, fra le decisioni C_2 e C_3, non è meglio scegliere quella ritenuta preferibile, oppure, se non si hanno motivi di preferenza, sce-

Tabella 3.

Gioco tra due giocatori, a somma nulla (cioè: tale che l'uno vince ciò che perde l'altro). Nel caso illustrato, la tabella ha sei righe e sei colonne (in generale il numero potrebbe essere qualunque).
I due giocatori scelgono (all'insaputa l'uno dell'altro) il primo una riga e il secondo una colonna. Nella figura sono scelte la riga R_4 e la colonna C3; il numero indicato nella casella all'intersezione di dette riga e colonna nell'esempio +2 nella casella C3/R4, indica il guadagno del I giocatore (a spese del II). Tale scelta è quella che si ha se entrambi i giocatori seguono la strategia del minimax; seguendo qualsiasi altra strategia si affidano alla sorte, nel senso che possono sia guadagnare di piú oppure perdere in confronto al minimax.
La strategia «minimax» è quella piú «conservativa», di chi cerca di mettersi al sicuro contro il rischio di risultati meno favorevoli pur rinunziando al miraggio di risultati piú vantaggiosi.
I dischetti bianchi e neri indicano il massimo e il minimo rispettivamente per la riga, se in alto a destra, e per la colonna, se in basso a sinistra.
Le colonne – C_1 C_2 C_3 C_4 C_5 C_6 – rappresentano le sei strategie disponibili per il giocatore I, mentre le righe – R_1 R_2 R_3 R_4 R_5 R_6+ quelle del II. In ogni riquadro della tabella è indicato il guadagno (positivo e negativo) del I giocatore (cambiando il segno si ha quello del II).
Il minimo dei massimi è +3 (ottenibile con decisioni C_5R_4); il massimo dei minimi è −2 (ottenibile con decisioni C_3R_2); il valore +2 è quello che si otterrebbe se i due giocatori scegliessero le strategie C_3 ed R_4, atte ad assicurare a ciascuno di non andare al di sotto del meno peggiore dei risultati che può assicurarsi con scelta autonoma (rispettivamente, appunto, minimo dei massimi o massimo dei minimi, che è la stessa cosa, scambiando il segno a seconda del riferimento al I o al II).
Lo schema ridotto (coi segni +, o, − per indicare «vittoria, pareggio, sconfitta») ha il medesimo significato quando si attribuisce importanza prevalente od esclusiva al risultato anziché al punteggio. Ciò vale ad esempio nelle partite di campionato di calcio, dove per la classifica conta solo il risultato, mentre la differenza-reti può avere rilevanza soltanto per dare una graduatoria tra squadre a pari punti.

gliere una «qualunque» delle due (senza preoccuparsi di sorteggi e di relative probabilità)?

Cosí si chiederebbe un profano e avrebbe ragione da vendere. Gli dànno torto solo dei «tabú», che inducono a (s)ragionare così: la decisione mista 70 per cento di C_2+30 per cento di C_3, nel caso di risposta R_1 dà $+8$ con probabilità 70 per cento e -1 con probabilità 30 per cento, il che equivale a $5{,}6-0{,}3=5{,}3$, ossia... 5,3 certi sotto la detta scelta (mentre si tratta o di $+8$ o di -1). Il che, in definitiva, è equivalente, ma resta l'autoinganno di considerare 5,3 come il risultato effettivo della scelta della strategia mista, anziché come valor medio fra i due risultati dei due casi.

Questa stravaganza porta però un grande beneficio teorico: permette di dire che la soluzione minimax esiste sempre purché la si cerchi nel piú ampio ambito delle strategie miste. È un bellissimo teorema di John von Neumann (si potrebbe dirlo: di algebra bilineare), quello che, nell'interpretazione al nostro caso, garantisce che se (invece di limitarsi a un certo numero di strategie pure, come le 6+6 del nostro esempio, $C_1, C_2, ..., C_6$ ed $R_1, R_2, ..., R_6$, con che, salvo casi sporadici di coincidenza, risulta minimax > maximin) consideriamo la totalità delle strategie miste (C misture delle C_i; R misture delle R_j; $i, j=1, 2, ..., 6$) si realizza sempre la coincidenza tra maximin e minimax, ossia esiste la soluzione minimax.

È chiaro che, in queste condizioni, in cui ciascuno è nella facoltà, in base alle regole del gioco, di impedire uno spostamento per lui sfavorevole rispetto al valore minimax, il risultato sarà sempre proprio il minimax, salvo sviste. Però...; nel caso che esso si raggiunga impiegando strategie miste, il risultato va interpretato – beninteso! – nel senso che il risultato coincide sí, *in previsione*, col minimax prima di conoscere la decisione altrui,... ma poi sarà quel che sarà: si scosterà dal minimax aleatoriamente in piú o in meno, equamente, di poco o di molto a seconda della variabilità tra i valori della tabella.

2.13. Osservazioni provvisoriamente conclusive.

Nell'esaminare preliminarmente il «groviglio di problemi» che ci si pone affrontando il problema delle decisioni in condizioni d'incertezza, abbiamo cercato, sí, di aprire la via all'impostazione naturale e corretta (o almeno che confido dovrà apparire tale), ma soprattutto di mettere in guardia contro i possibili malintesi. Ve ne sono di assai diffusi, nel momento attuale di sviluppo di varie concezioni, ed altri che ciascuno potrebbe crearsi anche motu proprio per la facilità di svisare – anche di poco, ma basta per dare magari effetti disastrosi – le cose che sente e apprende da altri (persone, libri, slogan) in cui molte terminologie sono ambigue o possono esserlo per chi non è abbastanza ferrato per interpretarle.

Importante è veder chiara la natura del rapporto fra i dati noti, o fin dall'inizio o aggiuntisi per successiva acquisizione di informazioni, e le previsioni: non un rapporto meccanico, ma dovuto al riflesso sul nostro complessivo stato d'informazione e alla revisione delle previsioni alla luce di essa. L'avvertenza nell'ultimo capoverso del § 2.6, e l'esemplificazione cui si riferisce, erano intese a dare sia pur solo un principio di chiarimento al riguardo. Ma la visione completa apparirà solo quando, poste alcune basi di teoria delle probabilità, tutto assumerà un aspetto piú sistematico e in particolare si vedrà cosa significhi il «valore di un'informazione» agli effetti di un problema di decisione.

Ciò che dovrebbe esser apparso già abbastanza nettamente è il divario di qualità tra metodi «ad hoc» (*adhockeries*) ed il metodo che si potrebbe dire «naturale», *coerente*. E su questo punto va aggiunta, in chiusura del presente § 2, un'ulteriore osservazione cui dà luogo un certo tipo di assimilazione (secondo gli oggettivisti) tra teoria dei giochi e teoria delle decisioni.

Non avremmo motivo di intrattenerci ulteriormente sull'argomento dei giochi (che potremmo considerare accennato al solo titolo di opportuno complemento culturale) se non fosse per accennare a certi tentativi di applicarlo nella teoria delle decisioni considerando, al posto delle «strategie dell'avversario», gli effetti delle varie ipotetiche situazioni che interferiranno con le decisioni che stiamo per prendere. Saranno decisioni altrui (ma, a differenza che nel caso di giochi, senza volere o senza sapere che possano interferire con le nostre), oppure fatti naturali, pensando al quale caso si usa anche dire «giochi contro la Natura» (e passi...) e si giunge anche a parlare di un'ipotetica «Natura malevolente»!

Tale idea – a parte il ridicolo sapore superstizioso – è intrinsecamente assurda per il fatto che la Natura dovrebbe, secondo tale veduta, comportarsi in modo antitetico agli interessi di ciascuno di noi, suoi competitori. Sarebbe come dire che essa sospinge il corso degli eventi in direzione diametralmente opposta a tutte quelle in cui si sforza di tirarlo, nel proprio interesse, ciascuno di noi che le stiamo tutt'attorno. Ciò che sarebbe manifestamente impossibile anche a Giano bifronte, «per la contradizion che nol consente» [*Inf.*, XXVII, 120]: occorrerebbe una Natura millantibifronte!

Ed era proprio su tale idea stravagante che mi sembrava valesse la pena di soffermarsi, perché costituisce probabilmente l'esempio piú raccapricciante delle storture che derivano dal concepire le metodologie matematiche e scientifiche come strumenti passibili di venire sperimentati e strapazzati in tutte le circostanze e in tutti i modi e in tutte le salse, anziché assimilarne intimamente i lineamenti concettuali ed intuirne le possibilità di effettiva rispondenza laddove tali lineamenti si addicono. È in tale maniera che nascono le peggiori adhoccaggini e tante altre cose ancor peggiori che non meritano neppure questo nome.

Occorre aggiungere però che di detto procedimento è stata data una giustificazione piú ragionevole, che respinge l'ammissione circa una «Natura malevolente», ma trova comprensibile che, essendo «sconosciuta» la «strategia della Natura», ci si preoccupi di vedere come le conclusioni variano a seconda delle valutazioni riguardo ad essa, e, come criterio cautelare, basarsi sulla piú sfavorevole. Questo è, in sintesi, il punto di vista di McKinsey [1952, pp. 277-90]: può interessarci di conoscere cos'è il peggio che la natura può farci, onde assicurarci almeno quel minimo che possiamo garantirci nell'ipotesi piú sfavorevole.

Tale ragionamento appare però solo in parte accettabile. Seguendolo sistematicamente alla lettera, dovremmo evitare ogni minimo rischio considerandoli tutti mortali: mai attraversare una strada per il pericolo di investimenti, mai dormire in una casa per timore di incendi e terremoti, mai mangiare un boccone che potrebbe essere avvelenato. Tutta la teoria delle decisioni ha senso soltanto se ciascuno valuta le probabilità secondo il proprio giudizio. Tuttavia, poiché tale valutazione non può essere che vaga, il suggerimento di McKinsey è opportuno se interpretato limitativamente: *conviene esplorare la situazione in relazione non ad un'unica valutazione delle probabilità che attribuiamo alle diverse ipotesi possibili considerate, ma osservando se e come e di quanto le prospettive si modificano ritoccando le valutazioni in varie direzioni*. Non però per adottare, «come misura di prudenza», la valutazione che risulta piú sfavorevole (sia pure nell'ambito dei «ritocchi»: peggio che mai adottando l'ipotesi piú sfavorevole di tutte), ma per riflettere ulteriormente se e in che senso ritoccare le valutazioni qualora un ritocco desse conclusioni sensibilmente diverse. Riflettere non significa però spostarsi nel senso di maggiore pessimismo (come «cautela») e meno che mai verso il massimo pessimismo, ma in un senso o nell'altro, visto che l'accuratezza non era adeguata all'importanza delle conseguenze.

Né la prudenza né la temerità, né l'avversione al rischio né il gusto del rischio rendono ragionevole in alcun caso basarsi su valutazioni di probabilità alterate rispetto alle proprie opinioni. Questa è un'idea totalmente priva di senso.

La prudenza, l'avversione al rischio, trovano in tutt'altra direzione la naturale via per esprimersi e per guidare nel comportamento: nella convessità della curva dell'utilità. Le probabilità non si toccano: il giudizio è soggettivo ma non modificabile a seconda delle operazioni scommesse decisioni che s'intende fare. Sono gli importi monetari invece, che, pur essendo in senso ovvio oggettivi, cambiano di valore, in termini di utilità, poiché successivi incrementi uguali avranno un'utilità sempre minore man mano che essi fanno crescere il grado di ricchezza del destinatario.

3. *Decisioni in condizioni di certezza.*

3.1. Qualche premessa.

Come risultava già chiaro dalle prime indicazioni riassuntive, gli aspetti piú rilevanti e complessi riguardanti le decisioni sono quelli che derivano dalla presenza di incertezza. Ciò non significa tuttavia che i problemi di decisione in caso di certezza non possano risultare interessanti e istruttivi, sia di per sé, sia come avvio alla trattazione dei casi piú completi ove interviene anche l'incertezza. Il fatto che manchi questo aspetto è anzi vantaggioso, in quanto permette di far risaltare altre circostanze che hanno la loro importanza, e su cui è bene meditare, senza che la presenza di un fattore predominante e piú ricco di problematicità le faccia relegare nella penombra.

In generale, i problemi di decisione sono, matematicamente, problemi di massimo (o minimo; è la stessa cosa salvo cambiare il segno): in versione economica, di massimo guadagno, di minimo costo, di minimo tempo, ecc. Vedremo su parecchi esempi come spesso ragionamenti semplici e piú o meno ingegnosi rendano chiara la soluzione e il suo perché, affinando l'intuizione.

Nel caso di certezza mancano quasi tutte le circostanze che nel caso generale complicano piú o meno le cose; in particolare, ad esempio, non serve distinguere valore monetario e utilità, perché l'uno è funzione crescente dell'altro cosicché non cambia nulla riferendosi all'uno o all'altro. (Cosí come è la stessa cosa, parlando di quadrati, dire «di massima area» o «di massimo perimetro», mentre ciò non vale se si parla di rettangoli). Mancano quasi tutte: rimane solo la necessità di scontare o capitalizzare gli importi, se dovuti in istanti diversi, per riferirli tutti a un medesimo istante (preso come «origine»; è indifferente la scelta, conviene naturalmente riferirsi all'istante in cui il conto va effettivamente regolato).

Importante è solo notare che, per tal motivo, le decisioni in condizioni di certezza (come del resto le altre) vanno distinte a seconda che siano immediate (in un solo istante) o con differimenti. (Va da sé che differimenti irrilevanti vanno in genere trascurati).

Presentiamo alcuni esempi, qualcuno sviluppandolo un po' piú ampiamente per rendersi conto dei ragionamenti, altri elencandoli e brevemente commentandoli, col solo intendimento di dare un'idea della varietà di questioni interessanti che si possono incontrare.

3.2. Un problema di scorte.

L'esempio piú semplice e pur interessante (ed anche ormai abusato, ma non c'è miglior scelta) è il problema delle scorte nel caso di una merce che viene consumata (venduta, oppure usata come fattore di produzione) ad un ritmo costante: quantità q per unità di tempo.

Occorrerà rifornire il magazzino di un quantitativo $Q=qT$ di detta merce a intervalli di tempo T. E il problema di decisione (ossia di ottimizzazione, di minimizzazione dei costi) è il seguente: converrà fare acquisti frequenti di partite piccole, o rari di partite grandi? (Cfr. fig. 1).

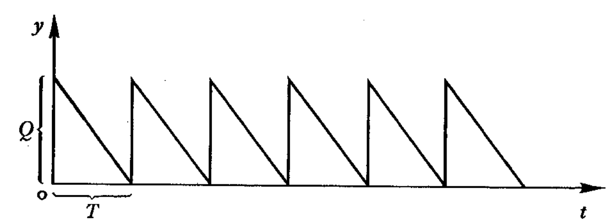


Figura 1.

Andamento delle scorte nell'ipotesi di consumo costante e di acquisto di una quantità fissa al momento in cui la scorta si esaurisce.

Naturalmente, dovremo fissare dei dati. Avremo evidentemente dei costi, e precisamente, supponiamo: un costo fisso K per ogni ordinazione (costo che si suppone indipendente dalla quantità); e un costo (di magazzinaggio, incluse spese di assicurazione, ecc.) proporzionale, per unità di tempo, al volume medio della scorta (che è $Q/2$), e sarà quindi $c(Q/2)$.

Il costo fisso, riferito a unità di tempo, risulta $K/T = K(q/Q)$ (poiché: $Q = qT$).

Complessivamente, il costo per unità di tempo sarà quindi $c(Q/2)+K(q/Q)$; al crescere di Q (ossia al diminuire di T) il primo addendo cresce e il secondo decresce, e la somma, cioè il costo complessivo, dapprima decresce fino a raggiungere il costo minimo $\sqrt{2K/qc}$ quando $Q=\sqrt{2Kq/c}$ (e $T=\sqrt{2K/qc}$) e di lí in poi cresce indefinitamente (cfr. figg. 2-3).

Si possono, naturalmente, considerare delle varianti, anche realistiche. Può darsi ad esempio che il costo c subisca delle variazioni al variare della quantità Q (ad esempio quando si superi la capacità normale del magazzino Q^*, e si deva ricorrere a ripieghi), oppure che sul prezzo d'acquisto vengano concessi sconti per ordinazioni superiori a certe quantità Q_1, Q_2, Q_3, ecc. L'andamento del costo totale, cioè l'iperbole della figura 3, viene modificato come indicato rispettivamente nelle figure 4 e 5: nel primo caso, superando Q^*, il ramo d'iperbole

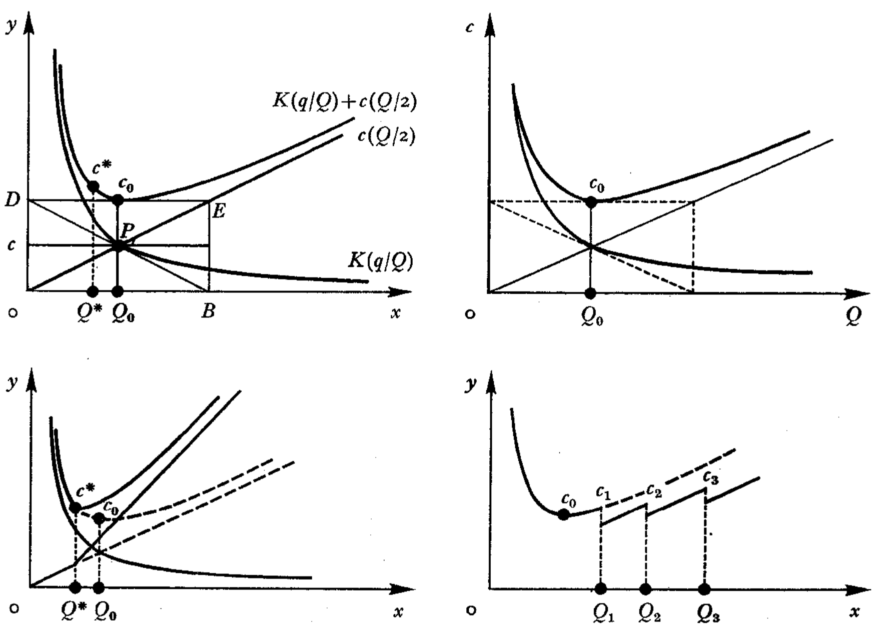


Figure 2-5.

Andamento del costo a seconda che i rifornimenti si effettuino a intervalli piú o meno lunghi e in presenza delle circostanze specificate nell'ultimo capoverso del § 3.2.

fa un angolo e cresce piú rapidamente; nel secondo caso, esso si abbassa di uno scalino ogni qual volta la quantità entra in un altro intervallo di prezzo.

3.3. Un esempio di «programmazione lineare».

Altro esempio è quello della programmazione lineare. Volendo limitarci a un cenno intuitivo, dovremo scegliere un esempio ridottissimo, con due soli prodotti. Consideriamo, precisamente, una fabbrica atta a fabbricare due prodotti diversi, A e B, che impegnano in misura diversa i reparti attraverso cui devono passare. Precisamente, il primo reparto potrebbe produrre (per unità di tempo: per esempio anno, o mese) o soltanto a_1 unità di A, o soltanto b_1 unità di B, oppure una mistura in proporzione qualunque: ad esempio, $1/2$ di a_1 e $1/2$ di b_1, oppure $1/3$ di a_1 e $2/3$ di b_1, ecc. In generale, la restrizione per le quantità prodotte a e b è data da $a/a_1+b/b_1 \leq 1$; geometricamente, sul piano x, y ove il pun-

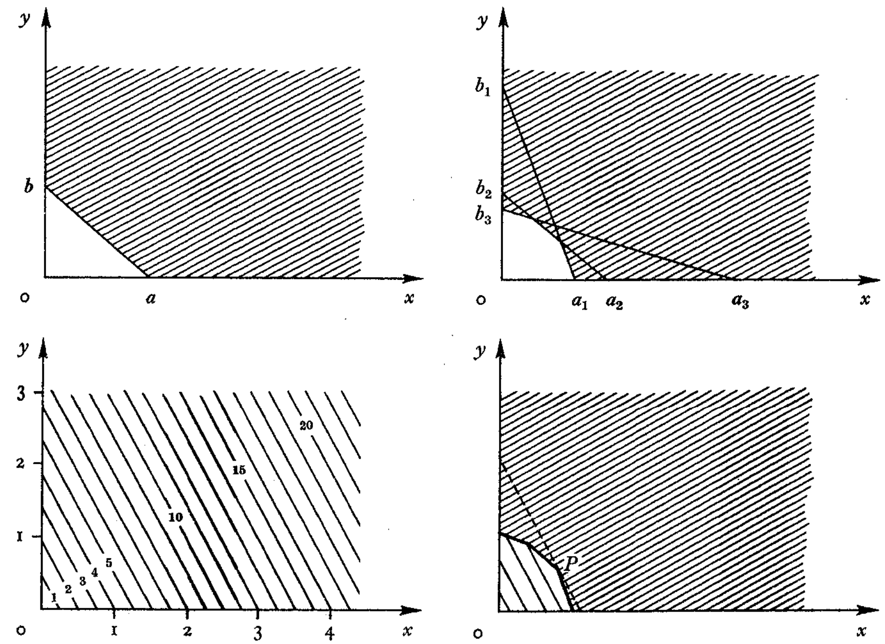


Figure 6-9.

In tutte le figure, un punto indica la quantità di produzione di due beni X e Y (risp. ascissa e ordinata). Nella figura 6 è indicato il triangolo delle produzioni compatibili con la potenzialità di un dato reparto; nella figura 7 sono indicate le limitazioni derivanti da tutti i reparti (nell'esempio, tre); la zona possibile è quella non tratteggiata. La figura 8 indica l'utile derivante da ogni decisione, e la figura 9 mostra come la massima convenienza si ha scegliendo il punto P ove la tangente al contorno è una di tali rette.

to $x=a$, $y=b$ rappresenta la produzione di rispettivamente a e b unità dei due prodotti, la zona delle produzioni possibili è data dal triangolo tra i semiassi positivi e la retta $b_1x+a_1y=a_1b_1$. Ciò per quanto dipende dalle possibilità del primo reparto; per gli altri avremo limitazioni del medesimo tipo, $b_hx+a_hy=a_hb_h$ ($h=1, 2, ..., n$); ciascuna delimiterà un triangolo come il precedente, e la parte comune a tutti sarà un poligono (di al piú $n+2$ lati: due sugli assi, gli altri sulle rette, ma non è detto che tutte contribuiscano a limitare l'area utile). (Cfr. figg. 6-7).

Il problema di programmazione lineare, in questo caso, consiste nel decidere il livello di produzione piú conveniente sapendo che l'utile derivante dalla vendita di ogni unità A o B è rispettivamente α e β (e quindi, per quantità x e y, è $\alpha x+\beta y$). Evidentemente, il massimo si ha nel vertice che si trova sulla piú lontana (dall'origine) fra le rette (parallele) $\alpha x+\beta y=$costante (retta che tocca il poligono solo in quel vertice, salvo il caso speciale che coincida con una delle rette che lo delimitano; in questo caso la scelta di uno qualunque dei punti di quel lato realizza il massimo, ed è quindi indifferente). (Cfr. figg. 8-9).

La ricerca del massimo, facile in questo caso, diventa evidentemente assai complessa aumentando il numero dei prodotti e il numero dei reparti. Avendo tre prodotti anziché due si passa dal caso del piano (due dimensioni) ora visto all'analogo nello spazio ordinario, e quindi, per piú prodotti, a spazi a 4, 5, ..., n dimensioni. Aumentando il numero dei reparti aumenta il numero delle corrispondenti rette nel caso visto, che diventano piani nello spazio, in generale iperpiani ($n-1$ dimensioni) nello spazio a n dimensioni; ciascuno dei quali può dar luogo a una faccia del poliedro.

3.4. Cenni su svariati esempi.

Ai due esempi precedenti, sviluppati in modo completo seppure succinto per dare un'idea anche dell'effettivo svolgimento dei calcoli, converrà aggiungerne parecchi altri limitandoci a brevi indicazioni essenziali, intese solo (ma è cosa importante) a dare un'idea della varietà di problemi e della possibilità di trovare spesso un'idea o una formulazione che rendano agevole pervenire alla soluzione, faticosa da raggiungere senza tale preliminare orientamento in una congerie di dati. (Questi esempi, come i due precedenti, si trovano sviluppati piú ampiamente, ma sempre in forma piana e con cura di mettere in luce il significato concreto e intuitivo, in De Finetti e Minisola [1961]).

1) Una ditta che effettua consegne a domicilio sa in quali periodi o giorni ha bisogno di impiegare un dato numero di furgoni; ogni furgone acquistato comporta un costo annuo dato; l'alternativa è noleggiare un furgone a un dato costo per giorno (sempre costo complessivo, incluso autista, benzina, ecc.). Quanti furgoni è conveniente possedere? – Si può fare il conteggio completo per ogni ipotesi sul numero di furgoni, ma la risposta si ottiene subito osservando che il costo annuo del furgone in proprio equivale, per esempio, al costo di quello noleggiato per cento giorni. Perché convenga di tenere (per esempio) quattro fur-

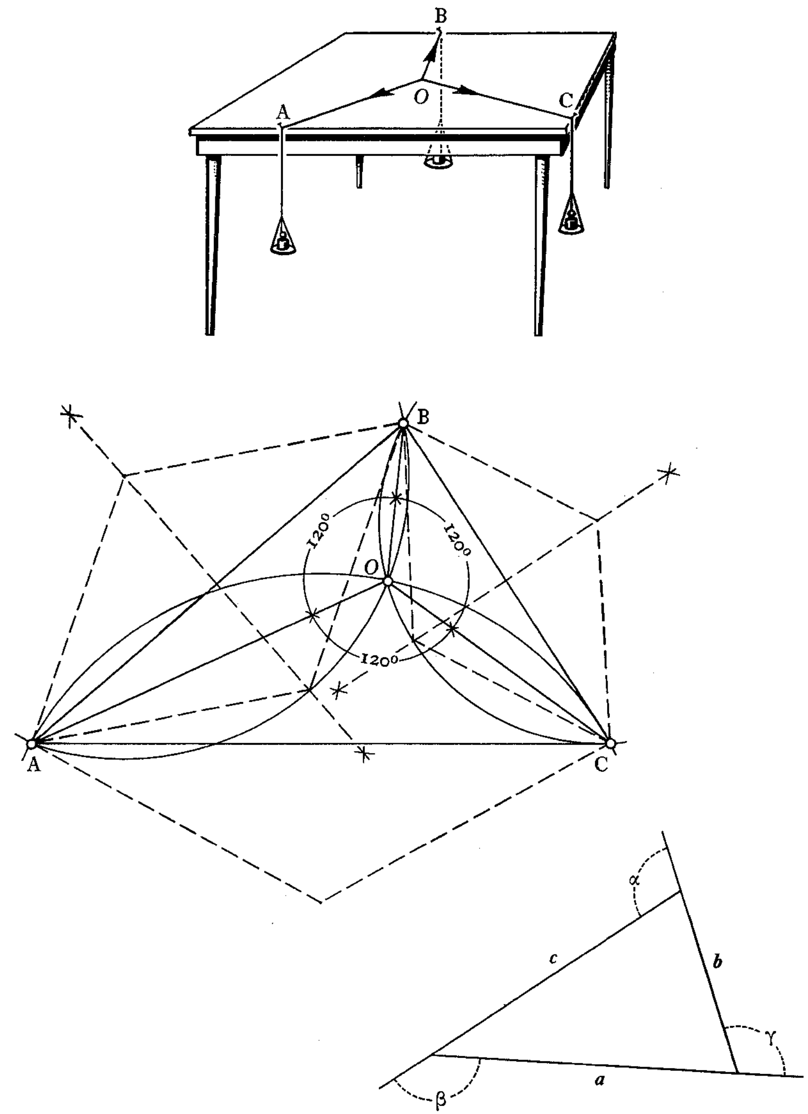


Figure 10-12.

Un punto (il nodo dei tre fili tesi verso punti determinati: nell'esempio, le pulegge all'orlo del tavolo) si sposta fino alla posizione in cui i tre angoli diventano uguali (120°); solo cosí, infatti, si ha la simmetria evidentemente necessaria per l'equilibrio (fig. 10).

Costruzione geometrica del punto di cui alla figura 10, nell'ipotesi di tre pesi uguali (fig. 11) e di diversità fra i tre pesi (fig. 12).

goni, occorre che in almeno cento giorni ne occorrano piú di tre, e che in meno di cento giorni ne occorrano piú di quattro.

II) Si vuol fare un collegamento (strada, conduttura, o altro) fra tre località, in modo che la lunghezza complessiva sia minima. – Soluzione: tre strade che si congiungono nel punto ove s'incontrano sotto angoli di 120° (se ciò è impossibile conviene un collegamento diretto lungo i due lati minori del triangolo). La soluzione è intuitiva in termini meccanici (e presenta il vantaggio che si presta a generalizzazione, se le tratte hanno «peso» diverso, ad esempio per diversa intensità del traffico). Le figure 10-12 spiegano la conclusione e mostrano la costruzione che dà il punto d'incontro dei tre tronchi.

III) Conviene un accordo che preveda, a scadenze diverse, pagamenti in un senso o nell'altro (in denaro, o in prestazioni, ecc. stimate equivalenti)? – Basta scontare tutti i valori al tasso che uno giudica rispondente alla propria propensione (al risparmio o meno), e vedere se il risultato è positivo o negativo. (Il risultato – quanto al segno – non cambia variando comunque la data di riferimento; non servirebbe quindi farne menzione).

IV) Problema del «commesso viaggiatore»: visitare – partendo e tornando alla sua località di residenza, O – un certo numero di altre località A, B, C, ...; combinando l'itinerario in modo che sia il piú breve possibile. La soluzione non si può in generale trovare che per tentativi, dopo esclusi molti percorsi visibilmente non convenienti (come andare a zigzag tra due lontani raggruppamenti di località vicine tra loro).

V) Analogo (ma esiste un'ingegnosa regola per risolverlo) il problema delle «destinazioni» (*assignement*). Si abbiano *n* individui, e per semplicità pensiamo a quattro, A, B, C, D, da attribuire ad altrettanti posti, 1, 2, 3, 4 (mansioni, sedi, macchine, o altro). Supponiamo di misurare il previsto rendimento (oppure gradimento, o altro) di ogni individuo se destinato a ciascun posto, e di cercare di effettuare l'assegnazione nel modo migliore. Naturalmente, l'ideale sarebbe mettere «la persona piú adatta al posto piú adatto», ma sarebbe un caso straordinario se ciò risultasse possibile. In generale ci saranno piú individui che risultano adatti piú ad un medesimo posto che ad altri, e piú posti che appaiono piú adatti ad un medesimo individuo.

Quel che si può fare è cercare di rendere massima la somma di punteggi, avendo dato dei punteggi al previsto rendimento di ciascun individuo se assegnato a ciascuno dei posti. Risulta una tabella (4 righe: una per ogni posto; 4 colonne: una per ogni individuo; all'incrocio il punteggio).

Il metodo per trovare la soluzione consiste nell'eseguire alterazioni nella tabella, per righe o per colonne, fino a ottenere degli zeri in quattro caselle (una per riga e una per colonna): sono quelle per cui la somma dei punteggi risulta massima, e pertanto definiscono, in tal senso, la migliore (o, comunque, la meno peggiore) delle soluzioni.

4. *Il caso d'incertezza, e la probabilità.*

4.1. Probabilità: di che cosa? e in che senso?

Fin dall'inizio abbiamo dovuto parlare di incertezza, dei problemi cui dà luogo, e, pur limitandoci a considerazioni preliminari, abbiamo dovuto usare il termine 'probabilità'.

Ma che significato ha questa parola?

È noto che sul modo di concepire e definire la probabilità si scontrano punti di vista diversissimi, dando luogo a una varietà pressoché illimitata di interpretazioni e sfumature di interpretazioni. Di tali questioni occorre discutere qui brevemente per evitare riflessi e incertezze nelle applicazioni attinenti al nostro argomento, cioè al problema delle decisioni in condizioni di incertezza.

Agli effetti del problema delle decisioni sono irrilevanti tutte le discussioni sull'esistenza o meno di probabilità in un senso o nell'altro «oggettive», «note» od «incognite», relative a singoli eventi o categorie di eventi, e via dicendo.

Ai nostri effetti, cioè nella teoria delle decisioni, il significato è tutto e solo nella seguente affermazione:

Dire che la probabilità di un evento E, secondo un dato individuo, vale p (ad esempio: $p=0,35$), significa dire che egli giudicherebbe equo scambiare il diritto a ricevere – se E risultasse vero – un importo generico S (positivo o negativo; non troppo grande) in cambio dell'importo pS (certo).

(Nell'esempio, con $p=0,35$, scambiare, in un senso o nell'altro, 350 lire certe con 1000 lire pagabili soltanto se E risulta vero. La limitazione «non troppo grande» allude alla divergenza tra valutazioni in termini monetari e in termini di utilità (§ 2.3), per cui sarebbe poco plausibile ripetere il precedente esempio elevando gli importi a 350 000 lire e 1 milione).

Vale la pena di accennare subito a una formulazione piú diretta per la richiesta, a un dato individuo, di esprimere le sue valutazioni di probabilità. Anziché parlare di scommesse pro o contro l'avverarsi di un evento, si può chiedere direttamente di indicare la probabilità che attribuisce a un certo evento indicando una regola di penalizzazione appropriata («proper scoring rule»), cioè tale che *nel suo stesso interesse, data la sua opinione, gli conviene esprimerla con sincerità perché solo in tal modo egli rende minima la sua «penalizzazione sperata».*

Riferiamoci a un esempio particolarmente espressivo, dato che riguarda un campo in cui sono state fatte molte sperimentazioni: il campo delle previsioni di risultati di calcio. Non nel senso del Totocalcio (di indovinare 1 o x o 2), ma nel senso di attribuire, per ogni partita, ai tre risultati, le probabilità corrispondenti alla propria opinione. Come base statistica, si può dire che (nei normali campionati) si ha circa un 50 per cento di vittorie in casa, un 30 per cento di pareggi, un 20 per cento di vittorie esterne; ma, naturalmente, pur tenendo presente tale dato «medio», è ovvio che le valutazioni se ne scosteranno secondo il rapporto di forza delle squadre e altri fattori, sia oggettivi e sia soggettivi, quali apprezzati da ciascuno.

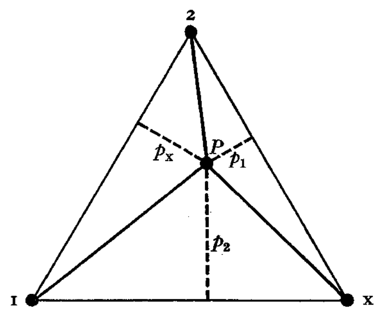


Figura 13.
Triangolo equilatero: per ogni punto interno P la somma delle tre distanze dai lati è costante (e precisamente è uguale all'altezza). Perciò, nel caso di tre eventi incompatibili ed esaustivi (come i risultati 1, x, 2 in una partita di calcio), ogni opinione sull'esito di una data partita (probabilità p_1, p_x, p_2) è rappresentabile come un punto del triangolo.

La regola di penalizzazione si chiarisce meglio sulla rappresentazione geometrica su di un triangolo equilatero (cfr. fig. 13) i cui vertici corrispondono ai tre risultati possibili: 1, x, 2. Di tali risultati dobbiamo indicare le probabilità secondo il proprio giudizio: p_1, p_x, p_2; a risultato acquisito ognuno avrà una penalizzazione data dal quadrato della distanza tra il punto P della sua previsione e il punto («1» o «x» o «2») del risultato. Sommando, per ogni partecipante al concorso di tali pronostici probabilistici, le penalizzazioni per tutte le partite di ogni giornata, e poi per tutto il campionato, si ha una graduatoria che premia sia la competenza (e... «oggettività», nel senso di non farsi troppo trascinare da simpatie) in campo calcistico, e sia il giusto apprezzamento della scala numerica delle probabilità. Un difetto comune è quello dell'attrazione verso «certo» e «impossibile»: nonostante il ripetersi di «sorprese», la probabilità di vederne ancora viene dai piú notevolmente sottovalutata. Ecco un insegnamento sul quale riflettere (riguardo alla già menzionata reazione di rigetto contro l'incertezza).

4.2. Perplessità, e come superarle.

Qualcuno riterrà forse di obiettare, a questo punto, che in tal modo viene imposta l'adozione della concezione soggettivistica (notoriamente e apertamente sostenuta dallo scrivente). Ma non è cosí.

Chiariamo il punto con una breve digressione.

La proprietà indicata, consistente nell'uso della probabilità come base di operazioni aleatorie («scommesse», «assicurazioni», affari rischiosi, ecc.) è accettata da tutti e applicata da tutti indipendentemente dalla concezione della probabilità cui aderiscono. La differenza sta soltanto nel modo di restringere il campo entro il quale si ritiene lecito o corretto o ragionevole (o come altro si voglia dire) l'uso della nozione di probabilità: solo per casi tipo dadi e lotterie (per gli

«oggettivisti» di un primo tipo), o solo per fenomeni con frequenza ritenuta «stabile» (per gli «oggettivisti» di un altro tipo), mentre i soggettivisti non pongono alcuna restrizione aprioristica. (Anzi – sia ben chiaro – essi non respingono neppure le valutazioni che si appoggiano sull'uno o l'altro di questi due criteri (o su qualche compromesso fra i due), ma caso per caso, a ragion veduta, e non come precetti formalistici da applicare ad occhi chiusi. E non, soprattutto, da imporre o da accettare apoditticamente col nome di «definizioni» (che, tra l'altro, sarebbero del tutto scorrelate e quindi contraddittorie tra loro).

Ciò posto, è chiaro che l'accettazione della teoria delle decisioni in condizioni di incertezza, in conformità all'impostazione sopra indicata, *non condiziona in alcun modo l'atteggiamento di ciascuna persona nei riguardi delle diverse concezioni della probabilità*. Ciascuno, infatti, potrà e dovrà applicare tali concetti in tutti i casi in cui, secondo la sua concezione, «esiste» una valutazione di probabilità «accettabile», e se ne dovrà astenere negli altri casi, dove, per lui, la probabilità «non ha senso», o «non esiste».

Cosa potrebbe recriminare, e perché? Sarebbe forse ragionevole, per fare un'analogia ineccepibile, che uno rifiutasse un biglietto di libera circolazione sull'intera rete ferroviaria italiana perché essa comprende regioni che egli non intende visitare? Ne ha facoltà, non obbligo.

Occorre però aggiungere un'avvertenza importante. Rinunziare a far uso del ragionamento probabilistico nei casi ove uno ha degli scrupoli o dei pregiudizi è sempre cosa che impoverisce la visione delle cose, ma si può trattare di danno relativo. Purtroppo, però, c'è di peggio: c'è il rischio che, anziché semplicemente astenersi da considerazioni corrette giudicate malcerte, qualcuno (anzi molti) non sappia resistere alla tentazione di arrovellarsi per inventare o scovare e adottare metodi aberranti, pur di eliminare l'indispensabile elemento soggettivo. Magari sostituendolo con cervellotiche e vuote costruzioni formalistiche prive di costrutto, o lasciandosi abbagliare da «superstizioni pseudostatistiche» che annebbiano e distorcono la visione dei problemi. Si giunge cosí, tra l'altro, alle famigerate *adhockeries*, di cui già si è fatto ampio cenno.

Ma è istruttivo meditare sul perché delle perplessità, e sulla loro reale consistenza. Molti fatti d'osservazione fanno ritenere infatti che la riluttanza ad esprimere un grado d'incertezza mediante l'indicazione numerica di una probabilità dipenda dal timore reverenziale che incute la probabilità in quanto termine «scientifico».

Una persona abituata a fare scommesse, o magari soltanto a usare il linguaggio degli scommettitori, è spesso pronta ad esprimere la sua opinione dicendo che «darebbe un certo evento a 3 contro 1, oppure a 2 contro 5»; il che è la stessa cosa che valutarne la probabilità a $3/(3+1)=3/4=75$ per cento o rispettivamente a $2/(2+5)=2/7=28,6$ per cento.

Una persona aliena dal fare scommesse sarebbe incapace od aliena anche da tale tipo di risposta (sempre astratta), ma se, invece che di una «scommessa» si parlasse di una «assicurazione», in un caso che lo riguarda o interessa, saprebbe probabilmente stimare quale premio giudicherebbe ragionevole pagare per l'eventuale risarcimento di un danno (per incendio, per infortuni, o qualsiasi altra cosa). Al riguardo, Lindley [1971, p. 25] osserva che molte persone, abituate a pensare in termini di premi d'assicurazione, sanno stimare ragionevolmente il premio – e quindi, indirettamente, la probabilità – in ogni caso che li riguarda, ma si troverebbero imbarazzate a pensare ed esprimersi astrattamente in termini di probabilità. E aggiunge un esempio ancora piú istruttivo (e che sarebbe utile seguire ove se ne veda la possibilità). Un ingegnere chimico si preoccupava per un rischio che sapeva esistere in un processo di cui era responsabile, ma si riteneva incapace di esprimere in numeri la relativa probabilità. Sapeva però specificare l'ordine di grandezza (in termini monetari) del danno che sarebbe risultato in quel caso. Lindley gli disse allora: «Se potessi offrirti un dispositivo capace di eliminare quel rischio, quanto saresti disposto a pagarlo?» Dopo qualche tira-molla (a vuoto, perché era pura ipotesi) venne fuori quale prezzo sarebbe stato accettabile, e cosí, indirettamente, la valutazione che con una domanda diretta appariva impossibile. Come osservazione generale, Lindley commenta che (data la consuetudine con contratti di assicurazione) basta a volte chiedere una valutazione in «premio di assicurazione» anziché in «probabilità», facendo pensare in moneta concreta anziché in numeri astratti, affinché la stessa cosa appaia piú semplice. È una verità che spesso si dimentica (o si vuol dimenticare per albagia matematica: quant'è piú vuoto chiedere quanto fa 350 diviso 7, anziché 350 chilometri diviso 7 ore impiegate a percorrerli: velocità di 50 km/h). È questa ambita vuotaggine che rende giustamente uggiosa la matematica ai ragazzi!

Molte svariate e approfondite discussioni su tali argomenti si trovano in Grayson [1960], di cui riparleremo anche nel § 5.6 a proposito della «utilità».

Qui ci riguardano soltanto le osservazioni su perplessità del genere circa la traduzione in termini di probabilità, da parte degli esperti (geologi, ecc.), delle loro opinioni circa l'esistenza (e in vari gradi di quantità) di petrolio che si potrebbe trovare perforando un pozzo in un dato punto. Certo, si tratta di una stima, una congettura (*guess*): ma – e qui sta il punto – è proprio in questo senso che il responsabile dell'impresa chiede lumi al geologo (o altro esperto). Nessuno può pretendere di piú. La valutazione di probabilità può essere inserita nei calcoli preventivi, consente il confronto tra diversi progetti, è una base «concreta» (nel senso in cui può esserlo) per ragionamenti non campati in aria. E quale è, altrimenti, la situazione? Un affastellamento di frasi un po' ottimistiche e un po' pessimistiche e sempre comunque generiche, atte piú a far confusione che aiuto al povero *decision maker*. (Di tali frasi – simili quasi ad oracoli delfici – Grayson ne incolonna una dozzina a p. 56, e non molto dissimili sono molte altre disseminate nel testo).

Il responsabile finanziario di un'importante impresa petrolifera riconosce che queste probabilità sono semplicemente stime, ma assicura che sono stime utili: «Sono fatte in base al giudizio tecnico del geologo, ed è proprio su questo che ad ogni modo il *decision maker* deve basarsi». E la traduzione di tale giudizio in una probabilità numerica permette calcoli altrimenti impossibili che aiutano nel guidare la politica d'investimenti dell'impresa (p. 62); e d'altronde, se si deve prendere una decisione, necessariamente, una qualche predizione dev'essere fatta, in modo esplicito od implicito (p. 241). Per il geologo è questione di farsi l'abitudine a valutazioni in termini di probabilità, di entrare nello spirito che le informa: allora troverà non solo piú facile ma anche piú rapido esprimere il suo giudizio in tale modo (p. 261). (È la stessa conclusione suggerita dagli esperimenti analoghi per le previsioni calcistiche, o, volendo, relative a qualsiasi altro campo; cfr. § 4.1).

4.3. La condizione di coerenza.

Accogliendo il significato delle probabilità quale indicato nel § 4.1, e cioè come «quote di scommessa», tutte le proprietà e i teoremi riguardanti le probabilità (e le previsioni, che tosto introdurremo come naturale estensione) appaiono automaticamente quali condizioni necessarie e sufficienti per soddisfare l'unica esigenza in cui tutto viene compreso: quella della coerenza.

L'esigenza della coerenza, riferendoci agli eventi, consiste semplicemente in questo: *uno può dare, alla probabilità di ciascun evento, qualunque valore che vuole (tra 0 e 1, conformemente al suo giudizio), badando però di non dare a un competitore la possibilità di guadagnare a colpo sicuro stipulando un'opportuna combinazione di scommesse alle condizioni da lui indicate.*

Se, ad esempio, uno stabilisce di accettare scommesse (in entrambi i sensi) sulla base di queste valutazioni: «L. 30 risp. 20 e risp. 60 contro 100 dovute risp. in caso di vittoria, in una certa gara, del concorrente A, o di B, o di uno qualunque dei due», è chiaro che chiunque può guadagnare a colpo sicuro 10 lire acquistando per $50=30+20$ lire due buoni da 100 lire risp. in caso di vittoria di A e di vittoria di B, e cedendo per 60 lire un impegno a pagare le stesse 100 lire in caso di vittoria di (A o B).

È stato detto (con un po' di ottimismo, o, viceversa, pessimismo) che un individuo incoerente è una *money pump*, o *money-making machine*: una macchina da cui spillare in perpetuo quattrini, se non gli accadrà (prima o poi: sperabilmente per lui, con disappunto per i profittatori) di accorgersi dell'errore, cioè della necessità di adeguarsi al «teorema delle probabilità totali».

In parole povere: un biglietto che dà i medesimi diritti, in ogni circostanza, di due biglietti del costo di 30 e 20 lire, non può avere né piú né meno che il medesimo valore, cioè lire $50=30+20$.

Non si può neppure dire, a rigore, che si tratti di un teorema di calcolo delle probabilità: è cosa che vale per oggetti quali si sia (ad esempio, una pentola, un coperchio, una pentola con coperchio) purché viga il diritto alla restituzione della merce alle stesse condizioni dell'acquisto, e il fatto che si tratti di biglietti relativi a scommesse, e che i prezzi siano fissati come valutazioni di probabilità, rientra come caso particolare in quell'ovvia proprietà generale.

Comunque (un po' di nomenclatura: non si può farne del tutto a meno!): nel calcolo delle probabilità si chiama «teorema delle probabilità totali» il seguente:

TEOREMA. *La probabilità che si verifichi l'uno o l'altro di due (o anche piú) eventi* INCOMPATIBILI *è la somma delle loro probabilità.*

Questa relazione – di additività, o linearità –, con opportuna interpretazione, risulterà valida nel caso piú generale, e sostanzialmente sufficiente a tradurre in forma matematica la nozione di coerenza.

Un accorgimento spesso usato per rendere intuitive certe situazioni e relazioni probabilistiche è quello consistente nel rappresentare gli eventi come figure in un quadrato di area unitaria, aventi aree uguali alle rispettive probabilità. Si può anche immaginare (come spesso viene suggerito, per associare alle figure un'interpretazione probabilistica, seppure artificiosa) che il quadrato venga colpito da una freccia in un punto «a caso», cioè in modo che aree uguali abbiano probabilità uguali di venir colpite (o, con altra locuzione, la densità di probabilità sia costante).

Per rammentare la finalità meramente indicativa e informale di tali figure useremo la denominazione familiare di «patate».

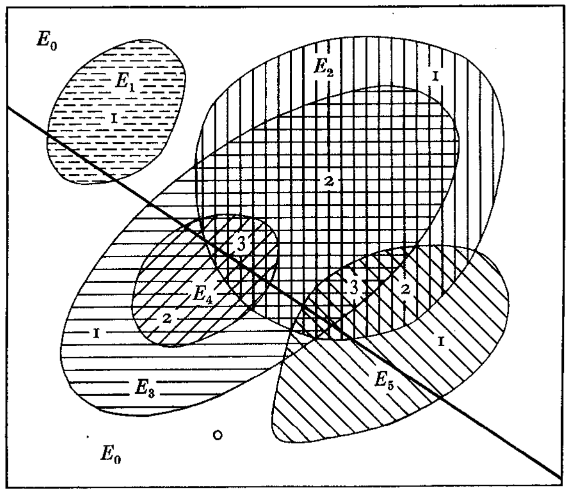


Figura 14.

Un esempio di rappresentazione di eventi mediante «diagrammi di Eulero-Venn», ossia figure (di solito dette «patate» per la forma comunemente usata). Si può pensare a un'interpretazione direttamente probabilistica intendendo l'evento E_5 come il fatto di colpire un punto di E_5 sparando sul quadrato come tiro a segno. E_5 è la figura con tratteggio obliquo vero il basso; E_0 è la parte bianca esterna a tutte le «patate»; le intersezioni di piú patate indicano il «prodotto logico» di due eventi (o piú) che esse rappresentano. (Si noti ad esempio il prodotto di $E_2E_3E_5$: lo spicchio con triplice tratteggio verticale orizzontale e obliquo decrescente. Per ogni pezzo, oltre alla designazione E_h ($h=1, 2, \ldots, 5$), è indicato il numero (0, 1, 2, 3) degli eventi E_h cui appartiene.

Si può cercare di disegnare le patate in modo che l'area di ciascuna (e loro intersezioni) corrisponda alla rispettiva probabilità. Ciò può valere però soltanto rispetto a un dato stato d'informazione, perché, se esso cambia, cambiano, naturalmente, anche le probabilità e i loro rapporti. Basti pensare, ad esempio, che una nuova informazione faccia escludere la parte al di sopra (o quella al di sotto) della retta in diagonale.

Di solito, purtroppo, per contagio di quella moda o mania che è la «insiemistificazione della matematica» (la tendenza, cioè, a vedere e tradurre tutto in termini di teoria degli insiemi) non si fa distinzione fra i molti casi ove ciò è appropriato e i moltissimi dove non lo è, e si ha inevitabilmente come risultato di rendere astruse le cose facili, e complicate – e spesso insensate – le cose semplici. È una stortura non meno deplorevole di quella criticata riguardo alle *adhockeries*, anche se in senso opposto, di raffinato iperlogicismo anziché di semplicismo.

L'immagine visiva (cfr. fig. 14) rende chiaro, ad esempio, che una patata racchiudente un'altra ne è una conseguenza necessaria (ossia l'altra ne è una condizione necessaria); che due patate disgiunte rappresentano eventi incompatibili, cioè escludentisi a vicenda; che la negazione di un evento è tutta la parte del quadrato che rimane escludendo la patata; che l'intersezione di due patate è l'intersezione (o «prodotto logico») dei due rispettivi eventi, e l'unione ne è l'unione (o «somma logica»), cioè, rispettivamente, l'evento consistente nell'esser veri entrambi, oppure almeno uno, gli eventi di partenza.

A proposito del termine 'patata', esso è introdotto pensando alle forme piú semplici che si usano nella rappresentazione geometrica (detta «diagramma di Venn»)... finché è possibile. Ma lo usiamo genericamente, per un'area di qualunque forma corrispondente a un evento, anche se di forma irregolare (come intersezione di piú patate originarie) od anche se è formato di piú pezzi staccati. Tutte queste circostanze non hanno alcuna importanza concettuale; unica circostanza che ne dipende è la maggiore o minore intelligibilità della rappresentazione grafica, che riesce di dubbia o nulla utilità se le complicazioni crescono e non si trovano accorgimenti per render tuttavia «visibile» ciò che conta. Parecchie cose diverranno piú chiare vedendole inquadrate in uno schema concettuale piú ampio, comprendente i «numeri aleatori» e gli «eventi subordinati».

4.4. Estensione ai numeri aleatori.

La probabilità di un evento non è che un caso particolare di una nozione piú generale: quella di previsione di un numero aleatorio.

Diciamo «numero aleatorio» un numero ben definito, X, di cui non conosciamo il valore x (o perché ancora non determinato, ad esempio il numero di nati a Roma nel mese prossimo, o perché non ancora accertato o pubblicato, o comunque non visto, od anche visto ma dimenticato, dalle persone che supponiamo chiamate a stimarlo).

In tale situazione si dirà previsione di X (secondo la valutazione di un dato individuo) il prezzo $\bar{x}$ che egli ritiene equo pagare (in lire, o qualsiasi altra unità) per ricevere X lire (cioè tante lire quanti risulteranno i nati ecc. ecc.); beninteso, sulla stessa base deve esser disposto a scommesse anche in senso inverso.

Si sarà notato che si tratta ovviamente, come detto, di un'immediata estensione della nozione di probabilità: la probabilità di un evento E qualunque altro non è che la previsione di un guadagno 1 se E si verifica. È anzi opportuno, formalmente, identificare addirittura l'evento E col numero aleatorio che vale 1 se E è vero e 0 se E è falso.

Spesso tale numero viene chiamato «indicatore di E», ma l'identificazione elimina un doppione fittizio e inutile (come se si distinguessero quali soggetti diversi «Giorgio» o «il signor Giorgio» o «la persona del signor Giorgio»), e consente di operare sugli eventi aritmeticamente per eseguire operazioni logiche: per esempio, $E_1+E_2+...+E_n=X$ è il «numero di successi», perché ogni evento vero vale 1 e gli altri 0; la negazione di E, che si indica anche con $\tilde{E}$, si esprime aritmeticamente scrivendo $1-\tilde{E}$ (complemento ad 1, cioè all'evento certo: $1-1=0$, $1-0=1$); il prodotto $E_1E_2...E_n$ è il prodotto logico degli n eventi, vero se sono tutti veri; e cosí via.

Grazie a questa identificazione dell'evento E col suo «indicatore», la nozione di probabilità si identifica (per riferimento agli eventi) con quella di previsione, e se ne trae ulteriore vantaggio usando senza distinzione il medesimo simbolo **P** per la probabilità $\mathbf{P}(E)$ di un evento E e per la previsione $\mathbf{P}(X)$ di un numero aleatorio X. La previsione è additiva: $\mathbf{P}(X+Y)=\mathbf{P}(X)+\mathbf{P}(Y)$ (sempre per lo stesso significato di «prezzo»); aggiungiamo subito, come osservazione particolare e utile, che se i valori possibili per X sono in numero finito: x_1, x_2, ..., x_n rispettivamente con probabilità $p_1, p_2, ..., p_n$, è pertanto $\mathbf{P}(X)=p_1x_1+p_2x_2+...+ +p_nx_n$, cioè media aritmetica degli x_i coi «pesi» p_i.

Analogia meccanica da rammentare e utilizzare mentalmente: $\bar{x}=\mathbf{P}(X)$ è il baricentro di masse p_i nei punti x_i: il baricentro di una distribuzione di masse, interpretando come tali le probabilità.

Per ora basti aggiungere qualche cenno generico. Se, come nel caso già indicato quale esempio, sappiamo che il numero aleatorio X ammette solo un numero finito di valori possibili, x_i, sui quali sono concentrate le probabilità (cosí come, nell'analogia meccanica, masse «puntiformi» o «concentrate»), la distribuzione si dice discreta. Tale può essere anche se tali masse sono in numero infinito. (Ad esempio si può ottenere «testa» per la prima volta, giocando a testa e croce, o al 1º colpo, o al 2º, al 3º, ..., all'n-esimo,... o anche mai; le rispettive probabilità sono $1/2$, $1/4$, $1/8$, ..., $1/2^n$, ..., 0. È opportuno notare che la probabilità del «mai» è 0, ma che ciò non significa impossibilità: le successioni infinite «sempre testa» o «sempre croce» sono altrettanto possibili che qualunque altra successione infinita, di cui una si dovrà pure verificare... se avessimo vita e

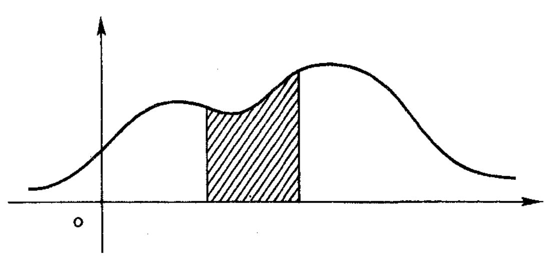


Figura 15.

Sul diagramma della densità (per distribuzioni continue) è $y=f(x)=$ densità nel punto x, e la probabilità di ogni intervallo è data dall'area tratteggiata che vi sta sopra.

pazienza bastevoli: ma questa eccezione, dipendente solo da contingenti limitatezze della vita terrena, è concettualmente irrilevante).

All'estremo opposto, abbiamo le distribuzioni continue (senza probabilità concentrate in alcun punto particolare); in pratica, dicendo «continue», si allude al caso molto piú restrittivo e «regolare» di distribuzioni che ammettono una densità (essa stessa, per lo piú, continua). Ciò significa, per esprimerci mediante una figura (fig. 15), che la probabilità che X cada in un dato intervallo è l'area sotto la curva nella parte ad esso sovrastante. L'ordinata sovrastante a un dato punto x si dice densità della probabilità ivi, in quanto la probabilità che X si trovi in un piccolo intervallo tra $x\pm\delta/2$ (piccolo dev'essere δ, che è la lunghezza di tale intervallo) è approssimativamente il prodotto della densità per δ (come se in quel tratto la curva fosse rettilinea, ferma restando l'ordinata centrale).

Fra questi due tipi estremi ne esistono altri di cui qui non c'è possibilità d'occuparci.

Vediamo invece anche qui alcune osservazioni su ciò che possono dirci le rappresentazioni grafiche tipo «patate». Abbiamo menzionato poco sopra, come esempio di numero aleatorio, il numero di successi $X=$numero degli E_i veri (ossia di valore 1); se gli E_i sono rappresentati con patate (naturalmente: in modo che ci sia rispondenza logica, cioè intersezioni vuote rappresentino eventi impossibili e non-vuote eventi possibili) ognuno dei pezzi dati da intersezioni dei contorni fra patate rappresenta il prodotto logico di tutti gli eventi alle cui patate è interno: se tali eventi sono h, in quella zona il numero aleatorio X, numero di successi, assume il valore h. (Nella figura 14 è indicato il valore di tale numero per ogni pezzo della partizione). Ma questo è solo un esempio banale; una generalizzazione (pur essa banale) si ha ad esempio supponendo che al verificarsi di ogni E_i corrisponda un guadagno (positivo o negativo) x_i: in tal caso, nella zona intersezione delle patate E_{i_1}, E_{i_2}, ..., E_{i_h} il guadagno (il numero aleatorio X in questa ipotesi piú complessa) avrà il valore $X=x_{i_1}+x_{i_2}+...+x_{i_h}$.

4.5. Estensione agli eventi e numeri aleatori subordinati.

Ancora un po' di pazienza: a un certo punto non si può fare a meno di introdurre dei termini e delle notazioni per semplificare l'esposizione nel seguito. Ma siamo quasi al termine, e forse si è già cominciato a vedere o intravvedere il tipo di vantaggi che si conseguono.

L'ulteriore estensione che ci resta da considerare è quella che riguarda gli «eventi subordinati», e quindi il caso di «scommesse sotto condizione».

Per fare un esempio, ci si può chiedere non solo quale probabilità attribuire alla vittoria di un certo concorrente, N. N., in un certo torneo, ma anche quale sia tale probabilità qualora la gara si svolgesse sotto la pioggia, o se un suo temibile avversario non vi partecipasse, o se egli fosse euforico, o viceversa depresso, per l'esito felice o deludente in una gara della vigilia.

In termini di «scommesse sotto condizione», si tratterebbe di pagare un certo importo pS per ricevere S se si verifica l'evento E (vittoria) ed anche la condizione H (la pioggia, oppure la rinunzia dell'avversario temuto, o l'euforia, o la depressione), mentre se H non si verifica (non si verifica la pioggia o quel che altro si era indicato come condizione) la scommessa viene annullata (cioè l'importo pS non viene pagato, o, se era stato pagato, viene restituito).

Indicheremo con $\mathbf{P}(E|H)$ la probabilità di E subordinatamente ad H (detta anche «probabilità dell'evento subordinato $E|H$), ossia la p delle scommesse di cui sopra. È chiaro che la probabilità di HE (prodotto dei due eventi, cioè del verificarsi di entrambi) deve essere uguale al prodotto $\mathbf{P}(H)\,\mathbf{P}(E|H)$, poiché pagando tale importo posso vincere, se H si verifica, l'importo $\mathbf{P}(E|H)$ mediante il quale posso vincere 1 se si verifica E (cosí come se avessi direttamente scommesso su HE). È, questo, il «teorema delle probabilità composte»:

$$\mathbf{P}(HE)=\mathbf{P}(H)\,\mathbf{P}(E|H) \qquad \text{od anche} \qquad \mathbf{P}(HE)=\mathbf{P}(E)\,\mathbf{P}(H|E).$$

È istruttivo riscrivere tali espressioni moltiplicando e dividendo risp. per $\mathbf{P}(E)$ e per $\mathbf{P}(H)$. Risulta

$$\mathbf{P}(HE)=\mathbf{P}(H)\,\mathbf{P}(E)\,\frac{\mathbf{P}(E|H)}{\mathbf{P}(E)}=\mathbf{P}(E)\,\mathbf{P}(H)\,\frac{\mathbf{P}(H|E)}{\mathbf{P}(H)}$$

$$\left(\text{quindi } \frac{\mathbf{P}(H|E)}{\mathbf{P}(H)}=\frac{\mathbf{P}(E|H)}{\mathbf{P}(E)}\right)$$

e mette in luce i tre casi che è opportuno distinguere in relazione al prodotto di due eventi, qui E ed H: se $\mathbf{P}(HE)=\mathbf{P}(H)\,\mathbf{P}(E)$, ossia $\mathbf{P}(E|H)=\mathbf{P}(E)$, ed anche $\mathbf{P}(H|E)=\mathbf{P}(H)$, i due eventi si dicono stocasticamente indipendenti (e spesso si dice «indipendenti» sottintendendo «stocasticamente», il che può farsi solo se tale precisazione è chiaramente sottintesa: altrimenti potrebbero sorgere equivoci con «indipendenza logica» o «lineare» e forse altro ancora).

Se invece $\mathbf{P}(HE)$ fosse maggiore del prodotto $\mathbf{P}(H)\,\mathbf{P}(E)$, ossia $\mathbf{P}(E|H)$ maggiore di $\mathbf{P}(E)$, e quindi anche $\mathbf{P}(H|E)$ maggiore di $\mathbf{P}(H)$, si direbbe che c'è correlazione positiva fra H ed E; nel caso opposto si direbbe che c'è correlazione negativa.

La nozione di correlazione (negativa, o nulla, o positiva) vale ugualmente, ed è importante, anche per numeri aleatori (ed anche – ci limitiamo a farne menzione – nel caso di distribuzioni statistiche: ad esempio, correlazione tra statura e peso in un gruppo di individui, o tra intensità del traffico e incidenti per una data rete stradale e periodo). Anche qui si tratta di distinguere se la previsione (o, nelle applicazioni statistiche, il valor medio) del prodotto XY è minore o uguale o maggiore del prodotto dei valori medi. Usando anche qui $\mathbf{P}$ (nella statistica si userebbe $\mathbf{M}$ (media), ma non è il caso qui di cambiare simbolo senza che formalmente ce ne sia motivo) i tre casi sono

$$\mathbf{P}(XY)<\mathbf{P}(X)\,\mathbf{P}(Y) \qquad \mathbf{P}(XY)=\mathbf{P}(X)\,\mathbf{P}(Y) \qquad \mathbf{P}(XY)>\mathbf{P}(X)\,\mathbf{P}(Y)$$

e indicano la tendenza di X e Y a crescere concordemente (primo caso) o discordemente (ultimo) oppure senza che alcuna delle due tendenze si manifesti o prevalga (caso centrale). Si badi che tale caso non va confuso con l'indipendenza (essa si avrebbe, nell'esempio statura-peso, se la distribuzione secondo peso fosse la stessa in tutti i raggruppamenti per statura – e quindi anche viceversa –, mentre una non-correlazione si potrebbe avere anche se, ad esempio, il peso tendesse a crescere con la statura fino a un certo punto e poi a diminuire. È utile richiamare l'attenzione su distinzioni del genere – e ce ne sarebbero molte! – per sviluppare una certa consapevolezza delle diversità di aspetti distinti, ciascuno idoneo per certi scopi e ricerche, che nel linguaggio corrente si considerano sinonimi e rischiano cosí di essere impiegati impropriamente o addirittura a sproposito conducendo a conclusioni erronee).

Volendo anche qui far menzione della rappresentazione mediante «patate» (che, del resto, riesce particolarmente felice al riguardo) basta dire che, data la rappresentazione di tutti gli eventi che interessano, compreso H, non occorre far altro che limitarsi a considerare quel che sta nella patata H (cosicché diventa inesistente tutto ciò che ne sta fuori). Le aree superstiti dei singoli E_i sono le rispettive intersezioni con H, e valgono quindi $\mathbf{P}(HE_i)$, ossia $\mathbf{P}(H)\,\mathbf{P}(E_i|H)$; le probabilità desiderate, $\mathbf{P}(E_i|H)=\mathbf{P}(HE_i)/\mathbf{P}(H)$, sono queste stesse salvo dividerle per $\mathbf{P}(H)$, ossia, praticamente, prendendo come unità di misura l'area rimasta, cioè l'area $\mathbf{P}(H)$ della patata rimasta, H.

Questo procedimento ha particolare interesse nei problemi di decisione, perché la richiesta di informazioni, che spesso ha fondamentale importanza per chi intenda decidere con maggiore cognizione di causa, equivale proprio a ragionare su di una «patata piú piccola». Se apprendiamo che è vero un certo evento H (rappresentato anch'esso come patata nello schema) e nulla piú, le probabilità degli E_i divengono le $\mathbf{P}(E_i|H)$, cioè le aree dentro la patata H prendendo l'area di questa come unità. Piú in generale, se l'informazione consiste nell'apprendere quale tra gli eventi (o, per il ruolo che qui hanno, «le ipotesi») di una partizione, $H_1, H_2, \ldots, H_s$, è vero – e sia H_j –, le probabilità delle E_i diverranno le $\mathbf{P}(E_i|H_j)=\mathbf{P}(E_iH_j)/\mathbf{P}(H_j)$, cioè le aree delle parti della E_i contenute nel pezzo rimasto H_j, la cui area si assume come unità.

4.6. Il ragionamento induttivo.

L'accenno al ruolo dell'acquisizione d'informazioni introduce ad una problematica fondamentale sotto il duplice aspetto teorico ed applicativo.

L'aspetto teorico consiste nell'analisi del ragionamento induttivo dal punto di vista logico e probabilistico, e ne diamo subito qui un cenno essenziale.

L'aspetto applicativo, riguardante principalmente la teoria delle decisioni, costituirà uno degli argomenti piú caratterizzanti e delicati del prossimo § 5, dedicato alle decisioni in condizioni di incertezza, dato che la questione implica congiuntamente il vaglio degli aspetti di carattere probabilistico e di carattere economico.

In certo senso, l'aspetto probabilistico potremmo dire di averlo già esaurito con l'indicazione or ora data, nel § 4.5, delle relazioni che legano le probabilità concernenti due eventi E ed H, e cioè $\mathbf{P}(E)$, $\mathbf{P}(H)$, $\mathbf{P}(EH)$, $\mathbf{P}(E|H)$, $\mathbf{P}(H|E)$, ma rimane da chiarire il ruolo che esse svolgono nel ragionamento induttivo, ed anche da illustrare precauzioni dubbi e critiche cui possono dar luogo, sia secondo il punto di vista qui sostenuto che da punti di vista diversi (di «oggettivisti»).

Come si ricava subito dall'uguaglianza delle due forme (simmetriche) in cui si è presentato ivi il «teorema delle probabilità composte», abbiamo:

$$\mathbf{P}(H|E)=\mathbf{P}(H)\frac{\mathbf{P}(E|H)}{\mathbf{P}(E)}.$$

È, questa, l'espressione del cruciale teorema di Bayes, fondamento unico e universale del ragionamento induttivo secondo i bayesiani (soprattutto soggettivisti), e oggetto di anatema per gli oggettivisti, che, pur di scansarsene, non si peritano di trastullarsi con *adhockeries*. I precedenti esempi e commenti – nei §§ 2.6-2.8 e *passim* in seguito – hanno già messo in luce i termini della contrapposizione di tesi.

Non è il caso di ripeterle, ma è forse opportuno tentare di integrarle in altro senso. Volendo azzardare una diagnosi della riluttanza degli oggettivisti a concepire il problema del ragionamento induttivo nella sua naturale immediatezza, si potrebbe ascrivere la colpa a quei preconcetti che in vari modi inibiscono di parlare e di fare uso della probabilità nel modo naturale.

Le principali remore del genere sono due.

Ecco la prima. Per applicare il teorema di Bayes occorre attribuire delle probabilità alle «ipotesi» prese in considerazione, e vedere poi come variano in seguito alle successive osservazioni e informazioni, e in base a ciò trarre le conclusioni. A ciò si oppone la preclusione dogmatica dei frequentisti contro l'attribuzione di una probabilità a un «caso singolo». Quale probabilità si doveva attribuire alla teoria di Wegener (sulla deriva dei continenti) all'epoca in cui i piú l'avversavano? E piú tardi? Tali questioni, per costoro, sarebbero state prive di senso, a meno di ammettere la possibilità di sperimentare molte ripetizioni della creazione del mondo e vedere in quale percentuale di casi i fatti si sviluppavano in accordo con la teoria di Wegener. Invece l'accumularsi di conoscenze meglio spiegabili accettando la teoria di Wegener anziché quelle dei suoi oppositori ha portato alla sua pressoché generale accettazione. Le *adhockeries* servono benissimo allo scopo di dare surrogati contorti di ragionamento per ottenere conclusioni convenzionali di «accettare» o «respingere» delle «ipotesi», nel senso di consigliare di comportarsi come se fossero risp. vere o false, pur avvertendo che ciò non significa ritenerle né vere o false né molto o poco probabili.

Ed ecco la seconda. Gli oggettivisti rifiutano addirittura di ammettere che il fatto rilevante agli effetti della scelta della decisione sia l'informazione acquisita in quanto tale; per loro non basta sapere ad esempio che in 20 estrazioni con reimbussolamento (e dovute precauzioni di rimescolamento, ecc.) si sia avuto 13 volte pallina bianca e 7 nera; per essi occorre conoscere il «disegno» dell'esperimento, perché una cosa è sapere che era prefissato il numero di 20 estrazioni, o di arrivare a 13 bianche, ed altra è il continuare fino a che la persona che l'esegue è stanca; e chi piú ne ha piú ne metta. Questi sono ingredienti delle *adhockeries*, e come tali sono sacri, e sarebbe sacrilegio rilevarne l'irrilevanza.

Perfino dopo che Abraham Wald aveva trovato e dimostrato che tutte le procedure tranne quelle seguenti – almeno formalmente – la procedura bayesiana erano inammissibili (cioè di certo oggettivamente peggiori di altre «ammissibili»), il pregiudizio continuò ad avere il sopravvento sul buon senso. Attualmente parecchi segni e sintomi inducono a sperare che una certa resipiscenza e conversione procedano, anche se finora ciò non avviene che faticosamente e lentamente.

Molto di piú specifico e preciso è difficile, ed anzi impossibile, aggiungere, oltre a tutto perché gli stessi oggettivisti – mancando di una qualsiasi base logica plausibile – scelgono varie scappatoie per sfuggire alla logica della soluzione bayesiana (o «ammissibile» secondo Wald), accettando o respingendo o intrudendo, ciascuno a suo modo e senza mai un reale costrutto, questo o quell'ingrediente. Per esempio: intervalli di confidenza, intervalli di fiducia, minimax, decisioni randomizzate, caratteristica operativa, no all'*optional stopping*, livello di significatività, principio della verosimiglianza (*likelihood*), intervalli di tolleranza, probabilità «fiduciale» (secondo Fisher), metodo «empirico-bayesiano»... e l'elenco è certamente ben lungi dall'essere completo.

Ma perché tanti sforzi, quando nel metodo bayesiano tutto è chiaro, lampante, irrefutabile?

Ebbene: nella sua effettiva applicazione si potrebbe ravvisare una certa faciloneria; però in un senso percettibile e inquietante soltanto per i soggettivisti (anzi: per i piú attenti e critici); non certo per gli oggettivisti. Il fatto è questo: dopo un'osservazione od esperimento che ha dato come risultato H_i (cioè una delle n «ipotesi» $H_1, H_2, ..., H_n$ prestabilite), diremo che, ora, nel nuovo stato di informazione, dovremo attribuire ad E, come probabilità, $\mathbf{P}(E|H_i)$ anziché la $\mathbf{P}(E)$ precedente (o «iniziale», rispetto all'informazione in oggetto). D'accordo, ma... la nuova informazione acquisita non sarà in genere limitata ad H_i, ma ad H_i con piú o meno numerosi e significativi particolari. Ad esempio, se l'informazione doveva riguardare l'esito di una partita di calcio, con le tre alternative «vittoria, pareggio, sconfitta» (e in base a ciò avremmo revisionato il pronostico per la giornata successiva), e apprendiamo che il risultato è vittoria, difficilmente l'informazione sarà tutta qui. Sapremo il punteggio (2-1 o 5-0 non è la stessa cosa!), sapremo da TV o radio o giornali i commenti e giudizi di giornalisti e tecnici; e non basta, ché altrettanto conterà sapere circa la prossima avversaria. Forse questo esempio è scelto in modo da rendere lampante una possibile trascuratezza altrove meno percettibile; e sia, ma l'importante è che, visibile piú o meno, è certo che esiste sempre.

Di ciò gli oggettivisti non potrebbero neppure far cenno, perché il loro schematismo astratto esclude ogni riferimento a quella parte di «realtà» che non è stata ritenuta degna a priori di far parte del modello teorico, dell'armamentario statistico-oggettivistico.

Possono certo dire molte cose, vantare molti pregi del loro atteggiamento «piú scientifico» perché bandisce il «deleterio» soggettivismo.

Possono vantare, ad esempio, di riuscire a dare risposte sicuramente oggettive, come quando basano una decisione sulla sola *likelihood*, cioè sulla sola

$\mathbf{P}(E|H)$, ignorando – anzi dichiarando priva di senso – la «probabilità dell'ipotesi», $\mathbf{P}(H)$. Ma questo è nient'altro che un procedimento bayesiano nel quale le diverse ipotesi H_i si considerino ugualmente probabili (sia pure applicandolo senza rendersi conto di aver fatto implicitamente questa assunzione aborrita).

E possono vantare una assai vasta collezione di metodi (piú o meno, *adhockeries*) messi insieme con ingegnosità e presentati con qualche piú o meno contorta pretesa di significatività (oltre che di originalità) da numerosi ricercatori.

Ben povera cosa, al confronto, è il metodo bayesiano, che non è né complicato né artificioso bensí semplice e naturale al massimo.

La migliore testimonianza al riguardo, tanto piú indubbiamente valida perché espressa come una *critica*, è quella di un valente statistico (e neppure dei piú settariamente avversi: Herman Chernoff). Tale teoria, secondo lui, «makes statistics dull»: rende cioè banale, ottusa, tediosa, inespressiva, la statistica (cioè: il metodo statistico in cui tale teoria si traduce).

Ed ha anche ragione, cosí come hanno ragione gli Inglesi di oggi (e come, da noi, i vecchi di quando io ero bambino) di trovare banale il sistema metrico decimale, che elimina non solo le traduzioni tra unità in uso in diversi paesi (o città), ma anche le suddivisioni in dodicesimi o ventiquattresimi (come once e carati) o in ventesimi e ventunesimi (come sterline e ghinee).

Mi pare indubbio però che (nel caso di Chernoff come negli altri citati) la comprensibile riluttanza ad abbandonare consuetudini inveterate e assimilate, vanto di una tradizione secolare e prestigiosa, non può essere che un fenomeno transitorio di disagio per l'adattamento.

È, del resto, un fatto comune in tutte le scienze l'alternanza di periodi in cui esse si accrescono piú o meno disordinatamente per acquisizioni e scoperte disparate, ed altri in cui affiora un naturale filo conduttore che piú o meno compiutamente le unifica, e semplifica la veduta d'insieme.

Perché non dovrebbe finalmente avvenire cosí anche nel campo di cui qui ci si occupa? E che il fatto di diventare apparentemente *dull* non significhi, per la statistica, il felice raggiungimento di una visione unitaria? Che una teoria cosí *dull*, bistrattata come il «brutto anatroccolo» della fiaba di Andersen, non venga finalmente accolta e riconosciuta come un cigno nel lago dei cigni?

5. *La decisione: neobayesiana, neobernoulliana.*

5.1. La Torre di Babele e l'Oasi.

Dobbiamo ora intraprendere, in forma piú sistematica e un po' anche tecnica, lo studio del problema della decisione, cioè del «come si dovrebbe» decidere: problema riguardo al quale abbiamo finora solo cercato di illustrare, in via preliminare e da un punto di vista concettuale, alcuni aspetti salienti e controversi. Tali aspetti si ripresenteranno – ovviamente – anche in questa nuova fase, perché le divergenze di punti di vista si ripercuotono ed estrinsecano nella costruzione e scelta di divergenti concezioni e conseguenti metodologie scientifiche e pratiche.

Ecco come lo dice uno dei massimi autori in argomento, Leonard Savage [1972, p. 2]: «È unanime il consenso sul fatto che la statistica dipenda in qualche modo dalla probabilità. Ma, riguardo a cosa la probabilità sia, e a come sia connessa con la statistica, c'è tale un completo dissenso e rottura di comunicazioni, quali raramente ce n'è stati dal tempo della Torre di Babele».

Però si può anche seguire, nel tempo, una serie di casi ove la visione che asseriamo «naturale» si è preservata o si è ripresentata o rinnovata, e viene spontaneo dire, per contrapposto alla famigerata torre, che essi costituiscono un'Oasi: un'oasi ecologica rimasta immune da inquinamenti.

Per dare qualche appropriata indicazione ed esemplificazione al riguardo, basterà limitarci ai due aspetti maggiormente rilevanti in tale contesto: il ragionamento induttivo e il criterio di preferenza.

Il ragionamento induttivo, del «come apprendere dall'esperienza», è antichissimo argomento di discussioni filosofiche da cui emerge il contrasto tra asserzioni apodittiche («cosí è avvenuto, e quindi cosí dovrà avvenire») e previsioni soggettive («cosí è avvenuto, e mi attendo che verosimilmente cosí abbia ad avvenire»).

Il conflitto di mentalità che emerge da tale contrapposizione è quello che Mises (col quale su questo punto concordo) condensa nelle definizioni di *great thinkers* 'grandi pensatori' (come Socrate e Hume, che vivificano il pensiero e ne stimolano il progresso), e di *school philosophers* 'filosofi cattedratici' (come Platone e Kant, che mortificano il pensiero e ne tentano la mummificazione). E infatti la spiegazione del ragionamento induttivo in Hume (benché non esposta matematicamente, e con la riserva che la mia interpretazione del suo pensiero potrebbe non essere conforme alla realtà, dato che esso, da altri, viene interpretato in modo diverso) mi appare del tutto consona alle vedute bayesiano-soggettivistiche; e ne traggo conferma dal sacro zelo sfoderato dal povero Kant in difesa della sacra ottusità minacciata dalla falla che Hume aveva aperta.

Il teorema di Bayes, il ragionamento bayesiano, costituisce nient'altro che la traduzione in formule di ciò che, concettualmente, è il ragionamento induttivo secondo Hume (se vale l'interpretazione qui sostenutane). Ed è perciò che il proposto metodo di decisione è stato indicato come neobayesiano («neo», perché non sono passati invano, benché il concetto fondamentale sia rimasto intatto, i due secoli e piú dal 1764, data della memoria di Bayes, non senza significato apparsa una ventina d'anni dopo l'opera filosofica di Hume).

Quanto al criterio di preferenza, la scelta è tra prefiggersi di massimizzare il guadagno in termini monetari o in termini di utilità. Ragionare in termini di utilità significa tener conto della avversione al rischio, rispondente a normale prudenza (ma può anche prevalere la tendenza opposta, la tentazione di rischiare, o in circostanze speciali o, per certuni, per mania). L'avversione al rischio si rivela ad esempio nel fatto che, normalmente, fra «1 milione certo» e «2 milioni se viene testa», si preferisce la prima soluzione (ed anche qualcosa di meno, per esempio 900 000 lire: tale «prezzo di 2 milioni con probabilità 50 per cento» è appunto l'utilità (mentre 1 milione sarebbe la previsione, o speranza matematica). Altra spiegazione, o dizione, equivalente: successivi introiti uguali in termini monetari hanno utilità (o, meglio, dànno luogo a un «aumento di utilità») decrescente.

Ebbene: la scelta, su questo punto, è a favore della misura in termini di utilità, ossia della teoria neobernoulliana (per cui, tenendo conto di entrambi gli aspetti, la teoria cui ci atteniamo si dice – come già indicato nel titolo del paragrafo – neobayesiana e neobernoulliana). Riguardo al «neo» valgono le stesse generiche ragioni dell'altro caso; quanto a «bernoulliano» occorre qualche precisazione. Nominando «Bernoulli» o «bernoulliano» si allude generalmente al piú celebre fra i numerosi matematici di questa famiglia: Giacomo, autore del primo trattato di calcolo delle probabilità (*Ars Conjectandi*, 1713) e allo «schema bernoulliano» (o «processo...»), di «prove ripetute ugualmente probabili e indipendenti» (come ad esempio a testa e croce, o simili).

Invece la denominazione si riferisce a Daniele Bernoulli (nipote di Giacomo) e alla sua nota memoria *Specimen theoriae novae de mensura sortis* (1738) che ha portato persuasivi argomenti alla tesi (già tempo prima affacciata da Gabriel Cramer) sulla decrescente utilità di successivi incrementi di ricchezza, e quindi sulla necessaria sostituzione delle valutazioni in termini di moneta con valutazioni in termini di utilità ai fini di scommesse e operazioni aleatorie. Una vittoria sullo spirito astrattamente e formalisticamente giuridico (o, se posso citare un neologismo personale, «giuridicolo») che dava un valore mitico, assoluto, all'aritmetica monetaria (una stortura che purtroppo imperversa tuttora). E vale la pena di citare un indovinato esempio di Daniele Bernoulli: forse quello che piú di tutti (piú del famoso ma poco realistico paradosso del gioco con vincita 2^n se è all'n-esimo colpo che una moneta dà testa per la prima volta) ha fatto capitolare le opposizioni: «Un povero possiede un biglietto di lotteria che, in base a un ultimo sorteggio, gli farà vincere o 20 000 ducati o nulla (con probabilità 1/2 e 1/2). Agisce in modo irragionevole se cerca di venderlo per 9000 ducati, mentre la speranza matematica è di 10 000?»

Fuori della Torre di Babele, ho detto, c'era anche un'Oasi. Non molto popolata, spesso frequentata da persone di passaggio, da non-conformisti che sfuggivano le adhoccaggini e le superstizioni di moda, punto di riferimento per i pratici che in vari campi affrontavano problemi seri e li esaminavano con innato criterio e non secondo regolette prefabbricate. I contorni sono sfumati: il soggettivismo è chiaro in De Morgan, e sopravvive piú sfumato in molti inglesi fino a Keynes, con finale riapparizione (pur con qualche incertezza) in Ramsey; il bayesianismo sopravvive vivace fino a Harold Jeffreys; la concretezza di visione e impostazione mantiene chiare le idee di persone intelligenti che si occupavano di problemi attuariali (come Bailey) o tecnici (come Molina e Fry), o dei precursori della ricerca operativa (come dei *polytechniciens* dell'Ottocento). Sono ricordi frammentari e vaghi (peccato non riuscire a far giustizia menzionando tutti!); forse soltanto Savage sarebbe stato in grado di riuscirvi: lui che aveva anche la dote (nonostante un grave difetto alla vista) di riuscire a leggere tutto e ricordare tutto.

5.2. Le decisioni in condizioni d'incertezza.

Abbiamo già spiegato i motivi per cui riteniamo dimostrato e accettato che l'impostazione esatta – cioè rispondente allo scopo (o *zweckmässig*, con parola che ad ogni occasione invidio ai tedeschi: *Zweck* significa 'scopo', e *mässig* 'adeguato, commisurato', da *Mass* 'misura') – sia quella neobayesiana neobernoulliana. Ci siamo soffermati alquanto sui singoli aspetti e sottoproblemi spiegando il perché di ogni passo. Ma ora bisogna coordinare le parti nel tutto, procedere al montaggio dell'intero meccanismo collocando al giusto posto ogni pezzo, rendendosi conto della sua connessione con gli altri, e giungendo cosí a capire nella sua semplicità sostanziale il funzionamento risultante del tutto.

Date queste premesse, apparirà certo strano che, come prima cosa, venga proposto al lettore di cominciare da una via di mezzo: neobayesiana sí, ma neobernoulliana no. Cioè, riferendoci sempre, dapprima, a misurazioni in termini di moneta anziché di utilità. Il motivo è questo: cominciare subito col caso piú complesso (utilità) raddoppia le complicazioni e potrebbe ingenerare confusioni e scoraggiamento; cominciare dal caso piú semplice (moneta) come se fosse l'unico e definitivo accrescerebbe le difficoltà del dopo, e potrebbe addirittura creare reazioni di rigetto verso complicazioni che disturbano il quadro fiduciosamente acquisito; cominciare dal caso piú semplice presentandolo però come un primo abbozzo grossolano che dovrà venire affinato sembra l'unico modo onesto e rassicurante.

Si può anche dire, mediante un'abbreviazione introdotta da Raiffa, che questo «primo abbozzo» si applica al caso che il *decision maker* sia un EMVer, mentre quello raffinato è richiesto per un non-EMVer; il significato di EMV è «expected monetary value» (previsione in termini monetari), e pertanto EMVer e non-EMVer significa individuo per il quale la curvatura dell'utilità non ha effetto (ad esempio, perché la sua ricchezza è tale, rispetto alla posta in gioco, che l'esito di un colpo non lo turba né punto né poco), e rispettivamente viceversa.

Con l'occasione, giova segnalare Raiffa [1968], che sviluppa abbastanza ampiamente, ma in forma chiara e relativamente semplice, concetti ed esempi in senso conforme alla presente esposizione (EMVer vi è introdotto a p. 134). In forma piú discorsiva, e con osservazioni spesso stimolanti, cose in parte analoghe ma seguenti un diverso scopo piú orientativo si trovano in Lindley [1971].

La schematizzazione piú semplice di un problema di decisione in situazione d'incertezza è quella già vista per i giochi (e potremmo addirittura riferirci alla tabella 3), pensando che le colonne C_1, ..., C_6 riguardino le diverse ipotesi o situazioni possibili (ed esclusive), mentre le decisioni accessibili al giocatore (o a Te) sono indicate dalle righe. Ivi si vede cosa si vince o si perde scegliendo una data riga, nelle eventualità consistenti nelle colonne. Spesso si usa dire (specie in inglese) che esse indicano la «scelta della Natura»; si rammenti (cfr. § 2.13) l'avvertenza [De Finetti e Emanuelli 1967, p. 89], contro interpretazioni superstiziose o stravaganti: sarebbe meglio, per non dar adito ad equivoci, evitare di parlare di «Natura», e comunque, semmai, spiegare che per Natura s'intende tutto, anche fatti causati da altri individui, però non supponendo che agiscano espressamente per avvantaggiarsi a nostro danno, come nello schema dei giochi.

Il *decision maker*, che supponiamo sia un EMVer, se vuole o deve decidere subito (senza possibilità di acquisire, prima, opportune informazioni), non potrà che stimare le probabilità $\mathbf{P}(C_i)$ delle diverse ipotesi (colonne) C_i, in base ad esse calcolare la previsione di guadagno di ogni decisione, $\mathbf{P}(R_j) = \mathbf{P}(C_1)\,G_{1j} + \mathbf{P}(C_2)\,G_{2j} + \ldots + \mathbf{P}(C_n)\,G_{nj}$ (media aritmetica dei numeri nella riga j-esima ponderata coi pesi $\mathbf{P}(C_i)$, probabilità delle diverse «ipotesi»).

Beninteso, qui si suppone siano note le circostanze che distinguono le diverse «ipotesi» C_i: circostanze di fatto riguardo alle cui probabilità avrà senso cercare di farsi un'idea. (È questa per l'appunto la differenza radicale dal caso dei giochi, non sempre sufficientemente compresa e sottolineata da parte degli oggettivisti. E non senza spiegazione: se di probabilità si vuol parlare solo con riferimento a percentuali di palline bianche e nere o a frequenze osservate, le precedenti considerazioni non varrebbero che in quegli esempi puerili).

E diciamo subito che, in questo primo semplice caso, nulla cambierebbe, come schema di calcolo, se ci si volesse riferire a valutazioni in termini di utilità anziché monetarie (cioè a un non-EMVer). Basterebbe indicare, nella tabella, i valori delle utilità in luogo dei guadagni monetari. Vedremo infatti che l'utilità è additiva rispetto alle *misture* («strategie miste» o «randomizzate»: vi si è accennato nel § 2.12), mentre non lo è se si tratta di sommare dei guadagni: la chiave per dissolvere i dubbi sta tutta qui.

5.3. Esempi con tabelle esplicative.

Le tabelle 4 e 5 hanno lo scopo di illustrare su esempi numerici come funziona lo sviluppo dei calcoli richiesti per confrontare la preferibilità tra varie decisioni.

In entrambe le tabelle, il riquadro in alto a sinistra contiene i dati del problema: le colonne rappresentano le decisioni possibili (quattro nella 4, tre nella 5: D_1, D_2, D_3, D_4), e le righe i tre eventi possibili (E_1, E_2, E_3) di cui uno e uno solo può verificarsi. All'incrocio della colonna D_i con la riga E_j è indicato il guadagno S_{ij} se la decisione scelta era la D_i e l'evento avveratosi è E_j. (Ad esempio, se si sceglie la decisione D_1, il guadagno risulta o $+8$, o 0, o -3 a seconda dell'evento che si verifica; per la prima riga ciò vale sia per la tabella 4 che per la 5).

Le successive elaborazioni sono diverse per le due tabelle.

Nella prima viene messo in luce come varia la preferibilità tra le diverse decisioni a seconda del modo in cui persone di mentalità diversa, o in possesso di informazioni diverse, o comunque sia, valutano diversamente le probabilità dei vari eventi. Sono considerate sei ipotetiche valutazioni di probabilità, e per ciascuna è calcolato il valore di ogni decisione (e la piú favorevole, il valore massimo della riga, è indicata in *corsivo*).

Si noti come la scelta della decisione migliore vari a seconda della valutazione di probabilità, ossia a seconda delle opinioni dell'individuo. Due individui che scelgono diversamente (in quanto hanno diverse opinioni) sono entrambi

Tabella 4.

Ricerca della decisione preferibile, in dipendenza della valutazione delle probabilità di eventi rilevanti al riguardo. (Per spiegazioni piú sostanziali cfr. § 5.3).

	Decisioni				*Valutazioni di probabilità (%)*					
Eventi	D_1	D_2	D_3	D_4	*1*	*2*	*3*	*4*	*5*	*6*
E_1	8	4	1	5	70	30	10	80	40	20
E_2	0	0	1	−1	10	20	85	15	40	20
E_3	−3	9	1	2	20	50	5	5	20	60
1	*5,00*	4,60	1,00	3,80						
2	0,90	*5,70*	1,00	2,30						
3	0,65	0,85	*1,00*	−0,25						
4	*6,25*	3,65	1,00	3,95						
5	2,60	*3,40*	1,00	2,00						
6	−0,20	*6,20*	1,00	2,00						

Tabella 5.

Ricerca della decisione preferibile, in dipendenza della valutazione delle probabilità di eventi rilevanti al riguardo, con la complicazione che è necessario considerare diverse ipotesi (circostanze incerte) che hanno rilevanza per la decisione.

	Decisioni			*Valutazioni di probabilità (%)*							
					subordinate alle informazioni						
Eventi	D_1	D_2	D_3	*non subordinate*	H_1' *(25%)*	H_2' *(25%)*	H_3' *(50%)*	H_1'' *(50%)*	H_2'' *(50%)*	H_1''' *(60%)*	H_2''' *(40%)*
E_1	8	4	1	30	15	10	80	30	30	43½	10
E_2	0	0	1	35	20	85	15	50	20	1½	85
E_3	−3	0	1	35	65	5	5	20	50	55	5
Non sub.	1,35	*4,35*	1,00	*4,35*							
H_1'	−0,75	*6,45*	1,00		25%	25%	50%				
H_2'	0,75	0,85	*1,00*		di +	di +	di				
H_3'	*6,25*	3,65	1,00		6,45	1,00	6,25				
						= *4,98*					
H_1''	1,80	*3,00*	1,00					50%	50%		
H_2''	0,90	*5,70*	1,00					di +	di		
								3,00	5,70		
								= *4,35*			
H_1'''	1,826	*6,683*	1,00							60%	40%
H_2'''	0,65	0,85	*1,00*							di +	di
										6,683	1,00
										= *4,41*	

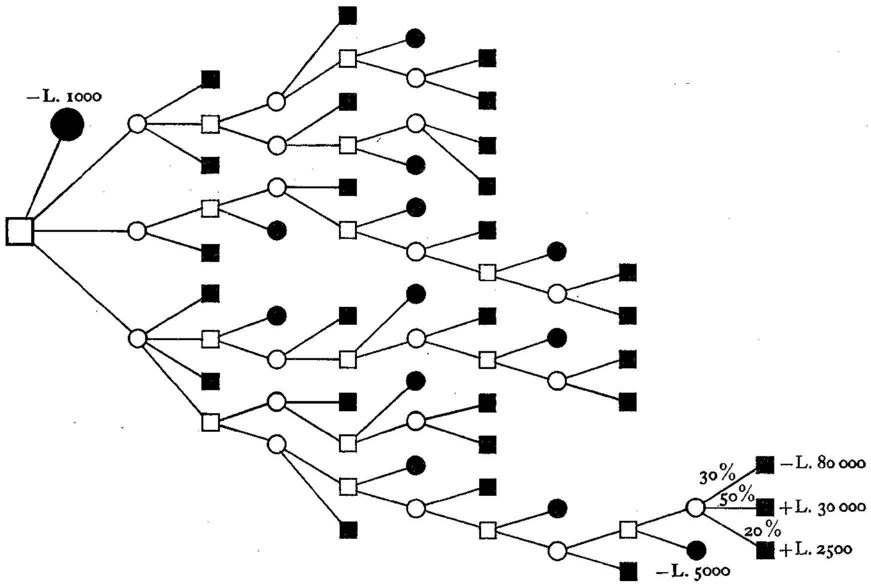


Figura 16.

Un esempio di problema di decisione. Nella situazione iniziale (quadretto a sinistra) il giocatore può scegliere tra quattro alternative: rinunziare al gioco (tratto verso l'alto che porta alla penalizzazione di L. 1000), o parteciparvi secondo una qualunque delle tre altre vie che conducono a situazioni d'incertezza (sorteggio fra due o piú vie su cui proseguire).

Tali situazioni, in cui il proseguimento del cammino dipenda non piú dal giocatore ma da un sorteggio (o, comunque, da eventi casuali, o, se si vuol usare tale dicitura, dalla «Natura»), vengono indicate con cerchietti, e si trovano ai passi dispari: 1°, 3°, ecc.; in particolare, un cerchietto può anche essere un punto terminale del gioco e allora è pieno con l'indicazione dell'importo che spetta al giocatore. (Nella figura, per evitare complicazioni superflue, tale indicazione dell'importo – come quella, di cui tosto parleremo, delle probabilità – è stata limitata a pochi casi, a titolo esemplificativo).

Anche il sorteggio può dar luogo a un numero qualunque di proseguimenti (secondo lo schema; nella figura il numero è basso per non complicarla, ma si può pensare senza difficoltà a casi con scelte assai piú numerose); ogni scelta porta a un quadratino, che a sua volta può dare un risultato certo, e concludere il processo con un pagamento finale (positivo o negativo), e allora è pieno; oppure offre di nuovo una scelta al giocatore tra varie vie. Tra esse, una può essere un risultato certo: cerchietto pieno; pensarne piú di una sarebbe vano, perché, trattandosi di pagamenti certi, è ovvio che nessuno ne sceglierebbe altri che il piú favorevole. Analoghe restrizioni valgono anche per pagamenti incerti: se (riferendosi all'unica scelta del quinto ordine: il quadratino da cui si dipartono le ultime diramazioni a destra complete di indicazioni di valori e probabilità) l'importo certo (qui: perdita di L. 5000) fosse invece un guadagno di L. 30 000 o piú, oppure una perdita di L. 80 000 o piú, la scelta non comporterebbe dubbi.

coerenti: le valutazioni di probabilità 1ª e 4ª rendono preferibile la decisione D_1, la 2ª, 5ª e 6ª la decisione D_2, la 3ª la D_3, nessuna la D_4.

Nella seconda, le probabilità dei tre eventi sono fisse, ma si considerano tre distinte possibilità di chiedere informazioni: la prima con tre alternative (H_1', H_2', H_3'), la seconda e la terza con due (H_1'' e H_2'', risp. H_1''' e H_2'''). E si vede come variano i risultati delle tre decisioni, D_1, D_2 e D_3, non subordinate ad informazioni ulteriori, oppure avendo l'informazione circa quale delle H' è vera, oppure l'analogo per le H'' o per le H'''. Il conteggio nei riquadri in diagonale (verso il basso a destra) indica la previsione del guadagno al momento di chiedere l'informazione sugli H' ma prima di averla avuta.

Ad esempio, col chiedere l'informazione H' (cioè di sapere quale delle tre ipotesi H_1' o H_2' o H_3' è vera) egli potrà fare la scelta piú appropriata per ciascun caso, assicurandosi 6,45 nell'ipotesi H_1' (di probabilità 25 per cento), risp. 1,00 (probabilità 25 per cento) o 6,25 (probabilità 50 per cento) per H_2' e H_3'; in previsione riceve 4,98, che, come si voleva dimostrare, è maggiore di 4,35 ottenibile senza informazione.

5.4. Decisioni sequenziali.

Spesso le decisioni sono «sequenziali», o «a piú stadi». Ciò è, in un certo senso teorico, inessenziale: nel senso, precisamente, in cui si è detto – nel § 2.11 – che una partita in qualunque gioco complicato (tipo dama, o scacchi, ecc.) comportante numerose mosse potrebbe ridursi ad un'unica «mossa» (da parte di ciascuno) consistente nella scelta di un atlante in cui sia indicata la mossa prefissata per ogni possibile situazione della scacchiera. Un prontuario direbbe subito il risultato della contrapposizione di due tali «atlanti». Ciò vale anche se l'incertezza dipende non da scelte dell'avversario ma – per usare il termine abituale anche se non troppo felice – dalla «Natura».

I diversi sviluppi possibili di tale processo di decisione si diramano a forma di albero (fig. 16). Dal punto iniziale si dipartono due o piú vie, fra cui il giocatore (chiamiamolo cosí) deve scegliere. Ciascuna porta ad un punto del secondo stadio di decisione, che spetta al «caso» (o «Natura») e si può immaginare realizzata mediante un'estrazione o roulette o simile aggeggio. La terza mossa (come la prima, e tutte quelle di posto dispari anche in seguito) spetta al giocatore; la quarta, e le altre pari, al «caso» (o «Natura»). Il numero delle mosse può essere fissato in un numero pari qualunque (un'ultima mossa da parte del giocatore sarebbe superflua, perché ovviamente sceglierebbe il valore piú alto). Al punto terminale di ogni cammino si trova indicato il valore del guadagno (positivo o negativo) per il giocatore.

Nella figura sono indicati (seguendo la convenzione usata da Raiffa [1968]) con quadratini i punti di decisione del giocatore (dai quali escono le vie tra cui può scegliere) e con cerchietti i punti di decisione (sorteggio) della «Natura» (dai quali escono le vie tra cui si effettua la scelta a sorteggio, per ciascuna delle quali è nota – e indicata – la probabilità).

In questa forma astratta e arbitraria, lo schema invita solo a pensare alla continua interferenza di previsioni di rischi o di opportunità sulle nostre decisioni e dei fatti effettivi sul risultato definitivo. In situazioni effettive, con minor numero di passi ma piú diretta visione e significatività delle ipotesi, tale continuo intersecarsi di scelte meditate e di fatti casuali può ben rappresentare un comportamento che cerchi continuamente di adeguarsi nel miglior modo ai fatti via via accertati e alle prospettive del prossimo futuro.

In tal senso questo cenno a decisioni sequenziali può dare un senso piú concreto alla precedente spiegazione generica e astratta.

5.5. Valutazioni in termini di utilità.

Il presente passaggio – dalle valutazioni di convenienza fatte in termini monetari a quelle fatte in termini di utilità – è certamente (almeno a mio avviso) il punto piú delicato e piú meritevole di attenzione fra quanti abbiamo avuto occasione di incontrare parlando di decisioni. Ma non solo per la sua rilevanza in questo contesto, bensí per quella che ha (e piú dovrebbe avere) in tutta l'economia.

Seguendone un po' le vicende storiche (ad esempio nell'attenta e densa esposizione di Georgescu-Roegen [1968]), appare chiara la molteplicità di aspetti e di interpretazioni, attraverso il tempo (da Galiani, 1750, ai nostri giorni), attraverso le diverse mentalità e preoccupazioni di specialisti nei campi piú disparati (filosofi, matematici, economisti, e anche ingegneri come l'italiano Antonelli e il francese Dupuit), nonché riflettendo sulle precise osservazioni dell'autore, verrebbe quasi paura a doverne parlare, sia pure limitatamente al nostro tema: le decisioni in condizioni d'incertezza.

Tale limitazione ci dispensa comunque dal toccare gli aspetti piú sfuggevoli dell'argomento (quali si presentano riferendosi specificamente al possesso o consumo di beni ecc. in date quantità e circostanze), ammettendo soltanto che, in un certo senso (e secondo un qualche criterio), ad ogni data situazione possa attribuirsi una valutazione in termini monetari. Quindi anche ogni cambiamento di situazione (in particolare quelli dovuti a risultati aleatori di scommesse, operazioni rischiose, ecc.) si traduce in un aumento o in una diminuzione di tale valutazione.

Per semplicità non abbiamo detto (riferendoci alla situazione di partenza) che anch'essa conterrà elementi di rischio (che esistono inevitabilmente, sempre): per semplicità li pensiamo ora inesistenti per parlarne piú espressamente con riferimento a incertezze e rischi inerenti a nuove decisioni dal risultato aleatorio.

Indichiamo con S la valutazione della situazione iniziale (in termini monetari), e supponiamo che, in base a variazioni certe ed incerte, la situazione in un tempo successivo possa assumere uno dei valori S_1, S_2, ..., S_n, con probabilità risp. p_1, p_2, ..., p_n; quale valore dobbiamo attribuire a questa «situazione contenente incertezza»? La risposta piú «normale» sarebbe quella data dalla media aritmetica (o «previsione», o «speranza matematica»), cioè $S=p_1S_1+p_2S_2+\ldots+p_nS_n$; ma è ben noto che, per l'avversione al rischio, possedere un milione soggetto ad una decisione a «lascia o raddoppia» (2 milioni o zero, con probabilità 1/2 e 1/2) vale meno che possederlo senza rischi. È ciò che si è già osservato, e indicato come base del criterio neobernoulliano.

La conclusione di queste (pur inevitabilmente lacunose) premesse si riduce proprio a ribadire e precisare la necessità di un'impostazione che tenga conto della «avversione al rischio», della «curvatura» dell'utilità.

Ciò significa che, in luogo della S (data, come sopra indicato, dalla media aritmetica delle S_i, coi pesi p_i), occorrerà calcolare un valore piú basso: sarà ancora una media delle S_i coi pesi p_i, ma non la media aritmetica bensí una media associativa che dia valori piú bassi.

Formalmente, le medie associative altro non sono che le «trasformate» della media aritmetica, ottenibili cioè operando la media aritmetica su di una «scala» alterata. Ad esempio, la media quadratica, la media armonica, la media geometrica, si ottengono facendo la media aritmetica, rispettivamente, sui quadrati o sui reciproci o sui logaritmi dei dati anziché sui dati stessi, e ritornando quindi alla scala di partenza (rispettivamente, prendendo la radice o il reciproco o l'esponenziale).

Quest'ultimo caso, della media geometrica, corrisponde proprio alla prima formula proposta per l'utilità già da Daniele Bernoulli, come logaritmo della ricchezza in termini monetari: $u=\log x$ (il che significa giudicare eque scommesse consistenti nel «raddoppia o dimezza», mentre «lascia o raddoppia» sarebbe equo per $u=x$, identificando cioè utilità con valore monetario). Non bisogna però pensare alla nozione di utilità come a qualcosa di oggettivo, per cui si ponga la questione di quale sia (o dovrebbe essere) la «vera» $u=f(x)$: non si tratta di verità o di norme ma di atteggiamenti, diversi non solo da individuo a individuo ma da momento a momento e da circostanza a circostanza; ne sono un esempio le interpretazioni date per $u=\log x$ e per $u=x$, e le diverse propensioni ad azzardare che uno può avere non soltanto o sempre o mai, ma solo negli affari, o solo nel gioco, o solo in particolari stati d'animo.

E veniamo finalmente a indicare il modo di costruire la funzione di utilità di un dato individuo, in base a risposte circa sue preferenze. Supponiamo egli abbia la tendenza (che ben possiamo dire «normale») all'avversione al rischio; comunque il procedimento è sempre valido, e non farebbe differenza supporre che la risposta fosse in tal senso o in quello opposto.

Consideriamo due livelli di ricchezza, x_0 ed x_1 ($x_1>x_0$), e chiediamo all'individuo di cui vogliamo esplorare la curva dell'utilità di indicare quale valore $x_{1/2}$ (evidentemente, compreso tra x_0 ed x_1) considera equivalente al diritto ad avere o x_0 o x_1 con probabilità uguali (per esempio a testa o croce). Ciò significa, ovviamente, che, per lui, $u(x_{1/2})=(1/2)[u(x_0)+u(x_1)]$; abbiamo cosí diviso l'intervallo inizialmente scelto in due parti di uguale incremento di utilità. In modo analogo possiamo costruire intervalli di uguale incremento di utilità esterni, proseguendo cioè la suddivisione con punti x_2, x_3, x_4, ecc. presi in modo che $u(x_1)-u(x_0)=u(x_2)-u(x_1)=u(x_1)-u(x_0)$, ecc. (ossia in modo che per ogni x_i sia $u(x_i)=1/2[u(x_{i-1})+u(x_{i+1})]$).

Probabilmente la figura 17 rende piú facile capire il procedimento: le u crescono di tratti uguali quando le x crescono di tratti (sempre piú lunghi, procedendo

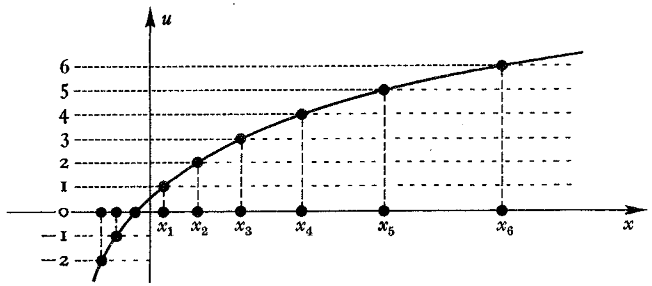


Figura 17.
Curva dell'utilità (per un individuo che attribuisce «utilità» decrescente a successivi incrementi uguali di ricchezza in termini monetari).

da sinistra verso destra!) cui corrispondono uguali incrementi di utilità come dimostra l'indifferenza tra la scelta di un punto divisorio (per esempio x_1) o la scelta «a caso» di uno dei due adiacenti di uguale dislivello (qui, x_0 e x_2).

5.6. L'«utilità» nella pratica: gli EUVer.

I non-EMVer, e cioè coloro che, nel decidere, si basano su confronti di utilità e non di valore monetario, li chiameremo (seguendo Grayson) EUVer, da EUV abbreviazione di «expected utility value» (con locuzione perfettamente simmetrica ad EMV per «monetary»). Ma abbiamo informazioni sulla psicologia degli EUVer, sul modo in cui essi effettivamente affrontano i rischi, amano e temono i rischi?

Interessanti tentativi in questo senso sono stati fatti, per l'appunto, da Grayson, interrogando imprenditori e specialisti nel campo delle ricerche petrolifere, e i risultati sono esposti da Grayson [1960] come parte di un completo studio dei molti problemi afferenti tutte le decisioni in quel campo: ad esempio, convenienza o meno della spesa per una prospezione sismica prima di decidere, e poi assumere in proprio tutto il rischio o dividerlo con altri e in che modo e misura, ecc.

Nelle figure riportate in quel libro si vede una curva che concorda con l'andamento logaritmico: si potrebbe indicarne l'equazione con $u=a\cdot\log(1+x/a)$ (dove a è la «massima perdita ammissibile», nel senso che per essa l'utilità va a «meno infinito»; dalla figura risulterebbe $a=200\,000$ dollari); è la curva che rappresenta le propensioni di uno dei proprietari di un'impresa. Le altre due, di un altro proprietario (e fratello) e di un geologo, sono molto meno spiegabili (e lo stesso può dirsi in genere di quelle riprodotte in altre figure del libro). Probabilmente certi cambiamenti bruschi in alcuni punti riflettono situazioni casuali e momentanee (come limiti oltre i quali una perdita renderebbe necessario il ricorso a misure particolarmente sgradevoli oltre che onerose). Anche tali riflessioni sono però istruttive, perché è bene non dimenticare che «ci sono molte

piú cose al mondo di quante ne contenga» non solo la filosofia (come bene ha detto Shakespeare) ma ogni teorizzazione, per quanto profonda e accurata, della realtà.

Comunque, si può forse concludere, paradossalmente, dicendo che, pur non esistendo forse nessun uomo vero che somigli all'uomo medio, la curva dell'utilità bernoulliana risponde abbastanza bene al comportamento dell'astrazione che chiamiamo «uomo medio». Ciò sarebbe anche conforme al concetto informatore della legge di Fechner, secondo la quale le differenze percettibili sono quelle che superano una certa misura in proporzione, non come differenza. (Come è stato confermato, nel modo piú evidente, constatando che proprio su tale circostanza – naturalmente ignota agli astronomi dei secoli precedenti – si basa la classificazione da essi fatta delle stelle secondo «grandezza»).

5.7. La non-additività dell'utilità.

Le precauzioni occorrenti (come abbiamo preannunziato nel § 5.2) per passare, nella trattazione di problemi in condizioni d'incertezza, dal caso di valutazioni in termini monetari a quello di valutazioni in termini di utilità, ossia dai criteri adatti agli EMVer a quelli occorrenti per gli EUVer, dipendono dal fatto che le utilità non sono additive.

Precisiamo meglio. Tutto dipende dal fatto che l'utilità di un guadagno certo, x (positivo o negativo), non dipende solo da x, bensí anche dalla «ricchezza» precedente (sia x_1): l'utilità di un incremento x, dato che si parte da x_1, è infatti $u(x|x_1)=u(x_1+x)-(ux_1)$, ossia l'incremento di utilità, $u(x_2)-u(x_1)$ detto x_2 lo stato di arrivo ($x_2=x_1+x$). Da quanto detto, è chiaro che tale incremento diviene sempre minore al crescere della ricchezza di partenza (per la concavità della funzione u). Se consideriamo due incrementi, x'e x'', non possiamo calcolarne separatamente le utilità (come se l'altro non ci fosse) e sommarli, perché otterremmo $[u(x_1+x)-u(x_1)]+[u(x_1+x'')-u(x_1)]=u(x_1+x')+u(x_1+x'')-2u(x_1)$ anziché $u(x_1+x'+x'')-u(x_1)$; a parte le formule, la differenza è questa: l'aggiunta del secondo incremento va valutata tenendo conto del primo: cioè, si parte non ancora da x_1 bensí da x_1+x', e l'incremento di utilità dovuto ad x'' risulta minore. (Però, si badi, l'ordine è indifferente: si poteva prima sommare x'' e poi x' partendo però, naturalmente, da x_1+x''). L'importante è vedere l'incremento di u tra x_1 e $x_1+x'+x''$ (in un sol passo, o con tappa in x_1+x' o in x_1+x'', non cambia nulla).

La cosa è un po' complicata (benché piuttosto in modo apparente che in realtà: piú che nel «non capire» la difficoltà sta nel dovere «non confondersi»).

Esistono però due modi infallibili per trarsi d'impaccio (il primo esatto, il secondo approssimato). Il metodo esatto consiste nel fare tutti i calcoli coi valori monetari, riferendosi sempre all'intera «ricchezza» (cioè: x_1=«ricchezza» iniziale e x_2=«ricchezza» finale); avremo naturalmente $x_2=x_1+x'+x''+\ldots$ (eventuali altri incrementi (e/o decrementi) nel periodo considerato), $u(x_2)-u(x_1)$ sarà l'utilità complessiva degli incrementi (o decrementi). Non avrà però senso chiedere l'incremento di utilità dovuto ad ogni singola operazione a meno di non fissare un ordine (per esempio cronologico): cambiando l'ordine, il totale non cambia, ma i singoli addendi sí.

Il metodo approssimato potrebbe consistere nel ridurre proporzionalmente tutti gli incrementi x', x'', ... (positivi o negativi) del periodo considerato (ad esempio, anno), per ottenere i corrispondenti incrementi di utilità, kx', kx'', ..., con k=rapporto (incremento in utilità / incremento in valore monetario) per l'intera «ricchezza». Ciò è tanto piú prossimo all'esattezza quanto piú la «ricchezza» complessiva è grande rispetto al movimento: in tal caso infatti il rapporto tra utilità e valore monetario rimane praticamente costante, e circa uguale alla derivata $u'(x)$ $(=du(x)/dx)$ in $x=x_1$ (e in tutto $x_1<x<x_2$).

Rammentiamo, per contrasto, che l'utilità è additiva riguardo alle misture: esprimendo in modo piú completo quanto accennato nel § 5.2, diciamo che la relazione ivi scritta $\mathbf{P}(\mathrm{R})=\mathbf{P}(\mathrm{C}_1)\mathrm{G}_1+...+\mathbf{P}(\mathrm{C}_n)\mathrm{G}_n$ vale anche per le utilità interpretando sempre le $\mathbf{P}(\mathrm{C}_i)$ come probabilità di n casi di una partizione, le G_i come numeri aleatori, ed R come la G_j che corrisponde all'evento (o «ipotesi») che si realizza.

5.8. Un esempio semplice: il «problema del giornalaio».

Gli esempi semplici sono indubbiamente i piú istruttivi (beninteso, purché non siano banali). Particolarmente semplice e istruttivo è il cosiddetto «problema del giornalaio», perché porta in modo naturale a vedere l'importanza di certi concetti e di saperli connettere.

Che si tratti di giornali è inessenziale: si tratta del rischio di chi compra e vende merce deperibile, che domani è da buttare. Il giornale è l'esempio tipico (supponendo, però, che non esista, come in Italia, la «resa», e il prezzo d'acquisto delle copie invendute costituisca quindi una perdita per il giornalaio).

Il problema è: quante copie gli conviene acquistare?

Dipenderà, naturalmente, dalla previsione riguardante le vendite, ma non in modo banale (come pensando a una qualunque «media», ad esempio, per pigrizia o abitudine, la media aritmetica, basata sull'esperienza dei giorni o mesi precedenti). Intanto, il problema ha carattere economico, e quindi bisogna tener conto del prezzo (d'acquisto a, e di vendita, v); ovviamente, quanto maggiore è il margine di guadagno, tanto piú conviene acquistare un maggior numero di copie pur col rischio che restino invendute.

Ebbene: basta esplicitare questo discorsetto ovvio per dare subito la risposta esatta. (Naturalmente, si potrebbe fare una enorme tabella con colonne per le decisioni: D_0, D_1, ..., D_{100} (o fino a 200, 1000, ...) e altrettante righe per gli eventi: E_1, E_2, ..., E_{100} (ed eventualmente ecc.), indicando per ogni E_h la probabilità che gli attribuiamo, e riportando il guadagno in ogni casella (D_n, E_m) (copie acquistate n, richieste m; quindi vendute m se $m<n$ e altrimenti tutte n). Il guadagno è $mv-na$, col massimo di $nv-na$ se $m>n$; d'ora in poi supporremo $m\leq n$, ossia trascureremo le richieste che non potevano esser soddisfatte). Come si fa a trovare il bandolo della matassa, arrivando cosí alla conclusione in modo elegante e immediato?

Dimentichiamo tutto, tranne l'ovvia osservazione iniziale.

Basta una domanda semplice: fino a quando c'è vantaggio a comperare una copia in piú, $n+1$ anziché n, ad esempio 101 anziché 100? L'acquisto costa a, la vendita ci dà v con probabilità $\mathbf{P}_{101}$, se indichiamo con $\mathbf{P}_m$ la probabilità che il numero delle richieste sia uguale o maggiore di m; esso è anche, in particolare, la probabilità di vendere la m-esima (nell'esempio, la 101-esima) copia. Comperare la 101-esima copia ci dà quindi un guadagno sperato $v\mathbf{P}_m-a$ (qui per $m=$ $=101$, ma ovviamente vale in generale), e quindi il suo acquisto è vantaggioso (il guadagno sperato è positivo) se $\mathbf{P}_m>a/v$. Ad esempio, $a/v=2/3=67$ per cento se il prezzo di vendita è quello di acquisto maggiorato del 50 per cento; è $1/2=50$ per cento per maggiorazione 100 per cento; è $1/3=33$ per cento per maggiorazione 200 per cento, ecc.; e ciò vuol dire che, nei tre casi citati ad esempio, il numero di copie da acquistare è quello per cui la probabilità di venderle tutte diviene risp. 67, 50, 33 per cento (e cosí per ogni possibile rapporto a/v).

Se si pensa alla distribuzione di probabilità (per il numero di copie richieste), quelli indicati sono dei valori di posizione o quantili: valori che dividono la distribuzione in parti preassegnate (mediana, a metà; analogamente quartili, decili, ecc.).

A parte il risultato dell'esempio particolare, va meditato il fatto che esiste una risposta univoca strettamente legata al significato economico, e che è lontana dal tipo di risposta che verrebbe in mente a prima vista. Non si tratta di una media aritmetica, e neppure di una media, bensí di un valore di posizione, e neppure la mediana che sembrerebbe l'unica tra esse degna di esser presa in considerazione, ma una «qualunque», però richiesta dal problema tenendo conto di tutti i dati rilevanti, non lasciando nulla al capriccio o a «mode».

E sfruttiamo ancora lo stesso esempio per parlare del «valore di un'informazione»; sia l'informazione che «il giornale di domani porterà un articolo di particolare interesse». In base ad essa il giornalaio valuterà ad m' anziché ad m il numero di copie che gli dà probabilità a/v di venderle tutte, e ne comprerà m'. Il suo maggior guadagno sarà $(m'-m)(v-a)$ se effettivamente le venderà tutte, o minore se la previsione si avvererà solo in parte; avrà una perdita se ne vende meno di $m'a/v$. Questi dati e queste considerazioni non bastano però per rispondere riguardo al «valore dell'informazione»: esso dipende dalla previsione del numero di copie vendute in piú, e vale cioè (indicando con sopralineature le previsioni): $(\bar{m}'-\bar{m})v-(m'-m)a$. (Infatti: compera $m'-m$ copie in piú al prezzo a, aumenta la previsione del numero di copie vendute da $\bar{m}$ ad $\bar{m}'$, con prezzo v). Tali previsioni si potrebbero ottenere in base alla conoscenza di tutta la distribuzione (per entrambi i casi), e non per solo qualche valore (come i tre rapporti a/v dell'esempio). Tanto per dare una risposta (sia pure con un'ipotesi grossolana), supponiamo che in entrambi i casi (cioè per la previsione m e la m', risp. per giornate normali e la circostanza speciale) la previsione sia dell'80 per cento del numero ordinato, m risp. m'. Allora la previsione di guadagno in base all'informazione, ossia il «valore dell'informazione», sarebbe l'80 per cento di $(m'-m)(v-a)$.

5.9. Cenni su qualche altro esempio.

Su qualche altro esempio basteranno dei cenni, intesi a dare un'idea della varietà di problemi e di ragionamenti, pur senza entrare in questioni e dettagli di natura piú tecnica e matematica.

Riprendiamo il problema delle scorte (cfr. § 3.2) tenendo conto che il tempo in cui si esaurisce la quantità Q acquistata ad ogni rifornimento non si esaurirà esattamente nel tempo T teoricamente previsto (in base a $Q=qT$, q consumo per unità di tempo), ma o prima o dopo. Il sistema piú noto e naturale per tener conto di ciò sta nel provvedere all'ordinazione di un nuovo rifornimento (sempre del quantitativo Q) non col dovuto anticipo t rispetto alla fine del periodo T in corso, bensí quando la scorta abbia raggiunto quel livello minimo Q_0 che si reputa necessario per la pratica garanzia che il rifornimento arrivi prima che esso si esaurisca del tutto. Comunque, si verifica sempre, piú o meno sensibilmente, uno dei due inconvenienti: o la scorta si esaurisce prima e c'è un periodo di mancanza di scorta (cfr. fig. 18), o rimane una scorta all'arrivo della nuova ordinazione e per un certo periodo è superata la capacità normale del magazzino (cfr. fig. 19). Si dovrà scegliere il livello minimo Q_0 in modo che i due rischi, complessivamente, diano un costo minimo, ossia abbiano valori marginali uguali ed opposti.

Altro esempio importante e istruttivo è quello dei collaudi in forma sequenziale, consistenti cioè nel procedere sottoponendo alla prova un pezzo dopo l'altro, scegliendoli «a caso», arrestandosi e decidendo per l'accettazione o il rifiuto della partita quando l'informazione raggiunge il livello sufficiente per decidere. Il metodo è dovuto ad Abraham Wald [1947]; la sua naturale interpretazione in termini bayesiani è stata subito notata dai soggettivisti, come Savage; quanto allo stesso Wald, egli aveva apportato notevoli innovazioni nelle impostazioni oggettivistiche che in parte corrispondevano ad esigenze dei soggettivisti, senza

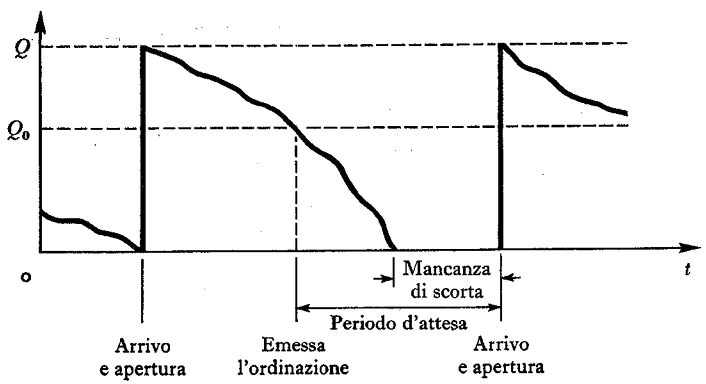


Figura 18.
Emissione ritardata dell'ordinazione = periodo di mancanza di scorta.

però portarle fino in fondo. A parte le avversioni allora prevalenti da affrontare, egli era però effettivamente ancora mentalmente assai lontano dal superare l'ultimo tratto di strada.

Circa il procedimento, basti guardare la figura 20. L'esito delle prove sui successivi pezzi estratti per il collaudo viene indicato, sullo schema, con successivi passi verso destra partendo da o, e verso l'alto o il basso a seconda che l'esito è favorevole o sfavorevole. Si prosegue finché il cammino esce dalla striscia tra le due parallele (che è la zona entro la quale permane il dubbio) entrando o nella zona superiore (e allora la partita va accettata) o in quella inferiore (e allora va rifiutata). Concettualmente, si tratta di aggiornare, in base ad estrazioni successive, le opinioni sulla composizione di un'urna come quella dell'esempio nel § 2.8; la differenza concettuale tra l'interpretazione soggettivista e oggettivista sta nel considerare anche l'opinione iniziale (ad esempio 1/3, 1/3, 1/3, oppure 3/10, 5/10, 2/10, come ivi supposto come esempi) o escluderla (e quindi basarsi soltanto sulle *likelihood*). Ciò corrisponde esattamente alla prima delle due valutazioni indicate come esempio di valutazioni soggettive (ma gli oggettivisti respingono tale necessario complemento preferendo usarla formalisticamente come *adhockery*, ma... immacolata).

5.10. Decisioni e interesse collettivo.

L'aspetto piú importante e piú arduo della teoria delle decisioni è quello riguardante l'interesse collettivo. Ogni decisione di ciascuno di noi ha conseguenze (piccole o grandi, dirette o indirette) su tutti gli altri; tanto piú ne hanno le decisioni prese da persone o da comitati o da istituzioni cui spetta deliberare nel modo che appare meglio rispondente all'interesse collettivo.

Non v'è dubbio che la nozione di «interesse collettivo» è assai vaga, ma ancor meno dubbio mi sembra che chi insiste nel sottolineare tale «vaghezza» appartenga al numero di coloro che ritengono di non dovere preoccuparsi se non

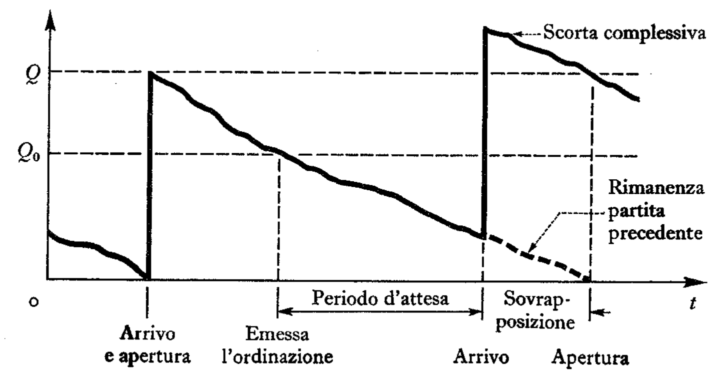


Figura 19.
Emissione prematura dell'ordinazione = periodo di eccesso di scorta.

dell'interesse proprio. Minore disuguaglianza e maggiore sicurezza per tutti, nessuno che debba soffrire l'inedia e nessuno che possa permettersi lo spreco, abolizione di tutte le idiozie burosadiche (come le chiamò Louis Armand) che avvelenano senza motivo e costrutto alcuno la vita di tutti, giorno per giorno. Chi può non essere d'accordo?

E, forse, nessuno dice di non essere d'accordo. Il guaio è che non basta «non *non* essere d'accordo» e neppure «essere d'accordo». Occorre una profonda convinzione dell'urgenza di porre rimedio ad una situazione sempre piú insostenibile, assurda, spaventosa. Se un grido di responsabile allarme c'è stato, fu per iniziativa illuminata di un singolo uomo coscienzioso e consapevole, Aurelio Peccei; a un congresso internazionale sul Futuro (Frascati, 1974) partecipò, attivamente, Senghor, presidente del Senegal (ma nessun capo di stato o di governo di altri paesi!)

Dovrebbe ormai apparire fin troppo evidente l'importanza delle economie e diseconomie esterne: dei danni irreparabili inflitti agli altri e alla collettività e alla natura mediante speculazioni o produzioni o decisioni economiche di qualunque tipo comportanti disastrose conseguenze di carattere ecologico, ambientale, estetico, sanitario; dalla cecità di chi vede solo il proprio guadagno immediato e non i disastri futuri, di chi considera lo spreco come diritto ed esibizione di opulenza anziché delitto ed esibizione di deficienza. Al riguardo, rileggere e rimeditare Veblen [1899]!

Naturalmente, si ripropongono a questo punto le difficoltà segnalate fin dall'inizio (§ 2.2) riguardo alle preferenze collettive, e soprattutto va deprecata l'insufficienza (per non dire assenza) di tutela dell'interesse pubblico da parte delle autorità «competenti (?)», troppo spesso inerti o deboli se non succubi e complici dei piú squallidi e illeciti e dannosi interessi privati. Se non una soluzione al problema di una migliore difesa della cosa pubblica (il che dovrebbe significare «nostra, di tutti noi», non «di nessuno»!), c'è tuttavia un suggerimento [Lindley 1971, p. 179] inteso a rendere piú chiare e impegnative le tesi e le intenzioni dei diversi partiti che si presentano alle elezioni. Si tratterebbe di chiedere l'abbandono di slogan e di fraseologie sibilline, presentando invece concreti propositi nella forma qui illustrata. Egli vorrebbe che «i partiti politici che si presentano alle elezioni abbandonino i loro slogan elettorali e rendano pubbliche, invece, le loro funzioni di utilità; è questo ciò di cui abbiamo realmente bisogno per una decisione coscientemente democratica; se noi sapessimo che il partito A attribuisce grande utilità alle spese militari e il partito B a quelle per il problema degli alloggi, una scelta sarebbe piú facile».

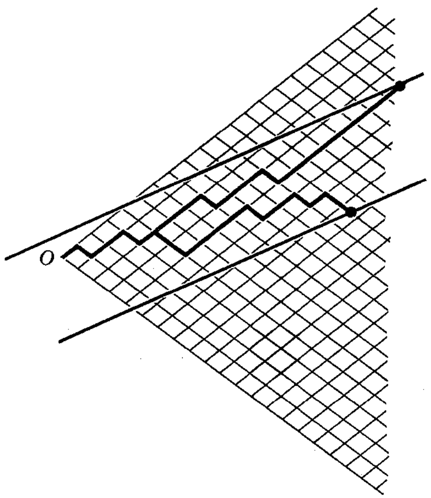


Figura 20.

Collaudo «sequenziale». Partendo dall'origine O, si muove un passo verso l'alto per ogni pezzo «buono», un passo verso il basso per ogni pezzo «difettoso»: quando viene raggiunta la retta inclinata superiore si ha *accettazione*, quando viene raggiunta la retta inclinata inferiore si ha *rifiuto*.

Dell'esigenza di «sicurezza» si può dire qualcosa di piú preciso riferendoci al grafico (fig. 20) a proposito dei collaudi in forma «sequenziale» (secondo Wald). Il medesimo schema (ma con la «fascia utile» orizzontale anziché inclinata) serve a illustrare il problema della «rovina dei giocatori». Abbiamo due giocatori, I e II, che all'inizio possiedono c' e c'' lire; partendo dall'origine O, seguiamo il grafico che dà il guadagno di I nel tempo, cioè man mano che si susseguono i colpi. Se la linea sta al di sopra dell'asse è in vantaggio il I e se sta sotto lo è il II; quando attraversa (o tocca) l'asse si è in parità; infine – e questa è la cosa che interessa – quando la linea tocca una delle due rette che delimitano la striscia vuol dire che uno dei due giocatori è rovinato e il gioco ha termine. Infatti, livello $-c'$ significa perdita dell'intero capitale iniziale c' da parte di I, ed analogamente per II se viene raggiunto per primo, invece, il livello c''.

La probabilità che il gioco continui eternamente (senza toccare mai, cioè, le rette-limite) è nulla; il gioco è equo; quindi le probabilità q' e q'' di rovina di I e di II, dovendo risultare $q'c'=q''c''$ $(q'+q''=1)$, sono $q'=c''/(c'+c'')$, $q''=c'/(c'+c'')$; in breve, a parole, la probabilità di rovina è proporzionale alla ricchezza dell'*altro* (e quella di vincere, alla propria), come era pressoché ovvio.

Il tempo o, piú precisamente, il numero di colpi per giungere alla rovina (indifferentemente di quale dei due giocatori) è dato, come valor medio o previsione, dal prodotto $c'c''$ (ad esempio, occorrono «in media» 1000 colpi per giungere alla rovina partendo da: $1+1000$, $2+500$, $4+250$, $5+200$, $8+125$, $10+100$, $20+50$, $25+40$: sembra strano soprattutto il 1° caso, dove c'è probabilità $1/2$ di farla finita con un sol colpo!)

Morale: chi gioca ed insiste finisce per rovinarsi. E gli sta bene!

5.11. Riflessioni conclusive.

La teoria delle decisioni mostra come e in che senso la probabilità sia, e piú debba essere, la nostra guida nel pensare e nell'agire [De Finetti 1965].

Deve esserlo per ciascuno di noi singolarmente, ed è già cosa difficile, ma anche per decisioni collegiali (riguardanti una famiglia, un'azienda, una collettività, l'intera umanità, addirittura la biosfera: questo sottile rivestimento d'aria che rende possibile la vita sul nostro infinitamente piccolo pianeta).

È solo da poco tempo che ci si comincia a render conto della globalità di tutti i problemi, forse proprio perché l'attuale fase del «progresso» li ha fatti improvvisamente aggravare, esplodere, diventare evidenti nonostante l'assuefazione induca a diventare ciechi. E mai come oggi sarebbe necessario inquadrare ogni decisione anche minima nella visione d'insieme delle ripercussioni ad ogni successiva scala.

Può, la teoria dei giochi, essere di aiuto?

Probabilmente, purtroppo, sembra abbia ragione piuttosto Anatol Rapoport, secondo il quale «il piú importante apporto della *game theory* sta nel fatto che l'analisi della *game theory* rivela la sua propria limitatezza»; che essa rivela come «decisioni basate sul calcolato interesse egoistico possono condurre al disastro». Ma, disgraziatamente, «questo aspetto negativo della teoria è di gran lunga meno compreso che non l'aspetto positivo». «Talvolta dobbiamo imparare il significato della fiducia, o altrimenti entrambi, noi e i nostri avversari, saremo invariabilmente perdenti»; «talvolta dobbiamo essere in grado di convincere l'altro di giocare secondo certe regole o addirittura di giocare un altro gioco. Per convincere l'altro dobbiamo far sí che egli ci ascolti, e a ciò non è abitualmente possibile riuscire se noi stessi non siamo disposti ad ascoltarlo». «Tutte queste accorte riflessioni si riferiscono non alla sapienza ma alla saggezza» grazie alla quale «molti dei conflitti che gli esperti di strategia, nel loro zelo professionale, insistono a formulare come battaglie [...] verrebbero risolti di comune accordo» [Rapoport 1962].

Era l'epoca della *Pacem in Terris*, imperitura testimonianza del miracolo per cui un cristiano era divenuto papa: papa Giovanni. E concordanti motivi di speranza venivano da Kennedy, da Kruščëv, da U Thant; ma poco dopo tutti questi personaggi scomparvero dalla scena che ripiombò nel buio.

Il «caso», la «fatalità», svolgono sempre un ruolo pauroso nel determinare il divario tra previsioni e fatti, e in particolare tra conseguenze volute o previste o «prevedibili» di una decisione e conseguenze effettive. Sull'estrema «imprevedibilità» che regna in tale campo ha particolarmente richiamato l'attenzione Bertrand de Jouvenel: ad esempio, «chi, all'entrata in guerra di Austria e Russia in schieramenti contrapposti, nel 1914, avrebbe potuto prevedere che entrambe ne sarebbero uscite sconfitte?»; e «chi, allo scoppio della rivoluzione francese, avrebbe immaginato che essa sarebbe sfociata, con Napoleone, nella piú grande impresa militare della storia?»

Proprio con riferimento (particolarmente) a tale impresa esiste un libro intitolato *La probabilité dans l'histoire*; non vi si trovano vere e proprie analisi probabilistiche: piú che sulle probabilità, l'autore insiste sul fattore *caso* che spesso ha avuto un ruolo decisivo (il che è ben naturale, non foss'altro perché, ad esempio, a quell'epoca, Napoleone e Nelson apprendevano solo saltuariamente in quale punto del Mediterraneo l'altro fosse stato visto molti giorni prima).

Risulta però chiaramente, da descrizioni particolareggiate, che Napoleone

calcolava esattamente le forze e la loro dislocazione, le mosse ed eventuali contromosse, lo sfruttamento del successo, applicando in modo esemplare la teoria delle decisioni (o dei «giochi») ante litteram.

Ma in quel libro c'è anche qualcosa di piú consolante e profondo: una dettagliata (e, temo, poco nota) descrizione della politica con cui Napoleone, occupato l'Egitto, ne avrebbe voluto fare lo stato piú civile del mondo. Combatté soltanto i despoti e i mammalucchi al loro servizio; promosse la collaborazione su piede di uguaglianza tra i vari gruppi etnico-religiosi, coi Francesi soltanto presenti in qualità di garanti di tale cooperazione (e molti particolari sono stupefacenti: ad esempio, come un grande uomo sappia valutare l'importanza di cose cui i grandi politicanti o politologi neppure si degnerebbero di prestare attenzione per un attimo!)

Perché non dovrebbe esser possibile – non a un uomo d'armi, ma a un'umanità liberatasi dalla soggezione a grandi meschini interessi di cui non può non essere rabbiosamente indignata, schifata e stufa – di impegnarsi totalmente nel proponimento di arrestare il «progresso» verso la barbarie e di realizzare ovunque, in un mondo rinsavito, il sogno di «*quell'*Egitto»?

Se persisteremo – sordi all'accorato messaggio di Rapoport – a lasciarci trascinare sulla via della stupidità, si avvererà puntualmente la sua tragica profezia: «Noi e i nostri avversari saremo invariabilmente perdenti».

E ci starà bene: lo avremo voluto, o non avremo avuto sufficiente coraggio per impedirlo. Ricordiamo il detto: «Après moi le déluge!» ed il fatto che «il arriva bientôt». [B. D. F.].

Arrow, K. J.

1951 *Social Choice and Individual Values*, Wiley, New York.

Black, D.

1948-49 *On the Rationale in the Group-Decision Making*, in «Journal of Political Economy» (trad. it. in «Giornale degli Economisti», 1948).

1958 *The Theory of Committees and Elections*, Cambridge University Press, London.

Cohen, J.

1960 *Chance, Skill and Luck*, Penguin Books, Baltimore (trad. it. Giunti-Barbera, Firenze 1964).

De Finetti, B.

1965 *La probabilità: guida nel pensare e nell'agire*, Quaderno n. 11, Istituto Superiore Scienze Sociali, Trento.

1967 *L'incertezza nell'economia*, in B. De Finetti e F. Emanuelli, *Economia delle assicurazioni*, parte I, Utet, Torino.

1973 (a cura di) *Requisiti per un sistema economico accettabile in relazione alle esigenze della collettività*, Angeli, Milano.

De Finetti, B., e Minisola, F.

1961 *La matematica per le applicazioni economiche*, Cremonese, Roma.

Edwards, W.

1968 «Decision Making. Psychological Aspects», in *International Encyclopedia of the Social Sciences*, Macmillan, New York, vol. IV, pp. 34-42.

Georgescu-Roegen, N.

1968 «Utility», in *International Encyclopedia of the Social Sciences*, Macmillan, New York, vol. XVI, pp. 236-67.

Grayson, C. J. jr
1960 *Decisions under Uncertainty: Drilling Decisions by Oil and Gas Operators*, Harvard Business School, Division of Research, Boston.

Lindley, D. V.
1971 *Making Decisions*, Wiley, New York (trad. it. in preparazione presso Einaudi).

Luce, R. D., e Raiffa, H.
1957 *Games and Decisions*, Wiley, New York.

Marschak, J.
1968 «Decision Making. Economic Aspects», in *International Encyclopedia of the Social Sciences*, Macmillan, New York, vol. IV, pp. 42-55.

McKinsey, J.
1952 *Introduction to the Theory of Games*, McGraw-Hill, New York.

Raiffa, H.
1968 *Decision Analysis. Introductory Lectures on Choices Under Uncertainty*, Addison-Wesley, Reading Mass.

Rapoport, A.
1962 *The Use and Misuse of Game Theory*, in «Scientific American», dicembre.

Robinson, J. A.
1968 «Decision Making. Political Aspects», in *International Encyclopedia of the Social Sciences*, Macmillan, New York, pp. 55-62.

Savage, L. J.
1972 *The Foundations of Statistics*, Dover, New York 1972[2].

Schlaifer, R.
1959 *Probability and Statistics for Business Decisions*, McGraw-Hill, New York.

Veblen, Th.
1899 *The Theory of the Leisure Class*, Macmillan, London (trad. it. Einaudi, Torino 1949; nuova ed. 1971).

Wald, A.
1947 *Sequential Analysis*, Wiley, New York.

Le decisioni sono presenti quasi in ogni situazione dell'agire umano, e sono dunque analizzabili secondo vari punti di vista, da quello del comportamento singolo, secondo i piú vari condizionamenti (cfr. **comportamento e condizionamento**) sottoposto ai piú svariati controlli (cfr. **controllo sociale**), ricercando eventualmente i piú riposti motivi psicologici (cfr. **inconscio, desiderio**), a quelli pertinenti all'**economia** e alla **politica**.

Una serie di problemi illustra i legami con altri concetti: i paradossi delle decisioni collettive, la necessità della **coerenza**, il tentativo di eliminare l'incertezza con le medie (cfr. **distribuzione**, **probabilità**) il ruolo dell'**induzione statistica** (cfr. anche **induzione/deduzione**) e della **rappresentazione statistica**.

Il problema è poi affrontato dal punto di vista normativo, ricercando il miglior criterio per raggiungere certi risultati. Si esamina il caso semplice delle decisioni in condizioni di certezza (cfr. **certezza/dubbio**) che spesso si riducono a problemi combinatori (cfr. **combinatoria**), quello in cui l'**informazione** è minore, in condizioni di incertezza ed infine, brevemente, quello dell'incertezza competitiva (cfr. **giochi**) ove cioè esista un competitore che può influire con le sue scelte sul risultato di ogni nostra scelta.

Distribuzione statistica

1. *Premesse illustrative.*

1.1. Distribuzione: in che senso?

Occorre subito rispondere a questa domanda, perché il termine 'distribuzione' ha molti significati diversi. E, forse, conviene subito dire – per prima cosa – che *non* si tratta della distribuzione nel senso che probabilmente verrà per primo in mente a tutti, e cioè quello economico, in entrambi i sensi: dei fatti e problemi riguardanti il modo in cui particolari beni e merci raggiungono i consumatori, o quello in cui si genera una piú o meno disuguale distribuzione della ricchezza, del reddito, del carico fiscale, ecc., o, infine, le conseguenze politiche e sociali di tutto ciò.

La presente trattazione riguarda il concetto di distribuzione nei due significati – affini ma distinti – di 'distribuzione di probabilità' e di 'distribuzione di frequenza' (detta anche, piú brevemente, 'distribuzione statistica'); come modello concreto sarà anche tenuto presente quello di una distribuzione di masse.

L'oggetto della distribuzione può essere di qualsiasi natura (anche economica, anche attinente agli aspetti statistici dei fatti economici or ora citati), ma ciò rientra nelle interpretazioni applicative, mentre questa presentazione preliminare non intende che predisporre nozioni e strumenti interessanti, sia di per sé, concettualmente, e sia per le applicazioni di qualunque tipo e in qualunque campo. Di qui l'importanza di porre attenzione – oltre e piú che agli aspetti formali e tecnici di concetti e procedimenti – al significato e valore che essi possono avere in relazione alle conclusioni pratiche che si vogliono raggiungere e alle decisioni che in base ad esse potremmo dover prendere. *Respice finem!*

Una distribuzione può sempre venir pensata come una distribuzione di masse (nel caso piú semplice e abituale, di una sola dimensione: su di una retta, come pesi infilzati in uno spiedo; pesi staccati gravanti su singoli punti di esso – distribuzioni discrete – oppure tali da costituire un carico diffuso, ove piú e ove meno intensamente – distribuzioni continue). Piú che un'utile analogia, questa intuitiva immagine concreta, a tenerla sempre presente alla mente, diviene un valido ausilio per rendere intuitivi, e quindi facilmente comprensibili e applicabili, certi concetti piú o meno tecnici e certi ragionamenti che ne dipendono. Beninteso, in un'esposizione come la presente (non riservata a fisici o ingegneri) tali accostamenti saranno mantenuti nei limiti modesti di nozioni meccaniche elementarissime (e cercando di fare in modo che basti afferrarne il «succo»).

Per entrare nell'argomento specifico, conviene cominciare dal caso delle distribuzioni statistiche, dove le frequenze hanno un significato concreto,

oggettivo; da esse è poi agevole il trapasso all'interpretazione probabilistica. (In particolare, del resto, ogni distribuzione statistica dà luogo a una distribuzione di probabilità se si pensa ad una «estrazione a caso» (cioè con probabilità uguali) di uno fra gli individui o elementi di essa. Per fare un esempio concreto, nel sorteggio di un membro di giuria popolare, la probabilità che il prescelto sia un ferrotranviere è uguale alla percentuale di ferrotranvieri nell'elenco dei sorteggiabili).

Uno dei piú semplici e comuni compiti della statistica è quello d'indicare come un certo insieme d'individui (o oggetti, o fenomeni, o non importa che altro) si suddivide a seconda di qualche carattere.

Può trattarsi di caratteri qualitativi (come, riferendosi a persone, il sesso, lo stato civile, la professione, il comune di nascita, quello di residenza, il titolo di studio, ecc.), e si hanno altrettante «classificazioni». Il caso che qui piú specificamente interessa, e su cui dobbiamo intrattenerci, è invece quello di classificazioni riguardanti caratteri quantitativi (come età, statura, reddito, ecc.) per le quali piú propriamente si usa la denominazione di 'distribuzioni'.

Ed è bene distinguere subito anche i casi di grandezze di natura discreta o continua, e, analogamente, i casi di distribuzioni discrete o continue. Sono discrete le grandezze che possono assumere solo i valori di una particolare scala, ad esempio solo valori interi (come numero di figli, o di fratelli, o di vittorie in un determinato torneo, ecc.; oppure secondo tariffe, ad esempio postali, per lettere e pacchi, con scatti per classi di peso, ecc.); sono invece grandezze continue quelle che possono assumere qualunque valore (per esempio la statura, il peso, la pressione arteriosa, ecc., per un individuo; la temperatura, pressione atmosferica, grado di umidità, ecc., in un dato istante e luogo; e via dicendo).

Quanto alle distribuzioni, esse sono discrete quando si riferiscono a una grandezza discreta tenendo distinto ciascuno dei singoli valori possibili (ad esempio, numero di figli 0 o, o 1, o 2, ecc. ecc.); sono continue quando la grandezza è continua (e si distinguono quelle appartenenti a diversi intervalli, uguali oppure no: ad esempio comuni con altitudine (in metri sul livello del mare) da 1 a 100, da 100 a 300, da 300 a 600, da 600 a 1000, da 1000 a 1500, ecc.), od anche quando è discreta, ma tale carattere si attenua perché le suddivisioni (come nel caso precedente) raggruppano intervalli grandi in cui il preciso numero di unità scompare. E spesso infatti lo si indica arrotondato, ad esempio numero di abitanti di una città in migliaia o decine di migliaia.

Queste considerazioni si riconducono a quelle già fatte nel caso analogo di masse concentrate (in singoli punti) o diffuse; e l'analogia andrà ripresa e precisata per illustrare il grado di significatività e d'idealizzazione (e, viceversa, d'inevitabilmente presente grossolanità d'idealizzazione) in entrambi i casi.

1.2. Istogrammi.

Per proseguire nel nostro discorso, in forma piú espressiva di quella consistente nell'impiego di solo parole, è opportuno riferirci alle rappresentazioni grafiche delle distribuzioni. Sono esse infatti che costituiscono lo strumento piú idoneo per mettere in evidenza la «visione» della situazione concreta di cui ci s'interessa, ed anche per chiarire ulteriormente, in modo piú concreto e meglio percepibile, le distinzioni premesse in modo piuttosto formalistico.

Ci riferiremo, in genere, alla consueta rappresentazione grafica, dove la grandezza di cui si studia la distribuzione (per esempio, numero di fratelli, età) è indicata mediante la scala graduata sulla retta orizzontale di base (detta anche «asse delle ascisse», o asse x, perché le ascisse si designano usualmente con la lettera x). E notiamo subito il diverso modo in cui vengono trattate e rappresentate, rispettivamente, le grandezze di cui si considerano distribuzioni discrete o continue. Nel primo caso tutte le masse sono effettivamente concentrate nel punto iniziale di ogni singolo intervallo, e verranno rappresentate da segmenti innalzati da tale punto e di altezza misurante la numerosità rispettiva; nel secondo caso la massa è invece diffusa su tutto l'intervallo, e ciò si rappresenta sovrapponendovi un rettangolo la cui area misuri la numerosità rispettiva.

Si noti che, nel caso piú normale, in cui la suddivisione è fatta in intervalli uguali, anziché «area» si può dire «altezza», mentre ciò ovviamente non vale piú nel caso d'intervalli inomogenei (cfr. § 1.4).

Per fissare le idee su di un esempio «in piccolo», consideriamo una «popolazione» (come si dice, in gergo tecnico, l'insieme di persone o di casi o di avvenimenti, ecc. che sono soggetto di una certa indagine statistica), di cui si abbiano a considerare due distribuzioni: una qualitativa (discreta) e una quantitativa (continua): la prima per «numero di fratelli» e la seconda per «età» (in anni compiuti), pensando ad esempio (benché i dati siano inventati) che si tratti degli $N = 120\,000$ licenziati di istituti tecnici nel 1973-74.

L'indicazione di quella numerosità N cosí grande serve comunque solo a rendere piú accettabile di parlare – come presto faremo – di «densità», cosa che sarebbe del tutto fuori luogo nel caso di una numerosità troppo piccola (per esempio, se si trattasse di un'indagine limitata a una sola scuola).

Devo chiedere scusa, una volta per tutte, per la pedanteria di siffatte precisazioni (queste, ed altre, passate e future), ma sono necessarie: non c'è altro modo per evitare, in ispecie al lettore che consultasse trattati ed annuari che usano terminologie differenti, il rischio di non potercisi raccapezzare. Anche questioni di terminologia divengono spesso essenziali, in quanto si tratta di evitare ambiguità ma anche e soprattutto di far rammentare sempre (anche ove sarà sottaciuto) che nulla in realtà ha senso se non lo si rende rispondente a un reale motivo d'interesse e se non lo si integra e riveste con tutte le necessarie cautele atte a darne il senso in modo univoco.

Riferendoci ora al nostro esempio immediato, consideriamo la classificazione dei nostri 120 000 studenti per numero di fratelli e per età. Supponiamo per semplicità che bastino, per ciascuna tabella, cinque caselle:

Fratelli	*o*	0	*o*	1	*o*	2	*o*	3	*o*	4
Età	*o*	18	*o*	19	*o*	20	*o*	21	*o*	22

Suddividendo gli N individui nelle rispettive caselle, avremo:

A seconda del numero di fratelli	$N = f_0 + f_1 + f_2 + f_3 + f_4$
A seconda dell'età	$N = e_{18} + e_{19} + e_{20} + e_{21} + e_{22}$

Notiamo fin d'ora, incidentalmente, che si potrebbe anche considerare la classificazione doppia, per numero di fratelli e per età congiuntamente: N verrebbe diviso allora (in questo esempio) in $5 \times 5 = 25$ caselle secondo lo schema seguente:

$$
\begin{array}{l}
n_{0,18} + n_{0,19} + n_{0,20} + n_{0,21} + n_{0,22} = f_0 \\
n_{1,18} + n_{1,19} + n_{1,20} + n_{1,21} + n_{1,22} = f_1 \\
\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
n_{4,18} + n_{4,19} + n_{4,20} + n_{4,21} + n_{4,22} = f_4 \\
\hline
e_{18} \quad + e_{19} \quad + e_{20} \quad + e_{21} \quad + e_{22} \quad = \mathrm{N}
\end{array}
$$

Ma di tale argomento si tratterà piú avanti al § 3.5.

Dobbiamo intanto chiarire (e discutere) qualche aspetto terminologico.

Nel gergo statistico il numero N si dice esso stesso, a volte, 'popolazione', e i numeri come gli f, gli e e gli n dell'esempio 'frequenze assolute' dei rispettivi sottogruppi. Dividendo tali frequenze assolute per il numero totale N si hanno i dati in percentuale, e si dicono 'frequenze relative'.

Tale gergo è però assai scomodo e impreciso: basti pensare che spesso, per brevità, si deve omettere la specificazione «assoluta» o «relativa» e tutto diviene ambiguo. Inoltre, tale gergo è contrastante con quello dei matematici (in particolare probabilisti) nel quale 'frequenza' significa sempre frequenza relativa (o percentuale). Meglio pertanto chiamare 'numerosità' (della popolazione, dei sottogruppi) le N, f_i, e_i, e 'frequenze' *tout court* le percentuali (f_i/N, e_i/N).

Quanto al dire sempre percentuale (anziché, ad esempio 'frazione', che richiama la scrittura dell'operazione non eseguita: come 3/7 anziché 0,4286 ovvero 42,86 per cento) mi sembra sia l'unica soluzione pratica e non suscettibile di equivoci. In particolare – e sia detto incidentalmente – confesso di trovare strana e anacronistica l'abitudine di esprimersi a volte con frazioni, come 3/7, e, comunque, di privilegiare in certo senso (al di fuori che in certe impostazioni e indagini estremamente teoriche) i numeri «razionali», fosse anche solo per il semplice fatto di riservare loro un tale pomposo nome (anziché, per esempio, 'frazionari' o 'periodici'). Siamo ancora pitagorici? Archimede non è ancor nato? Fibonacci non ci ha ancora insegnato l'uso delle cifre arabe con la scrittura posizionale?

La differenza, già sottolineata, fra i casi di scala discreta e scala continua implica naturalmente un'analoga diversità per quanto riguarda la rappresentazione grafica di una distribuzione: è appunto questa la differenza che premeva rimarcare, e che a sua volta trova la migliore illustrazione facendone «vedere» il senso su tale raffigurazione.

Il numero dei fratelli è intero (0, 1, 2, 3, 4) e la distribuzione si concentra su questi cinque punti dell'asse delle ascisse: i punti $x = 0, 1, 2, 3, 4$. Volendo rappresentare la numerosità e la frequenza dovremo innalzare – *esattamente* in corrispondenza a dette ascisse – dei segmenti verticali di altezza proporzionale alla rispettiva frequenza. Soltanto per migliorare la visibilità – e con ciò anche l'efficacia e l'estetica – si può ingrossare un po' ogni segmento facendolo diventare una striscia (come nella figura 1), cosicché si ottiene un «istogramma discreto a canne d'organo». (È opportuno specificare 'discreto' per poter dire 'istogramma' (*tout court*, senza aggettivi) quello che sottende *intervalli* (cfr. oltre). Si noti, poi, come la cura dell'estetica contribuisca non solo alla gradevolezza delle immagini ma anche alla loro efficacia espressiva. D'altra parte, occorre tener presente che un tale allargamento (se lo si prendesse per vero) comporterebbe un'alterazione dei risultati: non sulla media aritmetica (a meno che le strisce non fossero ben centrate sul segmento inizialmente considerato), ma su altre medie o cose (per esempio scarto standard e simili).

La numerosità e la frequenza relative ai singoli valori si rendono piú co-

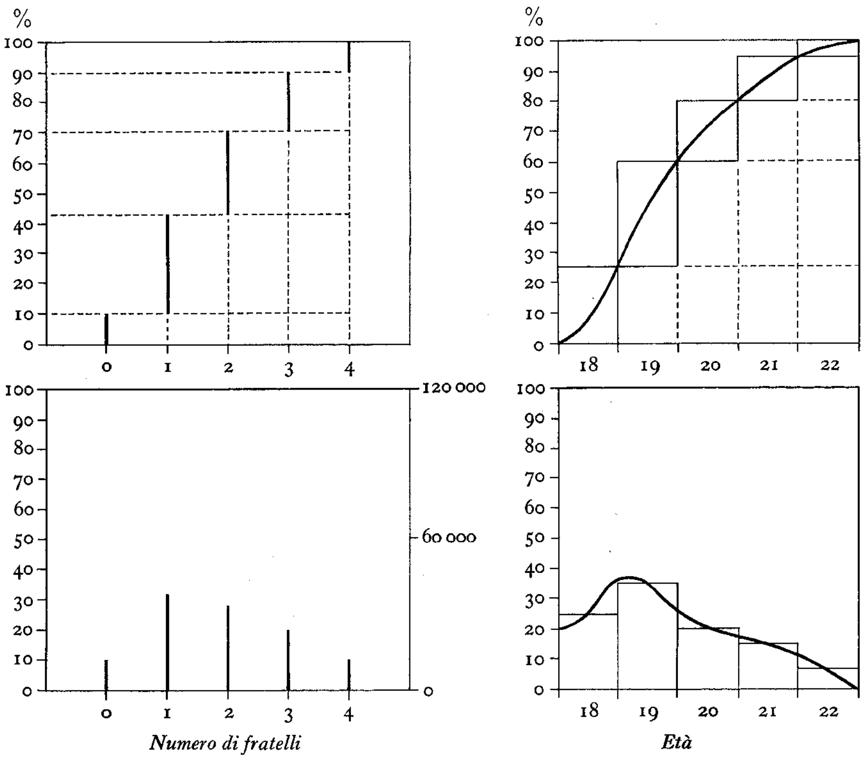


Figure 1-2.
Rappresentazioni grafiche di una distribuzione per numero di fratelli (a sinistra) e di una distribuzione per età (a destra): in basso i diagrammi di frequenza, in alto i diagrammi cumulativi.

modamente apprezzabili e confrontabili, leggibili, aggiungendo ai fianchi due scale di ragguaglio: una in numerosità e l'altra in frequenza (fig. 1).

L'età è invece una grandezza continua; perciò è giusto che l'istogramma si riferisca (si «appoggi») agli intervalli tra punti divisori (e non ai punti divisori stessi, come invece andava bene per il numero di fratelli).

Perciò, nell'esempio a fianco, la distribuzione per età avrà come rappresentazione grafica un istogramma in cui le frequenze saranno rappresentate da colonne aventi la larghezza di tutto l'intervallo su cui poggiano (nel nostro caso, anni di età), cosicché, insieme, formano una figura compatta. La numerosità di ogni classe sarà ancora data dall'altezza della corrispondente colonna, o, meglio, dalla sua area (il che è equivalente nel presente esempio, ma, come si vedrà nel § 1.4, è vero anche usando suddivisioni disuguali mentre allora la prima dizione cessa di essere valida).

La figura 2 mette in risalto la differenza fra il caso dell'età e quello del numero di fratelli, delle cui distribuzioni sono messi a confronto i grafici delle frequenze: quelle singole finora considerate nonché quelle cumulative di cui tosto diremo. Nel caso dell'età è indicata inoltre la curva che rappresenta la distribuzione esatta anche entro le singole frazioni di anno (come se si avessero per ciascuno le età esatte: non solo in anni ma in anni mesi giorni ecc.); oppure si può immaginarla come una ricostruzione ipotetica ottenuta «lisciando» il profilo (ma conservando l'area colonna per colonna); nel caso del numero di fratelli nulla di simile avrebbe senso perché i valori possibili sono soltanto gli interi 0, 1, 2, 3, 4.

La parte superiore della figura 2 mostra cosa significhino e come si costruiscano le (sopra accennate) frequenze cumulative: sono le frequenze (non per singole suddivisioni, bensí) di tutti gli individui al di sotto di un dato limite (ossia le successive somme parziali delle frequenze f_h); graficamente, si tratta di sovrapporre man mano le colonnine dell'istogramma, cosicché il profilo superiore dà, in corrispondenza ad ogni valore x, la frequenza dei casi con valore fino ad x. La funzione rappresentata da tale profilo, detta funzione di ripartizione, si suole indicare con $F(x)$, e ne parleremo piú espressamente fra poco, con osservazioni atte a prevenire dubbi o malintesi, a questo punto plausibili.

1.3. Effetto di suddivisioni piú o meno strette.

Sarà forse spontaneo a tutti chiedersi – e comunque è opportuno farlo – cosa avvenga a seconda che si scelgano suddivisioni piú o meno strette, e quindi come convenga sceglierle. Che cosa si guadagna o si perde?

In generale, come è facile prevedere anche senza ragionamenti sottili, un istogramma basato su larghi raggruppamenti mostrerà pochi larghi gradini con forti dislivelli al passaggio da ciascuno al successivo. È presumibile che passando a raggruppamenti piú stretti i salti divengano piú piccoli e il profilo risulti piú regolare, cosí da suggerire di «lisciarlo» tracciando una curva che rispetti (esattamente, o con qualche licenza) le aree entro ogni suddivisione.

Ma che cosa vuol dire «grande» o «piccolo»? e che cosa è (o appare) «liscio»?

Occorre riconoscere che si tratta di frasi che non hanno un senso assoluto e non ammettono risposte in senso assoluto. Hanno un senso soltanto in senso relativo, ossia in relazione alla dimensione delle circostanze di cui ci si occupa.

Se parliamo di variazioni, diciamo, della temperatura nel tempo, potremo pensare alle variazioni in periodi di tempo delle dimensioni di ere geologiche, oppure tra periodi storici piuttosto lunghi, oppure della variazione stagionale annua, o infine di quella diurna: per ogni livello abbiamo un certo senso da dare al concetto di piú o meno sensibile stabilità o differenza. E cosí è della nozione di «densità» (nella fisica, e cosí nel caso nostro, perfettamente analogo). Si può parlare della densità media di materia in una porzione della Via Lattea, dell'orbe terracqueo, del corpo di un elefante, di un panino, e, allo stesso modo, di densità di popolazione in un territorio (ampio come una nazione, o una regione, o una città, o un rione, ecc.).

L'analogia con la densità (pensando, come è spontaneo, a quella di un gas) può indurre facilmente all'idea di un concetto significativo nel senso piú stretto, sicuro: essa è il rapporto tra peso e volume in un piccolo spazio intorno a detto punto.

E l'analogia è anche esatta, ma il fatto non è semplice come questa formulazione indurrebbe a pensare: anche il gas, come la popolazione umana, è composto di atomi separati da (relativamente grandissimi) spazi vuoti: una porzione molto piccola di spazio ha grandissima probabilità di essere vuota, con densità nulla, mentre, se invece contiene per caso un atomo, la densità risulta immensa.

In entrambi i casi – è bene fissarselo bene in mente – la densità (per quanto corrispondente al senso che le dà il profano, nell'ambito dei fatti macroscopici che egli percepisce) è soltanto un'astrazione mentale che va sempre fatta alla scala adeguata al fine.

Come regola di comportamento, si tratta di scegliere nel modo migliore al fine di evitare, come meglio possibile, le irregolarità dovute ai due opposti fattori, tenendo conto di ciò che si sa essere (o si ritiene sia) significativo, nel fenomeno allo studio. È probabilmente utile e istruttivo indicare quali irregolarità sarebbero da attendersi in una statistica della natalità in Italia «per ogni minuto». In Italia, con la natalità degli ultimi anni, i nati, in media, sono circa 1,7 al minuto. Cosa ci si può attendere? Che in genere le nascite siano sempre 1 o 2? Forse sarebbe la risposta apparentemente piú logica (ché in genere si sopravvaluta, o meglio si miracolizza eccessivamente, lo slogan della «regolarità statistica»). Invece non è cosí.

Le ipotesi piú naturali di «casualità» sono infatti quelle cui risponde la distribuzione di Poisson, come si vedrà nel § 2.9, e, in base a detta media di 1,7 nati al minuto, le probabilità (in percentuale) che in un (qualunque) minuto si abbiano 0, 1, 2, 3, 4, 5, 6, 7 nascite (al di là sono trascurabili) risultano:

Numero di nascite	0	1	2	3	4	5	6	7	...
Probabilità (%)	18	31	26	15	7	2,5	0,7	0,2	...
Previsione nascite	0	0,31	0,52	0,45	0,28	0,12	0,04	0,01	...

(La somma dell'ultima riga dà 1,73, il che corrisponde al dato di partenza 1,7; la somma delle probabilità dà 100,4 (e manca il circa 0,03 per i valori oltre 7); ciò per gli arrotondamenti, utili da mettere in rilievo per far risaltare meglio l'essenziale).

Però... (è bene far riflettere alle ipotesi implicite che per solito molti fanno e accettano senza neppure accorgersene)... la frequenza delle nascite è soggetta a variabilità stagionale; sarà lecito non tenerne conto? non influirà sulla conclusione?

Un'ipotesi grossolana (ma non stravagante) è la piú adatta per rendersi conto se la risposta teoricamente corretta sia sí oppure NO; pensiamo quindi che la media 1,7 risulti dal fatto che la media sia 1 in un periodo pari al 30 per cento dell'anno, e sia 2 nel rimanente 70 per cento (circa mesi 3 ½ e 8 ½).

Ecco i risultati, col raffronto delle probabilità (in percentuale) corrispondenti a questa ipotesi (di variabilità stagionale) e alla precedente:

Numero di nascite	0	1	2	3	4	5	6	7	...
Probabilità									
senza variabilità stagionale	18	31	26	15	7	2,5	0,7	0,2	...
con variabilità stagionale	24	22	16	15	7	2,6	0,9	0,3	...

Particolarmente sensibile (e comprensibile) è l'aumento del numero probabile di minuti senza alcun nato (nati 0), perché nel periodo di bassa natalità (media = 1) la probabilità di un minuto senza nascite sale al 37 per cento (anziché 18 per cento), e non basta per compensare tale effetto la diminuzione dal 18 al 13,5 per cento nel periodo di alta natalità.

1.4. Il caso di suddivisioni inomogenee.

L'argomento degli istogrammi va ancora ripreso per tener presente che, spesso, la distribuzione è data non con riferimento ad intervalli costanti (tutti annuali, o tutti quinquennali, ecc.) bensí per intervalli di lunghezza diversa. Non che tale fatto introduca concetti nuovi, ma solo qualche complicazione che è bene conoscere per tenerne conto debitamente e correttamente. A parte ciò, le considerazioni che tale caso ci obbliga a fare possono contribuire a render sempre piú chiaro ciò che una forma di istogramma vuol dire e ciò che non vuol dire (in particolare, con l'andamento del contorno e coll'indicazione o mancata indicazione di linee divisorie verticali).

Possiamo considerare un esempio effettivo, il che è sempre desiderabile per concretezza e per maggiore facilità di comprensione (cfr. fig. 3).

L'esempio in oggetto è quello della tavola «Morti, per età» nel *Compendio statistico italiano* del 1975. La suddivisione per età vi è fatta secondo raggruppamenti di diversa ampiezza: 13 quinquennali (da 25-29 a 85-89), un ultimo illimitato (90 e oltre), e, prima dei 25 anni, con suddivisioni minori: 8 anni presi singolarmente (quelli da 0 a 5, e poi 14 e 20), 1 biennio (18-19), 1 triennio (15-17) e 3 quadrienni (6-9, 10-13 e 21-24); in totale 27. (Ignoro i motivi,

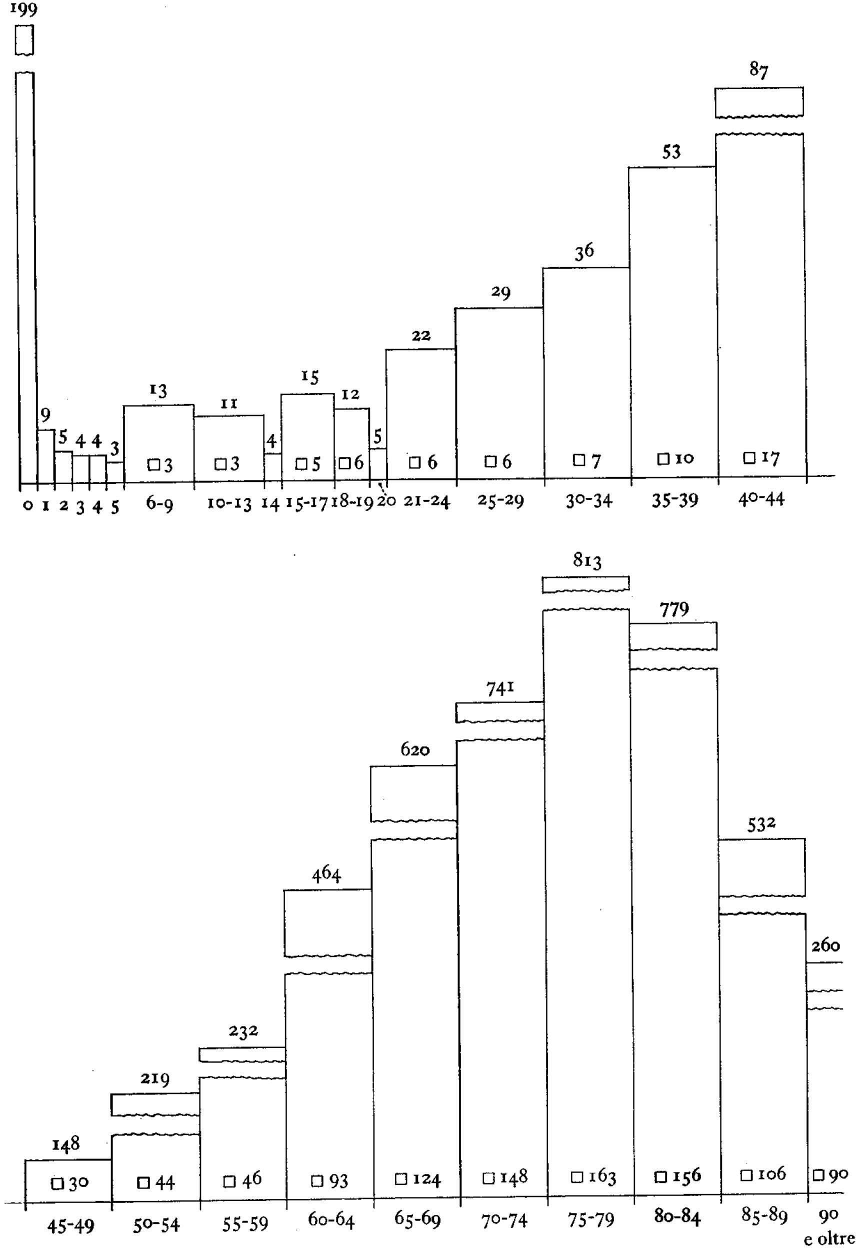


Figura 3.

Rappresentazione grafica di una distribuzione con intervalli non costanti (tavola «Morti, per età» del *Compendio statistico italiano* del 1975). La base del blocco rappresenta il numero di anni raggruppati, l'altezza il numero medio di morti nel gruppo di età (centinaia), il quadratino indica il numero di morti per anno di età nei raggruppamenti di piú anni.

ma sono intuibili: disporre dei dati per tutti i quinquenni con in piú certi dettagli per età di particolare interesse). Comunque, non c'interessano qui né la mortalità né i motivi della diversità d'intervalli, ma soltanto le osservazioni richieste per una corretta costruzione e interpretazione dell'istogramma in questa situazione.

È dato, ad esempio, il numero di morti – in cifra tonda 5000 – nelle età fra 30 e 34, cioè il totale dei cinque addendi (che nel detto schema di rilevazione restano incogniti) relativi ai singoli anni di età (30, 31, 32, 33, 34). Il modo piú corretto per segnalare questa situazione sta nel segnare un unico rettangolo di base 5 anni (fra il punto 30 e il punto 35, termine dell'intervallo «età 34») e altezza 1000 (1/5 del numero complessivo di morti di età 30-34). L'aspetto del diagramma (come contorno) è quello che si avrebbe se effettivamente il numero dei morti nei cinque diversi anni fosse stato uguale; per mostrare che non è detto sia cosí, basta evitare di dividere il rettangolo in cinque rettangolini uguali: lasciandolo intero, non tagliato da divisori intermedi, si avverte che l'indicazione ha soltanto valore globale (cfr. fig. 4).

Qualora poi, in base a congetture o altre indicazioni non certe, si ritenesse di poter accettare come sufficientemente plausibile una certa suddivisione,

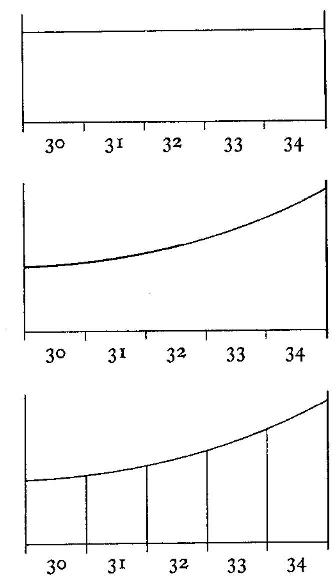


Figura 4.

Rappresentazioni grafiche del segmento 30-34 dell'istogramma della figura 3, rispettivamente nei casi in cui il totale è ripartito ugualmente sui cinque anni, in cui si presume (ma non si conosce) una variazione dell'andamento nel quinquennio, e in cui tale variazione sia effettivamente nota.

l'istogramma potrà essere disegnato in conformità di essa (come contorno), però sempre senza segnare i divisori intermedi anno per anno onde non far sembrare – come già in precedenza – che il contorno abbia un valore effettivo, accertato.

1.5. Ripartizione e densità; integrale e derivata.

Quanto finora abbiamo rilevato in forma discorsiva e descrittiva può e deve ora venir ripreso in forma piú teorica, introducendo le nozioni di funzione di ripartizione $F(x)$ e funzione di densità $f(x)$; con l'occasione potremo (e ci servirà) illustrare elementarmente anche i concetti matematici di derivata e di integrale (in quanto, nel nostro caso, $f(x)$ è la derivata di $F(x)$, la quale è, reciprocamente, l'integrale di $f(x)$).

Non c'è nulla di nuovo: si tratta solo di formulare e spiegare in modo piú preciso, e con un po' di terminologia matematica, ciò che implicitamente è già stato visto nella figura 1.

Nel caso di una distribuzione discreta (come nell'esempio del numero di fratelli) non ha senso la densità ma al suo posto abbiamo gli addendi f_h = numero degli individui aventi h fratelli, e $F(x)$ è la somma degli f_h con $h \leq x$ (cioè il numero d'individui aventi non piú di x fratelli). (Cfr. fig. 1).

Nel caso di una distribuzione continua il significato non cambia: $F(x)$ è il numero d'individui di età $\leq x$; qui però il diagramma di $F(x)$ cresce in modo praticamente continuo (benché in realtà a piccoli gradini, ciascuno in corrispondenza all'età esatta di ogni individuo). La densità $f(x)$ è, teoricamente, la «derivata» di $F(x)$ (ossia la pendenza della tangente al diagramma $F(x)$ nel punto x). Naturalmente, ciò non ha senso, dato l'andamento a gradini, a meno che non ci si riferisca (come è naturale) alla *curva* $F(x)$ «come la si

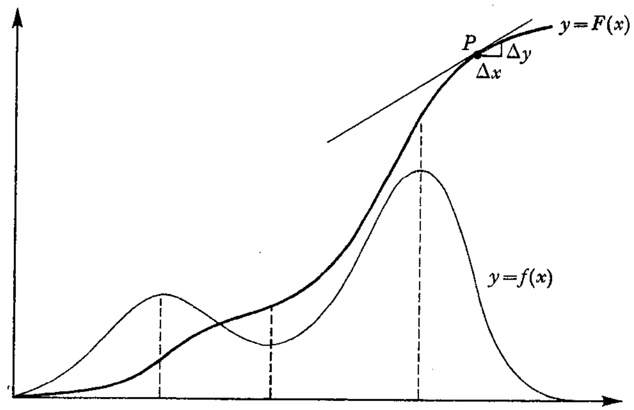


Figura 5.
Funzioni di ripartizione e di densità.

vedrebbe ad occhio nudo», cioè ignorando la dentellatura microscopica. (A questo riguardo, ricordiamo le osservazioni critiche sulla nozione di densità fatte nel § 1.3). In pratica, è lecito e conviene permettersi il lusso di ignorare tali sottigliezze, purché uno le abbia sufficientemente assimilate per sottintenderle istintivamente.

Allora si può anche dire (come si usa pensando nel «continuo») che $F(x)$ sia una funzione continua e derivabile, e che la $f(x)$ ne sia la derivata: $f(x) = dF(x)/dx$; geometricamente (per chi non lo sapesse o ricordasse) la derivata (di $F(x)$ in un punto $x = x_0$) è la pendenza della tangente al diagramma $y = F(x)$, e cioè il rapporto degli incrementi, $\Delta F(x)/\Delta x$, al limite per $\Delta x \to 0$. Ciò equivale – forse è utile dire anche cosí – a definire la tangente come caso limite della secante quando un punto si avvicina indefinitamente a quello dato (o ancora – in forma suggestiva, cara un tempo ai cultori di geometria algebrica, ma inaccettabile salvo come allusione ellittica al passaggio al limite – come «la retta passante per due punti della curva *infinitamente vicini*»).

Analogamente, data la densità $f(x)$ si può ottenere la funzione di ripartizione $F(x)$ col passaggio inverso: l'integrazione. La precedente relazione tra f ed F si può scrivere $dF(x) = f(x)\,dx$ (l'incremento di F è l'area di una strisciolina del diagramma della densità, di altezza $f(x)$ e base dx) e $F(x)$ è la «somma» (a rigore, nel senso di «limite», il che intuitivamente si può sottintendere) di tutti questi incrementi. Ciò si indica con $F(x) = \int f(x)\,dx$, «integrale» della funzione $f(x)$; il segno $\int$ di «integrale» – deformazione di S – indica la «somma» in tale senso speciale: «somma», per cosí dire, di infiniti addendi infinitamente piccoli; il significato geometrico e pratico dispensa da precisazioni teoriche che, in una prima introduzione di tali concetti per non-matematici, anziché chiarirli li farebbe apparire oscuri ed ostici, e fors'anche contraddittori.

1.6. Interpolazione e perequazione.

Abbiamo già notato come un diagramma – anche se in realtà varia per piccoli salti, derivanti da un fenomeno discreto – appare all'occhio come una curva. Guardandolo «ad occhio nudo» – dicevamo – scompare la «dentellatura microscopica» e l'andamento appare «continuo».

Questo lisciamento è di per sé un processo di «interpolazione» e/o «perequazione» fatto spontaneamente e inconsciamente; lo stesso «lisciamento» viene anche fatto espressamente, con metodi che accenneremo, con vari intendimenti. Può trattarsi del desiderio di eliminare irregolarità ritenute casuali, non significative, oppure di dare al diagramma una forma ritenuta teoricamente adeguata a spiegare la natura del fenomeno che essa rappresenta, o qualcosa d'intermedio.

In questi casi si parla – come già detto – di interpolazione (piú propriamente, quando non si alterano i dati osservati ma si «costruiscono» quelli mancanti: ad esempio, nota la popolazione in una successione di istanti, tracciando la curva che ne dà la numerosità in tutti gli istanti, anche intermedi) e di perequazione (piú propriamente, quando si ritoccano anche i valori osservati per dare all'andamento globale una maggiore regolarità: in certo senso, trascurando come «accidentali», «anormali», «non significativi» certi spostamenti da una linea piú armoniosa o piú plausibile).

Entrando in questo argomento, è appropriato (ed anzi è forse doveroso e inevitabile) affrontare un interrogativo inquietante riguardo alle manipolazioni di dati statistici (quali sono appunto, in particolare, interpolazioni e perequazioni). È lecito? È un delitto?

Come per tutte le questioni poste in forma di aut-aut, una risposta tipo SÍ-NO sarebbe avventata e irresponsabile. Indubbiamente, anzitutto, i dati osservati sono quelli che sono e non è lecito alterarli; se si dubita che siano affetti da errori bisogna controllarli (eventualmente correggendoli) *tutti* (e non solo quelli per cui il presunto e constatato errore sarebbe a sfavore dei propri interessi e delle proprie tesi, fosse pure scientifiche).

L'ottenimento di dati «corretti» ritoccando quelli osservati lascia sempre un certo margine di arbitrio ed è inevitabile che la scelta sia basata su apprezzamenti sostanzialmente soggettivi. Chi fa una tale scelta deve spiegarne le motivazioni (tecniche e fattuali) su cui si è basato, in modo che gli altri possano adeguarvisi (piú o meno) o dissentirne (piú o meno) in questo o quel senso (dando magari luogo a una discussione).

Non c'è mai certezza del sí o del NO, e neppure delle conclusioni di esperti per quanto autorevoli e attenti e obiettivi: qualunque procedura si segua (benché certe diano maggiore affidamento di altre) un margine di dubbio è sempre doveroso. Ma importa la fiducia reciproca (se meritata, beninteso!) nella sincerità, buona fede e imparzialità assolute (anche qualora costasse fatica per eventuali commistioni di interessi, personali o scientifici).

L'interpolazione e la perequazione di dati osservati sono appunto delle fasi in cui è inevitabile un certo arbitrio. Si tratta infatti di stimare (nel primo caso) dei dati mancanti (adattandoli all'andamento di quelli osservati) e (nel secondo) – oltre, eventualmente, a ciò – di ritoccare anche quelli osservati (in genere «lisciandoli» con l'eliminazione di irregolarità che sembrano attribuibili «al caso» o ad errori non identificabili).

Per tali operazioni, di interpolazione e perequazione, esiste una varietà illimitata di metodi di ogni tipo, e ne daremo dei cenni. Per quanto attiene alla preoccupazione discussa all'inizio del presente paragrafo – quella della «onestà» – non c'è niente da aggiungere in linea di principio. Va però osservato che spesso molti *cadono nell'*equivoco (vorrei non immaginare o insinuare che *giochino sull'*equivoco) di dire oggettivo il risultato da essi ottenuto in quanto ricavato applicando un metodo «oggettivo» (metodi che si trovano a bizzeffe in ogni manuale). Scegliere a priori quel metodo «oggettivo» che si prevede dia il risultato piú vicino a quello desiderato o «conveniente» – o, peggio, saggiarne parecchi e scegliere (dichiarandolo «il migliore») quello che di fatto ha dato il risultato *in tal senso* «migliore» – è altrettanto inescusabile che alterare i dati.

E vediamo di passare in rassegna (dal punto di vista concettuale, con un minimo di esemplificazioni a scopo di concretezza) la gran varietà di metodi in uso.

Occorre però avvertire che, nell'uso pratico, la distinzione tra i due termini

'interpolazione' e 'perequazione' non è cosí netta come nell'accezione indicata. Certamente tutti diranno perequazione il «lisciamento» di una successione di dati (come quella sul raccolto di grano anno per anno) ottenuto ad esempio mediante medie mobili (cioè: sostituendo ogni dato con la media (aritmetica) di tre o cinque successivi: l'anno stesso piú uno o due sia prima che dopo). E tutti diranno interpolazione il calcolo della popolazione a una certa data basandosi sui dati all'inizio e alla fine dell'anno e sulla supposizione che l'incremento annuo sia avvenuto nella stessa misura (1/365) ciascun giorno. Fin qui la distinzione (logica, o quanto meno etimologica) è rispettata. Ma si usa però dire 'retta interpolatrice' (non perequatrice, come vorrebbe la distinzione precedentemente indicata) una retta che passi abbastanza vicino a tutti i punti che indicano la popolazione ad ogni anno (ad esempio, determinata col metodo dei minimi quadrati; cfr. oltre).

Nel presente paragrafo sembra necessario, per non far confondere le idee, attenersi alla distinzione iniziale tra interpolazione e perequazione ed insistervi. In seguito sarà preferibile invece non scostarsi dall'uso corrente, che il lettore incontrerà un po' dovunque, e dal quale dovrà abituarsi da sé a non essere indotto in confusioni.

I procedimenti in uso (sia per l'interpolazione che per la perequazione, coi criteri adeguati all'uno o all'altro caso) sono comunque di tre tipi: metodi grafici, metodi meccanici, metodi analitici.

I metodi grafici consistono nel disegnare a occhio, o cercando di adattare un curvilineo (possibilmente tenendo conto di ipotesi ragionevoli sull'andamento dello specifico fenomeno, ecc.), una curva regolare che (nel caso della interpolazione) passa per i punti noti, oppure (nel caso della perequazione) se ne scosti di quanto occorre per eliminare «irregolarità» (attribuibili o ad errori di rilevazione, o a circostanze anomale o accidentali dal cui effetto si vuole prescindere, o altro). Nel caso di un istogramma, bisogna distinguere quello a canne d'organo (ciascuna dà la frequenza entro un intervallo) e il diagramma di ripartizione (ogni piede di scalino dà la frequenza complessiva *fino a* quell'ascissa; si ricordi la figura 2). Il disegnatore deve, nel primo caso, tracciare una curva che conservi l'altezza di ogni canna d'organo, nel senso di sostituire il tetto piatto con un tratto curvo (o anche no) che, comunque, lasci la medesima area al di sotto e al di sopra del livello iniziale (cfr. fig. 2). Un errore frequente e notevole – occorre perciò farne menzione e memento! – si può verificare specialmente in corrispondenza dei punti di massimo (o di minimo, ma è piú raro ne esistano): e ne è causa la tendenza a disegnare una curva che «non si scosti troppo» dal contorno dell'istogramma. Nella figura 6 la curva a tratto continuo marcato è il diagramma effettivo di una distribuzione (si tratta di una delle «curve di Pearson» (cfr. § 3.6) e precisamente del tipo IV); l'istogramma corrisponde ad essa (ogni «canna d'organo» ha area uguale a quella sotto il diagramma; in altre parole, i due «spicchi» che vanno rispettivamente tolto ed aggiunto da uno «scalino» per passare al tratto di curva corrispondente devono essere uguali di area). Chiedendo a un disegnatore di ricostruire la curva effettiva di partenza (naturalmente, senza che possa vederla:

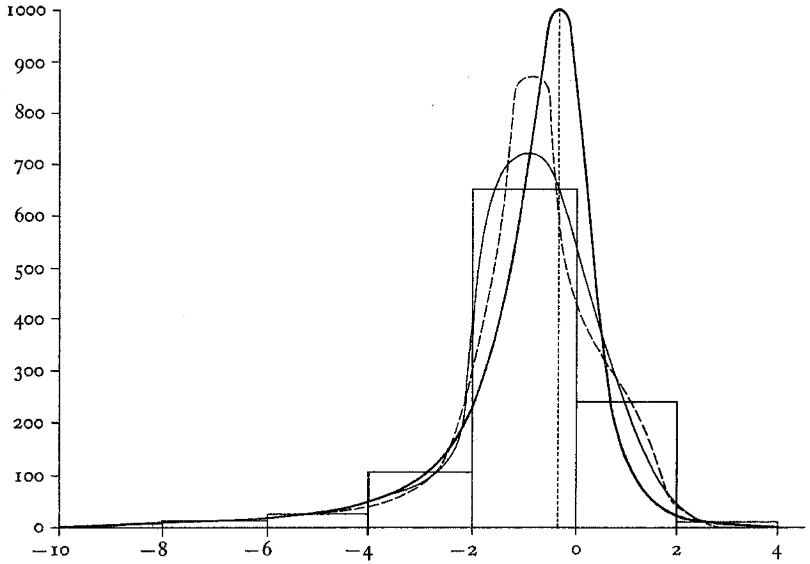


Figura 6.
Possibili errori nei metodi grafici di interpolazione e perequazione di un istogramma a canne d'organo. (Da Salvemini, *Ricerca sperimentale sull'interpolazione grafica di istogrammi*, in «Metron», XI (1934), 4, p. 159).

dandogli solo l'istogramma) è facile invece che il pinnacolo venga schiacciato e trasformato in una specie di larga piattaforma (come mostrano la linea continua sottile e quella punteggiata, dovute a due diversi disegnatori. Eseguendo il disegno sul diagramma di ripartizione il rischio di deformazioni cosí macroscopiche si riduce: si tratta infatti di far passare una curva per dei punti dati (i «piedi degli scalini»; cfr. fig. 2) e ciò guida l'occhio e la mano in modo assai piú vincolante e sicuro. Nel caso dell'esempio, quello che era il massimo per la densità diviene un punto di flesso per il diagramma di ripartizione: il punto cioè in cui la pendenza è massima, e dove l'andamento della curva, da concavo verso l'alto nel tratto precedente (a sinistra), diventa convesso alla destra. L'errore precedente si tradurrebbe qui nel disegnare una curva non abbastanza impennata verso l'alto in detto punto di flesso.

Occorre una certa sensibilità estetico-matematica per distinguere una linea goffa e probabilmente sbagliata da una linea elegante e probabilmente corretta: e ciò occorre non solo al disegnatore ma anche e in modo piú profondo per lo statistico e per chiunque debba «saper vedere» rappresentazioni grafiche di dati. Sull'importanza per tutti del «saper vedere» in matematica – troppo disattesa, in genere, sia dai matematici che dai profani in matematica – non si insisterà mai abbastanza.

1.7. I metodi analitici.

Sui metodi meccanici non vale la pena di aggiungere altro all'esempio (già citato nel § 1.6) delle medie mobili, salvo precisare che a volte (e forse piú ragionevolmente), anziché la media semplice dei tre o cinque dati si prende ad esempio il 50 per cento di quello centrale e il 25 per cento dei due adiacenti, oppure il 40 per cento di quello centrale piú il 20 per cento dei due adiacenti e il 10 per cento dei due piú esterni.

Quanto ai metodi analitici, essi consistono nel sostituire all'andamento empirico un andamento espresso mediante una relazione analitica (matematica), cioè con una funzione (in genere piuttosto semplice) che ne renda abbastanza fedelmente il comportamento.

Era opportuno esprimersi in termini molto generici perché il tema, enunciato cosí in generale, in astratto, è molto generico; al poco che si potrà premettere come generalità dovrà seguire qualche cenno molto piú specifico.

Dal punto di vista formalisticamente matematico si potrebbe schematizzare tutto col dire che, conoscendo i valori empirici $y_1, y_2, \ldots, y_n$ osservati in corrispondenza ai valori $x_1, x_2, \ldots, x_n$ di una certa variabile, ci proponiamo di suggerire una funzione $y=f(x)$ come «legame» tra le x e le y, nel senso che «teoricamente» le y_i «dovrebbero» avere come valore $f(x_i)$ ma sono affette in pratica da un certo (per cosí dire) «errore», $y_i-f(x_i)$.

Nel caso delle distribuzioni – il solo che espressamente rientra nella presente trattazione – l'esempio piú ovvio è quello in cui, per certi valori x_i, è nota $F(x_i)$: si conosce cioè la numerosità (o la frequenza: è cosa equivalente) in ogni intervallo tra successivi «traguardi» x_i. Un caso analogo piú artificioso consisterebbe nel pensare note le densità $f(x_i)$ in tali punti. Oppure si può cercare di trovare una distribuzione analitica $f(x)$ che conservi certi dati sintetici (medie, mediana, scarti, ...), cioè indici vari, piú o meno praticamente noti a tutti, ma su cui comunque ci s'intratterrà piú avanti (§§ 2.4, 2.6).

Ad ogni modo, per quanto riguarda l'aspetto matematico sul quale dobbiamo ora soffermarci, importa solo chiarire un po' gli elementi essenziali: quelli, cioè, di cui anche un profano dovrebbe avere un'idea chiara e corretta seppure a livello intuitivo. È quindi inessenziale distinguere se si tratti di applicazioni alla statistica o a qualsiasi altro argomento (come l'esempio già menzionato di distribuzioni di masse).

Riprendiamo quindi il discorso dalla formulazione in termini astratti: vogliamo determinare una funzione $f(x)$ che per dei dati valori della x (siano $x_1, x_2, \ldots, x_n$) assuma rispettivamente i valori $y_1, y_2, \ldots, y_n$: esattamente (caso dell'interpolazione) o approssimativamente (caso della perequazione).

Bisogna naturalmente, per prima cosa, precisare in quale ambito vogliamo scegliere la funzione $f(x)$, ossia, pensando al senso geometrico, quale tipo di curve vogliamo prendere in considerazione come idonee rappresentanti di possibili andamenti del fenomeno, rispettivamente nel senso di curve interpolatrici o perequatrici.

Conviene avvertire subito che l'interpolazione (in senso stretto: di far passare una curva esattamente per punti prefissati, dati da misure empiriche) è praticamente impossibile e comunque sconsigliabile. Basti riflettere, ad esempio, sul caso piú spesso trattato (e considerato, in certo senso, come metodo standard, «naturale»): l'interpolazione mediante polinomi (necessariamente: di grado n se sono dati $n+1$ punti: una retta ($n=1$) per due punti, una parabola ($n=2$) per tre punti, e cosí via).

Riferiamoci al caso di cinque valori corrispondenti ad ascisse equidistanti (senza perdere in generalità, siano $x=0, \pm 1, \pm 2$) e vediamo cosa succede se il valore centrale si altera un po' (sia per un'effettiva casuale irregolarità nel fenomeno studiato, sia per una piccola inesattezza della misura, o addirittura soltanto per il necessario arrotondamento). L'alterazione è la stessa del caso in cui i cinque valori siano nulli ma per errore quello centrale sia (diciamo) 1; vedremo che le ripercussioni «per 1» di errore al centro sono grandissime (certo assai piú di quanto penserebbe un profano o anche uno specialista che non vi avesse mai posto attenzione).

Il polinomio interpolante è $K(x^2-1)(x^2-4)=K(x^4-5x^2+4)$ e, perché il valore per $x=0$ diventi 1 basta prendere $K=1/4$, ottenendo $1+(x^4-5x^2)/4$.

Cosa avviene per $x=\pm 3$?, per $x=\pm 4$?, per $x=\pm 10$? L'errore risulta (sempre con lo stesso segno dell'errore in $x=0$, ma) ingigantito: moltiplicato rispettivamente per 10, per 45, per 2376. (Un errore – piccolo – in senso opposto si ha soltanto tra 1 e 2 (e tra -1 e -2: ovvia simmetria) con massimo (del valore assoluto) $-0{,}56$ per $x=\pm\sqrt{5/2}=\pm 1{,}58$.

Forse era eccessivo sviluppare qui questi semplici calcoli e queste considerazioni, ma, d'altronde, non è certo male che almeno ci si renda conto di quanto sia illusoria l'esattezza di certi risultati ottenuti, sí, con grande accuratezza di calcoli, da ottimi calcolatori umani ed elettronici, perfettamente programmati per eseguirli, ma altrettanto immuni, spesso, dal sospetto di dover anche «saper vedere» qualcosa piú in là. (Sempre il «saper vedere!»)

Nel caso della perequazione (il concetto è piú duttile!) il rischio di tali alterazioni è minore; comunque occorre sempre evitare di affidarsi ciecamente a un «metodo», come se fosse valido per virtú propria, e giudicare invece in modo critico, col massimo di obiettività sia pure basata su un fondo essenzialmente soggettivo. Il metodo piú usato (e spesso giustificato con considerazioni che è per lo meno eccessivo ritenere valide in tutti i casi pratici) è quello dei minimi quadrati (indubbiamente appropriato nelle applicazioni all'astronomia, alla geodesia e in genere a misure di precisione, ripetute per maggiore accuratezza, tra cui scegliere il valore piú attendibile). Nel caso di misure ripetute di una stessa grandezza, il metodo porta ad assumere, come «valore vero», quello dato dalla media aritmetica delle misure, eventualmente ponderata, nel senso di dare maggior peso alle misure fatte in condizioni migliori (con un apparecchio piú perfetto, in circostanze piú favorevoli, da una persona piú esperta, ecc.). Si vedrà la giustificazione teorica nel § 2.6, parlando di medie e scarti.

Il caso piú semplice di perequazione è quello mediante una retta, che (come già avvisato, e come è comune anche in casi analoghi) si dice retta interpolatrice (anziché «perequatrice»). Se si applica (come è consueto, e in certo senso «naturale») il metodo dei minimi quadrati, e si hanno i valori dei dati y_i in corrispondenza ai valori x_i della variabile (ascissa), si tratterà di determinare i parametri a e b della retta $y=ax+b$ in modo da rendere minima la somma dei quadrati degli scostamenti (in senso verticale) tra i punti e la retta.

Per ogni punto tale scostamento è $y_i-(ax_i+b)$; il quadrato è

$$y_i^2-2y_i(ax_i+b)+(ax_i+b)^2;$$

la somma dei quadrati è una funzione di a e di b (gli x_i e y_i essendo dati) ed è elementare trovare i valori di a e b, ossia la retta per cui tale somma di quadrati è minima. Limitiamoci a un'indicazione utile: tale retta passa per il centro di gravità dei punti dati; quanto alla pendenza, avremo occasione di accennarvi con maggior costrutto parlando di correlazione (§ 3.5).

Il procedimento usato per il caso della retta vale, senza sostanziali cambiamenti, per l'interpolazione (col metodo dei minimi quadrati) di funzioni (ossia curve) di qualunque tipo (purché formino un sistema lineare:

$$f(x;\, a, b, c, \ldots)=af_1(x)+bf_2(x)+cf_3(x)+\ldots);$$

il quadrato (e la somma dei quadrati) si presentano nella stessa forma precedente salvo sostituire (ax_i+b) con $af_1(x_i)+bf_2(x_i)+\ldots$

Limitiamoci ad accennare al caso piú semplice dopo quello della retta: quello dell'interpolazione di una parabola, $y=ax^2+bx+c$: al posto delle (ax_i+b) si avranno le $(ax_i^2+bx_i+c)$; il resto non cambia, tranne che i parametri da calcolare sono tre anziché due. E cosí per passare al terzo grado, ecc., nulla cambia salvo il crescere del numero delle equazioni (sempre lineari) e delle incognite (sempre uguale al grado piú uno).

E nulla sostanzialmente cambierebbe, come procedimento di calcolo, neppure qualora si cercasse di applicarlo con funzioni $f_i(x)$ qualsiasi in luogo delle potenze.

Per mostrare un caso di tipo diverso, in cui l'interpolazione è usata per trovare un andamento «teorico» (basato cioè su ipotesi di tipo demografico) dello sviluppo di una popolazione, menzioniamo ancora la curva *logistica*, o di Verhulst (che la propose in tale contesto). Essa corrisponde all'ipotesi che il tasso di accrescimento sia proporzionale al divario esistente tra la popolazione attuale e il livello massimo che potrebbe raggiungere (data la limitatezza dei mezzi di sussistenza) in un dato ambiente: territorio, o allevamento sperimentale, ecc. (circa la validità di quest'ipotesi le opinioni sono discordanti).

La funzione viene riportata abitualmente in questa forma:

$$f(t)=K/(1+ae^{-bt})$$

(o analoghe). Sarebbe piú semplice e significativo notare che si tratta della tangente iperbolica:

$$f(t)=\frac{1}{2}K\{1+\tanh[b(t-t_0)]\}.$$

Ricordiamo che la tangente iperbolica è il rapporto tra seno iperbolico e coseno iperbolico: $\tanh x = \sinh x / \cosh x = (e^x - e^{-x})/(e^x + e^{-x})$.

Si vede cosí subito che la logistica è simmetrica rispetto al punto di flesso $(t_0, K/2)$, che tale punto (ossia questi due parametri) e il coefficiente b bastano a determinarla (e, comunque, ne bastano tre). Se, poi, secondo la generalizzazione proposta da Pearl e Reed, si vuole aggiungere una costante positiva C (pensabile come «livello iniziale di una popolazione all'inizio di un nuovo ciclo di sviluppo secondo la logistica»: ad esempio per scoperta di nuove sussistenze), nulla cambia tranne che occorre dare C (oppure l'indicazione di un terzo punto).

2. *Delle medie (e altri indici sintetici).*

2.1. Un esempio pratico, a scopo introduttivo.

Tutti siamo abituati a dire, e a sentir dire, e a ritenere di sapere cosa vogliano dire, frasi come «la statura degli abitanti di quella regione è di 168 cm», oppure «tra 160 e 175 cm», o «di 168 ± 8 cm», e simili. E, pressappoco, è anche vero, e spesso il «pressappoco» basta, ma il pressappochismo come norma e come vizio è cosa deleteria.

Vediamo perciò, su un esempio effettivo, alcuni (e, di fatto, i piú significativi e usuali) dei significati precisi in cui tali frasi possono essere usate in modo non vagamente discorsivo ma esattamente informativo.

L'esempio è quello delle età degli sposi e delle spose in Italia negli anni 1969 e 1972 (dal già citato *Compendio statistico*, ove si possono vedere notizie piú dettagliate e riferentisi a parecchi anni):

	Età mediana	*Età modale*	*Età media*		*Scarto quadratico medio*	*Età media ± scarto q. m.*
1969						
sposi	26,80	25,50	28,28	±	7,14	21,14-35,42
spose	22,85	21,50	24,35	±	6,44	17,91-30,79
1972						
sposi	26,43	25,50	28,93	±	8,75	20,18-37,08
spose	23,10	21,50	25,06	±	7,92	17,14-32,98

Per età mediana s'intende quella di confine tra la metà dei piú giovani e la metà dei meno giovani; ad esempio, le spose del 1972 erano per metà di età inferiore e per metà di età superiore a 23,10 (cioè: 23 anni, 1 mese e 7 giorni).

Per età modale s'intende quella di massima frequenza: i dati indicano che (in entrambi gli anni) la classe di età piú rappresentata fra gli sposi era quella «25» (ossia tra 25 e 26 anni esatti) e per le spose quella «21» (ossia tra 21 e 22 esatti).

Per età media s'intende la media aritmetica (somma delle età di tutti gli individui considerati, divisa per il loro numero). Si vede ad esempio che tale età media è cresciuta di circa 8 mesi (tra il 1969 e il 1972) tanto per i maschi (+0,65) che per le femmine (+0,71).

Per scarto quadratico medio (nel *Compendio* l'intestazione dice genericamente «Indice di variabilità»; la definizione di sc. q. m. vi si trova in nota) s'intende la media quadratica degli scarti (cioè: degli scarti fra l'età di ogni individuo e la media, si calcola il quadrato; di tali quadrati si fa la media (aritmetica); di tale media si calcola la radice). Spiegazioni piú approfondite e significative si vedranno nel § 2.6; questo esempio e i cenni esplicativi non pretendevano certo di risultare autosufficienti ma tendevano solo a familiarizzare un po' col significato pratico di dati, tabelle e confronti che anche ogni profano incontra ormai come notizie e che sarebbe bene tutti potessero capire e utilizzare sia pure al livello del «profano sufficientemente aggiornato».

Per completare un po' queste indicazioni preliminari, dobbiamo avvertire che le tre nozioni presentate sull'esempio di sposi e spose non sono che casi particolari di altre piú generali. L'età mediana – e diciamo meglio la mediana (per non rimanere legati al particolare esempio) – è un caso particolare dei valori di posizione, o quantili, di definizione perfettamente analoga: il 1° e il 3° quartile sono i valori al confine tra il 1° quarto e i tre successivi (risp. fra il 3° e l'ultimo) nell'ordine di grandezza (per età, se pensiamo all'esempio precedente, o secondo quel qualsiasi altro carattere di cui si tratti). Analogamente, si dicono sestili (1°, 2°, ..., 5°) i valori di confine fra l'analoga suddivisione in sei gruppi di uguale numerosità e in ordine di grandezza crescente per il carattere che interessa. (Pare che tale suddivisione in sesti riesca utile in antropologia).

Piú in uso, conformemente alla generale adozione della numerazione decimale e del sistema decimale per misure, monete, ecc. (ormai, finalmente, anche in Inghilterra!), è l'uso dei decili (e, a volte, dei centili). È chiaro che, dando tutti i centili, piú che fornire dati sintetici si descrive la distribuzione in modo praticamente completo. Si può infatti costruire la funzione di ripartizione per punti, salvo che, invece di partire da una suddivisione in (ad esempio 100) parti uguali della base e segnando in corrispondenza l'altezza dell'ordinata, si suddivide in parti uguali l'altezza e si trova su ogni orizzontale il punto della curva di ripartizione corrispondente al quantile. In parole povere: la curva di ripartizione è quella che è; in un caso la si traccia individuandone le intersezioni con le verticali, nell'altro con le orizzontali, della quadrettatura del quadrato dato (ad esempio in 10×10 o 100×100 quadratini).

Sulla moda c'è poco da aggiungere; da sola dice poco; occorrerebbe almeno dire (in piú) se la densità ha un unico massimo o parecchi, e altri dettagli: per esempio se la moda è data da un «pinnacolo» o se invece è su un tratto pianeggiante. Si ricordino le osservazioni sull'interpolazione grafica (§ 1.6), utili anche per consigliare di risalire alle fonti in caso di legittimi sospetti su inconvenienti del genere.

Quanto a medie (e poi anche a scarti) il discorso che occorre è molto piú ricco, interessante, illuminante. E lo iniziamo subito.

2.2. La nozione di media (Oscar Chisini: un dubbio fecondo).

Da tempo immemorabile gli uomini (scienziati e profani) facevano uso di medie senza rendersi bene conto del concetto unitario e delle esigenze effettive sottostanti a tale nozione. Erano delle *adhockeries* ante litteram.

E tali forse sarebbero ancor oggi se... se Oscar Chisini non fosse stato inviato un anno come commissario per gli esami in un istituto tecnico. Grazie alle riflessioni suggeritegli da quegli esami, Chisini va menzionato qui come lo «scopritore del concetto di media». Potrebbe sembrare uno scherzo (come la favola di Campanile sull'«inventore del cavallo»), ma non si tratta di invenzione di qualche cosa di nuovo, bensí del chiarimento di un concetto fino allora confuso, mancante di una caratterizzazione unitaria e profonda.

L'aneddoto, o la storia, di quegli esami merita di essere narrato, perché è istruttivo sotto molti punti di vista. Chisini, professore universitario, presta attenzione alle domande stereotipate sulle medie (media aritmetica, geometrica, armonica, ecc.); le trova stucchevoli, ma, anziché distrarsi, si appassiona a cercare se e quale sia il concetto sottostante a tante nozioncine staccate.

Ci pensa non da «matematico puro», ma (merito ben maggiore!) da persona intelligente che è *anche* un matematico. E si chiede il perché; le medie: perché? perché sono nate? perché SERVONO? (Quanti non inorridirebbero a porsi o sentir fare una simile domanda: «matematica che *serve*?»... ma allora non è scienza, per lo meno non è Scienza!)

Chisini invece rifletté all'argomento e trovò la risposta che lo soddisfaceva, rispondente ad ogni esigenza sia matematica che pratica e filosofica; ma non poteva tuttavia valutarne appieno l'importanza. Era capitato per caso a contatto con una problematica cui era estraneo, aveva risposto a un suo intimo bisogno di chiarezza, e si limitò a trarne un articoletto didattico per il «Periodico di Matematiche» (1929), che sottolineava il significato relativo e funzionale, «rispondente a un dato scopo», della nozione generale di media.

Ne tratteremo tra poco, ma occorre ancora qualche complemento alla storia. L'articoletto di Chisini, apparso in una rivista di didattica, sarebbe forse passato inosservato se, fra i suoi allievi, non ci fosse stato uno che aveva cominciato a interessarsi alla probabilità e alla statistica, e al quale fu facile cogliere l'importanza dell'idea di Chisini e divulgarla e applicarla nel campo per cui piú appropriatamente era fatta e in cui doveva essere feconda. Ne derivò una trattazione abbastanza ampia e sistematica, che, per un punto importante, ebbe nuovo impulso grazie a un risultato ottenuto indipendentemente, proprio in quel torno di tempo, dal russo Antonij Kolmogorov e dal giapponese Mitio Nagumo (il «teorema di Nagumo-Kolmogorov»; cfr. § 2.5).

Prima di entrare in argomento in termini tecnici conviene forse indugiare ancora un po' ad illustrarlo in termini intuitivi, o, almeno in una certa accezione, «filosofici».

L'idea generale della definizione data da Chisini si può esprimere benissimo in parole, nel linguaggio comune. E cosí forse molte altre cose, usualmente riservate ai «competenti», agli «iniziati» (!), potrebbero venir spiegate a tutti o molti, nella loro esatta essenza, pur senza offuscare il concetto informatore coi tecnicismi richiesti per un lavoro da specialisti. Sarebbe una grande conquista (nonostante il pericolo di vederla guastata, magari ridicolizzata, da «volgarizzatori» da strapazzo). Ma perché non potrebbe piú esserci qualcuno che sappia esprimere in modo piano, preciso, appassionato, poetico, fatti e teorie della scienza, come fece, ad esempio, Michael Faraday in quel piccolo capolavoro che è il gruppo di conferenze per ragazzi *The Chemical History of a Candle*?

Forse il guaio è che siamo troppo specializzati, isolati, unilaterali, chiusi nei rispettivi compartimenti stagni da cui è difficile spaziare con lo sguardo sulle molte cose semplici ed essenziali al cui contatto potremmo forse riacquistare la naturale intelligenza e spontaneità del bambino!

2.3. Le «definizioni» e «la» definizione.

Le «definizioni» che Chisini sentiva chiedere dagli esaminatori e declamare pappagallescamente dagli studenti – e che lo facevano giustamente inorridire – non erano che le singole «ricette» per calcolare varie medie. («Dati n numeri, la loro media aritmetica è la loro somma divisa per n; quella geometrica, la radice ennesima del loro prodotto; quella armonica, è n diviso per la somma dei reciproci; quella quadratica, la radice della media aritmetica dei quadrati» e via dicendo).

Queste sono definizioni esatte (per il lettore che dovesse apprenderle andrebbero benissimo), ma esse sono puramente formali, non spiegano *il perché*, che è la cosa essenziale. (È questo il tipo di «indottrinamento» a vuoto grazie al quale – nell'esame comparativo circa l'apprendimento della matematica nei diversi paesi promossa dalla IEA – si concluse che gli studenti italiani «sanno tutto, ma non a che serve!»; c'è dell'esagerazione in ambo i sensi, ma fondamentalmente il guaio del nozionismo a vuoto è il nostro handicap che annulla ogni capacità di risultato concretamente educativo).

Occorrerebbe «pensare alla Chisini»: con riferimento al nostro caso (ma ciò vale per tutto), quel che occorre è chiedersi, con spirito critico, quale sia *il significato* del concetto di media, il che vuol dire analizzare i motivi profondi ed essenziali che hanno costituito, sia pure inconsciamente, lo scopo per cui quel concetto è stato introdotto e che spiegano la ragione intima della sua utilità.

«Non si eseguisce certamente tale analisi – come osserva il Chisini – quando si pretende di definire "media fra piú quantità date una nuova quantità compresa tra la piú piccola e la piú grande delle quantità considerate", e la si evita o trascura quando si preferisce definire direttamente, volta per volta, le singole specie di medie che s'incontrano abitualmente, "facendo cosí opera bensí esatta, ma puramente formale e antifilosofica, che può servire, e male, solo per un uso empirico"».

Bisogna cominciare invece, come appunto dice e fa il Chisini, mettendo in rilievo che la ricerca di una media ha come scopo quello di semplificare una data questione «sostituendo in essa, a due o piú quantità date, una quantità sola che valga a sintetizzarle, senza alterare la visione d'insieme del fenomeno considerato», e si noterà allora anzitutto che «non ha senso parlare di media di due (o piú) quantità, ma ha senso parlare della media di esse all'effetto della valutazione sintetica di un'altra grandezza che ne dipende».

Chisini traduce tale concetto in forma di definizione nel modo seguente: se, di n grandezze omogenee $x_1, x_2, ..., x_n$ interessa considerare la funzione (simmetrica: tale cioè che non varia cambiando l'ordine delle variabili) $f(x_1, x_2, ..., x_n)$ e x è il valore per cui $f(x, x, ..., x)$ (x ripetuto n volte) dà il medesimo valore (ossia se, agli effetti del calcolo della funzione f, tutto va *come se* le n variabili x_i avessero tutte quel medesimo valore x), si esprimerà tale fatto dicendo che x è la media di $x_1, x_2, ..., x_n$ agli effetti del calcolo di f. (Oltre al valore concettuale, e concreto, anziché formalistico e astratto, della definizione, appariva affascinante anche il carattere relativo, pratico, pragmatico, del concetto informatore, e la presenza del «come se», riecheggiante la «filosofia dell'*als ob*» di Veihinger).

La molteplicità e peculiarità delle medie (ciascuna a suo modo, a seconda del fine) è cosa da ricordare sempre per evitare conclusioni semplicistiche ed erronee. L'esempio dato da Chisini nella sua nota è molto pratico, semplice, e facile da ricordare: si tratta di un viaggio in automobile, alla velocità di 60 km/h per due ore e di 105 km/h per un'ora; velocità media 75 km/h. Però (secondo una formula empirica citata da Chisini) il maggior consumo di benzina nel tratto piú veloce non è compensato dal minor consumo nel tratto piú lento, per cui il consumo corrisponde a una velocità media di 80 km/h. In altri termini (giova ripeterlo e ribadirlo, benché in vari modi sia già stato detto e ripetuto) il «come se» non ha un valore universale né estensibile per apparenti analogie, ma è proprio legato all'ipotesi specifica cui risponde quella certa media e nessun'altra.

2.4. Le medie nell'ambito delle distribuzioni.

Finora si è parlato di medie «di n numeri», ed era il modo migliore di esprimersi per considerazioni introduttive elementari. Ma, in realtà, come concetto generale, le medie sono grandezze che si riferiscono a distribuzioni (proprio nel senso di cui stiamo trattando). In realtà, anziché dire, come finora, «la media di $x_1, x_2, ..., x_n$», si sarebbe dovuto dire «la media della distribuzione di n masse uguali ($1/n$) collocate nei punti $x_1, x_2, ..., x_n$».

Ciò è tanto piú indispensabile in quanto tale caso – di masse uguali – non è che un caso particolarissimo: in generale le masse collocate in quei punti saranno diverse, con diversi «pesi» $p_1, p_2, ..., p_n$. Ed anche questo non è che un caso particolare e particolarmente elementare, perché nel caso generale potremo avere una distribuzione qualsiasi, discreta o continua.

Fra i due casi non c'è nessuna differenza sostanziale: il concetto è sempre il medesimo, e, tecnicamente, la differenza sta solo nell'indicare l'operazione di «somma» col segno $\sum$ (di sommatoria) o quello $\int$ (di integrale: ed anzi l'$\int$

vale sempre, in entrambi i casi, se lo s'interpreta – come accenneremo verso la fine di questo paragrafo – nel senso di Stieltjes).

Anche con riguardo ai «pesi» risultano anzitutto necessarie delle avvertenze atte (come quelle del paragrafo precedente) a richiamare l'attenzione su esigenze di appropriatezza, sugli insegnamenti racchiusi in semplici esempi come quello or ora citato del Chisini.

Esso è istruttivo non soltanto in quanto indica la differenza tra medie riferite a diverse conseguenze (tempo e consumo) ma anche, sol che si rifletta a cose importanti benché banali, alla necessità di specificare bene cosa si assume come «peso». Se, anziché specificare che i due tratti di percorso avevano durate risp. di due ore e un'ora, uno avesse detto «un terzo e due terzi del viaggio», si sarebbe potuto interpretare esattamente (in tempi) ma anche, per esempio, in lunghezze di percorsi, e ovviamente il risultato sarebbe stato diverso. (Per completare l'esempio: percorrendo 210 km alla velocità di 60 km/h e 105 km alla velocità di 105 km/h, il tempo complessivo è di ore $3\frac{1}{2}+1=4\frac{1}{2}$ per percorrere 315 km, e la velocità media è di 70 km/h (anziché 75).

Prima ancora che all'apprendimento di nozioni e strumenti tecnici, l'abito mentale «educato» statisticamente è prezioso per indirizzare a individuare ed esprimere compiutamente tutto ciò che è necessario specificare, volta per volta, volendo che contenga tutto e solo ciò che serve a rendere univoca, né manchevole né sovrabbondante, la conoscenza di una situazione, dei presupposti di un problema.

Ancora un esempio per sottolineare tale esigenza di non confondere situazioni di cui la diversità può sfuggire. Se vogliamo stimare, in base ai dati di un sondaggio (o semplicemente ai dati relativi a casi osservati) quale sia il numero medio di persone per famiglia, ci si potrà basare su un campione «scelto a caso» di famiglie, oppure di individui. Il secondo modo di «scelta a caso» è equivalente al primo? è un metodo idoneo per «scegliere a caso» le famiglie?

La risposta è NO, ed è facile spiegarlo ma è anche probabile che a prima vista uno pensi che sí. Mentre nel primo caso ogni famiglia ha la stessa probabilità di venire sorteggiata, ed è ragionevole attendere che le percentuali di famiglie con 1, 2, 3, 4, ... componenti siano prossime a quelle vere, e quindi il numero medio dei componenti nel campione risulti abbastanza vicino a quello esatto (in Italia: 3,97), nel secondo procedimento una famiglia ha tante probabilità di venir sorteggiata quanti sono i suoi componenti e la media risulta ovviamente piú alta (in Italia: 5,15). Anche questo dato ha un significato, ma diverso: si può dire che ogni cittadino ha in media 4,15 altri membri nella stessa famiglia. (Il metodo statisticamente corretto per eliminare in via presuntiva tale causa di distorsione consisterebbe nel sommare, al numero di appartenenti a «famiglie di un solo componente», la metà di quelli appartenenti a famiglie di due, un terzo per quelli di tre, e cosí di seguito).

Riprendendo – dopo queste osservazioni critiche, opportune per segnalare facili equivoci – il discorso «tecnico» sulle principali medie, dovremo darne esplicitamente l'espressione con riferimento ai diversi casi: caso «continuo» (in senso stretto), con densità $f(x)$; caso discreto (un numero finito, o un'infinità numerabile di «masse concentrate» in singoli punti); caso generale (masse di entrambi i tipi precedenti, piú altro caso «singolare»).

Qualunque sia la distribuzione, è sempre possibile rappresentarla con la funzione di ripartizione $F(x)$; nel caso piú generale potranno esservi: 1) delle masse concentrate (e in corrispondenza ad esse la $F(x)$ avrà un salto); 2) delle masse diffuse regolarmente con densità $f(x)$, che sarà la pendenza del diagramma della $F(x)$ nel punto x; 3) infine possono esistere masse distribuite in modo piú «irregolare» (che ci limiteremo a illustrare con un esempio). Cominciamo subito da tale esempio per «levarci il pensiero».

L'esempio classico è quello della distribuzione di Cantor, che conviene presentare in forma un po' modificata per riferirci al piú familiare sistema di numerazione decimale (anziché a quello in base 3, piú adatto per l'esempio). Basta dire allora che la distribuzione che consideriamo è quella in cui, del segmento (0, 1) (pensiamo: 1 metro), conserviamo solo i punti la cui ascissa, scritta in cifre decimali, non contiene nessun 5. Ne vengono, cioè, tolti, il decimetro tra 50 e 60 cm; poi da tutti gli altri decimetri il centimetro tra 5 e 6 cm; da tutti gli altri centimetri il millimetro tra 5 e 6 mm, e cosí via. In altre parole, si conservano soltanto i punti la cui ascissa non contiene mai una cifra 5. Poiché, ad ogni ripetizione di tale eliminazione di 1/10 dell'insieme residuo, la lunghezza complessiva (o «misura») si riduce del 10 per cento, dopo un passo rimarranno 90 cm, dopo due 81 $(0{,}90)^2$, dopo tre 72,9 $(0{,}90^3)$ e cosí via, tendendo a zero; perciò l'insieme «tipo Cantor» dell'esempio (come quello

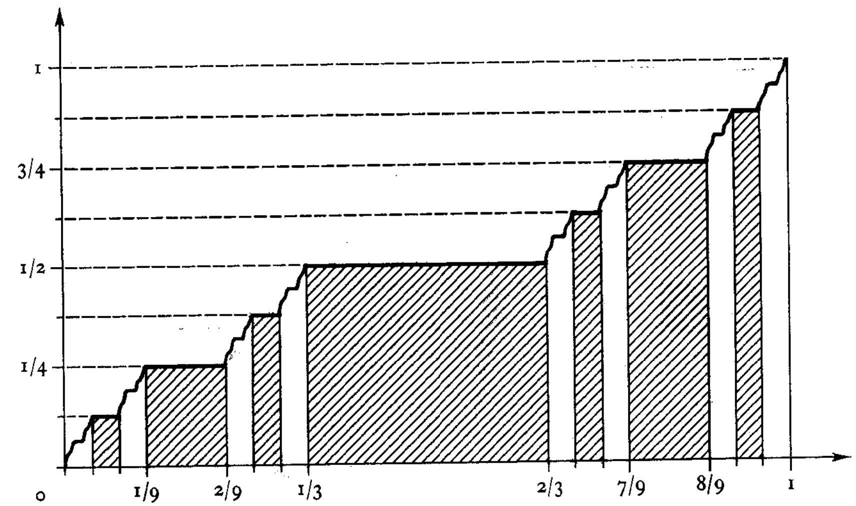


Figura 7.
Distribuzione di Cantor.

originale) ha misura nulla. (È, cioè, racchiudibile in un insieme di segmenti di lunghezza complessiva piccola quanto si vuole).

Come figura riesce piú armonica quella della distribuzione di Cantor (fig. 7), perché in essa viene tolto ogni volta il terzo centrale degli intervalli via via ottenuti. Si vedono (tratteggiati) i pezzi tolti: quello centrale tra 1/3 e 2/3, con altezza 1/2 perché la massa rimane divisa a metà fra il 1° e il 3° «terzo»; poi i due «noni» (parte centrale del 1° e del 3° «terzo»); poi i quattro «ventisettesimi», parte centrale dei quattro «noni» rimasti, e cosí il procedimento continua (all'infinito). Il bordo superiore di tali riquadri (completato al di sopra dei «ventisettesimi» rimasti bianchi con degli zig-zag che a rigore dovrebbero venir disegnati con infiniti zig-zag sempre piú piccoli), rappresenta la funzione di ripartizione della distribuzione di Cantor.

Usualmente, le distribuzioni sono del tipo continuo (a rigore: «assolutamente continuo») con densità $f(x)$, oppure discreto (con «pesi concentrati», p_h nel punto x_h, $p_h = F(x_h^+) - F(x_h^-)$; oppure combinazione dei due tipi (ed eventualmente anche del terzo tipo (Cantor)).

Tutto ciò occorreva dire, a questo punto, per entrare piú tecnicamente nell'argomento delle «medie»; e riprendiamolo da principio con l'intendimento di trattarlo in generale, per distribuzioni di qualunque tipo e medie di qualunque specie.

Conviene tuttavia cominciare dal caso piú semplice, quello della media aritmetica, che del resto darà l'idea valida per altri casi simili, e poi per altri ancora, meno simili.

La media aritmetica (semplice: pesi uguali!) è la somma degli x_h divisa per il loro numero, n: $m = (x_1 + x_2 + \ldots + x_n)/n$; se alcuni dei valori sono uguali (x_h ripetuto n_h volte), la stessa somma si scriverebbe $m = (n_1x_1 + n_2x_2 + \ldots + n_kx_k)/n$ $(n = n_1 + n_2 + \ldots + n_k)$, oppure $m = (p_1x_1 + p_2x_2 + \ldots + p_kx_k)$, con $p_i = n_i/n$ («normalizzando» i «pesi» p di modo che la loro somma sia già 1).

Questa somma si può interpretare e scrivere come un integrale di Stieltjes mediante la funzione di ripartizione $F(x)$ che nel presente caso (di distribuzione discreta), non varia che per i salti p_h nei punti x_h (come detto poco sopra).

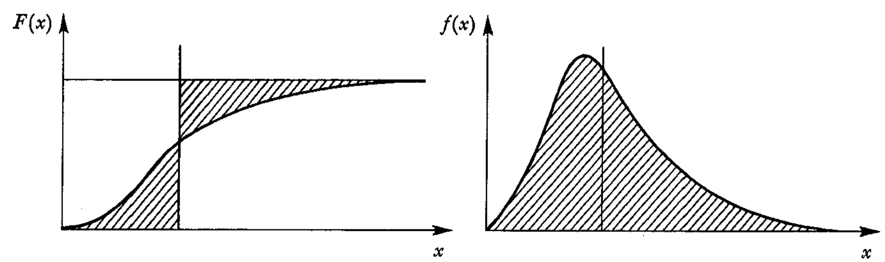


Figura 8.

Rappresentazioni grafiche della media aritmetica: nel primo caso, diagramma di ripartizione, essa è visualizzata dall'uguaglianza delle due aree tratteggiate; nel secondo, diagramma di frequenza, dal baricentro dell'area.

Nel caso di una distribuzione continua, con densità $f(x)$, la scrittura con l'integrale ordinario sarebbe $m = \int x f(x)\,dx$, che, scrivendo $f(x)\,dx = dF(x)$, equivale all'integrale di Stieltjes $\int x\,dF(x)$. La figura 8 (nelle due parti: inferiore e superiore) illustra visivamente il concetto informatore.

Tutto ciò era ovvio o quasi (e forse per questo stesso fatto difficile a seguirsi); ma ora abbiamo lo strumento per proseguire in modo agevole ad occuparci dei piú interessanti concetti e problemi sulle medie.

2.5. Le medie associative (teorema di Nagumo-Kolmogorov).

Le medie piú comuni, di cui abbiamo fatto cenno nel § 2.3 (come geometrica, armonica, quadratica, oltre, naturalmente a quella aritmetica), sono esempi di medie associative: di medie, cioè, che godono di questa comoda proprietà: la media non cambia se a gruppi di dati si sostituisce la loro media (sempre, beninteso, in quello stesso senso) attribuendole un peso uguale alla somma dei pesi dei valori in essa riuniti. Per fare un esempio banale: il tempo impiegato per percorrere un dato itinerario non varia se variano le velocità (per esempio, in due tratte in cui il percorso venga suddiviso), purché rimanga invariata la media armonica delle velocità, da cui dipende il tempo complessivo.

Tale proprietà è tanto ovvia che il rischio non è di non capirla bensí di ritenere ovvio che valga sempre. Per mostrare che ciò non è, basta un esempio molto concreto: quello della media antiarmonica, che, esprimendoci a parole, è la «media aritmetica di n grandezze positive x_i prendendo come *pesi* gli stessi valori x_i»; beninteso, la locuzione è scorretta, ma il senso è esatto. (Aritmeticamente: è la somma dei quadrati divisa per la somma dei valori, ossia anche la media dei quadrati divisa per la media aritmetica). Come interpretazione fisica, possiamo dire che essa dà la «lunghezza ridotta» di un pendolo composto, cioè la lunghezza di un pendolo semplice ideale (una massa puntiforme tenuta a distanza l dal fulcro con un'asta di peso trascurabile) che oscillasse con la medesima frequenza. E non è vero che, collegando due pendoli, il loro moto congiunto abbia la frequenza data dalla media antiarmonica delle rispettive lunghezze ridotte. (Al contrario dei due «ingredienti» di essa: la distanza fulcro-baricentro e il momento d'inerzia).

Quale sia la forma generale delle medie associative, lo dice il già menzionato teorema dimostrato indipendentemente e quasi contemporaneamente dal giapponese Mitio Nagumo e dal russo Antonij Kolmogorov: esse sono tutte e sole le «trasformate» della media aritmetica. Tali sono ad esempio la media geometrica, in quanto è la radice del prodotto: il prodotto dei valori $x_1, x_2, \ldots, x_n$ non varia se ad essi si sostituisce sempre la media geometrica x, cioè considerando x^n; la media armonica, perché il suo reciproco è la media dei reciproci; la media quadratica (per valori x_i positivi), perché la media dei quadrati x_i è il quadrato della media quadratica; e via dicendo.

Il concetto di media associativa è molto largo, perché si può considerare una tale media con riferimento a una qualunque funzione $\gamma(x)$ crescente (al

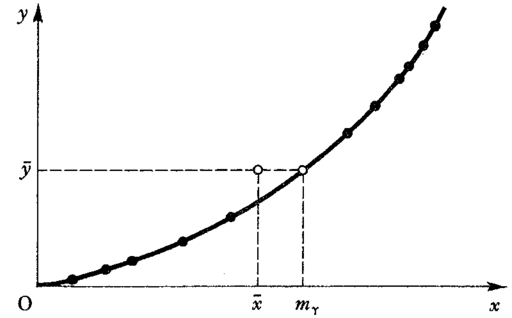


Figura 9.
γ-media e media aritmetica di una funzione $y=\gamma(x)$.

posto della $\gamma(x)=x$, prendendo ad esempio $\gamma(x)=1/x$, $\gamma(x)=x^2$, $\gamma(x)=\log x$, ecc., che dànno luogo alle medie menzionate: aritmetica, armonica, quadratica, geometrica). Della dimostrazione vale la pena di dare un'illustrazione geometrica che la rende intuitiva, senza ricorso a sviluppi matematici.

Nella figura 9, la curva è il diagramma $y=\gamma(x)$ della funzione rispetto a cui si vuole considerare la media (diciamo: «la γ-media»). Le masse p_i della distribuzione sono collocate sulla curva in corrispondenza alle ascisse x_i (per cui automaticamente le loro ordinate sono le $y_i=\gamma(x_i)$). Il baricentro (di coordinate $\bar{x}$ e $\bar{y}$, medie aritmetiche delle x_i e delle y_i coi pesi p_i) si trova naturalmente dalla parte all'interno della concavità della curva (se è sempre nello stesso senso). La γ-media delle x_i (secondo la definizione data) è invece il valore indicato in figura con m_γ, proiezione sull'asse x del punto della curva all'altezza del baricentro. È chiaro dalla figura (e ci limitiamo a farlo osservare senza ulteriori ragionamenti) che la γ-media è minore o maggiore della media aritmetica a seconda che il diagramma $y=\gamma(x)$ volga la concavità verso l'alto (come in figura) o viceversa. (Se vi fossero tratti concavi e convessi, si dovrebbe esaminare caso per caso).

Lo stesso criterio varrebbe per l'analogo confronto fra due medie associative relative a due diverse funzioni γ, diciamole γ_1 e γ_2. Basta disegnare la curva di equazioni parametriche $\xi=\gamma_1(x)$, $\eta=\gamma_2(x)$, e fare lo stesso confronto. Oppure ricondursi al caso precedente ponendo $\gamma(x)=\gamma_2(\gamma_1^{-1}(x))$.

Anche senza il confronto grafico, basta tener presente che «maggiore concavità relativa» corrisponde (localmente) a maggior valore del rapporto fra la derivata seconda di $\gamma(x)$ e la prima, $\gamma''(x)/\gamma'(x)$ (se, nell'intervallo che interessa, non s'inverte). In particolare, si può ricordare che, tra le «medie di potenze» – quelle ottenute da $\gamma(x)=x^n$ – la media cresce con l'esponente n, per cui in particolare valgono le disuguaglianze armonica < geometrica < aritmetica < quadratica < cubica (ecc.).

2.6. La variabilità.

L'indicazione di una media può dire «tutto ciò che interessa» al piú con riferimento a quell'unica circostanza con riguardo alla quale quel tipo di media è stato prescelto; lo si era detto fin dal principio ma è bene ripeterlo. Ripeterlo per continuare il ragionamento, e vedere cos'altro occorra secondo i casi, e in particolare nei casi piú abituali.

Dire che in un dato periodo tutti mangiano un pollo a testa (sottintendendo «in media») non può non apparire sarcastico al personaggio di Trilussa che si lamenta perché il pollo che lui non mangia «entra ne la statistica lostesso, perché c'è un altro che ne magna due». Ma, a parte ciò, se *effettivamente* tutti avessero il medesimo valore medio (per esempio, di statura, di robustezza, ecc.) la situazione sarebbe ben diversa e peggiore, perché la diversificazione di compiti e mestieri richiede qualità peculiari per ciascuno.

Una media non basta; possono aggiungere indicazioni utili delle medie diverse (che in certo modo dicono qualcosa di piú: ad esempio conoscendo sia la media aritmetica che quella quadratica se ne trae una misura della «variabilità»); precisamente una conoscenza della variabilità è la prima cosa per arricchire l'indicazione data da una media (anche se fosse quella piú significativa rispetto al carattere in questione). Un esempio lo abbiamo già visto nel § 2.1: età sposi 1972, anni $28,93 \pm 8,75$: «età media»$\pm$«scarto quadratico medio».

E completiamo subito l'accenno mostrando che è equivalente aggiungere all'indicazione della media aritmetica *o* la media quadratica *o* lo scarto quadratico medio. La media quadratica, indichiamola con m_q, è per definizione tale che $m_q^2 = \sum_h p_h x_h^2$, mentre lo scarto quadratico medio σ è tale che $\sigma^2 = \sum_h p_h (x_h - m)^2 = \sum_h p_h (x_h^2 - 2mx_h + m^2) = \sum_h p_h x_h^2 - 2m \sum_h p_h x_h + \sum_h p_h m^2 = m_q^2 - m^2$. Si può rendere visivamente la relazione ricordando il teorema di Pitagora: la media quadratica è (geometricamente) l'ipotenusa di un triangolo rettangolo di cui i cateti sono il valor medio m e lo scarto quadratico medio σ: $m_q^2 = m^2 + \sigma^2$.

Esistono anche altri «indici di variabilità», di cui non vale la pena di dire molto. Accenniamo soltanto alla «differenza media» nelle due varianti *con* ripetizione o *senza* ripetizione. Si tratta di fare tutte le differenze a due a due, $|x_i - x_j|$ in valore assoluto (n^2 se «con ripetizione», cioè se si contano anche le differenze nulle tra ogni valore e se stesso, $x_i - x_i$, e allora si divide la somma per n^2; altrimenti basta dividere la medesima somma (i termini che si omettono sono nulli) per $n(n-1)$). Si parla anche di differenza media quadratica (con o senza ripetizione); questi, peraltro, non differiscono da σ^2 se non per coefficienti fissi.

Oltre alla variabilità, ulteriori caratteristiche qualitative di un certo interesse sono quelle riguardanti l'asimmetria (*skewness*) e la *kurtosis*, che distingue (detto alla buona) il caso di addensamento in un pinnacolo o di tratto prolungato di densità abbastanza elevata. Come indici per tali due qualità vengono usati risp. i momenti 3° e 4° (rispetto al baricentro) rapportati alla variabilità, ossia dividendo il momento 3° per il cubo e risp. il momento 4° per la quarta

potenza dello scarto quadratico medio (preso – si può dire – per unità di scala): in tal modo infatti i detti «indici» risultano invarianti per modificazioni di «scala», e tali quindi da caratterizzare la *forma* in relazione alle due caratteristiche menzionate. (Si veda nel § 3.6, e graficamente sulla «mappa Boetti», una classe di distribuzioni la cui forma varia in base a tali parametri).

2.7. Alcune distribuzioni discrete.

Ci riferiamo ora particolarmente al caso in cui la variabile è un numero intero (come nell'esempio del numero di figli o di fratelli); ma potremmo anche avere altre successioni piú o meno irregolari di valori possibili discreti, come ad esempio le cilindrate dei modelli di auto attualmente in fabbricazione in un dato stabilimento, oppure la tassa di circolazione a seconda delle diverse categorie di cilindrata.

Può anche interessare (a chi piú e a chi meno) conoscere certi tipi di distribuzioni; piú importante è rendersi conto del diverso senso che hanno a seconda della natura dei fatti che considerano e delle conoscenze cui devono contribuire (magari, ed è il caso piú completo e complesso) nel consigliare per le decisioni da prendere in condizioni d'incertezza.

Per tener conto del ***respice finem***, occorrerà, sí, prestare attenzione ai dati e agli aspetti formali, ma soprattutto saper vedere cosa essi significhino o non significhino, cosa essi dicono o non dicono. In ciò si mescolano considerazioni metodologiche e probabilistiche che è impossibile presentare prima di entrare in argomento, ma che occorrerà introdurre man mano che se ne presta la necessità e l'occasione per evitare facili fraintendimenti in senso confusionario.

L'esempio piú semplice e significativo è forse quello in cui si chiede la frequenza (o la probabilità: al momento non sottilizziamo su distinzioni di cui sottolineeremo l'importanza) che la prima cifra (significativa, non zero) di un numero «qualunque» sia 1 o 2 o ... o 9. È chiaro che se pensiamo a tutti i numeri da 1 a 9 o da 1 a 99 o... da 1 a 999 999, ecc. la risposta è sempre 1/10 (per 1, come del resto per ogni altra cifra). Però... possiamo pur riferirci invece a un segmento diverso di numeri, per esempio da 1 a 1 999 999, e i numeri che cominciano per 1 sono piú della metà!

Qual è la risposta esatta? quale la spiegazione della non-univocità della risposta? Nulla, a priori, è «esatto» o «sbagliato», ma ogni risposta è piú o meno adeguata ad una certa problematica. La risposta piú «ragionevole», nel nostro caso, è quella che dà per la prima cifra le seguenti probabilità (differenze dei logaritmi decimali):

30,10%	per l'1	9,69%	per il 4	5,80%	per il 7
17,61	2	7,92	5	5,14	8
12,50	3	6,69	6	4,55	9

Perché è la risposta «piú ragionevole»? Perché è «invariante rispetto alla scala», considerandovisi «ugualmente probabili» gradini uguali «in percentuale». Il gradino dall'1 al 2 (raddoppio) non può essere ugualmente considerato che quello dal 4 al 5 (aumento del 25 per cento) bensí a quello dal 4 all'8 (e infatti risulta 9,69+7,92+6,69+5,80=30,10).

A titolo di curiosità: sembra che tale andamento della distribuzione della cifra iniziale sia stato notato dapprima come «stranezza» nella preferenza delle costanti fisiche per valori con cifra iniziale piccola. Fatto che non è «stranezza», ma cosa conforme a «naturalezza».

Il caso visto precedentemente è un esempio di come la scelta di una particolare distribuzione possa avvenire in base a un'intuizione probabilistica (di presumibile invarianza rispetto alla scala), conducendo direttamente a una distribuzione di probabilità. Si potrebbe realizzare una distribuzione statistica conforme ad essa in modo esatto (a parte l'arrotondamento a interi) con una sperimentazione «esaustiva», oppure approssimata in senso probabilistico (con una sperimentazione «casuale»).

Due esempi banali: la distribuzione uniforme e quella triangolare coi 100 numeri da 00 a 99. Con la sperimentazione esaustiva si prendono tutti i numeri e si distribuiscono nelle caselle da 0 a 9 a seconda della cifra delle unità (oppure delle decine): ne vanno 10 in ciascuna, realizzando in modo certo ed esatto la distribuzione uniforme. Analogamente, distribuendoli nelle caselle da 0 a 18 a seconda della somma delle cifre si ha la distribuzione triangolare. È istruttivo «vederlo» pensando i numeri scritti in quadrato (10 righe per decine 00, 10, ..., 90 di numeri incolonnati per cifra delle unità 0, 1, ..., 9): la numerosità maggiore (10) si ha per la somma 9 perché data dai numeri sulla diagonale (09, 18, 27, 36, 45, 54, 63, 72, 81, 90), le somme da 0 a 8 hanno numerosità crescente da 1 a 9 e quelle da 10 a 18 simmetricamente decrescente da 9 a 1 (trovandosi su parallele alla diagonale principale).

Ancora un esempio di «distribuzione statistica» ottenuta artificialmente come problema (se cosí si vuol dire) di «statistica aritmetica» (in cui cioè contiamo i casi considerati equiprobabili senza farne oggetto di estrazioni casuali).

Consideriamo i 100 000 numeri di cinque cifre da 00 000 a 99 999 e classifichiamoli a seconda di quante volte contengono una data cifra (per esempio il 9), oppure tre date cifre (per esempio il 4, 5 oppure 6); nel primo esempio le cifre specificate sono 1 contro 9, nel secondo sono 3 contro 7. Ecco le tabelle che indicano quanti dei 100 000 numeri da 00 000 a 99 999 contengono 0, 1, 2, 3, 4, 5 volte risp. *una* cifra prescelta (il 9, o qualsiasi altra) oppure le tre cifre prescelte, complessivamente (4, 5 oppure 6, o qualsiasi altra terna):

Cifre 9 contenute nel numero	
0	$1\cdot 9^5 = 59\,049$ (59,049%)
1	$5\cdot 9^4 = 32\,805$ (32,805)
2	$10\cdot 9^3 = 7\,290$ (7,290)
3	$10\cdot 9^2 = 810$ (0,810)
4	$5\cdot 9^1 = 45$ (0,045)
5	$1\cdot 9^0 = 1$ (0,001)

Cifre 4, o 5, o 6 contenute nel numero	
0	$1\cdot 7^5\cdot 3^0 = 16\,807$ (16,807%)
1	$5\cdot 7^4\cdot 3^1 = 36\,015$ (36,015)
2	$10\cdot 7^3\cdot 3^2 = 30\,870$ (30,870)
3	$10\cdot 7^2\cdot 3^3 = 13\,230$ (13,230)
4	$55\cdot 7^1\cdot 3^4 = 2\,835$ (2,835)
5	$1\cdot 7^0\cdot 3^5 = 243$ (0,243)

Notiamo con l'occasione (è bene allenarsi a vedere le possibili interpretazioni di un medesimo esempio o risultato in contesti diversi) che la distribuzione incontrata è quella stessa delle probabilità che un campione di cinque pezzi scelti a caso da una grande quantità in cui quelli difettosi siano il 10 per cento, ne contenga nessuno, oppure uno, o due, ..., o tutti e cinque.

2.8. Due digressioni.

Prima di passare ad altri esempi e riprendere la tematica avviata, è opportuno interromperla per due digressioni. I due esempi precedenti, in cui la numerazione va rispettivamente da 00 a 99 e da 00 000 a 99 999, vorrebbe anche richiamare l'attenzione sull'incongruenza di cominciare le numerazioni con l'1 anziché con lo zero (perdendo cosí uno dei pregi della numerazione in cifre arabe!) Infatti, a causa di tale slittamento, le centinaia *non* sono caratterizzate dalla cifra delle centinaia (da 000 a 099, da 100 a 199, ..., da 2700 a 2799, ... ecc., come apparirebbe ovvio a chiunque non fosse digiuno di selezione automatica, o anche manuale), ma occorrono prodigi di illogicità acrobatica per imporre l'assurdo di voler dire, ad esempio, che il 1900 non è il primo anno del «Novecento», bensí... l'ultimo dell'«Ottocento»!

Piú assurde ancora (e del tutto insulse) le complicazioni derivanti da una numerazione progressiva siffatta, ad esempio per banconote, buoni del Tesoro, repertori, codificazioni, classificazioni, ecc.

L'altra osservazione riguarda i tenaci e strani sofismi tendenti a far ritenere che un evento di probabilità molto piccola sia *impossibile*.

Nel precedente esempio, la tabella mostrava che la probabilità di estrarre un numero con cinque cifre tutte «9» (cioè il 99 999) era piccolissima (0,001 per cento); questo caso – direbbero molti – si può escludere perché è praticamente impossibile (o, piú semplicisticamente, «impossibile»). E magari lo si direbbe anche per quelli con quattro cifre «9» (con probabilità 0,045 per cento).

Quindi (si potrebbe concludere) se uno vuole comperare un biglietto di una lotteria faccia attenzione che il numero non sia «speciale» (con cifre tutte uguali o con qualunque cosa che uno veda come «peculiare») perché allora «non può uscire»! L'esempio piú classico e tenace di tali superstizioni è visibile nel lotto, con la preferenza per i numeri «ritardati» (non saprei dire se come corollario o controesempio) e con l'esclusione di giocate «troppo *speciali*» quali la cinquina 1-2-3-4-5, oppure «bruciate» (?!) come «la cinquina uscita sabato scorso» (perché... «è impossibile che la stessa cinquina esca due volte di seguito!»)

Un ragionamento del tutto analogo (non so se qualcuno lo faccia davvero) sarebbe che volendo comperare un biglietto della lotteria conviene acquistarlo a Roma perché i biglietti «fortunati» sono stati venduti in buona parte a Roma, e nessuno di quelli acquistati «nel suo paesello» invece ha vinto. Non sarebbe infatti un «miracolo» che il premio andasse a un villaggio dove fosse stato acquistato un solo biglietto?

Il sofisma consiste nel pensare che la «piccola probabilità» derivante a un gruppo di biglietti dal fatto di essere pochi si rifletta su ciascuno di essi facendolo meno probabile di quanto lo siano gli altri biglietti. Sarebbe come pensare che un individuo, cui fosse stato imposto un nome raro, per esempio Asdrubale, avesse per tale fatto minore probabilità di divenire capo dello Stato che se si chiamasse Luigi o Giuseppe o Giovanni. Quella che è minore è soltanto la probabilità che lo divenga un altro personaggio dal medesimo nome.

2.9. Il processo di testa o croce.

Nello stesso spirito dei casi esaminati nel § 2.7, consideriamo ora la distribuzione binomiale nel particolare caso di testa o croce. Anche qui (per sottolineare la distinzione, in questo caso e insieme per gli altri già visti) parliamo intanto solo di «numero di percorsi», per distinguere «sperimentazione esaustiva» da «sperimentazione statistica» e da «valutazione di probabilità».

Lo schema della figura 10 indica i 64 percorsi che rappresentano i 64 possibili andamenti di guadagni e perdite in 6 colpi a testa o croce per chi vince o perde 1 lira ad ogni colpo a seconda della faccia.

Il percorso di ogni freccia indica un guadagno $(+1)$ se verso destra e una perdita (-1) se verso sinistra. I numeri nello schema indicano quanti percorsi portino ad ogni punto di incrocio (e sono dati, evidentemente, dalla somma dei due sovrastanti: le dieci vie che portano ad un guadagno $+1$ dopo cinque colpi sono le sei che in quattro colpi avevano portato ad un guadagno 0 (seguite da una vincita) piú le quattro che avevano portato a un guadagno $+2$ (seguite da una perdita).

Nel caso di testa o croce (probabilità uguali: 1/2 e 1/2) tutti i percorsi sono ugualmente probabili; perciò la probabilità di un guadagno 6 è 1/64, di un guadagno 4 è 6/64, di un guadagno 2 è 15/64 (e lo stesso per le uguali perdite) mentre 20/64 è la probabilità di parità (3 colpi a favore e 3 contro).

Questa è la distribuzione bernoulliana (da Bernoulli) nel caso piú semplice (quello di probabilità 1/2 e 1/2 ad ogni colpo, indipendentemente dal risultato degli altri). Se le probabilità fossero diverse, diciamole p e $q=1-p$, come nel caso di estrazioni (sempre con reimbussolamento) da un'urna che contiene

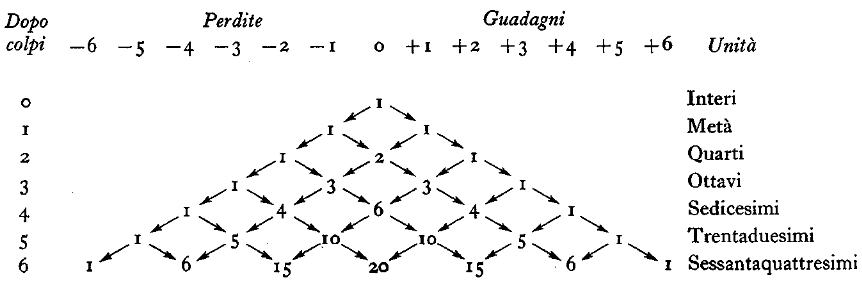


Figura 10.
Distribuzione binomiale nel caso particolare di testa o croce.

palline bianche e nere in proporzione diversa, ogni passo verso destra (nello schema) avrebbe probabilità p ed ogni passo verso sinistra probabilità $q=1-p$.

Le due tabelle a p. 1206 illustrano un esempio dei due casi.

È interessante considerare subito anche un caso limite di tali distribuzioni bernoulliane: la distribuzione di Poisson (già incontrata nelle osservazioni circa le variazioni dei nati di giorno in giorno, nel § 1.3); essa viene anche chiamata «legge degli eventi rari». Consideriamo ad esempio fatti di cui ogni anno se ne verificano in genere nessuno o pochi: supponiamo circa 5 in media, senza motivi di diversità che spieghino le variazioni, da considerarsi perciò «accidentali». Quali probabilità attribuiremmo al fatto che, in un anno fissato, se ne verifichino 0, o 1, o 2, ecc.? La probabilità che se ne verifichino h è $p_h=e^{-5}\cdot 5^h/h!$, ossia $p_0=e^{-5}=6{,}74$ per mille, $p_1=5e^{-5}=33{,}7$ per mille, $p_2=25e^{-5}/2=84{,}2$ per mille, ecc. (ad ogni passo si moltiplica per 5 e si divide per h: quindi la probabilità è massima per $h=4$ e $h=5$ e poi diminuisce sempre piú rapidamente).

Tutti sapranno probabilmente che questo schema si chiama «triangolo di Tartaglia» (o di Pascal, ma la priorità appartiene ai Cinesi da secoli). Forse anche sapranno che i suoi elementi (chiamati «coefficienti binomiali», perché entrano nello sviluppo di $(a+b)^n$) si indicano col simbolo $\binom{n}{m}$: ad esempio $\binom{6}{2}$, o $\binom{6}{4}$, è il 15 che si trova nella *riga* 6 al *posto* 2° e 4°. E aggiungiamo, dato che ci servirà comunque introdurre il «fattoriale», che il coefficiente binomiale $\binom{n}{m}$ è dato da $n!/m!(n-m)!$ (ad esempio $\binom{6}{2}=6!/2!\,4!=720/(2\times 24)=$ $=15$; dove per «fattoriale di n» s'intende il prodotto di tutti gli interi fino ad n: $1!=1$, $2!=2$, $3!=6$, $4!=24$, $5!=120$, $6!=720$, e cosí via, in progressione sempre piú rapida (circa come $2{,}5(0{,}37\,n)^n$: formula di Stirling in forma «arrotondata»).

Ad esempio, $10!=3\,628\,800$ è il numero di numeri che si possono scrivere usando una e una sola volta le dieci cifre (si badi che lo zero può anche essere al primo posto, il che significa che vanno compresi i numeri «di nove cifre tutte diverse ed escluso lo zero»; volendo escluderli, essi sono $9!=362\,880$, cioè, come era ovvio, il 10 per cento: altrettanti sono infatti quelli che cominciano per 1, 2, ..., 9).

La frequenza (o percentuale) di numeri con cifre tutte diverse fra quelli di dieci cifre (o meno: nel senso che contiamo da 0 000 000 000 a 9 999 999 999) è dello 0,36288 per 1000. (Con le solite riserve già ripetute, questa è anche la probabilità di estrarre una e una sola volta ogni pallina in dieci estrazioni con reimbussolamento da un'urna che ne contiene dieci, con le cifre 0, 1, 2, ..., 9).

Utilizziamo, infine, quest'altro esempio, per considerazioni che si ricollegheranno a un problema su testa o croce. Dei 10^{10} numeri non c'interessano ora le cifre ma solo il fatto che esse siano pari o dispari; pari significhi testa e dispari croce.

In questa collezione di tutti i risultati possibili, ove tutti i risultati differenti che corrispondono ad un medesimo risultato di testa o croce sono nel medesimo numero (5^{10}), quanti sono quelli che corrispondono a testa nel primo colpo? Ovviamente, la metà. E, tra gli altri, quanti corrispondono a testa nel secondo colpo? ovviamente, la metà di detta metà, cioè un quarto del complesso. E poi, per il terzo, quarto colpo, ecc., se si prosegue finché testa appaia per la prima volta, ciò avverrà sempre nella metà dei casi residui, e cioè, rispetto al complesso, in misura ogni volta dimezzata: $1/2$ al primo colpo, $1/4$ al secondo, $1/8$ al quarto, e poi via via $1/16$, $1/32$, e via dicendo. (Nell'esempio, ciò vale fino al decimo colpo, ma il ragionamento si può continuare all'infinito pur di proseguire con decimi, centesimi, millesimi, ecc. finché si vuole). Abbiamo un esempio di distribuzione *geometrica* (cioè in progressione geometrica): qui di ragione $1/2$, ma potrebbe essere qualsiasi; se, ad esempio, invece di interpretare «testa» con cinque cifre ne avessimo associate solamente due, o invece sette, la progressione sarebbe stata di $(2/10)^h$ risp. $(7/10)^h$. Pensando che le estrazioni si succedano a intervalli di tempo uguali, quello che abbiamo considerato come «numero di colpi fino al successo» si può interpretare (secondo il termine corrente) come «tempo d'attesa» fino al primo successo.

2.10. Alcune distribuzioni continue.

La distinzione, cosí netta dal punto di vista matematico, tra caso discreto e caso continuo, lo è molto meno in pratica, come già rilevato fin da principio (cfr. §§ 1.2, 1.3 e *passim*), in particolare discutendo del significato di 'densità'. Nello stesso spirito, senza dilungarci, va intesa qui la nozione di distribuzione continua.

La distribuzione uniforme discreta (n valori equidistanti con «peso» $1/n$),

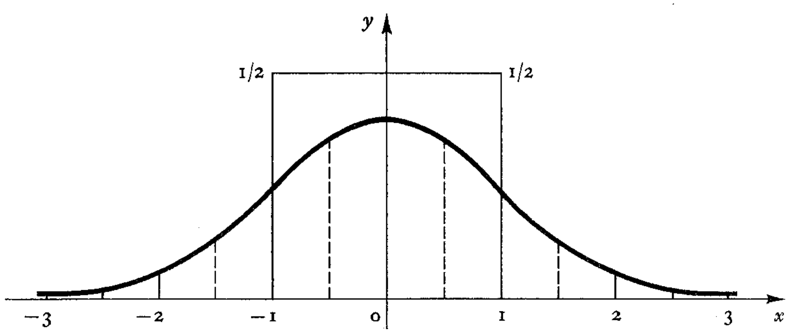


Figura 11.
Distribuzione normale ridotta ($m=0$, $\sigma=1$): curva della densità. Le suddivisioni indicate (0, ± 1, ± 2, ± 3) corrispondono a σ, 2σ, 3σ; in ± 1 si hanno due flessi, tra i quali il profilo della densità è convesso, mentre al di fuori è concavo. Il rettangolo di altezza $1/2$ mostra, per confronto, la distribuzione uniforme sul tratto $(-1, +1)$. La scala verticale è stata quadruplicata per evitare che la curva appaia (come in realtà è) molto appiattita, con andamento poco percettibile.

se n è molto grande sarà naturale considerarla uniforme continua (spesso sarebbe impossibile o illusoria anche la precisione necessaria per individuare il valore arrotondato piú prossimo). In tal caso si considera costante la densità sull'intervallo interessato, sia (a, b), avendosi $f(x)=1/(b-a)(a\leq x\leq b)$. Un'espressione come $(a\leq x\leq b)$ significa 1 quando è vera e 0 quando è falsa; nel presente caso, ad esempio, dice che è $f(x)=1/(b-a)$ quando x è tra a e b, *ed è* $f(x)=0$ *al di fuori*, dove $(a\leq x\leq b)=0$.

Dalla distribuzione uniforme (e sia, per semplicità, sull'intervallo $(0, 1)$, cioè $a=0$, $b=1$; quindi $f(x)=1$ in $(0, 1)$ ed $f(x)=0$ fuori), moltiplicando la densità per x (o $1-x$) e normalizzando (moltiplicando per il coefficiente K che rende $=1$ l'area – o probabilità – totale) si ottiene una distribuzione triangolare; moltiplicando per x e per $(1-x)$ si ha la distribuzione di densità $f(x)=Kx(1-x)$ (un arco di parabola, nulla agli estremi); piú in generale, con esponenti qualunque, $f(x)=Kx^\alpha(1-x)^\beta$, si hanno le distribuzioni Beta. In particolare, per $\alpha=\beta=1$ ($f(x)=Kx(1-x)$) la curva è un arco di parabola; per $\alpha=\beta=1/2$ è una semicirconferenza; ecc. Gli esponenti possono essere anche negativi (fino a -1 escluso): particolarmente interessante il caso $\alpha=\beta=-1/2$ (distribuzione «Arcoseno»): è la proiezione sulla diagonale di «un punto "scelto a caso" sulla circonferenza», che interviene in problemi d'interesse teorico e pratico (cfr. anche § 3.6).

La distribuzione binomiale, se si fa crescere n e si modifica la scala opportunamente, si avvicina sempre piú alla distribuzione normale (o gaussiana): $f(x)=Ke^{-x^2/2}$, con $K=V\sqrt{2\pi}$ (fig. 11). Questa è la formula per il caso di media $=0$ e scarto quadratico medio $=1$; nel caso generale (media $=m$, sc. q. m. $=\sigma$) basta sostituire x con $(x-m)/\sigma$ e il K di conseguenza. Avvertiamo qui, una volta per sempre, che il valore di K nelle varie formule è sempre quello che occorre per la normalizzazione (cioè perché la massa totale sia 1; o analoghe condizioni ovvie per altri casi).

La distribuzione geometrica (valori 1, 2, 3, 4, ecc. con probabilità $1/2$, $1/4$, $1/8$, $1/16$, ecc. per «1ª testa al 1°, 2°, 3°, 4° colpo», ecc.), si trasforma analogamente nella distribuzione esponenziale se pensiamo che la diminuzione avvenga per scatti piú piccoli e piú frequenti (ad esempio 100 del 7 per mille circa).

L'analogo problema di quando uscirà la seconda testa (o la terza, ecc.), per il quale, nel caso discreto, la risposta era data da coefficienti binomiali, nel continuo riesce molto piú semplice: la distribuzione del tempo di attesa per la seconda, terza, n-esima testa ha la densità $f(x)=Kxe^{-x}$, Kx^2e^{-x}, in generale $Kx^{n+1}e^{-x}$ (distribuzione Gamma).

2.11. Diagramma di graduazione e di concentrazione.

Vi sono altri aspetti, in una distribuzione, che possono avere interesse a seconda dell'argomento cui si riferiscono, e che possono esser messi piú direttamente o piú efficacemente in luce mediante rappresentazioni geometriche appropriate.

L'esempio piú interessante è quello di due rappresentazioni grafiche, strettamente collegate tra loro, che si prestano, e sono impiegate, soprattutto ad illustrare in modo visivamente espressivo la distribuzione dei redditi (e a ciò faremo riferimento per concretezza).

Nella figura 12 il diagramma a destra (*b*), detto «curva di graduazione», ha un significato immediato se si pensa un momento di riferirci alla statura: è il profilo che si avrebbe disponendo tutti gli individui in riga ordinati secondo la statura in ordine crescente. Nel caso che c'interessa, nulla cambia tranne che l'altezza attribuita ad ogni individuo non è la statura bensí il reddito; anzi, qualcosa cambia, ma in meglio (come significatività) perché l'area (tratteggiata) indica il reddito totale, e la parte poggiante su un qualunque tratto della base indica il reddito totale di quel segmento di individui. Il reddito medio (livello m nella figura) è dato da quella orizzontale per cui è uguale l'area della parte tratteggiata che vi sta al di sopra e quella dell'area non tratteggiata che vi sta al di sotto; ciò significa, infatti, che distribuendo a coloro che non raggiungono la media ciò che per altri eccede la media si avrebbe l'uguaglianza assoluta.

Questa operazione, solo concettualmente possibile nel caso della ricchezza, non è neppure scherzosamente concepibile per la statura; nessuno direbbe che 1000 individui hanno «una statura complessiva» di (per esempio) 1673 m, per cui chi ha 4,58 cm in piú di 167,3 deve cederli e chi ha bisogno di 7,14 deve riceverli.

Il diagramma a sinistra (*a*), detto «curva di concentrazione» (di Lorentz), indica la frazione $G(y)$ del reddito complessivo posseduta dalla frazione y (per esempio, come in figura, verticale tratteggiata, circa $y = 73$ per cento) degli individui di minor reddito. Nella figura, si tratta degli individui con reddito inferiore alla media, indicata con m; si noti come, in corrispondenza a ciò, il punto corrispondente sulla curva di concentrazione abbia tangente inclinata di 45°; in ogni punto tale tangente ha inclinazione indicante il rapporto rispetto alla media del reddito degli individui che si trovano in quel tratto.

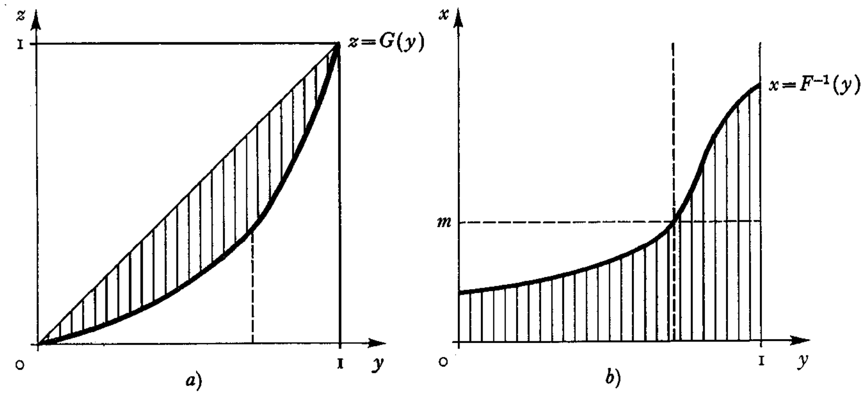


Figura 12.
Curve di concentrazione (*a*) e di graduazione (*b*).

Sull'argomento sono stati fatti, naturalmente, moltissimi studi, sia descrittivi ed empirici, sia come tentativi di sistemazioni teoriche. Informazioni sintetiche del primo tipo si esprimono mediante indici; quello piú espressivo (perché non legato a ipotesi specifiche sulla forma della distribuzione) è il *rapporto di concentrazione* (Gini), dato, in figura, dall'area tratteggiata tra la diagonale e la curva (prendendo come unità quella del triangolo sotto la diagonale). Tale rapporto è zero nel caso di equipartizione (allora infatti il diagramma coincide con la diagonale) ed è 1 nel caso di disuguaglianza massima (reddito nullo per tutti tranne uno: l'area tratteggiata comprende tutto il triangolo).

I tentativi di sistemazione teorica si basavano sull'idea che la forma della distribuzione dei redditi rispondesse ad esigenze e conseguenze del sistema economico traducibili in «leggi» o quasi e formulabili analiticamente trovando la forma matematica della funzione $G(y)$. La formula proposta da Pareto corrisponde all'espressione di G della forma $G(y) = 1-(1-y)^{\delta}$; Pareto usa però un altro indice (α, corrispondente teoricamente a $\delta/(\delta-1)$, ma in pratica alquanto diverso perché i valori numerici si ottengono con metodi d'interpolazione non identici). Dice al riguardo Feller che, al tempo di Pareto, si pensava (piuttosto *naïvely* da un punto di vista statistico moderno) che la distribuzione dei redditi dovesse avere una coda con densità dell'ordine di grandezza $Ax^{-\alpha}$ per $x \to \infty$.

L'ottocentesca fiducia in siffatte regolarità (attribuibili o meno a «mani invisibili») sembra però riscuotere sempre minore credito. Mani «invisibili» (fino a un certo punto) ce ne sono molte, ma non sembra affatto si preoccupino del benessere generale dell'umanità e della conservazione della vita nella biosfera.

3. *Distribuzioni di probabilità.*

3.1. Dalle frequenze alle probabilità.

Finora abbiamo sempre parlato di frequenze, cioè di dati oggettivi, accennando talvolta alla probabilità solo per avvisare che si tratta di qualcosa di diverso di cui si potrà parlare solo riprendendo il discorso (anche se, ed anzi appunto perché, frequenze e probabilità sono nozioni parallele che è facile ma rovinoso confondere). Proprio per illuminare tale differenza, con riferimento alle distribuzioni, gli esempi illustrati nei §§ 2.7 e 2.8 mostravano come certe distribuzioni potevano rappresentare distribuzioni di frequenze; però (attenzione!) nell'ipotesi *non* di una sperimentazione *casuale*, bensí di una sperimentazione *esaustiva* in cui si realizzasse una e una sola volta ciascuno dei risultati possibili (che circostanze di «simmetria» possono indurre a giudicare «ugualmente probabili»).

Ma... che cosa significa «ugualmente probabili»? Porsi questa domanda significa sostanzialmente chiedersi: «Che cosa è la probabilità?», e «Che cosa sono gli "eventi" cui la si riferisce?»

Di solito si vogliono far passare per «definizioni» la proprietà di additività (per eventi incompatibili), col corollario che la probabilità dell'unione di m tra gli n eventi di una partizione in eventi ugualmente probabili è m/n, oppure l'identificazione della probabilità con il «limite della frequenza» quando il numero delle prove si fa tendere all'infinito. Sono circoli viziosi: frasi che non hanno alcun senso se non si è già dato un senso a «probabilità», e che, quando lo si fosse dato, sarebbero espressioni manchevoli di teoremi che richiedono di venire formulati correttamente, non «alla carlona». Le due pseudodefinizioni corrispondono rispettivamente alla «concezione classica» e a quella «empirica» (o «statistica»), e si dicono «oggettivistiche» perché ignorano o negano l'apporto del giudizio umano nella valutazione.

La terza risposta – quella *soggettivista* che corrisponde alla concezione qui adottata – differisce dalle precedenti non solo in quanto riconosce e rivendica il carattere soggettivo della nozione di probabilità, ma anche per delle precisazioni conseguentemente necessarie per evitare ambiguità e fraintendimenti e nonsensi altrimenti inevitabili. Eccole, in sintesi.

Per «evento» intendiamo sempre «un caso unico ben specificato»: non si potrà parlare di «prove di uno stesso evento», ma si dovrà eventualmente dire che certi eventi sono «prove» di un medesimo «fenomeno» per far allusione a caratteri esteriori comuni, però senza con ciò implicare alcuna particolare ipotesi (come uguale probabilità, o indipendenza, od altro) che, se del caso, andrà esplicitamente dichiarata e specificata.

La probabilità di E non «esiste» di per sé, ma solo come misura del grado di fiducia (speranza, timore) nel suo avverarsi da parte di un dato individuo, in un dato istante.

Possiamo indicarla $P(E)$, ma sottintendendo tutto ciò, oppure $P_i(E|H)$ per dire che si tratta della valutazione dell'individuo i fatta subordinatamente ad H; ciò viene usualmente sottinteso se coincide con «tutto ciò che egli conosce attualmente», e va specificato se include condizioni ulteriori, o «ipotesi», nel senso che la scommessa andrebbe *annullata* qualora la condizione H non si fosse verificata.

Ed ecco il significato «operativo» che rende tale definizione non un *flatus vocis* bensí un'asserzione impegnativa: la probabilità di un dato evento E, per un dato individuo, è il prezzo $p=P(E)$ da lui giudicato «equo» per una lira (da pagare o ricevere) in caso che E risulti vero. In sostanza, uno o guadagna $1-p$ o perde p (il che corrisponde, nel gergo degli scommettitori, «dare a p contro $1-p$»). Tale significato (pratico e chiaro) è l'unico avente senso (mentre le valutazioni basate su giudizi di equiprobabilità o sull'osservazione di frequenze, se valide, non fanno che aiutare, in certi casi, nel fare tale valutazione).

Per evitare l'asimmetria di «scommesse» (a senso unico, sia pure scelte dall'«altro», con qualche garanzia ma anche con rischi) conviene applicare criteri basati su di una «regola di penalizzazione appropriata» (*proper scoring rule*) che rende vantaggioso per ciascuno esprimere sinceramente la propria opinione. Una regola «appropriata» è appunto una regola congegnata in modo che ciò avvenga. La piú semplice e nota regola del genere è quella di Brier: chi valuta $P(E)=p$ viene penalizzato di $(1-p)^2$ se E si verifica e di p^2 se si verifica «non-E» (e analogamente sono costruite le penalizzazioni per il caso di partizioni in tre o piú eventi).

Una volta dato – in tal modo (e altri non ce n'è!) – un significato operativo (cioè concreto, pratico, alieno da metafisicaggini) alla probabilità, una volta usciti dal pelago del vaniloquio e approdati sul solido terreno del pensiero pragmatico, del linguaggio atto a dire ciò che c'è da dire, tutto diventa chiaro, concreto, in senso buono «banale». Perfino le seppellite pseudodefinizioni, adottandole nel ruolo di criteri ausiliari talvolta utili, acquistano validità e significatività.

3.2. Rapporti tra probabilità e frequenze, e viceversa.

Tra le probabilità e le frequenze esiste una grande varietà di rapporti in entrambi i sensi: un complesso assai piú ricco e significativo di quella pretesa identificabilità che tutto confonde e contorce in una specie di «commedia degli errori».

Non è qui il luogo idoneo per approfondire l'argomento; è necessario, tuttavia, dare concisamente quelle indicazioni, e fare quelle osservazioni, che consentiranno una comprensione sia pur solo qualitativamente corretta della natura dei problemi, delle direttive che a tal fine è necessario seguire, della fallacia di certe idee distorte che spesso inavvertitamente s'insinuano, vuoi per contagio e vuoi per generazione spontanea: basta spesso un minimo fraintendimento per andare totalmente fuori strada, cosí come imboccando una corsia sbagliata su uno svincolo autostradale.

Forse, alla base di tutte le fallacie sta un abito mentale rozzo, insufficiente a far distinguere «previsione» da «predizione» nel modo netto che è dovuto e necessario. La stessa necessità di adeguatezza dei metodi alla natura e allo scopo di una ricerca o di una decisione (già sottolineata parlando delle medie secondo il concetto di Chisini) è esigenza ancor piú essenziale e piú delicata nei problemi ove interviene l'incertezza. Ragionare in condizioni d'incertezza non può significare altro che ragionare in termini di probabilità, in modo che è soggettivo e va riconosciuto come tale anche se tiene conto con la massima attenzione e obiettività dei fatti e dati oggettivi che si ritengono rilevanti. (Come è doveroso fare!)

Anche la probabilità di estrarre palla bianca da un'urna dipende non dalla proporzione effettiva di palline bianche e nere (se non è conosciuta con certezza) ma è la *previsione* di essa basata su quel che si sa (e quel che non si sa) riguardo al modo in cui l'urna è stata riempita e/o al risultato di estrazioni eventualmente già fatte e del cui esito si sia venuti a conoscenza. E l'effetto di ogni nuova estrazione è diverso a seconda del grado di affidamento che uno dà alla stima, momento per momento, della numerosità delle palline bianche e nere. Se egli ne è certo (e le estrazioni si fanno senza reimbussolamento), è certo che la probabilità del colore estratto diminuisce perché è rappresentato da una pallina in meno. Se la sua conoscenza è molto vaga, l'uscita di una pallina rafforza le ipotesi favorevoli a una maggiore presenza di palline di quel colore.

Nel caso di estrazioni con reimbussolamento, nel caso di composizione iniziale non nota (e supponendo uniforme la distribuzione iniziale di probabilità fra tutte le proporzioni possibili), è chiaro e ben noto che ci si avvicina, nella stima, alla frequenza osservata. Ma ciò non va inteso come un «criterio oggettivo», una conferma di una certa composizione (che pure di per sé è oggettiva), bensí come effetto di un ragionamento induttivo, cioè del ragionamento bayesiano. (Da Thomas Bayes che lo formulò in uno scritto che fu poi pubblicato nel 1764, tre anni dopo la sua morte).

Dopo alternanze di favore e sfavore (entrambi affetti da una certa superficialità), la concezione bayesiana sembra ora sulla via giusta per affermarsi.

3.3. Distribuzioni di probabilità e distribuzioni di frequenza.

L'argomento di cui abbiamo ora ad occuparci (e che potrà servire come esempio per tutte le altre situazioni analoghe) consiste nell'esaminare in quale misura è da attendersi che si scostino tra loro l'istogramma delle probabilità (quale considerato negli esempi del § 2.7) e quello delle frequenze.

Per fissare le idee, ci riferiremo all'esempio sulla frequenza di numeri di cinque cifre (da 00 000 a 99 999: a tali effetti anche gli zeri iniziali vanno contati) contenenti la cifra 9 risp. 0 volte, 0 1 0 2 0 3 0 4 0 5. Le frequenze che si avrebbero in un'estrazione «esaustiva», cioè estraendo ogni numero una e una sola volta, sono quelle indicate nel § 2.7. Se però peschiamo ogni volta «a caso» uno dei 100 000 numeri (o li «generiamo» con un programma «casuale»), le frequenze di numeri con 0, 1, 2, 3, 4 o 5 cifre «9», pur approssimandosi per solito a quelle «teoriche», se ne scosteranno piú o meno, in piú o in meno (il totale rimanendo fissato in 100 000). Ma di quanto?

Tutto può accadere (perfino che si peschino sempre numeri «uguali» quanto a numero di «9»), ma in genere è prevedibile che gli scostamenti delle frequenze «sperimentali» dalle frequenze «teoriche» (del caso esaustivo) abbiano un ordine di grandezza «ragionevole». Ciò lo dice, probabilmente, quel «senso comune» ormai assimilato (forse anche in misura esagerata, vicina a una nuova specie di superstizione), grazie soprattutto alla pratica di sondaggi, collaudi su campioni, ecc.

Cerchiamo di chiarire la conclusione senza sviluppi e ragionamenti teorici, ma indicando i risultati e presentando un'espressiva immagine analogica.

Cominciamo da quest'ultima, che è atta a dare una visione intuitiva della differenza tra soluzione nel caso esaustivo e nel caso di sorteggio (nonché dell'influenza della numerosità del campione sorteggiato).

Nell'istogramma, le aree corrispondono alle probabilità, nonché alle numerosità, dei sei tipi di numeri; si potrebbe rendere visibile tale numerosità riempiendo ogni colonna di tanti puntini (disposti secondo uno schema di quadrettatura) rappresentanti ciascuno un «individuo». Cosa succede se, invece, i punti vengono scelti «a caso», o vi cadono «a caso» come gocce di pioggia con uguale densità di probabilità su tutto l'istogramma ma con le irregolarità o fluttuazioni dovute alla casualità del fenomeno? (Il fatto che in aree piú o meno piccole ci si possano attendere «irregolarità» è esattamente il medesimo illustrato parlando nel § 1.3 del numero di nascite per minuto).

Per tale motivo c'è da attendersi – come di fatto in genere avviene in casi di questo tipo – che la distribuzione ottenuta da scelta «casuale» risulti, sí, grosso modo simile a quella delle probabilità (o alla scelta esaustiva), però con divergenze (naturalmente, ove in piú e ove in meno, dato che il totale è per ipotesi invariato).

Qualitativamente, l'apparire di scostamenti delle frequenze dalle rispettive probabilità (fatto ovvio di per sé) viene illuminato dal paragone della pioggia; ma, quantitativamente, si può dire di quale entità abbiano ad essere tali scostamenti?

Possiamo farlo (per ora) a titolo di notizia. Se la numerosità della popolazione (totale) è N, e la probabilità per ciascun individuo di appartenere allo scompartimento (o colonna) h-esimo è p_h, in media (o «in previsione») il numero degli individui ivi appartenenti è Np_h; ma quello che interessa ora è proprio l'entità dello scostamento «casuale» in piú o in meno, dovuto all'aleatorietà delle estrazioni o alla negligenza di Giove Pluvio che non si cura di far cadere le gocce secondo una quadrettatura perfetta.

La conclusione è graziosa: lo scarto (in piú o in meno) ha «ordine di grandezza» $\sqrt{Np_hq_h}$ (ove $q_h = 1 - p_h$), e tale formula ha un'espressiva interpretazione geometrica in quanto $\sqrt{p_hq_h}$, disegnando il semicerchio $y = +\sqrt{x(1-x)}$, ne è l'ordinata nel punto $x = p_h$ (come in quello simmetrico, $x = q_h$).

Collegando tali calcoletti con l'immagine concreta delle gocce di pioggia, si potrebbe dire che tale regola è quella che Giove Pluvio dovrebbe seguire per «imitare il Caso» se... non si fidasse del Caso. Ma forse si fiderebbe ancor meno di se stesso, rendendosi conto, o avendolo appreso da probabilisti come Henri Poincaré ed Emile Borel, che «c'est très difficile de imiter le hasard!» (Occorre aggiungere che queste frasi sul «Caso», un tempo comuni nelle discussioni sulle probabilità e venate di tinte metafisicheggianti, sono ormai giustamente obsolete. Tuttavia, in pratica, si dicono «casuali» le successioni, per cosí dire, «che godono di certe regolarità generiche ma di nessuna regolarità specifica»; volendo considerare tale terminologia bisogna contentarsi di accettarla come cosa assai vaga cui non si potrebbe dare un senso meno impreciso senza ucciderla).

3.4. Probabilità valutate in base a frequenze. Scambiabilità.

Secondo certe vedute la probabilità viene addirittura definita come frequenza, o come «valore teorico» della frequenza, o «limite della frequenza al crescere del numero delle prove» di un «medesimo evento» (in questa terminologia 'evento' viene inteso in senso generico, quello per cui qui, ad evitare ambiguità, usiamo il termine 'fenomeno', e si sottintende, in genere, che tali «prove» siano «indipendenti» e di probabilità inizialmente «incognita», ma «costante», che si imparerà a valutare e aggiornare in base alle frequenze via via osservate.

Sarebbe un nonsenso voler criticare questi nonsensi: il procedimento, se interpretato correttamente, è corretto, ma l'interpretazione no: è *metafisica.* Fa sembrare che esista una «probabilità assoluta» (l'Incertezza della «Natura»?!), a noi sconosciuta, e di cui le nostre definizioni e procedure tentano di scoprire il valore arcano e inaccessibile. Invece non solo si deve, ma *si può*, giungere alle stesse conclusioni, e «naturalmente», senza cioè avventurarsi nel vuoto della metafisica.

Basta basarsi sulla nozione di «scambiabilità».

La nozione di scambiabilità sostituisce fedelmente (e tuttavia dandole un'interpretazione ineccepibile) la disgraziata e contraddittoria dizione sopra riferita («eventi indipendenti con probabilità costante ma incognita»). Il modo piú semplice di esprimerla è il seguente: in questo momento (con le informazioni che abbiamo) diamo a tutti gli eventi E_i che consideriamo («prove di uno stesso fenomeno», se vogliamo dire una frase innocua, forse utile per «fissare le idee») una stessa probabilità; non solo, ma lo stesso vale per tutti i prodotti a due a due $p(E_iE_j)$, e cosí per i prodotti a tre a tre, ecc.

Bastano queste premesse (immuni da ogni metafisicaggine quale la «probabilità costante ma incognita») per condurre esattamente alle stesse conclusioni, e perfino a quella criticata «probabilità incognita» intesa non in tale senso privo di senso bensí come «il limite della frequenza». Se ne determina la distribuzione di probabilità in base alla successione di valori $P(E_i)$, $P(E_iE_j)$, $P(E_iE_jE_k)$, ecc. (che ne costituiscono i «momenti»: cfr. § 2.6).

Per dire meglio, sopprimiamo anche la dizione (priva di senso) di «limite della frequenza» (dizione che presuppone di poter non solo fare ma concludere una infinità di prove constatando poi se la frequenza abbia effettivamente ottemperato all'obbligo di tendere a un limite, e quale sia!)

Ciò che dobbiamo indicare è la distribuzione di probabilità che attribuiamo alla frequenza su un gran numero di prove (asintoticamente, al limite per $n \to \infty$; praticamente, per un n ritenuto sufficientemente grande); sia $F(x)$ la funzione di ripartizione, e supponiamo (cosa inessenziale, utile solo perché consente di esprimerci in modo piú semplice) che sia derivabile; cioè esista la densità $f(x)$. (Possiamo allora scrivere $dF(x)$ nel modo piú familiare $f(x)\,dx$).

Ciò posto, tutto diventa semplice e automatico. Ad ogni risultato favorevole, la $f(x)$ viene moltiplicata per x (e all'opposto per $1-x$ ad ogni risultato sfavorevole); quindi, dopo n osservazioni con r risultati favorevoli ed $s=n-r$ sfavorevoli, la $f(x)$ viene moltiplicata per $x^r(1-x)^s$ (e, naturalmente, anche se finora lo avevamo sottinteso, per la costante K necessaria per la normalizzazione: cioè, affinché la probabilità complessiva risulti uguale a 1, come dev'essere).

Un caso classico e particolarmente semplice è quello in cui la distribuzione iniziale della (per cosí dire) «probabilità incognita» è uniforme: $f(x)=1$. Allora infatti essa si moltiplica per x dopo ogni prova con risultato favorevole e per $1-x$ nel caso opposto; in definitiva, dopo r risultati favorevoli ed s sfavorevoli, la densità di probabilità è $f(x)=Kx^r(1-x)^s$. (Si chiama distribuzione Beta: cfr. § 2.10).

Per tale caso vale una semplice regola di «ragionamento induttivo», oggetto

d'interminabili dispute dall'epoca di Bayes e Laplace che le davano una posizione privilegiata e forse non ancora del tutto superata. Si tratta della «regola di successione», valida nell'ipotesi di distribuzione iniziale $f(x)\equiv 1$ in $(0, 1)$, ritenuta applicabile quanto «non si sa nulla» (frase alquanto vuota). Comunque, tale regola (conforme a dette premesse) conduce a valutare la probabilità di una prova futura (o, è ciò che conta, di cui non conosciamo l'esito) mediante il rapporto $(r+1)/(n+2)$; a parole, si tratta della frequenza «corretta» aggiungendo 1 sia al numero di risultati favorevoli che a quello degli sfavorevoli.

3.5. Distribuzioni in due (o piú) dimensioni.

Prima ancora di proseguire in una rassegna includente altre distribuzioni a una dimensione, conviene allargare la visuale comprendendovi cenni sul caso (già menzionato nel § 1.2 per un accenno preliminare) di due o piú dimensioni. Soltanto in tale ambito infatti certe significative proprietà di certe distribuzioni a una dimensione acquistano un senso. Ci limiteremo in genere al caso di due dimensioni (distribuzioni sul piano), con cenni su estensioni a tre o piú dimensioni soltanto dove risultano, non solo significative e interessanti, ma anche facilmente comprensibili o almeno intuibili. Possiamo pensare indifferentemente che si tratti di distribuzioni di probabilità o di frequenza, a seconda dell'interpretazione richiesta in singoli esempi. Per basarsi su un'immagine piú concreta, nulla vieta (ed è opportuno, per chi crede gli giovi) di pensare a una distribuzione di masse anziché di probabilità o frequenze.

Un esempio concreto, da cui potremo derivare intuitivamente diverse osservazioni e generalizzazioni, consiste anche qui nella statistica sull'età di sposi e spose. Quando ne abbiamo riportato alcuni dati (§ 2.1) essi riguardavano statistiche separate a una dimensione (una sugli sposi, un'altra sulle spose) con relative medie e scarti; ciò nulla ci diceva però riguardo alle coppie: non risultava ad esempio quante fossero le coppie di età (25, 22) e neppure quante quelle con differenza d'età $+3$, 0, 0 -1 (nei riguardi dello sposo).

Non occorrerebbero, tuttavia, ulteriori indicazioni qualora si sapesse che le distribuzioni (o i due caratteri: età dello sposo ed età della sposa) sono indipendenti. (Piú specificamente: «statisticamente indipendenti», o, se si tratta di distribuzioni di probabilità, «stocasticamente indipendenti»: 'stocastico' è l'equivalente di 'aleatorio' con radice greca). Ciò significa che (nel caso statistico, e riferendoci al nostro esempio) la percentuale di matrimoni tra coppie di età (25, 22) è il prodotto delle percentuali di sposi in età 25 e di spose in età 22 (ad esempio, se 20 su 100 sposi hanno età 25, e 25 delle 100 spose età 22, le coppie di età (25, 22) sono 5 perché $25\% \times 20\% = 5\%$).

Salvo questo caso particolarissimo, occorre indicare tutti i dati in dettaglio, mediante una «tabella a doppia entrata».

Una «tabella a doppia entrata», nel nostro caso con caselle indicanti il numero di matrimoni tra sposi di età (25, 22) (25 lo sposo e 22 la sposa) darebbe tali informazioni piú dettagliate (mentre quelle riassuntive per sposi (indipendentemente dall'età della sposa), e viceversa, sarebbero date dai totali per righe e per colonne). Ciò è mostrato dalla tabella 1; per avere una visione analoga a quella dei diagrammi a una dimensione, nel caso di due, anziché «colonne» nel senso di strisce, appoggiate su segmenti dell'asse x si dovrebbero pensare «colonne» come pilastri poggianti su ogni quadratino, e di altezza pari al numero delle coppie di sposi corrispondenti a quella coppia di età. Ne risulterebbe una specie di collina allungata, con la cresta lungo la retta che corrisponde alle coppie di età con la differenza di età piú «normale».

Anche qui, per sintetizzare il comportamento essenziale, nelle grandi linee, trascurando le irregolarità ritenute «accidentali» o comunque non interessanti, si può procedere ad una perequazione; la figura 13 mostra una tale perequazione fatta con una distribuzione normale a due dimensioni. La superficie che la rappresenta ha come linee di livello (di costante densità bidimensionale $f(x, y)$) delle ellissi concentriche e simili tra loro, mentre il profilo di ogni sezione (verticale) è il diagramma della distribuzione normale. Delle proprietà di tale distribuzione (e altre) si troveranno cenni nel § 3.6; per il momento limitiamoci a osservare che il punto di massimo della densità, su di ogni retta, è – naturalmente – quello di tangenza con una linea di livello (qui: ellisse). In particolare, i punti ove la tangente all'ellisse è parallela all'uno o all'altro dei due assi (e che si trovano su due rette – dette «rette di regressione» – uscenti dal centro) sono (evidentemente) i punti di altezza massima per chi percorresse un cammino rettilineo parallelo, ed indicano anche come varia (in media) l'età delle spose di sposi di età data (e viceversa).

L'inclinazione positiva delle due rette indica che, in media, a sposi di età piú alta corrispondono anche spose di età piú alta, e ciò si esprime dicendo che questi due caratteri hanno correlazione positiva (si avrebbe correlazione negativa se le inclinazioni fossero in senso inverso, e non-correlazione se fossero l'una orizzontale e l'altra verticale). Si ha, in particolare, «non-correlazione»

Tabella 1.
Matrimoni (per 1000) secondo l'età degli sposi (1958).

Età dello sposo	*Età della sposa* 13-15	15-18	18-21	21-24	24-27	27-30	30-33	33-40	40-50	oltre 50	*Totale*
15-18	0,2	0,9	0,6	0,2	0,1				—	—	2,0
18-21	0,8	9,9	15,9	5,9	1,8	0,6	0,2	0,1		—	35,2
21-24	0,6	12,3	44,7	39,9	13,0	3,5	1,1	0,5			115,6
24-27	0,3	10,0	63,7	106,9	65,0	16,8	4,1	1,7	0,1		268,6
27-30	0,1	4,0	36,7	79,3	77,1	37,0	9,7	3,8	0,3	0,1	248,1
30-33		1,3	13,8	36,2	44,2	34,7	15,9	7,0	0,5		153,6
33-40		0,5	5,0	15,6	24,3	27,4	21,2	19,5	2,0	0,1	115,6
40-50			0,4	1,2	2,3	4,3	5,9	14,3	7,3	0,9	36,6
Oltre 50	—		0,1	0,1	0,3	0,6	1,0	3,7	8,9	10,0	24,7
Totale	2,0	38,9	180,9	285,3	228,1	124,9	59,1	50,6	19,1	11,1	1000,0

nel caso di *indipendenza* («statistica», o – in altri casi – «probabilistica»), ciò è però solo sufficiente, non necessario.

Per il momento, basti introdurre il «coefficiente di correlazione» e spiegarne significato e proprietà. Per ciascuna delle n coppie di sposi indichiamo risp. con x_i ed y_i ($i=1, 2, \ldots, n$) l'età dello sposo e quella della sposa, e siano $\bar{x}$ ed $\bar{y}$ risp. l'età media degli sposi e delle spose. La media dei prodotti $(x_i-\bar{x})(y_i-\bar{y})$, divisa per $\sigma_x\sigma_y$ (gli «scarti quadratici medi»: cfr. § 2.1) è un numero compreso tra ± 1, che si indica $r_{x,y}$ (o semplicemente r, quando non c'è ambiguità), e si chiama coefficiente di correlazione tra le due grandezze. Nei casi estremi, $r=\pm 1$, i punti (x_i, y_i) sono allineati (su una retta crescente se $r=+1$ e decrescente se $r=-1$; se $r=0$ (non-correlazione) vuol dire che tra le x e le y non prevale la tendenza a crescere né in senso concorde né in senso discorde. Geometricamente, se disegnamo un'ellisse (come in figura 13) di una distribuzione normale avente le stesse caratteristiche (σ_x, σ_y, $r_{x,y}$), le due rette di regressione hanno come pendenza (coefficiente angolare) risp. r moltiplicato per il rapporto σ_x/σ_y o, viceversa, per σ_y/σ_x).

Anche nel caso di due dimensioni una distribuzione può venire descritta dandone la funzione di ripartizione, $F(x, y)$, od anche (quando esiste) mediante la densità, $f(x, y)$. (Analogamente in tre o piú dimensioni).

Con $F(x, y)$ si indica la probabilità (o, nel caso di distribuzione statistica, la frequenza; o, nell'interpretazione fisica, la massa contenuta nel quadrante in

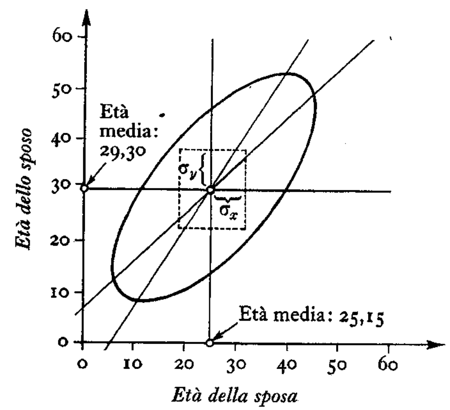


Figura 13.

Distribuzione schematica dei matrimoni secondo l'età dello sposo e della sposa: ogni matrimonio sarebbe indicato come un puntino di coordinate $x=$età della sposa ed $y=$età dello sposo. L'ellisse indica, con la sua pendenza e allungamento, in che misura vale la tendenza a matrimoni con differenza di età piú o meno larga. La distribuzione normale a due dimensioni, corrispondente ad una teorica regolarizzazione di quella effettiva (conservando le caratteristiche di valori medi, scarti e correlazione), sarebbe data da una superficie (diciamo: una collina allungata) col massimo nel centro dell'ellisse e con curve di livello tutte simili ad essa e concentriche. Il «profilo» della collina – secondo una qualunque direzione, passante per il culmine – è sempre quello della distribuzione normale, con «larghezza» data dall'intersezione con l'ellisse del disegno.

basso e a sinistra rispetto al punto (x, y) (o «a sud-ovest», SW, di esso, con espressivo riferimento alle carte geografiche).

La densità (quando esiste) è data dalla «derivata seconda mista» della funzione di ripartizione: $f(x, y) = \partial^2 F / \partial x \, \partial y$; una rappresentazione grafica delle linee di livello della densità, $f(x, y) =$ costante, dà l'idea della distribuzione immaginabile come una montagna (o altra forma qualunque) di cui dette linee di livello rappresentano le isoipse. Si noti che, nel caso d'indipendenza (e soltanto allora), si ha $F(x, y) = F_1(x) \cdot F_2(y)$ e (se esiste la densità) anche $f(x, y) = f_1(x) \cdot f_2(y)$.

Sembra preferibile non entrare in questioni o precisazioni piú tecniche, facilmente accessibili in qualunque libro sull'argomento ma difficilmente presentabili in forma discorsiva senza rischio di risultare confuse od equivoche (seppure il limite, o «livello di guardia», non sia già stato superato).

3.6. Distribuzioni che «occorre conoscere».

Vi sono delle distribuzioni che «occorre conoscere», o per interessanti proprietà matematiche (e significative per le applicazioni), o per la loro attitudine a rappresentare la forma (piú o meno idealizzata) di distribuzioni statistiche osservate in natura; a volte per «ragioni storiche».

La prima è, naturalmente, la distribuzione normale, atta a rappresentare spesso gli scarti in piú o in meno da un «valore vero» (o da un valore «medio») di misure ripetute (teoria degli errori di osservazione), di dati come statura, peso, ecc. in un gruppo di coetanei (per esempio coscritti), ecc., oppure la vincita (o perdita) su un numero sufficientemente grande di colpi a testa o croce o altri giochi analoghi. Dal punto di vista teorico, è notevole la proprietà di cui gode di essere stabile (e anzi l'unica tra quelle a scarto quadratico medio finito); «stabile» significa che se X e Y hanno distribuzione normale (e sono indipendenti), anche la loro somma $Z = X + Y$ ha distribuzione normale.

Esistono infiniti altri tipi di distribuzioni stabili (non esprimibili con formule esplicite tranne un paio di casi); vale la pena di menzionarne qui soltanto una: la distribuzione di Cauchy, con $f(x) = K/(1 + x^2)$, $F(x) = (1/2) + K \operatorname{arctg} x$. Un esempio: se, da un punto d'osservazione qualunque, guardiamo una strada rettilinea infinita puntando il cannocchiale verso una direzione «qualunque» (scelta cioè con uguale probabilità tra i 180° rivolti verso la strada), i tratti osservati grado per grado sono sempre piú lunghi quanto piú ci si avvicina agli estremi, e il tempo di osservazione per unità di lunghezza varia in modo inversamente proporzionale, secondo la formula della distribuzione di Cauchy.

Entrambe queste distribuzioni, e parecchie di quelle che menzioneremo, rientrano nella famiglia delle «distribuzioni di Pearson» (chiamate cosí perché il loro studio sistematico risale a Karl Pearson; la numerazione dei «tipi» sarà quella data da Elderton e seguita da Kendall). La proprietà che le caratterizza non dice gran che (la derivata logaritmica della densità, cioè $f'(x)/f(x)$, deve valere $x/(ax^2 + bx + c)$); la classificazione è alquanto «obsoleta» (come dice anche Feller), ma è comunque sorprendente come una grande varietà di curve

effettivamente interessanti in statistica e teoria delle probabilità rientrino in tale «famiglia».

La «mappa» della figura 14 mostra in che modo il «tipo» cambia al variare dei due parametri ξ ed η (rispettivamente momento terzo e quarto, prendendo come unità di misura lo scarto quadratico medio). (Anzi, nella presente versione, la mappa, dovuta a Boetti (1964), è stata da lui stesso resa meglio «leggibile» indicando come ordinata, in corrispondenza alle ascisse ξ, non la η bensí $\eta - 3(1 + (\xi^2/2))$. Si evita cosí che le curve fuggano tutte verso l'alto, creando suddivisioni mal distinguibili).

Possiamo osservare subito, intanto, che la distribuzione normale corrisponde all'unico punto, «N» ($\xi = 0$, $\eta = 3$), e cosí quella di Cauchy all'unico punto (improprio), «C» ($\xi = +\infty$, $\eta = 0$); altri casi del genere sono quelli della distribuzione esponenziale, $f(x) = e^{-(x+1)}$ $(x \geq -1)$, punto «E» ($\xi = 2$, $\eta = 9$), e infine (chi mai se la sarebbe aspettata nell'ambito di questa famiglia?!) la distribuzione rettangolare. Eppure... era naturale che tutti i sottocasi del tipo I, nonché quelli dei casi di transizione tra essi, confluissero verso il punto «R» ($\xi = 0$; $\eta = 9/5 = 1{,}8$) quando si fa tendere l'esponente a zero.

A proposito della «E», distribuzione esponenziale, ricordiamo che essa rappresenta, tra l'altro, la distribuzione del «tempo di attesa» da un istante dato qualsiasi al prossimo verificarsi di un «fenomeno casuale» (stessa intensità di probabilità ad ogni istante, indipendentemente dal passato: in particolare, indipendentemente dal tempo trascorso dall'ultimo suo verificarsi). Il tempo di attesa per la seconda, terza, n-esima, ripetizione di quel fenomeno ha densità Kxe^{-x}, Kx^2e^{-x}, ..., $Kx^{n-1}e^{-x}$ (con massimo in $x = n - 1$ e valor medio in $x = n$). Queste distribuzioni *non* rientrano nello schema di Pearson, ma andavano menzionate per la loro derivazione dalla esponenziale, loro «capostipite», e per riferimento ad applicazioni esse pure dipendenti da essa.

Risulta a vista dalla mappa che altri tipi corrispondono solo a delle linee divisorie, e soltanto i tipi I, IV e VI occupano zone del piano. I casi rappresentati sulle linee divisorie sono, in certo senso, casi di «transizione» tra due tipi (o sottotipi); i «sottotipi» sono distinti – entro un medesimo tipo – dall'andamento che può essere «campanulare», «C» (densità crescente fino a un massimo e poi decrescente); ad «U» nel caso opposto (densità infinita agli estremi, decrescente fino a un minimo e poi crescente); a «J» se la densità tende ad infinito soltanto in uno degli estremi.

La distinzione fra questi tre casi si può vedere nel modo piú immediato pensando alle distribuzioni Beta: $f(x) = Kx^\alpha(1 - x)^\beta$ $(\alpha, \beta > -1)$. Come già detto, gli esponenti α e β possono avere valori qualsiasi purché maggiori di -1; è chiaro che se sono entrambi positivi siamo nel caso «C» (campanulare); ad esempio per $\alpha = \beta = 1/2$ (oppure $= 1$) si ha $f(x) = K\sqrt{x(1 - x)}$ (semiellisse), risp. $f(x) = Kx(1 - x)$ (parabola); se sono entrambi negativi, ad esempio per $\alpha = \beta = -1/2$, $f(x) = K/\sqrt{x(1 - x)}$, la densità tende ad infinito per entrambi gli estremi (e lo si può chiarire dicendo che «è la stessa cosa che scegliere "a caso" un punto sul semicerchio e proiettarlo sul diametro»: è chiaro che sugli estremi si proietta un tratto di arco molto piú lungo (al limite infinitamente piú

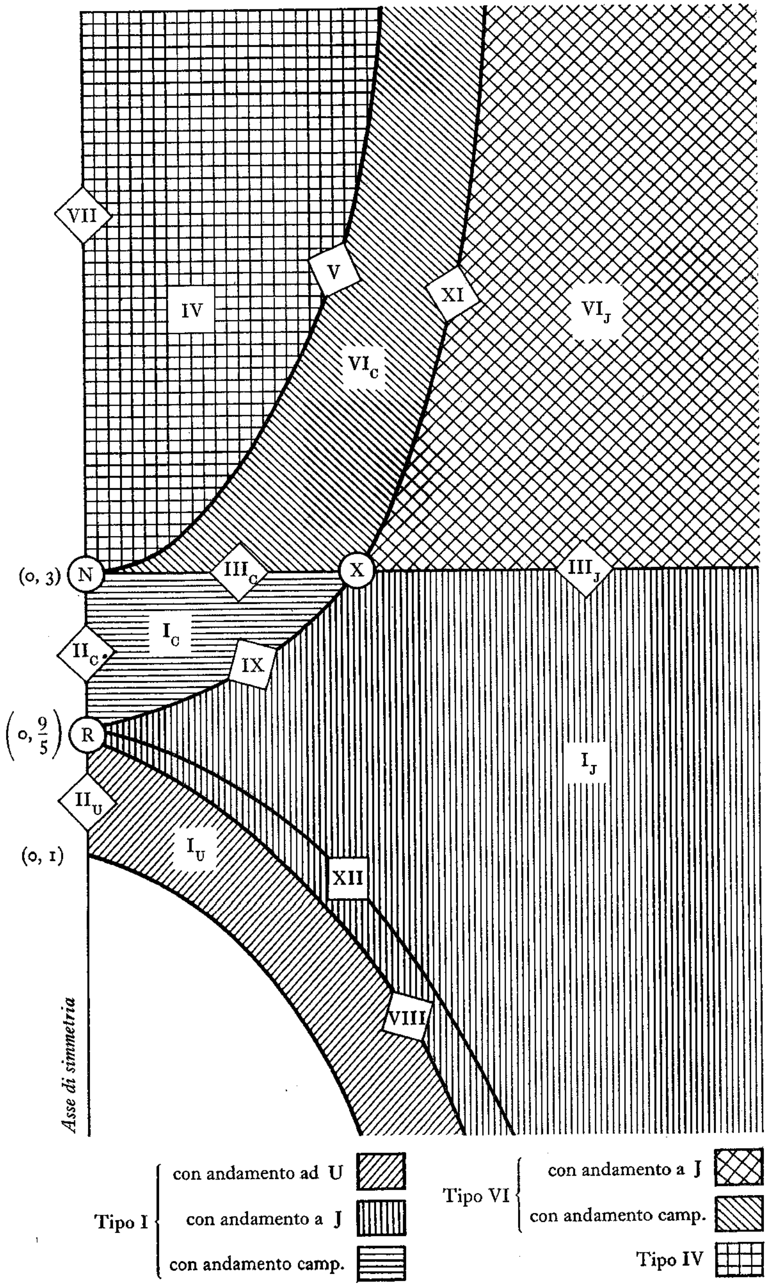


Figura 14.
La «mappa Boetti».

lungo) della sua proiezione sul diametro). Prendendo $\alpha = 1/2$, $\beta = -1/2$, si avrebbe l'andamento a «J» ($f(x) = K\sqrt{x/(1-x)}$: $f(0) = 0$, $f(1) = \infty$).

3.7. Osservazioni finali.

Pur avendo curato di far sempre riferimento – in modo piú o meno diretto – alle applicazioni, è ovvio che solo prestando attenzione alla forma di distribuzioni nell'ambito di situazioni concrete e di analisi importanti e significative che se ne vogliano fare si comincia ad afferrare e apprezzare il valore di conoscenza presentato in quella precisa forma.

Informazioni del genere, piú o meno interessanti e importanti, vengono spesso trasmesse in forme diverse che è bene saper «leggere» e ricordare: suddivisione in percentuali indicate mediante «settori circolari» (o spicchi) piú o meno larghi; intensità di fenomeni in diverse regioni (o comuni, ecc.) mediante colorazioni (o sfumature di bianco-nero) in appropriata gradazione; confronti di numerosità di «popolazioni» (abitanti di vari paesi, maestranze di diverse ditte, studenti di diverse facoltà, ecc.) mediante omini di diversa grandezza, ecc. Affinché messaggi del genere risultino correttamente interpretabili, «a vista», occorre che le differenze – vuoi d'intensità di colore, vuoi di scala – siano tali da non indurre (volutamente o sbadatamente) in errore. Basti un esempio: se si disegnano due omini di forma uguale ma di statura nel rapporto da 1 a 2, l'effetto visivo (per chi guarda il disegno) passa al quadrato: da 1^2 a 2^2 ossia da 1 a 4; se poi, nell'immaginazione, uno si raffigura non l'immagine appiattita bensí un corpo nelle sue tre dimensioni, i rapporti salgono al cubo, e cioè nell'esempio da 1^3 a 2^3 ossia da 1 a 8. Analoghe riflessioni vanno fatte per le gradazioni di colore (di intensità piú o meno regolarmente crescenti), e, quando si usano diversi colori, sulla maggiore visibilità, maggiore forza di attirare l'attenzione, di certi colori (e tonalità) rispetto ad altre. [B. D. F.].

Il concetto di distribuzione è trattato nei due significati di distribuzione di **probabilità** e di distribuzione di frequenza, e non è invece trattato nel senso economico (cfr. **produzione/distribuzione**).

Lo studio delle distribuzioni è facilitato dall'uso di rappresentazioni grafiche (cfr. **sistemi di riferimento**) diversificate a seconda che si tratti di distribuzioni discrete o continue (cfr. **continuo/discreto**). Per una trattazione teorica sono di giovamento gli strumenti del calcolo differenziale (cfr. **differenziale, funzioni**).

Alle distribuzioni si connettono degli indici sintetici importanti, quali le medie, le quali tuttavia possono essere un indice abbastanza grossolano. Per supplire maggiori informazioni si ricorre anche ad altri elementi, quali la variabilità, ecc. (cfr. anche **decisione**).

La conoscenza delle distribuzioni è poi di grande importanza per la trattazione dei **giochi**. Gli intimi rapporti tra distribuzioni di probabilità e distribuzioni di frequenza sono poi ben noti (anche se spesso lasciati nel vago).

Giochi

La parola 'giochi' oggi non indica solo certe forme del comportamento dell'individuo o del gruppo, tale nozione rappresenta un'arma del pensiero, uno strumento concettuale. Questa promozione del gioco si è sviluppata in due direzioni radicalmente distinte, l'una sul fronte della scienza, l'altra su quello della filosofia. A partire dall'idea di gioco si sono infatti costruiti dei modelli astratti di *decisione razionale* in situazioni in cui intervengono sia la competizione fra i partecipanti che elementi d'incertezza. Il punto di partenza di tale teoria, detta teoria dei giochi, è costituito dal libro di Oskar Morgenstern e John von Neumann *Theory of Games and Economic Behavior* [1947], in cui essa, fin dall'inizio presentata in forma matematica evoluta, viene immediatamente applicata alle situazioni economiche di concorrenza e di monopolio. Successivamente ne furono tentate altre applicazioni, per modellizzare processi psicologici o sociali.

La nozione di gioco appare invece come modello del funzionamento della comunicazione e, piú in generale, dei simboli, nell'altra direzione, aperta dal filosofo Wittgenstein con l'idea di «gioco linguistico» (*The Brown Book* [1933-1935]; *Philosophische Untersuchungen* [1941-49]). Non si tratta qui di costruire schemi astratti matematizzabili, ma di considerare la comune esperienza del gioco, nelle sue forme piú diverse, come prototipo suscettibile di guidarci nell'elucidazione dei problemi filosofici posti proprio dagli usi molteplici del linguaggio, considerato pertanto come «forma di vita».

Non si deve credere che questa duplice elaborazione dell'idea di gioco conduca ad una definizione unica e ad una descrizione piú completa. Essa testimonia invece di quel fondamentale procedimento del pensiero, già verificatosi in altri campi, che consiste nel passaggio dalla descrizione dell'oggetto in esame, alla sua trasformazione in operatore, in strumento utilizzato a sua volta per descrivere, analizzare, spiegare altri oggetti o per formulare con maggiore esattezza problemi rimasti confusi. Con riferimento all'analisi del gioco proposta da Roger Caillois [1967, p. 135] nelle quattro componenti di *competizione*, *simulacro*, *fortuna* e *vertigine*, la prima delle elaborazioni sopra ricordate recepisce come fondamento intuitivo gli aspetti del combattimento e del rischio. Nell'elaborazione di Wittgenstein ci si deve piuttosto riferire ad un aspetto non contemplato nell'analisi di Caillois, ma che è tuttavia dominante in ogni riflessione sul gioco, quello cioè delle *regole*; il filosofo si pone proprio, con riferimento al linguaggio, il problema della natura delle regole. Peraltro il quesito cui cerca di rispondere la teoria dei giochi, con il suo apparato astratto e matematicizzato, potrebbe essere cosí formulato: determinare le regole di comportamento ottimale, in ben definite situazioni di conflitto o di rischio, o perlomeno descrivere globalmente le condizioni di fine del gioco con il relativo risultato, supponendo che i giocatori tengano un comportamento razionale. Il primo, assai ambizioso, obiettivo è in generale so-

stituito, almeno in un primo momento, dal secondo, seguendo cosí quella prassi semplificatrice del pensiero scientifico che introduce la statica prima di affrontare la dinamica. In ogni caso per raggiungere tali obiettivi la teoria dei giochi ha innanzitutto affrontato il concetto intuitivo di alea in una prima elaborazione in cui la nozione di gioco ha effettivamente un posto secondario; al giorno d'oggi la nozione di conflitto è diventata essenziale e si definiscono allora i concetti di «equilibrio» e di «soluzione» di un gioco, con o senza cooperazione tra i partecipanti.

1. *I giochi d'azzardo e il concetto di probabilità.*

I concetti fondamentali del calcolo delle probabilità si sono formati per riflessione sui giochi d'azzardo. Pascal e Fermat verso la metà del XVII secolo si interessano alla previsione del successo nel gioco dei dadi e sono condotti a precisare l'idea di probabilità come quoziente del numero dei casi favorevoli per il numero totale delle combinazioni possibili. Essi esaminano inoltre il problema, particolarmente significativo della pratica del gioco, delle «parti» e cioè della ripartizione equa della posta nel caso in cui i partecipanti interrompano il gioco nel corso della partita. Nella corrispondenza scambiata con Fermat fra luglio e ottobre 1654 e nel *Traité du triangle arithmétique* [1654, ma pubblicato solo nel 1665], Pascal determina il «valore di ciascuna parte», e cioè di quella frazione della posta che tocca a ogni giocatore dopo una qualunque partita, nel gioco dei punti a *n*-vittorie, in funzione delle vittorie che mancano a ognuno dei giocatori. L'idea della dimostrazione è di cominciare dal caso in cui a un giocatore manchi una sola vittoria e all'altro due. Se il nostro giocatore vince la partita guadagna tutta la posta, se perde si trova in parità con l'avversario e dovrebbe dunque suddividere a metà il premio. È dunque certo di avere almeno tale metà poiché la ottiene anche in caso di sconfitta; a ciò si deve aggiungere la metà del resto, poiché, come gli fa dire Pascal «forse l'avrò, forse non l'avrò, la probabilità è uguale, dividiamo dunque...» (lettera a Fermat del 29 luglio 1654). Risolto questo primo caso nel modo descritto, ad esso ci si riconduce per induzione facendo variare il numero di vittorie che mancano a ogni giocatore.

Pascal introduce cosí tre nozioni fondamentali che saranno alla base del concetto di gioco come modello di comportamento razionale:

1) Definizione del valore di un avvenimento aleatorio mediante la speranza matematica, prodotto del valore attribuito all'avvenimento certo per la sua probabilità.
2) Schematizzazione dello svolgimento di un gioco come successione di punti di diramazione, costituenti ciò che piú tardi sarà definito un *albero*. Il gioco viene cosí sviluppato in tutte le sue eventualità e può essere esplorato all'indietro, a partire da una delle situazioni finali. Resa esplicita nell'ambito della teoria dei grafi (cfr. gli articoli «Combinatoria», § 2.3, e «Grafo» in questa stessa *Enciclopedia*), questa descrizione di un

gioco verrà piú tardi applicata ai giochi in cui interviene la destrezza dei giocatori, sotto il nome di *forma completa od estensiva.*

3) Introduzione della prima idea di ciò che diventerà piú tardi il concetto di *soluzione* e di *valore* di un gioco. Si cerca, nel problema delle parti, di sostituire il risultato effettivo del gioco con l'attribuzione di un guadagno a ogni giocatore, mediante un «regolamento... [che] deve essere talmente proporzionato a ciò che essi sono in diritto di aspettarsi dalla fortuna che ognuno di essi trova del tutto uguale prendere ciò che gli si dà o continuare l'avventura del gioco» [1654, ed. 1954 p. 115].

Beninteso si tratta sempre del gioco di puro azzardo nel quale non si può usare alcun stratagemma. Tuttavia gli stessi concetti di base potranno essere usati nella teoria dei giochi con strategia. In essi l'incertezza dei giocatori non proverrà solo da avvenimenti aleatori incontrollati, ma dall'ignorare i progetti dell'avversario. L'equità della ripartizione delle puntate dipenderà allora non solo dalle probabilità individuali di vittoria, ma anche dalla possibilità di alleanze.

2. *La nozione di «equilibrio» e il duopolio di Cournot.*

Lo sviluppo ulteriore di una teoria del comportamento nel gioco sarà proprio centrata su tale idea della *ripartizione equa*. Essa conduce a una soluzione semplificata e concisa del problema della ricerca delle regole di comportamento che definiscono una strategia nel dettaglio delle mosse. Come nel problema delle parti, si sostituisce al gioco effettivo una regola di ripartizione. Una strategia prudente, che non cerca cioè l'«ebbrezza» del gioco, è individuata alla fin fine dal suo risultato, e cioè da una ripartizione delle vincite stabile e accettabile, in certo modo, da tutti i giocatori.

Senza che la parola 'gioco' vi sia mai usata, le *Recherches sur les principes mathématiques de la théorie des richesses* di Cournot [1838] propongono un modello di comportamento duopolistico che fornisce un importante contributo alla esplicitazione di un concetto di equilibrio essenzialmente applicabile in una situazione di gioco.

Nel celebre esempio di Cournot, due persone possiedono ciascuna una sorgente di acqua minerale che permette di produrre a costo uguale (o nullo) uno stesso bene che essi sono i soli ad offrire sul mercato. Sapendo che il prezzo di vendita, necessariamente unico, di tale bene è legato alle quantità offerte – ed effettivamente acquistate – da una certa funzione della domanda, si vuole determinare quale quantità ciascuno, ignorando le intenzioni dell'altro, deve offrire per rendere massimo il proprio profitto.

Siano x_1 e x_2 le due quantità offerte dai duopolisti: si tratta, nel loro «gioco», delle rispettive strategie. Il profitto sarà per ognuno di essi il valore del gioco. La soluzione consiste in una coppia di strategie tali che nessuno degli antagonisti possa a colpo sicuro aumentare il proprio profitto.

Sia p il prezzo che si stabilisce in seguito alle loro offerte; la funzione della domanda può scriversi $p=F(x_1+x_2)$. Si devono massimizzare *contemporaneamente* i due profitti $x_1F(x_1+x_2)$ e $x_2F(x_1+x_2)$. Tale problema è insolubile nell'ambito dell'analisi classica, per la quale si dovrebbero annullare le derivate parziali di ognuna delle due espressioni, rispetto alle due variabili indipendenti x_1 e x_2, ottenendo in tal modo quattro equazioni per determinare due incognite.

Cournot immagina allora uno *schema di comportamento per successivi tentativi* dei duopolisti, schema che ha esattamente le caratteristiche di un gioco di strategia. Si supponga che il duopolista II abbia fissato e comunicato la quantità x_2 che vuole offrire; in tale caso il duopolista I determinerà la propria produzione in modo da massimizzare il proprio profitto, essendo fissato il valore di x_2. Ma il duopolista II, venendo a conoscenza del valore x_1, rivedrà il proprio piano di produzione per massimizzare anche lui il profitto, essendo ora fissato il valore x_1. E cosí via, fino a quando si ottenga una eventuale coppia di valori $\bar{x}_1$ e $\bar{x}_2$, tali che ogni revisione della propria strategia implichi una perdita di profitto per il duopolista (fig. 1). L'esistenza di una tale coppia dipende ovviamente dalle ipotesi fatte sul comportamento delle reazioni di ogni giocatore alla revisione della strategia dell'avversario. Le «curve di reazione» che collegano per il duopolista I un valore x_1 al valore x_2 immediatamente precedente si tagliano allora in un punto le cui coordinate $\bar{x}_1$ e $\bar{x}_2$ sono le «strategie» di equilibrio alle quali convergono i «tentativi» successivi dei duopolisti.

Anche se tale modello di Cournot è stato spesso criticato dal punto di vista degli economisti, qui interessa in quanto prototipo dell'idea di soluzione di un gioco. Esso introduce infatti in modo chiaro alcuni elementi essenziali:

1) La nozione di strategia come risposta adattata ad una mossa dell'avversario.
2) La nozione – fondamentale – di strategie in equilibrio: supposta fissata una di esse, ogni modifica dell'altra provoca una diminuzione di valore, o perlomeno non produce alcun miglioramento, a chi la effettua.

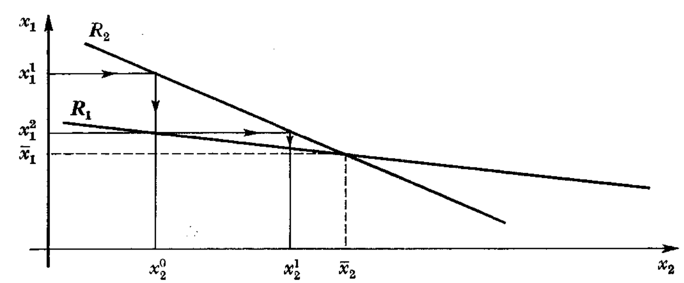


Figura 1.
Schema di comportamento per successivi tentativi in condizioni di duopolio.

3. *L'equilibrio nel gioco tra due persone, a somma nulla.*

Il concetto di strategia non verrà tuttavia esplicitamente introdotto ed utilizzato che un secolo piú tardi, nel trattato fondamentale di John von Neumann e Oskar Morgenstern [1947]. Tale concetto può essere presentato a due livelli di astrazione.

Si considerino n giocatori. A ogni «mossa» ognuno di essi sceglie – ovvero si trova imposta dal caso – una «tattica»; le scelte di tutti i giocatori ad ogni mossa determinano una situazione del gioco (per esempio, una certa distribuzione delle carte che rimangono in mano ai giocatori di bridge, oppure una disposizione dei pezzi negli scacchi). Dopo un numero prefissato di mosse, oppure in base ad una qualunque regola di arresto (ad esempio i giocatori hanno terminato la loro mano, oppure un re è messo sotto scacco), ogni giocatore ottiene un premio. Il gioco può quindi essere rappresentato come un albero orientato; in esso ogni nodo rappresenta una situazione del gioco e i rami che ne escono rappresentano le transizioni alle situazioni generate dalle diverse combinazioni delle tattiche dei giocatori; ogni cammino connesso che, partendo dalla radice dell'albero, conduce a un punto terminale descrive pertanto un'effettiva partita del gioco (cfr. fig. 2). Tuttavia tale grafo completo è l'immagine di ciò che può vedere un arbitro che conosca, per ogni partita, la storia esatta e completa delle tattiche applicate da ogni giocatore. Si può ipotizzare che un giocatore conosca solo approssimativamente tale storia e cioè che, a ogni mossa, egli sappia solo di trovarsi in una classe di nodi del grafo per lui indecifrabile. In caso contrario si tratta di un gioco con *informazione completa*: è questo il caso degli scacchi ma non del bridge.

Tuttavia, per edificare la teoria non è necessaria una schematizzazione cosí spinta del gioco. Si dirà «strategia» una pianificazione del gioco, presa nel suo complesso, considerando globalmente per ogni giocatore il *risultato* di una partita in corrispondenza delle diverse scelte operate a ogni mossa, e cioè «un piano che precisa quali scelte egli farà in tutte le possibili situazioni... in conformità al quadro che le regole del gioco gli forniscono per quel caso» [Neumann e Morgenstern 1947, p. 79]. Un gioco viene allora definito sotto forma ridotta, o *normale*, in modo assai semplice: assegnato a ogni giocatore l'insieme Σ_i delle sue strategie, il gioco è determinato assegnando n applicazioni G_i di

$$\prod_{i=1}^{n} \Sigma_i$$

nell'insieme dei numeri che misurano le vincite (o i guadagni); tali applicazioni fanno corrispondere a ogni giocatore l'ammontare della vincita in corrispondenza di ogni n-pla di strategie ognuna delle quali appartiene all'arsenale di uno degli n giocatori.

Nel caso semplice di un gioco tra due persone, senza possibili coalizioni, e nel quale ciascuno guadagna ciò che l'altro perde (*gioco a somma nulla*, in francese *duel de somme nul*) la forma normale è data da una matrice con n righe e m colonne, qualora i giocatori dispongano di n e rispettivamente di m strategie; il numero a_{ij} rappresenta la vincita del giocatore Noi, preso come riferimento, quando egli usa la strategia i contro la strategia j dell'avversario Loro. Si può allora definire una *soluzione* del gioco nell'ottica dell'equilibrio alla Cournot. Ciascun giocatore, ignorando la strategia dell'avversario, non può certamente ottenere il massimo guadagno indicato dalla matrice, ma può limitare il danno adottando la strategia per la quale la sua vincita minima è la piú grande. In altre parole il giocatore Noi sceglie la riga nella matrice il cui minimo è piú grande e il giocatore Loro sceglierà la colonna il cui massimo è il piú piccolo (fig. 3).

Se esiste un elemento che sia contemporaneamente minimo della riga e massimo della colonna, le due strategie degli avversari lo indicheranno. Esso è il valore del gioco per Noi e le due strategie sono allora in equilibrio. Infatti se Noi abbandona tale strategia egli, qualora Loro continui a giocare «razionalmente» e cioè con la strategia di prudenza, *rischia* di guadagnare di meno poiché tale elemento è un massimo della colonna. Peraltro ogni strategia alternativa usata da Loro rischia di accrescere il guadagno di Noi, poiché tale

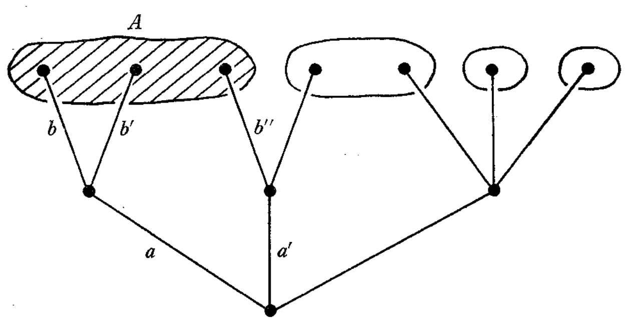


Figura 2.
Alla quarta mossa il giocatore che si trova in A ignora se la partita si è svolta secondo (a, b), (a, b') oppure (a', b'').

		Loro 1	2	3
Noi	1	0	1	4
	2	−1	3	1

Figura 3.
Matrice di un gioco a somma nulla fra due persone. Le frecce indicano le strategie ottimali adottate dai due giocatori.

elemento è un minimo della riga. Esattamente come nel duopolio, nessuno dei giocatori ha interesse ad abbandonare la propria strategia di prudenza; il valore del gioco è fissato da questo maximin del giocatore Noi.

Tuttavia, un elemento (o piú di uno) con tale proprietà può non esistere nella matrice di un gioco; esso esiste però nei giochi a somma nulla con informazione completa, quali gli scacchi (teorema di Zermelo-Kalmar). Se tale elemento non esiste, il gioco non ha soluzione nel senso sopra definito: nessuna strategia può permettere ai giocatori di limitare con sicurezza il danno (fig. 4).

Neumann e Morgenstern sostituiscono allora alla nozione di guadagno quella di *guadagno sperato*, ammettendo strategie «miste» – cioè combinazioni aleatorie di strategie pure – per ogni giocatore. L'idea di tali strategie miste era stata avanzata, sin dal 1921, da Emile Borel come «strumento per giocare in modo vantaggioso variando il proprio gioco» [1950, p. 257]. Si trattava di un'intuizione fondamentale e Borel viene giustamente considerato l'iniziatore della nuova teoria. In tali condizioni la speranza di guadagno per ogni coppia di strategie miste dipende bilinearmente dalle probabilità spontaneamente attribuite da ogni giocatore a ciascuna delle strategie pure, o tattiche. Orbene, si dimostra (teorema di Neumann, cfr. § 5) che, per ogni gioco avente un insieme finito di strategie, esiste almeno una coppia di strategie miste in equilibrio alla Cournot, che assicura a ogni giocatore la possibilità, con un'opportuna scelta dell'estrazione a sorte fra le tattiche, di massimizzare la sua speranza di guadagno minimo (e minimizzare la sua speranza di massima perdita). Ad esempio, nel gioco della figura 4 si dimostra che la soluzione per Noi è di giocare due volte su tre la prima tattica e una su tre la seconda, e allora il valore del gioco – la sua speranza di guadagno – diventa 7/3 qualunque sia peraltro la strategia mista dell'avversario. Se quest'ultimo vuole minimizzare la propria speranza di perdita, dovrà del resto giocare l'una o l'altra delle seguenti strategie miste: la prima tattica una volta su cinque e la terza quattro su cinque, oppure la prima tattica quattro volte su nove e la seconda cinque volte su nove; una qualunque di queste strategie miste gli assicura allora, qualunque sia la strategia di Noi, una speranza matematica di $-7/3$. Si osservi che il valore del gioco può non corrispondere ad alcun esito reale della partita: si tratta solo di una speranza, ossia di un valor medio, poiché ogni mossa giocata è una tattica o strategia pura. Tuttavia il carattere aleatorio del gioco protegge ogni giocatore contro l'astuzia dell'avversario, che, indovinando una scelta sicura, potrebbe allora adattarvi il proprio gioco.

		Loro 1	2	3 ↓
Noi	→ 1	4	1	2
	2	−1	5	3

Figura 4.

Il piú grande minimo per Noi è 1, il piú piccolo massimo per Loro è 3. Se però Noi gioca la sua strategia ottimale (1) e Loro la strategia ottimale (3), il valore del gioco è 2.

4. *L'equilibrio cooperativo.*

Pareto, nel tentativo di descrivere un'economia in cui venga raggiunto un benessere collettivo ottimale, propone il seguente criterio: vi è equilibrio quando è impossibile accrescere la soddisfazione di un qualunque agente senza diminuire quella degli altri. Questo criterio può essere adattato a un gioco nel quale coalizioni fra giocatori permettono di modificare a loro vantaggio la ripartizione dei guadagni. Si ha equilibrio quando tali coalizioni si bilanciano, nel senso che il miglioramento acquisito da taluni a detrimento di altri comporta, da parte delle vittime, la formazione di una alleanza antagonista. La teoria dei giochi di coalizione consiste essenzialmente nel precisare tale nozione di equilibrio paretiano.

Un gioco fra N persone viene definito mediante una *funzione caratteristica* che individua per ogni raggruppamento S di giocatori – e anche per i giocatori isolati – il guadagno globale $v(S)$ conseguente alla forza della coalizione nel gioco. Un arbitro può proporre al giocatore i il guadagno u_i che naturalmente verificherà le condizioni

1) $u_i \geq v(\{i\})$
2) $\sum_{i \in N} u_i = v(N)$,

giacché nessun giocatore accetterà meno di quanto può guadagnare con sicurezza giocando da solo (condizione 1) e l'arbitro non può distribuire piú della totalità della posta (condizione 2). Una tale ripartizione viene detta da Neumann e Morgenstern *imputazione*. Una qualsiasi imputazione non può certo essere considerata come una «soluzione» del gioco. Un'imputazione può essere resa inefficace se «bloccata» da una coalizione S grazie alla quale si ha una ridistribuzione w_i dei guadagni tra i partecipanti ad essa, tale che $w_i > u_i$ per ogni $i \in S$ e ovviamente $\sum_{i \in S} w_i = v(S)$.

In tale caso infatti i membri di S hanno interesse a rifiutare l'imputazione e a coalizzarsi. Si chiama *nucleo* o *cuore* (*core* per gli autori anglosassoni) l'insieme delle imputazioni che non sono bloccate da alcuna coalizione. Le imputazioni del nucleo possiedono una proprietà paretiana non appena si ammettano accordi di compensazione tra giocatori coalizzati. In effetti qualora si crei una nuova coalizione S', i guadagni distribuiti w_i non possono essere rispettivamente superiori a *tutti* gli u_i di una qualunque imputazione del nucleo, giacché essa non blocca nessuna delle imputazioni del nucleo. Se non si postula la libertà di ridistribuzione dei guadagni in S', si dovrebbe ridefinire il nucleo imponendo esplicitamente che le imputazioni siano paretiane e contemporaneamente non bloccate.

Neumann e Morgenstern hanno originariamente proposto una definizione piú larga della soluzione di un gioco. Si dice che un'imputazione *domina* un'altra se esiste una coalizione S che permette di realizzarla (tale che $\sum_{i \in S} u_i \leq v(S)$) e quindi che blocca l'altra. Una famiglia di imputazioni è soluzione se le imputazioni che la compongono non si dominano l'un l'altra e se ogni altra imputazione è dominata da almeno una imputazione della famiglia. È stato dimostrato da Lucas [1968] che esistono giochi senza soluzione in tale senso, ma non si conoscono condizioni generali di esistenza di una famiglia di soluzioni o di un nucleo (che fa allora parte di una soluzione) per un arbitrario gioco cooperativo. Si può riprendere da Neumann e Morgenstern l'esempio del gioco generico a somma nulla fra tre persone, la cui funzione caratteristica può essere definita nel modo seguente:

$$v(S) = \begin{cases} 0 & \text{se } S = \emptyset \\ -1 & \text{se } S \text{ ha un elemento} \\ 1 & \text{se } S \text{ ha due elementi} \\ 0 & \text{se } S \text{ ha tre elementi.} \end{cases}$$

Si dimostra che vi sono sempre due tipi di soluzione: una è costituita dalle tre imputazioni simmetriche $(-1, 1/2, 1/2)$, $(1/2, -1, 1/2)$, $(1/2, 1/2, -1)$ nelle quali si forma una coalizione di due giocatori con ripartizione uguale del guadagno. L'altro tipo comprende i tre insiemi infiniti di imputazioni ottenute per permutazione dei giocatori nella formula generale $(-c, a, c-a)$ con $-1 \leq a \leq 1+c$ e $1 \geq c > -1/2$. In tali soluzioni, due giocatori decidono di attribuire al terzo il guadagno $-c$ e di dividersi la posta restante c con regole che non intervengono nel gioco. Il valore $-c$ caratterizza in sostanza il potere congiunto dei due giocatori coalizzati per sfruttare il terzo ed il valore a dipende dalla capacità di dominio di uno dei due alleati sull'altro. Tali parametri non sono stabiliti dalle condizioni del gioco. Il nucleo di un tale gioco è ovviamente vuoto poiché in ogni imputazione vi sarà necessariamente una coppia di giocatori che otterranno meno di quanto assicurato dalla loro coalizione (e cioè 1) e che quindi bloccheranno l'imputazione qualunque essa sia. Si vede che l'idea di soluzione dei giochi cooperativi si riconduce alla determinazione di una *regola di ripartizione*; l'accettazione da parte dei giocatori di tale regola di ripartizione dipende dalle ipotesi fatte sulle convenzioni, o norme sociali di comportamento, alle quali sono soggetti i loro conflitti; infine sussisterà in generale una zona d'indeterminazione al cui interno è d'uopo supporre avvenga una «trattativa» dipendente eventualmente da un sottogioco da definire. Beninteso, altre ipotesi sulle norme di cooperazione, piú forti di quelle che conducono alla definizione di *soluzione* e di *nucleo,* permettono di precisare la determinazione degli equilibri (Luce, Shapley).

5. *La matematica della teoria dei giochi.*

Ci si limiterà a poche e rapide indicazioni di carattere generale.

5.1. Il teorema di John von Neumann.

Stabilisce che ogni gioco finito a somma nulla fra due persone ha almeno una soluzione mista. Si consideri la matrice delle vincite $[a_{ij}]$ con n righe e m colonne. Una soluzione mista è costituita dalla coppia formata da una n-pla $\{x_1, x_2, ..., x_n\}$ e da una m-pla $\{y_1, y_2, ..., y_m\}$ di numeri positivi o nulli tali che

$$\sum_{i=1}^{n} x_i = 1 \qquad \sum_{i=1}^{m} y_i = 1$$

e inoltre tali che

$$\max_{(x)} \min_{j} \sum_{i} a_{ij} x_i = \min_{(y)} \max_{i} \sum_{j} a_{ij} y_j.$$

La dimostrazione introduce gli spazi vettoriali topologici delle strategie a m e n dimensioni nei quali una strategia di un giocatore è rappresentata da un punto di coordinate uguali ai guadagni determinati dalle strategie pure dell'altro. La realizzazione di strategie miste richiede la considerazione di tutti i punti dell'***inviluppo** convesso* del poliedro delle strategie pure di ogni giocatore. La dimostrazione si basa sulla separazione dei convessi mediante iperpiani (teorema di Hahn-Banach). Si dimostra inoltre che, se esistono piú strategie ottimali per ogni giocatore, esse sono interscambiabili in ogni soluzione.

5.2. Il teorema di Nash.

Generalizza il punto di vista del gioco finito a somma nulla tra due persone al caso di un gioco tra n giocatori, ove il giocatore i dispone di un insieme Σ_i di strategie σ_i^j e per ipotesi Σ_i è un insieme convesso e compatto di uno spazio vettoriale topologico. Il teorema fornisce condizioni sufficienti per l'esistenza di un equilibrio:

a) se le funzioni delle vincite G_i da $\prod_i \Sigma_i$ in $\mathbf{R}$ sono continue;

b) se l'insieme dei punti di Σ_i in cui la funzione parziale della vincita $\tau_i^j \to G_i(\sigma_i^1, \sigma_i^2, ..., \tau_i^j, ..., \sigma_i^n)$ raggiunge il massimo è convesso per tutti i σ e tutti gli i.

La dimostrazione utilizza le proprietà di *punto fisso* della corrispondenza che associa a ogni punto di $\prod_i \Sigma_i$ la parte di $\prod_i \Sigma_i$ prodotto dei convessi su cui ogni funzione parziale $\tau_i^j \to G_i(\sigma_i^1, ..., \tau_i^j, ..., \sigma_i^n)$ raggiunge il suo massimo (teorema di Kakutani). I punti fissi di tale corrispondenza sono ovviamente equilibri alla Cournot del gioco. Ma le strategie ottimali di ogni giocatore non sono piú interscambiabili.

5.3. Giochi e programmazione lineare.

La programmazione lineare studia la soluzione del seguente problema: Determinare n variabili positive o nulle soddisfacenti agli m vincoli

$$\sum_{i=1}^{n} a_{ij} x_i \geq b_j \qquad 1 \leq j \leq m$$

e minimizzanti la forma lineare $C = \sum_i c_i x_i$.

Il programma *duale* associato è il seguente: Determinare m variabili positive o nulle y_j soddisfacenti agli n vincoli

$$\sum_{j=1}^{m} a_{ij} y_j \leq c_i \qquad 1 \leq i \leq n$$

e massimizzanti la forma lineare $B = \sum_j b_j y_j$.

Si dimostra il teorema di dualità: se un problema di programmazione lineare ammette soluzione, lo stesso accade per il suo duale, e gli estremi delle due forme lineari hanno lo stesso valore.

Orbene, un gioco finito a somma nulla tra due giocatori può essere considerato come particolare problema di programmazione lineare il cui duale corrisponde al gioco dell'avversario. In effetti il giocatore che adotta la strategia mista $\{x_i\}$ con $\sum_i x_i = 1$ si assicura il guadagno

$$\min_j \sum_{i=1}^{n} a_{ij} x_i = v'.$$

Assumendo come nuove variabili $x'_i = x_i/v'$, si può riscrivere tale condizione di minimo nella forma di m vincoli

$$\sum_{i=1}^{n} a_{ij} x'_i \geq 1 \qquad 1 \leq j \leq m$$

e il giocatore cerca di massimizzare v', cioè di minimizzare la forma lineare $1/v' = \sum_i x'_i$. Si tratta di un problema di programmazione lineare ove le b_i e le c_i sono uguali a 1. Il gioco duale dell'avversario consiste nel massimizzare $\sum_i y'_j$ sotto gli n vincoli

$$\sum_{j=1}^{m} a_{ij} y'_j \leq 1 \qquad 1 \leq i \leq n.$$

Il teorema di dualità generalizza allora il teorema di John von Neumann. Le tecniche di ricerca effettiva delle soluzioni di un gioco sono in questo modo ricondotte a quelle della programmazione lineare.

6. *Applicazioni della teoria dei giochi.*

Considerata come *modello* descrittivo ed esplicativo di un'azione umana, la nozione di gioco con strategia consiste essenzialmente nel neutralizzare il carattere di *processo* di quest'ultima per mettere l'accento sul suo risultato. Il problema centrale di una teoria dei giochi diventa cosí la definizione precisa e la determinazione di un *equilibrio*. Tuttavia, nella versione originale di Neumann e Morgenstern, i parametri che costituiscono la soluzione di un gioco a somma nulla tra due giocatori sono interpretati come i coefficienti di probabilità volontariamente attribuiti a una scelta; essi reintroducono dunque sotto forma operativa il carattere dinamico dell'azione umana. È anche vero che li si può interpretare in modo non probabilistico, come avviene nella teoria, matematicamente equivalente, della programmazione lineare ove i coefficienti corrispondono a ripartizioni statiche di fattori di produzione oppure, nel problema duale, ai costi marginali delle materie prime.

D'altra parte, come si è visto, la costituzione di un modello di gioco richiede una «funzione delle vincite», che presuppone quindi una scala di valori attribuiti dai partecipanti agli avvenimenti presi in considerazione. Le ipotesi che vanno allora fatte esplicitamente non sono affatto banali, e la loro analisi costituisce l'oggetto di una teoria della misura.

Infine, le norme di ottimizzazione individuale o collettiva che sono alla base della costituzione di un concetto di equilibrio prefigurano un'idea del comportamento «razionale» che non può certo essere uniformemente accolta per ogni fenomeno umano. I diversi tipi di soluzione di un gioco adottati nella teoria coincidono con l'idea generale di un'azione volta a una massimizzazione prudente e pessimista dei desideri dell'agente. È tuttavia possibile schematizzare anche le norme di un «gusto del rischio» o di una prudenza piú ottimista. Inoltre, una tale nozione di ottimalità non conviene affatto, come è stato sottolineato da Rapoport [1960], a situazioni di conflitto senza quartiere, in cui lo scopo perseguito è la punizione o la distruzione dell'avversario.

Per tutti questi motivi la teoria dei giochi nelle sue forme attuali, malgrado gli inizi assai promettenti e un'ispirazione innovatrice, ha avuto applicazioni abbastanza limitate. L'intenzione dei suoi iniziatori era di rifondare per suo tramite la teoria economica. Nella forma piú generale della programmazione lineare, i modi di ragionare e le tecniche matematiche della teoria dei giochi sono state effettivamente applicate alla definizione neoclassica dell'equilibrio generale. A ben vedere il beneficio essenziale è stato però un raffinamento concettuale e un maggior rigore, anziché un sostanziale rinnovamento.

Nell'ambito sociologico, ove dovrebbe fornire schemi di conflitto e spiegare rapporti tra gruppi, la teoria dei giochi non è stata quasi applicata. Lo stesso vale in psicologia, ove alcuni modelli di gioco sono stati tentati per spiegare le leggi della percezione e dell'apprendimento. In questo caso, il gioco non è piú interpretato come lotta fra centri di decisione, ma come adattamento economico delle reazioni di un sistema organizzato di fronte a un ambiente ancora inesplorato; la struttura del gioco e la definizione di equilibrio forniscono allora una forma operativa precisa all'idea antropomorfa dell'«ostilità» o perlomeno dell'estraneità di tale ambiente.

Al momento attuale, la teoria dei giochi è forse piú efficace nelle sue applicazioni a semplici e circostanziati problemi di razionalizzazione dell'azione. La «ricerca operativa» ne fa grande uso. Ma i modelli del comportamento umano che la teoria dei giochi permette di costruire non sono solo schemi di razionalità per *coloro che decidono*. Cosí come i sistemi di equazioni algebriche ed analitiche usate nelle scienze della natura servono a formulare una razionalità, che non può certo essere interpretata come uno degli scopi della natura stessa, parimenti i modelli di gioco serviranno forse a definire una razionalità immanente del comportamento umano da non confondere con il perseguimento di un obiettivo.

7. *I giochi linguistici.*

La nozione di «gioco linguistico» (*language game*, *Sprachspiel*) introdotta in filosofia da Wittgenstein, appartiene a un universo completamente diverso: essa non è mai stata presentata come un modello astratto dei comportamenti umani e ancor meno come uno schema strutturabile matematicamente. Va tuttavia sottolineato come a partire dallo stesso significato complesso e superdeterminato evocato dalla parola 'gioco', il pensiero contemporaneo abbia cercato di costruire sia uno strumento d'indagine scientifica sia uno strumento d'indagine filosofica.

Il problema filosofico centrale posto da Ludwig Wittgenstein potrebbe in definitiva essere formulato nel modo seguente: cosa è significare? Il *Tractatus logico-philosophicus* [1922] costituisce una prima risposta a tale quesito, posto allora in forma meno esplicita. Il linguaggio vi è descritto come *rappresentazione* per «immagini» del mondo. La sua «grammatica» è una logica, le cui regole possono essere *esibite* ma non veramente *espresse* con formule del linguaggio stesso. Questa concezione presenta difficoltà peraltro già riconosciute nel *Tractatus*. Piú tardi Wittgenstein ne illustrerà la limitazione, il carattere «unilaterale», anche se non ne ha mai rifiutato il nucleo essenziale.

La funzione rappresentativa del linguaggio, che privilegia i simboli nominali, corrisponde solo ad uno dei suoi aspetti. Wittgenstein scopre, fin dai primi anni '30, una prospettiva piú ampia: si deve concepire il linguaggio come una molteplicità di *giochi di comunicazione*. Significare e comprendere vuol dire attenersi a certe *regole*. Come nel *Tractatus*, tali regole non vanno pensate come esprimibili in modo definitivo, né suscettibili di essere sistematizzate in un tutt'unico. Il filosofo tuttavia, togliendo all'esercizio linguistico le sue finalità concrete, mettendolo per cosí dire «in vacanza», può costruire modelli ridotti di giochi di comunicazione. Tali sono i celebri 73 giochi linguistici descritti all'inizio del *Brown Book* [1933-35], il primo dei quali consiste per chi parla nel pronunziare un nome indicante un oggetto e per chi ascolta nel rispondere semplicemente portando l'oggetto indicato. Un tale schema, afferma Wittgenstein, non è un linguaggio mutilato, ma un linguaggio altrettanto completo, ancorché piú semplice, del linguaggio usuale. Ciò significa che esso mostra, mediante «una finzione grammaticale», il funzionamento stesso dei simboli attraverso i comportamenti; soltanto si tratta di una situazione in cui le regole sono il piú ridotte possibile. Complicandole, variando la situazione del gioco, Wittgenstein desidera arrivare a descrivere la significazione nell'indefinita molteplicità dei suoi aspetti. Non si tratta del resto di dare una formula generale che la definisca attraverso l'enunciazione dei caratteri comuni a tutti questi aspetti. La loro unità è per Wittgenstein dovuta a una certa «aria di famiglia», o ancora è quella della corda tenuta insieme dall'intrecciarsi dei fili, nessuno dei quali la percorre dall'inizio alla fine.

Se il gioco linguistico è essenzialmente l'applicazione di regole, queste ultime non sono però assimilabili a regole empiriche (di cucina o di una qualsiasi tecnica) concepite in vista di un risultato. Al contrario, è detto nella *Philosophische Grammatik* [1932], la grammatica non ha da render conto alla realtà. Le sue regole sono in un certo senso totalmente arbitrarie, proprio come quelle di un gioco. Non vi è dubbio che quel gioco particolare presentato nel *Tractatus*, consistente nella descrizione del mondo, ha come grammatica la logica. Ma esso rimane arbitrario poiché non è il solo e unico modo di significare, cosí come la scienza, sovrana ed inflessibile nei limiti del suo gioco, non è il solo modo disponibile per esprimere il mondo. Il linguaggio, nella sua totalità mai compiuta, è una «forma di vita», è cioè qualcosa «al di là del giustificato e dell'ingiustificato; dunque per cosí dire, come un che di animale» [1949-51, trad. it. p. 57]. Ogni uomo vive dunque nel linguaggio, e il filosofo, che ne intravede la complessità e gli enigmi apparenti, esercita, grazie a quel particolare gioco linguistico che è la stessa filosofia, un'attività supplementare che deve liberarlo facendogli capire che i veri problemi non sono realmente dei problemi, poiché essi rimangono sottintesi, inesprimibili e non costituiscono gli elementi di alcun gioco.

Una concezione cosí strana ed apparentemente cosí nuova dell'analisi filosofica risulta dunque in conclusione profondamente dipendente dalla nozione stessa di gioco. Essa è stata ripresa, talvolta in modo caricaturale anche a parere di Wittgenstein, da taluni rappresentanti della cosiddetta filosofia «analitica». Essa costituisce comunque una delle espressioni piú originali del pensiero contemporaneo e la sua ispirazione, in modo imprevedibile, continuerà a stimolare a lungo.

D'altro canto, l'idea di gioco linguistico è servita d'avvio allo sviluppo di una «pragmatica» linguistica. Opposta ed associata a una sintassi, la pragmatica, il cui progetto risale a Peirce, e la cui definizione precisa è dovuta a Morris (1938), insiste nello studio del linguaggio, sul rapporto parlante-uditore. Wittgenstein, associando a tale rapporto di comunicazione le condizioni della significazione, ha ispirato i fondatori di una linguistica filosofica – Grice, Austin e Searle – che esercita una positiva influenza sulla linguistica propriamente detta. [G.-G. G.].

Borel, E.
1950 *Eléments de la théorie des probabilités*, Michel, Paris.

Caillois, R.
1967 *L'uomo e il gioco*, in *L'avvĕntura umana*, vol. IV, Vallardi, Milano, pp. 133-38.

Cournot, A.-A.
1838 *Recherches sur les principes mathématiques de la théorie des richesses*, Hachette, Paris; ed. Rivière, Paris 1938.

Lucas, W. F.
1968 *A game with no solution*, in «Bulletin of the American Mathematical Society», LXXIV, pp. 237-39.

Neumann, J. von, e Morgenstern, O.
1947 *Theory of Games and Economic Behavior*, Princeton University Press, Princeton N.J. 1947².

Pascal, B.
[1654] *Traité du triangle arithmétique, avec quelques autres petits traitez sur la mesme matière*, Desprez, Paris 1665; ora in *Œuvres mathématiques*, IV, Gallimard, Paris 1954.

Rapoport, A.
1960 *Fights, Games and Debates*, University of Michigan Press, Ann Arbor Mich.

Wittgenstein, L.
1922 *Tractatus logico-philosophicus*, Kegan Paul, Trench and Trubner, London (trad. it. Einaudi, Torino 1974).
[1932] *Philosophische Grammatik*, Blackwell, Oxford 1969.
[1933-35] *Preliminary Study for the Philosophical Investigations Generally Known as the Blue and Brown Books*, Blackwell, Oxford 1958.
[1941-49] *Philosophische Untersuchungen*, Blackwell, Oxford 1953 (trad. it. Einaudi, Torino 1967).
[1949-51] *On Certainty*, Blackwell, Oxford 1969 (trad. it. Einaudi, Torino 1978).

Con la teoria dei giochi la **scienza** moderna ha costruito uno strumento d'indagine di alcune forme del comportamento individuale e sociale (per questo cfr. **modello**, **comportamento e condizionamento**, **gruppo**, **società**) che, attraverso i metodi della programmazione lineare, fornisce un maggior rigore concettuale allo studio di situazioni economiche di concorrenza e di monopolio (cfr. **economia**). Piú in generale, la teoria si occupa di tutte le situazioni di competizione fra contendenti e della ricerca di regole ottimali di comportamento (cfr. anche **agonismo**, **conflitto**, **guerra**, **tattica/strategia**).

L'analisi dei giochi d'azzardo ha permesso il sorgere del moderno **calcolo** delle **probabilità** oltre che lo sviluppo di metodi statistici e combinatori (cfr. **caso/probabilità**, **combinatoria**, **grafo**, **distribuzione statistica**). Nonostante i tentativi di applicazione ai problemi sociologici e psicologici, i migliori risultati della teoria dei giochi si hanno nell'ambito decisionale, con la «ricerca operativa», vale a dire con i tentativi di costruire schemi di razionalità per le decisioni umane (cfr. **decisione**, **certezza/dubbio**, **razionale/irrazionale**, **ragione**).

Ma a partire dall'idea di gioco il pensiero moderno ha costruito anche uno strumento d'indagine filosofica nel quale il gioco appare come un modello funzionale della **comunicazione** (cfr. anche **linguaggio**, **competenza/esecuzione**, **atti linguistici**). Con riferimento ai «giochi linguistici» e all'intero sistema del linguaggio inteso come molteplicità di giochi si presenta poi l'aspetto normativo delle regole grammaticali e della loro natura (cfr. **grammatica**, **codice**, ma anche **logica**, **convenzione**).

Induzione statistica

L'induzione statistica è la risposta scientifica al problema filosofico dell'induzione sollevato da Hume nel *Treatise of Human Nature*. Siamo di fronte a un bigliardo. «Una palla viene messa in moto; ne tocca un'altra che immediatamente si mette in moto». Questa è la scena che Hume invita ad osservare. E continua: «Se un uomo fosse creato, come Adamo, nel pieno vigore della sua intelligenza, egli senza esperienza non sarebbe in grado di inferire dal movimento ed impulso della prima il movimento della seconda... Ma se egli avesse visto un numero sufficiente di casi di questo tipo... allora inferirebbe sempre senza esitazione il movimento della seconda» [1740, trad. it. pp. 676-77]. Hume ne conclude che tutti i ragionamenti basati sull'esperienza dipendono dal principio secondo cui «*i casi dei quali non abbiamo avuto nessun'esperienza debbono somigliare a quelli dei quali l'abbiamo avuta, e il corso della natura continua sempre uniformemente lo stesso*» [1739, trad. it. p. 102]. Ma, si chiede Hume, «che ragione abbiamo di supporre che il futuro sia conforme al passato?» E risponde: «Questa conformità è una *questione di fatto*. E, se deve essere provata, non ammetterà altra prova che non sia quella tratta dall'esperienza. Ma la nostra esperienza del passato non può provare nulla per il futuro, se non in base alla supposizione che ci sia somiglianza tra passato e futuro. Perciò, questo è un punto che non ammette affatto prova di sorta e che noi diamo per concesso senza prova alcuna» [1740, trad. it. p. 677].

Questo è allora il problema filosofico dell'induzione: si è razionalmente giustificati nel ragionare da casi ripetuti di cui si sia avuta esperienza a casi di cui non si sia avuta esperienza? La risposta di Hume è seccamente negativa. Né – dopo Hume – chi ha raccolto la sua sfida è riuscito a far molto meglio. Broad, nel 1926, battezzò il problema irrisolto dell'induzione «lo scandalo della filosofia» e Russell, nel 1946, sostenne che senza una soluzione «non si dà alcuna differenza intellettuale tra normalità e pazzia».

1. *La speranza.*

Benché riconoscesse che il ruolo principale del ragionamento induttivo è quello di costituire «la guida della vita», Hume eluse il problema della *forma specifica* in cui il ragionamento induttivo contribuisce alla formazione di decisioni *pratiche*. Tuttavia, se si riconosce che in *questo* contesto *la pratica è primaria*, allora è proprio questo il problema principale. Cosí, si sarà razionalmente giustificati a ragionare da esempi di cui si è avuta esperienza ad esempi di cui non se ne è avuta, se, e solo se, questo tipo di ragionamento contribuisce a determinare decisioni ottimali.

Già negli anni in cui Hume scriveva il suo *Treatise*, alcuni degli scienziati di punta della rivoluzione scientifica culminata nella pubblicazione dei *Principia* di Newton (1688) avevano affrontato precisamente questo problema abbozzandone una soluzione. Nel racconto di Leibniz: «I matematici del nostro tempo hanno cominciato a stimare le probabilità per quanto concerne i giochi. Il cavaliere di Méré... uomo di acuto intelletto, che era giocatore e filosofo, dié occasione a ciò formulando problemi sulle probabilità, per sapere quel che valesse il gioco interrotto a questo o a quel punto. Con che invogliò Pascal, suo amico, a esaminare un po' queste cose. La questione fece rumore e dette occasione a Huygens di scrivere il suo trattato *De alea*... Il fondamento sul quale si costruí si riduce... a prendere una media aritmetica tra piú proposizioni ugualmente attendibili» [1703-704, trad. it. II, p. 232]. Pascal stesso estese questo approccio dal caso dei giochi d'azzardo al caso generale di decisioni in condizioni d'incertezza nel pensiero *Infini-Rien: Le Pari*. Le conseguenze di questa estensione le trassero Arnauld e Nicole nel capitolo XVI della loro *Logique, ou l'Art de Penser* affermando che si dovrebbe sperare o temere un evento non solo in proporzione al guadagno o alla perdita, ma anche alla sua probabilità di accadere.

In conclusione, il tipo di problema che questi ricercatori intendevano risolvere era il seguente. Dato un agente che debba scegliere una linea d'azione tra varie linee d'azione a lui aperte, diciamo $a_1, ..., a_m$, in circostanze tali che 1) i risultati della sua scelta saranno a lui piú o meno favorevoli in funzione di quale stato di cose si realizzerà in un certo istante futuro, e 2) egli non sappia quale stato di cose si realizzerà in quell'istante, qual è la decisione ottimale? Se si denotano con $H_1, ..., H_n$ gli n stati di cose possibili (per semplicità, assumiamo qui che siano in numero finito) e con g_{ij} il guadagno che l'agente otterrà nello stato di cose H_i se sceglie a_j, allora la risposta avanzata da questo gruppo di ricercatori è che la scelta ottimale è quella della linea d'azione che rende massima la somma di prodotti $p_1 g_{1j} + ... + p_n g_{nj}$, dove p_i è la probabilità di H_i, detta *speranza matematica* associata ad a_j.

Questo primo modello della *pratica* induttiva era appena nato ed era già falsificato. Può sorprendere l'applicazione di questo termine a un modello di carattere apparentemente prescrittivo. Cosí lo intendeva certamente Arnauld. Non si trattava per lui di descrivere la pratica induttiva, bensí di «riorientarla». Il fatto è che – almeno in questo caso specifico – la dicotomia prescrittivo-descrittivo non regge. Ci si domandi infatti sotto che condizioni un modello del genere deve essere considerato *in*adeguato. Una condizione sufficiente sembra essere la seguente: che esso prescriva in alcuni casi *paradigmatici* – in cui cioè già *preteoricamente* è chiaro qual è la scelta ottimale – una scelta diversa. In altri termini, il minimo che si può richiedere perché un modello del genere sia adeguato è che almeno nei casi paradigmatici esso prescriva come ottimali le stesse scelte che si sarebbero fatte preteoricamente. Questo è certo un requisito molto lasco nella misura in cui la nostra pratica induttiva non è completamente determinata, nel senso che vi sono molti casi in cui siamo incerti su quale decisione prendere. Allora, un modello che descriva sufficientemente bene le classi di casi in cui essa è completamente determinata, organizzandoli in una struttura semplice e coerente, tenderà a farci prendere decisioni in conformità con esso anche nei casi del primo tipo. Le descrizioni funzioneranno quindi come prescrizioni. In conclusione, questo significa che se si riesce a formulare regole generali che si accordino bene con la pratica induttiva in tutti i casi paradigmatici, si tenderà a seguire quelle regole anche negli altri casi. È dunque il modello a dover essere riadattato alla pratica se il conflitto è sui casi paradigmatici, e la pratica al modello sui casi non-paradigmatici.

Appunto un conflitto su un caso paradigmatico portava alla luce il famoso paradosso di Pietroburgo, formulato da Nicola Bernoulli e pubblicato nell'appendice dell'*Essai sur les Jeux d'Hazard* di P.-R. de Montmort (1714). Il caso è il seguente. Pietro propone a Paolo il seguente gioco d'azzardo. Egli si impegna a pagargli 1000 lire se esce testa al primo lancio di una moneta equilibrata, 2000 se esce anche al secondo, ..., 2^{i-1} mila lire se esce anche all'i-esimo. La speranza matematica della linea d'azione consistente nell'accettare la proposta è:

$$\left(\frac{1}{2}\right)1000+\left(\frac{1}{4}\right)2000+\ldots+\left(\frac{1}{2^i}\right)(2^{i-1})1000+\ldots$$

assumendo che i vari lanci siano indipendenti con probabilità 1/2. Non è difficile vedere che il valore di questa somma è infinito. Questo significa che, *qualunque* sia il prezzo che Pietro chiede a Paolo per partecipare al gioco, il modello implica che la scelta ottimale per Paolo è accettare. Ma questa conseguenza è in contrasto con il fatto che per ciascuno di noi esiste un prezzo al di sopra del quale la scelta preteoricamente ottimale è rifiutare.

«Sembra allora chiaro che non tutti gli uomini possono servirsi della stessa regola per valutare il gioco. La regola stabilita deve perciò essere scartata». Cosí Daniele Bernoulli concludeva la sua analisi del paradosso [1738, § 3] e continuava: «Ma chiunque consideri il problema con perspicacia e interesse comprenderà che il concetto di *valore* che abbiamo impiegato in questa regola può essere definito in modo da rendere l'intero procedimento universalmente accettabile senza riserve. Per far ciò la determinazione del *valore* di un bene non deve fondarsi sul suo *prezzo*, ma sull'*utilità* che produce» [*ibid.*]. Il nuovo modello proposto da Bernoulli si basava sulla seguente ipotesi: «Se l'utilità di ogni possibile previsione di profitto è moltiplicata per il numero di modi in cui questo può verificarsi e dividiamo la somma di tali prodotti per il numero totale dei casi possibili, abbiamo un'utilità media, e il profitto che corrisponde a questa utilità uguaglierà il valore del rischio in questione» [*ibid.*, § 4]. Esso implicava allora che la scelta ottimale è quella che massimizza non la somma dei prodotti dei guadagni per le loro probabilità, bensí la somma dei prodotti delle loro utilità per le loro probabilità, e cioè la *speranza morale*.

2. *La coerenza.*

Per applicare il nuovo modello a specifici problemi decisionali, e quindi controllarne l'adeguatezza nel senso specificato, era tuttavia necessario fare assunzioni specifiche circa la forma delle funzioni di utilità e di probabilità. Ma assunzioni valide soltanto per casi particolari vennero generalizzate in modo arbitrario; nel caso della probabilità, si trattava dell'assunzione della distribuzione uniforme, e nel caso dell'utilità, dell'assunzione della concavità. Dopo la grande sintesi di Laplace (1812), la comunità scientifica – invece di tentare di superare queste limitazioni – abbandonò gradualmente il modello sotto l'influenza degli standard di una piatta filosofia empiristica.

Tali limitazioni furono superate soltanto tra il 1926 e il 1954 grazie ai lavori di Ramsey [1926], De Finetti [1937] e Savage [1954]. Il risultato principale dello «slittamento creativo di problema» operato da Ramsey - De Finetti - Savage consistette essenzialmente nel mostrare che un agente *necessariamente* agisce in accordo con il modello della speranza morale, purché il suo ordinamento di preferenza sia *coerente*. Questo implica in particolare che – se si assume la coerenza – la forma specifica delle funzioni di probabilità e di utilità dipenderà esclusivamente dall'ordinamento di preferenza dell'agente considerato.

Intuitivamente, dati m atti entro cui scegliere, la condizione di coerenza impone essenzialmente che le preferenze siano *transitive* (se a è preferito a b, e b a c, allora a è preferito a c), e *connesse* (o a è preferito a b, o b è preferito ad a, o b è indifferente ad a) e inoltre soddisfino il cosiddetto principio della cosa sicura. Si rappresenti l'atto a che ha la conseguenza x nello stato di cose H_1, e y in quello complementare, nel modo seguente:

$$a\left\langle\begin{matrix}x\\ y\end{matrix}\right.$$

Si può ad esempio pensare ad a come all'acquisto del biglietto H203 alla lotteria di Merano, e ad H_1 come all'eventualità che venga estratto proprio il biglietto in questione; allora, è possibile identificare x con un certo guadagno monetario e y con zero. Allora, il principio della cosa sicura afferma:

Se si preferisce l'atto $a_1\left\langle\begin{matrix}c\\ d\end{matrix}\right.$ all'atto $a_2\left\langle\begin{matrix}c'\\ d\end{matrix}\right.$ allora si deve preferire l'atto $a_3\left\langle\begin{matrix}c\\ x\end{matrix}\right.$ all'atto $a_4\left\langle\begin{matrix}c'\\ x\end{matrix}\right.$ qualunque sia x.

In altri termini, il principio in questione impone – nel confrontare due atti rispetto alla preferenza – di non tenere conto delle loro conseguenze «comuni». Se la prima condizione (transitività e connessione) pecca indubbiamente di idealizzazione – ma questo è un peccato comune nelle scienze esatte da cui dipendono molte delle loro virtú –, la seconda sembra invece interamente ovvia, una semplice conseguenza del significato di «preferenza». Si tratta invece, come si vedrà nell'ultimo paragrafo, di un'ipotesi ricca di contenuto empirico, e non del tutto corroborato.

Se queste condizioni sono soddisfatte dall'ordinamento di preferenza dell'agente considerato, diciamolo Tu, come d'uso, è possibile allora procedere a definire la relazione «l'evento A è piú probabile dell'evento B per Te», dove un evento può qui essere semplicemente identificato con un sottoinsieme dell'insieme di stati di cose considerati, lungo le linee seguenti. Si supponga di voler accertare se un individuo considera piú probabile che al Gran Premio di Monza del '79 vinca Lauda oppure Brambilla. Gli si offra la scelta seguente: 10 000 lire se e solo se vince Lauda, oppure 10 000 lire se e solo se vince Brambilla. Sembra del tutto plausibile affermare che l'individuo in questione considera piú probabile la vittoria di Lauda se preferisce la prima opzione. Si conviene perciò di dire che l'evento A è piú probabile dell'evento B per Te se Tu preferisci ricevere il premio c se e solo se A si realizza al ricevere lo stesso premio se e solo se B si realizza. Assumendo che questo ordinamento non dipenda dall'entità del premio, e che esistano per Te partizioni dell'universo in un numero di eventi equiprobabili arbitrariamente grande, si dimostra [cfr. Savage 1972, pp. 33-43] l'esistenza di un'unica funzione di probabilità, Prob, e cioè di una funzione che, per ogni coppia di eventi A e B appartenenti a una classe di eventi sufficientemente ricca, soddisfa le seguenti tre condizioni:

A.1. $\mathrm{Prob}(A) \geq 0$
A.2. Se A e B sono eventi incompatibili, $\mathrm{Prob}(A \cup B) = \mathrm{Prob}(A) + \mathrm{Prob}(B)$
A.3. $\mathrm{Prob}(S) = 1$, dove S è l'evento necessario;

dove A è piú probabile (per Te) di B se e solo se $\mathrm{Prob}(A) > \mathrm{Prob}(B)$.

Si è già detto che i particolari ordinamenti di preferenza sopra considerati debbono essere invarianti rispetto all'entità del premio. In generale, tuttavia, varieranno nel caso si diano ulteriori informazioni a chi deve fare la scelta. Ad esempio, nel caso già considerato, può ben darsi che si sia indotti a cambiare l'ordinamento se la scelta offerta è tra 10 000 se e solo se Lauda vince, *supponendo che* Brambilla abbia vinto i tre precedenti gran premi, e 10 000 se e solo se Brambilla vince, facendo la stessa supposizione. Come si può ridurre questa nozione di preferenza *condizionale* alla precedente nozione?

Si supponga che Tu preferisca a ad a', nell'ipotesi che l'evento D si realizzi. Questo – per coerenza – significherà che Tu preferisci qualunque atto b – che abbia in D le stesse conseguenze di a – a qualunque atto b' – che abbia in D le stesse conseguenze di a' –, purché naturalmente b e b' abbiano le stesse conseguenze nel complemento di D. La seguente definizione si presenta allora come estremamente naturale: a è preferito ad a', dato D, se e solo se presi due atti b e b', tali che 1) hanno le stesse conseguenze di a e a' rispettivamente in D, e 2) b e b' hanno le stesse conseguenze nel complemento di D, b è preferito a b'. Si assume qui naturalmente che D (l'evento *condizionante*) non sia *virtualmente impossibile*, dove D è un evento virtualmente impossibile se e solo se per tutti gli a e a', a è preferito ad a', dato D, se e solo se cioè le conseguenze di ogni atto in D sono irrilevanti per ogni decisione.

Alla luce di questa definizione, e della precedente di «A è piú probabile di B», si dimostra allora facilmente che a è preferito ad a', dato D (non virtualmente impossibile) equivale a dire che l'intersezione tra A e D è piú probabile di quella tra B e D, dove a è l'atto che dà il premio c se e solo se A si realizza, e a' lo stesso premio se e solo se B si realizza. Questo teorema giustifica allora la seguente definizione: A è piú probabile di B, *dato* D, se e solo se $A \cap D$ è piú probabile di $B \cap D$.

Si dimostra allora che se esiste una funzione di probabilità, Prob, tale che A è piú probabile di B se e solo se $\mathrm{Prob}(A) > \mathrm{Prob}(B)$, esiste anche, per ogni D non virtualmente impossibile, un'unica funzione di probabilità $\mathrm{Prob}(\ |D)$, tale che A è piú probabile di B, dato D, se e solo se $\mathrm{Prob}(A|D) > \mathrm{Prob}(B|D)$.

D'altra parte, si dimostra anche che – data la relazione «A è piú probabile di B, dato D» – la funzione $\mathrm{Prob}(A \cap D) \,/\, \mathrm{Prob}(D)$, per D costante, è tale che A è piú probabile di B, dato D, se e soltanto se $\mathrm{Prob}(A \cap D) / \mathrm{Prob}(D) > \mathrm{Prob}(B \cap D) / \mathrm{Prob}(D)$. Dunque, in vista dell'unicità di $\mathrm{Prob}(\ |D)$, ne segue che:

$$(1) \qquad \mathrm{Prob}(A|D) = \frac{\mathrm{Prob}(A \cap D)}{\mathrm{Prob}(D)}.$$

Questa identità, che nelle impostazioni usuali è una stipulazione piú o meno arbitraria del significato di «probabilità subordinata (o condizionata)», nell'impostazione di Ramsey - De Finetti - Savage risulta dunque una conseguenza naturale della coerenza dell'ordinamento di preferenza dell'agente considerato. Questo fatto risulta estremamente importante alla luce della seguente interpretazione di $\mathrm{Prob}(A|D)$. Si supponga che Tu sia un agente con un ordinamento di preferenza coerente. Allora, per Te, esiste, per ogni dato evento A, $\mathrm{Prob}(A)$. Si supponga che Tu abbia osservato l'evento D. *Dopo* questa osservazione, come deve cambiare la Tua valutazione iniziale della probabilità di A, e cioè $\mathrm{Prob}(A)$? Come si è appena visto, la coerenza implica che essa deve cambiare *per condizionalizzazione*, e cioè che la Tua valutazione finale della probabilità di A deve essere $\mathrm{Prob}(A|D)$. Dato tuttavia che $\mathrm{Prob}(A|D) = \mathrm{Prob}(A \cap D) / \mathrm{Prob}(D)$, la Tua valutazione finale risulterà interamente determinata dalla Tua valutazione iniziale. Il ruolo dell'«esperienza» (qui, dell'osservazione di D) consiste allora essenzialmente nel consentirTi di scontare – nella Tua valutazione della probabilità dell'evento considerato – le possibilità escluse dalla realizzazione dell'evento osservato. Se ad esempio si è interessati a valutare la probabilità che al terzo lancio di una moneta esca testa, sapendo che nei primi due è uscita testa, sarà sufficiente valutare la probabilità dei due eventi $E_1 \cap E_2 \cap -E_3$, e $E_1 \cap E_2 \cap E_3$, dove E_i è l'evento «l'i-esimo lancio dà testa» e $-E_1$ è il suo complemento: l'esperienza si è limitata dunque a ridurre il campo delle possibilità pertinenti alla valutazione, escludendo ad esempio la possibilità $-E_1 \cap E_2 \cap E_3$.

Si è visto che la coerenza è sufficiente a garantire l'esistenza e l'unicità sia delle probabilità iniziali sia delle probabilità subordinate. È anche sufficiente a garantire l'esistenza e l'unicità di una funzione U, che si dice di utilità, tale che l'atto a è preferito ad a' se e solo se

$$U(c_1)\mathrm{Prob}(H_1) + \ldots + U(c_n)\mathrm{Prob}(H_n) > U(c'_1)\mathrm{Prob}(H_1) + \ldots + U(c'_n)\mathrm{Prob}(H_n),$$

dove c_i (risp. c'_i) è la conseguenza di a (risp. a') nello stato di cose H_i?

Si dimostra [cfr. Savage 1972, pp. 69-75] anzitutto che se una funzione U del genere esiste, ne esistono infinite altre U' soggette alla condizione che $U'=rU+s$, dove r e s sono numeri reali, e r è positivo. Si dimostra d'altra parte che questa è l'unica relazione possibile tra ogni coppia di funzioni di utilità, e cioè che se U e U' sono funzioni di utilità, allora esistono un reale r (positivo) e un reale s tali che $U'=rU+s$. Dunque, ogni funzione lineare crescente di una utilità è una utilità, e viceversa. Tutto questo è perfettamente naturale se si pensa che la scelta di r e s equivale alla scelta di un'unità di misura e di un'origine. Ne consegue che se esiste una funzione di utilità, essa è unica, a meno della scelta di una unità di misura e di un'origine, e cioè a meno di trasformazioni lineari crescenti. La prima dimostrazione che esiste – nelle circostanze specificate che si riducono all'assunzione della coerenza – una funzione di utilità risale a Neumann e Morgenstern (1944). Un'ulteriore assunzione non piú problematica delle precedenti – nella misura in cui può essere intesa come una generalizzazione del principio della cosa sicura – consente di estendere questo risultato ad atti che non necessariamente abbiano un numero finito di conseguenze, come quelli fin qui considerati [cfr. *ibid.*, pp. 76-82].

È cosí conclusa l'illustrazione del risultato principale di questo paragrafo. Esso può essere riassunto dicendo che se Tu hai un ordinamento di preferenza *coerente*, allora, dato un insieme di atti $a_1, \ldots, a_m$, necessariamente esistono un'unica funzione di probabilità Prob, e una funzione di utilità U, unica a meno di trasformazioni lineari crescenti, tali che Tu preferisci a_i ad a_j $(i \neq j)$ se e solo se

$$\sum \mathrm{Prob}(H_k)U(c_k) > \sum \mathrm{Prob}(H_k)U(c'_k)$$

dove entrambe le sommatorie sono comprese tra $k=1$, e $k=n$. Questo significa che se un agente ha un ordinamento di preferenza coerente, allora necessariamente agisce in modo da massimizzare la speranza morale, o, in una terminologia piú moderna, l'*utilità prevista*. Il risultato può naturalmente essere esteso al caso di atti con un insieme infinito di conseguenze.

3. *Il teorema di Bayes.*

Nel precedente paragrafo si sono omesse dimostrazioni e altre sottigliezze matematiche perché la teoria ivi considerata ha una rilevanza essenzialmente *concettuale* per il problema dell'induzione statistica. [Chi fosse interessato può vedere, oltre al già ripetutamente citato Savage 1972, Jeffrey 1965 e 1978 e Domotor 1978].

Il punto di partenza è stata la sfida di Hume: siamo razionalmente giustificati a ragionare da casi di cui si è avuta esperienza a casi di cui non si è avuta esperienza? la questione cioè della giustificazione del ragionamento induttivo. Si è sostenuta (§ 1) l'opportunità di affrontare tale questione in rapporto alla *pratica* induttiva, intesa come la *scelta* di una certa linea d'azione (tra varie possibili) *in condizioni d'incertezza*. È stata quindi delineata (§ 1) l'ascesa – tra Pascal e Laplace attraverso Huygens, Leibniz e Daniele Bernoulli – e la caduta – dopo Laplace – di un modello di queste scelte, il modello dell'utilità prevista. I risultati del § 2 hanno allora una importante conseguenza per la sua valutazione. Rendendolo infatti indipendente da dubbie assunzioni circa la forma delle funzioni di probabilità e di utilità, consentono di supporre la sua adeguatezza, nel senso specificato nel § 1, a descrivere la pratica induttiva, e dunque ne costituiscono una giustificazione nella misura in cui tale supposizione è almeno *approssimativamente* corretta. Infine, l'evidenza sperimentale disponibile non contraddice la correttezza *approssimata* di questa supposizione. (Si veda però il § 7, soprattutto alle pp. 425-27).

C'è naturalmente un senso in cui la condizione di coerenza non ha meno bisogno di una giustificazione del modello che essa giustifica. Tuttavia, questa osservazione è corretta nella misura in cui non implica la tradizionale ricerca filosofica di giustificazioni ultime e definitive. È meglio infatti rendersi conto che è in linea di principio impossibile costruire giustificazioni di questo tipo e che di conseguenza le migliori «giustificazioni» di cui si dispone hanno piuttosto il carattere delle usuali *spiegazioni* scientifiche. Il loro merito principale non consiste nel dare certezze, ma nell'organizzare in una struttura semplice e intelligibile, nel ridurre a leggi generali, una massa di dati che precedentemente lasciava perplessi. In tal modo, la condizione in questione risulta «giustificata» dalla sua conformità al modello dell'utilità prevista, almeno nella misura in cui esso costituisce una descrizione approssimativamente adeguata della pratica induttiva, e tale modello risulta a sua volta «giustificato» dalla sua conformità a tale condizione generale. Tutto questo sembra palesemente circolare. Ma, come ha notato Goodman, si tratta di un circolo *virtuoso*. Il punto è che la giustificazione per entrambi sta proprio nel loro reciproco adattamento.

Dal nostro specifico punto di vista, questo argomento risulta cruciale, poiché se i risultati del § 2 costituiscono davvero una giustificazione del modello, come si è sostenuto, essi costituiscono in particolare una giustificazione della relazione (1) tra probabilità iniziali e probabilità finali stabilita in quel paragrafo. Ma nel caso in cui A descriva «un esempio di cui non si è avuta esperienza» (un evento futuro, ad esempio), la (1) determina appunto il modo in cui si deve ragionare «dagli esempi di cui si è avuto esperienza a quelli di cui non si è avuto esperienza», e costituisce quindi una risposta alla sfida di Hume. Si è in tal modo riusciti a giustificare – alla luce della sola condizione di coerenza – la forma piú generale di ragionamento induttivo, che, per gli sviluppi successivi, è opportuno riscrivere sotto la forma

$$\mathrm{Prob}(A|D) = \frac{\mathrm{Prob}(D|A)\,\mathrm{Prob}(A)}{\mathrm{Prob}(D)}$$

o, piú in generale,

$$\mathrm{Prob}(H_i|D) = \frac{\mathrm{Prob}(D|H_i)\,\mathrm{Prob}(H_i)}{\sum_j \mathrm{Prob}(D|H_j)\,\mathrm{Prob}(H_j)}$$

dove H_i ($i = 1, ..., n$) sono eventi incompatibili a coppie la cui unione è l'evento necessario.

Questa conseguenza immediata della (1) è nota come teorema di Bayes (1764). Esso asserisce che la probabilità *finale* di A, dato D, e cioè la probabilità di A, dopo aver osservato D, è proporzionale alla probabilità *iniziale* di A moltiplicata per il fattore Prob$(D|A)$, usualmente detto *verosimiglianza per D di A*, e quindi si scriverà usando il simbolo di proporzionalità, $\propto$:

$$\mathrm{Prob}(A|D) \propto \mathrm{Prob}(D|A)\ \mathrm{Prob}(A).$$

Esso stabilisce dunque *il* modo in cui Tu devi elaborare nuove informazioni (quelle espresse da D) per passare da un'opinione iniziale (Prob(A), la cui esistenza è garantita dalla coerenza del Tuo ordinamento di preferenza) ad una opinione finale (Prob$(A|D)$, la cui esistenza è pure garantita dalla coerenza del Tuo ordinamento di preferenza, purché D non sia virtualmente impossibile).

L'intera induzione statistica, nella sua accezione corrente, non è che un caso particolare del teorema di Bayes. Piú in particolare, la maggior parte delle applicazioni corrispondono al caso in cui sia A sia D hanno una struttura particolarmente semplice, e cioè D è una sequenza di prove «indipendenti» (in un senso che si vedrà nel § 4) soggette a una stessa legge di carattere «statistico», la cui forma – specificata dalla funzione di verosimiglianza – si assume nota a meno di uno o piú parametri incogniti, di cui A specifica il valore. Ad esempio D potrebbe descrivere l'esito di n lanci di una moneta, la cui probabilità, p, di dare testa è «incognita». In questo caso, si assume usualmente che la forma della legge cui obbediscono gli eventi in questione sia quella della *distribuzione bernoulliana*, con p che svolge il ruolo di parametro incognito. Questo significa assumere che

$$\mathrm{Prob}\Big(\bigcap_{i=1}^{r} E_i \bigcap_{j=1}^{s} -E_{i+j} \mid p = \vartheta\Big) = \vartheta^r (1-\vartheta)^s,$$

dove E_i è l'evento «l'i-esimo lancio dà testa», e $-E_i$ il suo complemento. Il problema è allora quello di «stimare» il valore di questo parametro ignoto. Un altro esempio tipico è quello in cui D descrive il risultato di n misurazioni di una data grandezza fisica, il cui valore vero, ϑ, è ignoto. In questo caso, si assume usualmente che la forma della legge cui obbediscono gli errori di misurazione sia quella determinata dalla *distribuzione normale*, con ϑ, e cioè *la previsione* della distribuzione, ignota, e la *varianza*, e cioè l'inverso della precisione dello strumento di misura, nota. Il problema è nuovamente un problema di stima del parametro ignoto. Si riaffronteranno particolareggiatamente in seguito (cfr. §§ 5.1, 5.2) entrambi questi problemi.

Il punto che va chiarito fin da ora è che non è possibile alcuna soluzione senza specificare – oltre alla verosimiglianza per D (il campione osservato) del valore del parametro incognito (specificato da A) – anche la probabilità iniziale di A. È questa la via per tenere conto della totalità d'informazioni di cui si dispone – prima di osservare il campione – circa il parametro incognito. Trattandosi in una certa misura di conoscenza «tacita», è ovviamente piú complesso codificarla sotto forma di una funzione di probabilità ben definita. Ma, in linea di principio, non si tratta di un problema di *natura* diversa da quello consistente nella specificazione della verosimiglianza. In entrambi i casi, si tratta di una Tua valutazione. La differenza è solo di grado. Nel secondo caso, *di fatto* la Tua valutazione concorderà con quella altrui piú frequentemente che nel primo.

Questo si spiega naturalmente col carattere maggiormente «pubblico» dei dati sottostanti alla valutazione delle probabilità subordinate. Ma, in nessuno dei due casi, la valutazione è imposta dai dati. Anzi, essi diventano significativi solo alla luce di una opinione iniziale; in breve, senza opinioni iniziali, niente dati, e senza dati nessun *mutamento* (nel senso già specificato) di opinioni. Come ha suggerito Suppes, è proprio questa specifica capacità umana di elaborare sotto forma di opinioni iniziali vaste masse d'informazioni non completamente verbalizzabili, non esplicitabili cioè sotto forma di proposizioni canoniche, che distingue gli esseri umani dai moderni calcolatori elettronici. Non sembra nemmeno una congettura molto ardita quella di Ramsey-Keynes secondo cui «la base dei nostri gradi di credenza, – o le probabilità a priori, come si era soliti chiamarle, – fa parte del nostro bagaglio umano, ci è forse conferita soltanto in virtú di una selezione naturale» [Keynes 1933, trad. it. p. 297]. Una tale ipotesi, e cioè che questa capacità sia una caratteristica *genetica* degli esseri umani (e, aggiungeremmo, di *tutti* gli organismi animali), costituisce in ogni caso una buona spiegazione del fatto che nella maggior parte dei casi interessanti le opinioni iniziali non differiscono in maniera tale da rendere inconciliabili le opinioni finali, dopo una sequenza di osservazioni comuni sufficientemente lunga. Si vedrà meglio nel § 4 di quale classe di casi si tratta. Essa costituisce inoltre una «giustificazione» per questa classe di opinioni, nel senso che – sotto quest'ipotesi – la classe di opinioni iniziali che ha determinato le decisioni migliori «avrà teso a predominare nella selezione naturale. Le creature ostinatamente in errore nelle loro induzioni hanno una tendenza patetica ma lodevole a perire prima di aver riprodotto la loro specie» [Quine 1969, p. 126]. È chiaro che allo stadio attuale delle conoscenze scientifiche è questa un'ipotesi di carattere del tutto speculativo; tuttavia, sembra questa una linea di ricerca molto piú feconda di quella – perseguita da Leibniz a Jeffreys fino al primo Carnap – consistente nel «giustificare» una particolare classe di opinioni iniziali alla luce di considerazioni a priori di carattere «logico». (Si veda però il § 5.1).

Infine, non è difficile dimostrare – entro l'impostazione «bayesiana» – che, sotto opportune condizioni, l'opinione finale di diversi agenti non è significativamente influenzata dalle loro opinioni iniziali. Si vedrà nel § 4 che, date due opinioni iniziali che soddisfano condizioni estremamente naturali, esiste un numero di osservazioni (appartenenti a un classe ragionevolmente ricca) sufficientemente grande perché le opinioni finali si approssimino l'una all'altra a qualunque grado desiderato di precisione.

È opportuno anticipare questo teorema generale con un esempio che ne dà il succo senza complicazioni matematiche. Pietro e Paolo trovano sul pavimento di un negozio che vende articoli per prestigiatori una moneta e iniziano a lanciarla senza preoccuparsi di esaminarne le facce. Si supponga che inizialmente nessuno dei due abbia opinioni *estreme* sulla moneta in questione, non sia cioè praticamente certo che sia una moneta equilibrata oppure che le due facce siano identiche. Si hanno allora tre stati di cose possibili: H_1 «La moneta è equilibrata», H_2 «La moneta ha due teste» e H_3 «La moneta ha due croci»; in vista dell'ipotesi del carattere non-estremo delle opinioni di Pietro e Paolo, si può ad esempio assumere che le distribuzioni iniziali di Pietro (Prob) e Paolo (Prob′) siano rispettivamente:

$$\mathrm{Prob}(H_1)=\mathrm{Prob}(H_2)=\mathrm{Prob}(H_3)=\frac{1}{3}$$

$$\mathrm{Prob}'(H_1)=\frac{2}{10}$$

$$\mathrm{Prob}'(H_2)=\mathrm{Prob}'(H_3)=\frac{4}{10}.$$

Si denoti con $E(r, s)$ l'osservazione di r lanci che dànno testa e s lanci che dànno croce, dove $r+s=n$. Si supponga allora che Pietro e Paolo concordino sulla verosimiglianza, e pongano in particolare:

$$\mathrm{Prob}(E(r,s)|H_i)=\mathrm{Prob}'(E(r,s)|H_i)=\begin{cases} 1 & \text{se } i=2,\ s=0 \\ 1 & \text{se } i=3,\ r=0 \\ 0 & \text{se } i=2,\ r=1 \\ 0 & \text{se } i=3,\ s=1 \\ \left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^s=\left(\frac{1}{2}\right)^n & \text{altrimenti.} \end{cases}$$

È anzitutto chiaro che l'osservazione dell'evento $E(1, 1)$ sarebbe *definitiva* nel senso che si avrebbe $\mathrm{Prob}(H_1|E(n, 0))=\mathrm{Prob}'(H_1|E(n, 0))=1$; in questo caso, dunque, le distribuzioni finali coinciderebbero dopo due sole osservazioni. Si consideri allora il caso piú interessante in cui l'evento osservato è $E(n, 0)$. (Naturalmente, $E(0, n)$ andrebbe ugualmente bene). In questo caso, già per $n=1$ si ha $\mathrm{Prob}(H_3|E(n, 0))=\mathrm{Prob}'(H_3|E(n, 0))=0$, in base al teorema di Bayes. Piú in generale,

$$\mathrm{Prob}(H_1|E(n, 0))=\frac{1}{3}\,2^{-n}/\mathrm{Prob}(E(n, 0));$$

$$\mathrm{Prob}'(H_1|E(n, 0))=\frac{2}{10}\,2^{-n}/\mathrm{Prob}'(E(n, 0)).$$

Naturalmente, $\mathrm{Prob}(H_2|E(n, 0))=1-\mathrm{Prob}(H_2|E(n, 0))$, e analogamente per Prob′. D'altra parte, per la condizione A.2 di p. 388 (detta anche principio delle probabilità totali) e per la (1),

$$\mathrm{Prob}(E(n, 0))=\mathrm{Prob}(H_1)\,\mathrm{Prob}(E(n, 0)|H_1)+ +\mathrm{Prob}(H_2)\,\mathrm{Prob}(E(n, 0)|H_2)=\frac{1}{3}\,2^{-n}+\frac{1}{3};$$

$$\mathrm{Prob}'(E(n, 0))=\frac{2}{10}\,2^{-n}+\frac{4}{10}.$$

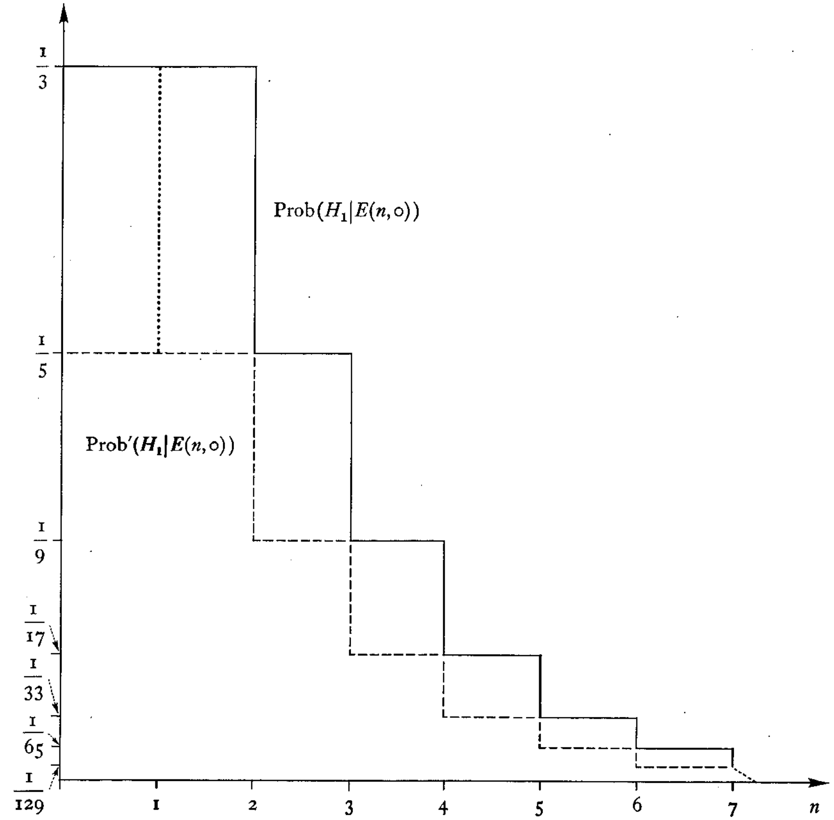


Figura 1.

In ascissa, è stato rappresentato il numero di osservazioni, n. In ordinata, $\mathrm{Prob}(H_1|E(n, 0))$ e $\mathrm{Prob}'(H_1|E(n, 0))$. Si può prendere l'area dei successivi rettangoli come una misura della differenza tra le opinioni finali. Quindi, l'area del primo rettangolo rappresenterà la differenza dopo 0 osservazioni, dèl secondo dopo 1, e cosí via. Naturalmente **Prob′** è rappresentata dalla spezzata tratteggiata.

Dunque:

$$\mathrm{Prob}(H_1|E(n, 0))=\frac{1}{1+2^n};$$

$$\mathrm{Prob}'(H_1|E(n, 0))=\frac{1}{1+2^{n+1}}.$$

Nella tabella 1 sono riportati i valori di queste due funzioni per alcuni valori di n, e la differenza tra tali valori; la figura 1 rappresenta la stessa situazione. Non è difficile vedere che dopo solo sette lanci la differenza di opinioni iniziali si è praticamente annullata, e che aumentando sufficientemente il numero

Tabella 1.
Alcuni valori di $\mathrm{Prob}(H_1|E(n, 0))$ e $\mathrm{Prob}'(H_1|E(n, 0))$.

	$n=0$	1	2	3	4	5	6	7
Prob	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{9}$	$\frac{1}{17}$	$\frac{1}{33}$	$\frac{1}{65}$	$\frac{1}{129}$
Prob′	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{9}$	$\frac{1}{17}$	$\frac{1}{33}$	$\frac{1}{65}$	$\frac{1}{129}$	$\frac{1}{257}$
Prob − Prob′	0,13	0,13	0,11	0,058	0,030	0,015	0,0070	0,0038

di osservazioni, le opinioni finali possono essere portate l'una vicina all'altra quanto si vuole.

Si è sottolineato che il teorema di Bayes costituisce la forma piú generale di ragionamento induttivo. Si è perciò sostenuto che la «giustificazione» proposta in termini di coerenza costituisce una soluzione al problema filosofico dell'induzione: si è razionalmente giustificati a ragionare da esempi di cui si è avuto esperienza ad esempi di cui non si è avuto esperienza? Risulta tuttavia chiaramente dalle precedenti citazioni che Hume intendeva riferirsi a una forma piú specifica di ragionamento induttivo, e cioè a quella forma in cui la probabilità di un evento futuro è tanto maggiore quanto maggiore è la frequenza osservata di eventi «analoghi».

Fortunatamente, nessuna regola di questo genere è implicata dalla condizione di coerenza. Dunque, questa specifica forma di ragionamento induttivo non ha alcuna «giustificazione» alla luce di tale condizione. In effetti, questa regola non è in generale valida, come ha mostrato in modo definitivo il paradosso di Goodman [cfr. ad esempio Goodman 1955, pp. 59 sgg.; Jeffrey 1965, pp. 175-76]. Ma, anche indipendentemente da tale paradosso, non è difficile vedere che essa determina decisioni inaccettabili ad esempio nel caso di scommesse su sequenze binarie periodiche.

Tuttavia, se pure è impossibile – alla luce della sola condizione di coerenza – una giustificazione *globale* di questa regola, è possibile una sua giustificazione *locale*, aggiungendo ai postulati che caratterizzano la condizione di coerenza una sola nuova ipotesi estremamente chiara e semplice, l'ipotesi di *scambiabilità*. In tal modo, questa forma particolare di ragionamento induttivo risulterà valida in tutti e soli quei casi in cui Tu valuterai un dato insieme di eventi scambiabili, in un senso che si vedrà nel prossimo paragrafo.

4. *Il teorema di rappresentazione di De Finetti.*

È questo uno dei pochi risultati profondi – sia da un punto di vista statistico che da un punto di vista filosofico – di tutta la letteratura statistica. Val perciò la pena di esaminarlo in dettaglio.

Si può cominciare riformulando il problema come segue: perché nella maggioranza dei casi si è indotti a valutare le probabilità sulla base delle frequenze osservate? Si tratta allora di capire un meccanismo particolare, ma estremamente importante, di *apprendimento dall'esperienza*.

È opportuno in primo luogo considerare in quali casi è *escluso* l'apprendimento dall'esperienza. Dati n eventi, $E_1, ..., E_n$, e una funzione di probabilità Prob, si dice che tali eventi sono *indipendenti* rispetto a Prob se e solo se $\mathrm{Prob}(E_{i1} \cap E_{i2} \cap ... \cap E_{ij}) = \mathrm{Prob}(E_{i1})\,\mathrm{Prob}(E_{i2}) \,...\, \mathrm{Prob}(E_{ij})$, per $j = 1, ..., n$; se cioè la probabilità dell'intersezione di ogni j eventi scelti da quelli dati è uguale al prodotto delle loro probabilità. Nel caso di $n = 2$, si ha in particolare $\mathrm{Prob}(E_1 \cap E_2) = \mathrm{Prob}(E_1)\,\mathrm{Prob}(E_2)$.

Non è allora difficile vedere che l'assunzione dell'indipendenza esclude l'apprendimento dall'esperienza. Infatti, per la (1), si ha $\mathrm{Prob}(E_2|E_1) = \mathrm{Prob}(E_2 \cap E_1)/\mathrm{Prob}(E_1)$. Ma l'assunzione dell'indipendenza implica $\mathrm{Prob}(E_2 \cap E_1) = \mathrm{Prob}(E_2)\,\mathrm{Prob}(E_1)$, e dunque $\mathrm{Prob}(E_2|E_1) = \mathrm{Prob}(E_2)$, e cioè l'identità tra probabilità finale e iniziale. L'osservazione dell'evento E_1 non ha influenzato in alcun modo l'opinione iniziale, cosicché l'opinione finale coincide con quella iniziale. L'indipendenza è dunque un modo per immunizzare le nostre opinioni iniziali dall'esperienza, per non mutarle *qualunque cosa accada*.

Il caso interessante – dal punto di vista dell'induzione statistica – è perciò quello della *dipendenza*. L'interesse del caso d'indipendenza consiste essenzialmente nella possibilità di rappresentare il caso della dipendenza in termini di misture opportune di casi d'indipendenza.

Quindi, condizione *necessaria* perché si dia apprendimento dall'esperienza è che l'opinione iniziale *non* sia caratterizzata dall'indipendenza. Si consideri allora la seguente condizione, detta di *scambiabilità*. Dati n eventi, $E_1, ..., E_n$, si dicono n-costituenti le intersezioni di s ($s = 0, ..., n$) di tali eventi, con ciascuno dei complementi degli $n-s$ ($=r = 0, ..., n$) rimanenti, e cioè gli eventi: $(-)E_1 \cap ... \cap (-)E_n$, dove «$(-)$» va rimpiazzato con «$-$» o con nulla in tutte le combinazioni possibili. È chiaro che il numero di n-costituenti è 2^n, e di quelli caratterizzati da un numero fissato r di «insuccessi» è $\binom{n}{r}$. La condizione di scambiabilità equivale allora all'assunzione che ciascuno di questi $\binom{n}{r}$ n-costituenti ha la stessa probabilità, e cioè che per n fissato la probabilità di un arbitrario costituente dipende soltanto dal numero di «insuccessi» r che lo caratterizzano, e non dal loro ordine. Per dirla con Savage, «ogni storia finita h [n-costituente] ha [sotto la condizione di scambiabilità] la stessa probabilità di ogni altra storia finita h' della stessa lunghezza e con altrettanti successi e insuccessi» [1973, pp. 425-26].

Se si denota la probabilità dell'unione di tutti gli n-costituenti con un r fissato con «$\omega_r^{(n)}$», allora

$$(2) \qquad \mathrm{Prob}(E(n-r, r)) = \frac{\omega_r^{(n)}}{\binom{n}{r}}$$

dove naturalmente $E(n-r, r)$ denota uno qualunque di tali $\binom{n}{r}$ costituenti equiprobabili.

Un modello semplice di questa situazione è costituito dall'esempio già trattato della moneta truccata. Qui, gli n-costituenti sono tutte le sequenze di esiti di n lanci di una data moneta, per E_i = l'i-esimo lancio dà testa. Non è difficile controllare che entrambe le funzioni di probabilità considerate soddisfano su tali eventi la condizione di scambiabilità.

Il problema è allora: è la scambiabilità una condizione sufficiente perché si dia apprendimento dall'esperienza? Il teorema di De Finetti dà una risposta positiva (a meno di una condizione addizionale del tutto banale) precisamente a questa domanda. Per formularlo, occorre tuttavia ancora una nozione tecnica, la nozione di funzione di distribuzione (per cui si veda anche l'articolo «Distribuzione statistica», in questa stessa *Enciclopedia*). Una funzione F definita sull'insieme dei reali si dice funzione di distribuzione se e solo se soddisfa le seguenti tre condizioni:

1. F è non-decrescente (cioè se $x<y$, allora $F(x)\leq F(y)$)

2. *a*) $\lim\limits_{x\to+\infty} F(x)=1$

2. *b*) $\lim\limits_{x\to-\infty} F(x)=0$

3. $\lim\limits_{y\downarrow x} F(y)=F(x)$ (cioè, F è continua a destra).

Identificando l'insieme di stati di cose possibili con l'insieme dei reali, $F(a)$ s'identifica con la probabilità dell'evento che x sia minore o uguale ad a, e cioè con la probabilità del sottoinsieme di $\mathbf{R}: \{x \mid x\leq a\}$. Due casi sono di particolare interesse.

Il primo è quello nel quale la probabilità è *concentrata tutta su un punto*, in cui cioè:

$$(3)\qquad F(x)=\begin{cases} 0 & \text{se } x<a \\ 1 & \text{se } x\geq a \end{cases}$$

È il caso in cui si è certi che $z=a$, in cui cioè ciascun evento ha probabilità 1 oppure 0 secondo che contenga o meno il punto $z=a$. Il secondo caso è quello in cui F è ovunque continua e differenziabile in un intervallo (a, b) tale che $F(b^-)-F(a)=1$, dove $F(x^-)=\lim\limits_{y\uparrow x} F(x)$, in cui cioè si è certi che $z\in(a, b)$. In tal caso, vale che

$$(4)\qquad F(x)=\begin{cases} 0 & \text{se } x<a \\ \int_a^x f(v)\,dv & \text{se } a\leq x<b \\ 1 & \text{se } x\geq b \end{cases}$$

dove $f(x)=dF(x)/dx$ viene detta funzione di densità.

Si è ora in grado di formulare il seguente teorema:

TEOREMA DI RAPPRESENTAZIONE (De Finetti). *Se E_1, E_2, E_3, ... è una successione indefinitamente proseguibile di eventi scambiabili rispetto a una data funzione di probabilità* Prob, *allora esiste un'unica funzione di distribuzione F concentrata sull'intervallo tra* 0 *e* 1 *(tale che cioè $F(1)-F(0)=1$) tale che*

$$\text{Prob}(E(n-r, r))=\int_0^1 x^{n-r}(1-x)^r\,dF(x).$$

Se ne dà ora un abbozzo di dimostrazione. Posto $\mu_0=1$, si denoti con «μ_n» il valore di $\text{Prob}(E(n, 0))$. Dunque, in base alla (2)

$$\mu_n=\omega_n^{(n)}\Big/\binom{n}{n}=\omega_n^{(n)}.$$

D'altra parte, A.2 implica: $\text{Prob}(E(n, 0))+\text{Prob}(E(n-1, 1))=\text{Prob}(E(n-1, 0))$. Dunque:

$$(5)\qquad \text{Prob}(E(n-1, 1))=\mu_{n-1}-\mu_n.$$

L'operatore Δ^s viene definito nel modo seguente:

a) $\Delta^1\mu_n=\mu_{n+1}-\mu_n$

b) $\Delta^s\mu_n=\Delta(\Delta^{s-1}(\mu_n))$,

dove $\Delta=\Delta^1$. Si ha allora, in base alla (5),

$$\text{Prob}(E(n-1, 1))=-\Delta\mu_{n-1}=(-1)^1\Delta^1\mu_{n-1},$$

da cui segue per induzione:

$$\text{Prob}(E(n-r, r))=(-1)^r\Delta^r\mu_{n-r}.$$

Ma A.1 implica $\text{Prob}(E(n-r, r))\geq 0$. Dunque:

$$(-1)^r\,\Delta^r\mu_{n-r}\geq 0.$$

Si dice allora che la successione (μ_n) è *completamente monotona* (o totalmente decrescente). È possibile allora applicare il teorema di Hausdorff [cfr. Hardy 1949, pp. 258-59] alla successione (μ_n). Esso infatti afferma che se (μ_n) è una successione totalmente decrescente, allora esiste un'unica funzione di distribuzione concentrata sull'intervallo tra 0 e 1, tale che $\int_0^1 x^n\,dF(x)=\mu_n$. Ne segue allora che

$$(6)\qquad \text{Prob}(E(n, 0))=\int_0^1 x^n\,dF(x).$$

È immediato a questo punto passare dalla (6) alla conclusione voluta. Da A.2 segue infatti che

$$(7)\qquad \text{Prob}(E(s+1, r))+\text{Prob}(E(s, r+1))=\text{Prob}(E(s, r)).$$

Partendo dalla (6), e utilizzando successivamente la (7), non è allora difficile dimostrare per induzione su r che

$$(8)\qquad \text{Prob}(E(n-r, r))=\int_0^1 x^{n-r}(1-x)^r\,dF(x).$$

Non è a prima vista ovvio come questo teorema si connetta con il problema da cui si sono prese le mosse, e cioè il problema dell'apprendimento dall'esperienza, o ancora del rapporto tra frequenza e probabilità. È bene cominciare allora con il dare una versione piú precisa del problema. Una data funzione di probabilità Prob consente *apprendimento dall'esperienza relativamente a una successione data di eventi,* $E_1, E_2, E_3, \ldots$ se e solo se

$$\mathrm{Prob}(E_{n+2}|E(s+1, r)) > \mathrm{Prob}(E_{n+1}|E(s, r));$$

se e solo se cioè la probabilità che si realizzi un evento di un dato tipo è tanto maggiore quanto maggiore è la frequenza osservata degli eventi dello stesso tipo.

Dal teorema di rappresentazione di De Finetti segue allora il seguente corollario [dimostrato in Humburg 1971]:

COROLLARIO. Se la successione $E_1, E_2, E_3, \ldots$ è scambiabile rispetto a Prob, allora

$$\mathrm{Prob}(E_{n+2}|E(s+1, r)) \geq \mathrm{Prob}(E_{n+1}|E(s, r)). \tag{9}$$

La sua semplice dimostrazione servirà a mettere in luce la condizione addizionale sotto cui è possibile escludere l'identità tra le due espressioni. In primo luogo

$$\mathrm{Prob}(E_{n+2}|E(s+1, r)) = \frac{\mathrm{Prob}(E(s+2, r))}{\mathrm{Prob}(E(s+1, r))}$$

$$\mathrm{Prob}(E_{n+1}|E(s, r)) = \frac{\mathrm{Prob}(E(s+1, r))}{\mathrm{Prob}(E(s, r))}$$

Quindi, la (9) vale se e solo se

$$\mathrm{Prob}(E(s+2, r))\mathrm{Prob}(E(s, r)) \geq (\mathrm{Prob}(E(s+1, r)))^2. \tag{10}$$

In vista dell'ipotesi di scambiabilità, è possibile applicare a queste espressioni il teorema di De Finetti. Si ottiene allora che la (10) vale se e solo se

$$\left[\int_0^1 x^{s+2}(1-x)^r\, dF(x)\right]\left[\int_0^1 x^s(1-x)^r\, dF(x)\right] \geq \tag{11}$$

$$\geq \left[\int_0^1 x^{s+1}(1-x)^r\, dF(x)\right]^2.$$

Si ponga $d\varphi(x) = x^s(1-x)^r\, dF(x)$. φ è cioè la distribuzione che si ottiene a partire dalla distribuzione iniziale F, *dopo* aver osservato s «successi» e r «insuccessi». La differenza tra φ e F riflette dunque «l'accumularsi dell'esperienza». Dunque la (11) vale se e solo se

$$\left[\int_0^1 x^2\, d\varphi(x)\right]\left[\int_0^1 d\varphi(x)\right] \geq \left[\int_0^1 x\, d\varphi(x)\right]^2. \tag{12}$$

Ma questa disuguaglianza è soltanto un caso speciale della disuguaglianza di Cauchy-Schwarz.

A questo punto, dunque, il problema di sapere sotto quali condizioni vale il segno di uguaglianza nella (9) si riduce all'analogo problema nella (12). Si può dimostrare che esso vale se e solo se esiste un numero reale a tale che $x = a$ quasi ovunque nell'intervallo tra 0 e 1, relativamente a F. In altre parole, F è interamente concentrata su un punto dell'intervallo. Si tratta precisamente del caso (3) considerato a p. 398. Ma tutto questo è estremamente naturale. Significa infatti che Prob, anche se «scambiabile», escluderà apprendimento dall'esperienza, se Tu, prima di ogni sperimentazione, sei praticamente certo dell'esito che otterrai. Dunque, la scambiabilità è una condizione sufficiente di apprendimento dall'esperienza se e solo se F non rappresenta un'opinione iniziale «estrema».

Piú in particolare, sotto queste condizioni,

$$\mathrm{Prob}(E_{n+1}|E(s, r)) = \frac{\int_0^1 x^{s+1}(1-x)^r\, dF(x)}{\mathrm{Prob}(E(s, r))} = \int_0^1 x\, dG(x),$$

dove

$$dG(x) = \frac{x^s(1-x)^r\, dF(x)}{\mathrm{Prob}(E(s, r))},$$

$$G(x) = \frac{1}{\mathrm{Prob}(E(s, r))}\int_0^x v^s(1-v)^r\, dF(v).$$

In tal modo, l'effetto dell'accrescersi dell'esperienza sulla Tua opinione circa un qualunque evento non osservato (rappresentato qui da E_{n+1}) è interamente rispecchiato dalla differenza tra le funzioni di distribuzione iniziale e finale, F e G. Prima dell'osservazione degli s «successi» e degli r «insuccessi», la probabilità di un successo è: $\int_0^1 x\, dF(x)$. Dopo, essa risulta uguale a $\int_0^1 x\, dG(x)$. Si può allora dimostrare che al crescere di $s+r$, G si approssima a una funzione di distribuzione interamente concentrata sul punto $x = s/(s+r)$.

Si riconsideri ora – alla luce di questi risultati – l'esempio della moneta del § 3, abbandonando l'ipotesi restrittiva che gli unici valori possibili della probabilità, p, che essa dia testa, siano 0, 1, e 1/2. Si ha quindi che $p = \mathrm{Prob}(E_1) = \mathrm{Prob}(E_2) = \ldots$, dove p è un parametro «incognito». Se si assume inoltre che i vari lanci della moneta siano prove indipendenti di uno stesso esperimento, allora, subordinatamente all'ipotesi che il valore «vero» di p sia ϑ, la verosimiglianza per $p = \vartheta$ di $E(s, r)$ sarà determinata dalla distribuzione bernoulliana $\vartheta^s(1-\vartheta)^r$. Si ha allora, banalmente,

$$\mathrm{Prob}(E(s, r)) = \int_0^1 \vartheta^s(1-\vartheta)^r\, dF(\vartheta),$$

dove F è una funzione di distribuzione concentrata sull'intervallo compreso tra 0 e 1.

Il teorema di rappresentazione di De Finetti stabilisce precisamente il converso di questo ovvio risultato. Esso implica dunque che la nozione d'indipendenza subordinata è «traducibile» nella nozione di scambiabilità. Questo dà la possibilità di evitare di parlare della probabilità, p, che la moneta in questione dia testa, come di un parametro «incognito». Questo modo di parlare è in contrasto con il metodo con cui entro la impostazione di Ramsey - De Finetti - Savage si costruisce la nozione di probabilità. Esso infatti esclude che possano esservi probabilità «incognite». La scambiabilità è invece una nozione perfettamente compatibile con tale impostazione, e, alla luce del teorema di rappresentazione di De Finetti, consente di affrontare i casi di scambiabilità *come se* fossero casi di «prove indipendenti di uno stesso evento a probabilità costante, ma incognita». La rappresentazione sarà «effettiva, significativa, se ogni caso di indipendenza corrisponde a un'"ipotesi" oggettivamente definita (come ad es. la percentuale di palle bianche in un'urna di composizione ignota), mentre la rappresentazione è solo formale se cosí non è (come ad es. nel caso di una moneta deformata)» [De Finetti 1970, p. 595]. Ma proprio questa possibilità di traduzione consente di continuare a parlare di probabilità «incognita» senza pericolo.

Si supponga – nell'esempio in questione – di considerare equiprobabili tutti i valori di p. Questo significa allora assumere come distribuzione iniziale la distribuzione *uniforme* (o di Bayes-Laplace), a porre cioè $F(\vartheta)=\vartheta$. Il grafico di F è allora quello rappresentato dalla figura 2. In tal caso,

$$\mathrm{Prob}(E(s, r)) = \int_0^1 \vartheta^s(1-\vartheta)^r\, d\vartheta = \frac{s!r!}{(s+r+1)!}$$

e dunque

$$\mathrm{Prob}(E_{n+1}|E(s, r)) = \frac{s+1}{s+r+2},$$

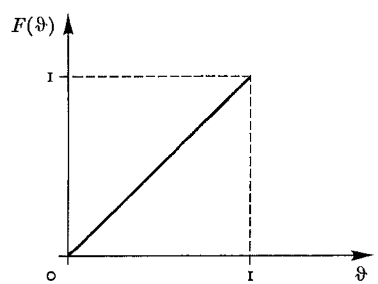


Figura 2.
Grafico della distribuzione $F(\vartheta)=\vartheta$.

che è la famosa *regola di successione* di Laplace. In questo caso è immediato vedere che – al crescere di $s+r$ – la probabilità di «successo» si approssima alla frequenza relativa osservata dei «successi».

Geometricamente, la situazione può essere rappresentata mediante le figure 3-5. Al crescere di $s+r$, la distribuzione iniziale uniforme F si sposta da sinistra a destra attraverso le distribuzioni determinate dai tre grafici nelle figure 3-5, approssimandosi dunque a una distribuzione finale interamente concentrata sul punto $x=s/(s+r)$. Se si pensa alla funzione $f(x)=dF(x)/dx$ come ad una distribuzione di masse sull'intervallo tra 0 e 1 (assumendo «massa totale» $=1$), in modo che $f(x)\,dx$ è la massa assegnata al punto x, quel che accade al crescere di $s+r$ è una ridistribuzione di masse «a favore» dei punti prossimi a $x=s/(s+r)$, che si approssima sempre piú a una distribuzione in cui la massa totale è interamente concentrata su questo punto.

È opportuno tornare ora al punto di partenza di questo paragrafo. Da allora, si è stabilito che l'ipotesi di scambiabilità – aggiunta alla condizione di coerenza – è in grado di «giustificare» quel caso speciale di ragionamento induttivo per cui la probabilità di un evento non-osservato è tanto maggiore quanto lo è la frequenza osservata di eventi «analoghi». Si è anche visto che in quei casi in cui l'ipotesi di scambiabilità è soddisfatta, le opinioni finali di agenti diversi tenderanno a convergere, per quanto divergano le loro opinioni iniziali (purché non siano estreme).

Un'ultima conseguenza importante di questo teorema. Esso – rendendo possibile una «traduzione» della locuzione «prove indipendenti di uno stesso evento a probabilità costante ma incognita» entro l'impostazione bayesiana – in termini della nozione di scambiabilità – mette quest'ultima nella posizione di affrontare in modo concettualmente soddisfacente l'intera gamma di problemi dell'induzione statistica, intesa secondo la cosiddetta impostazione «oggettivistica».

Tale paradigma della pratica statistica domina la scena scientifica a partire dagli anni '20-30, da quando cioè prima Fisher, e poi Neyman e Pearson, ne dettero (sia pure lungo linee diverse) una formulazione sistematica. Esso prese le mosse dalle critiche già avanzate verso la metà dell'Ottocento da Boole e Venn al paradigma bayesiano, centrate essenzialmente sull'inaccettabilità della distribuzione iniziale uniforme. Traendo da queste la (falsa) conseguenza che la nozione stessa di probabilità iniziale fosse senza significato, si sviluppò quindi in una serie di metodi ad hoc volti ad impostare l'induzione statistica indipendentemente dalla determinazione delle probabilità iniziali. Come ha notato De Finetti, si trattava di un completo abbandono «dell'idea di un'interpretazione sistematica e significativa dell'induzione statistica, per ridursi ad escogitare caso per caso dei "test" per "confermare" delle ipotesi, o dei metodi per "stimare" dei parametri» [1959, p. 20].

Si ometterà la riesposizione di tali metodi di stima e di test, e in generale del paradigma «oggettivistico» (di cui abbonda d'altra parte la letteratura statistica [un'ottima esposizione si trova ad esempio in Cox e Hinkley 1974]). Non perché la maggior parte dei suoi risultati non sia perfettamente accettabile. Ma piuttosto perché il modo in cui sono ottenuti è inaccettabile. «Ma

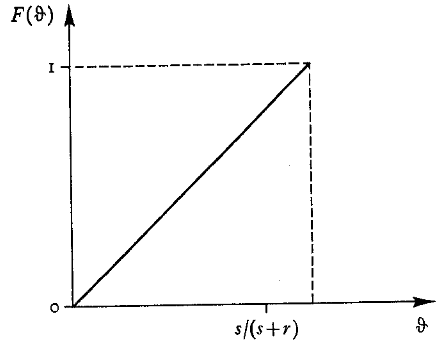


Figura 3.
Grafico della distribuzione iniziale: tutti i valori di ϑ sono equiprobabili.

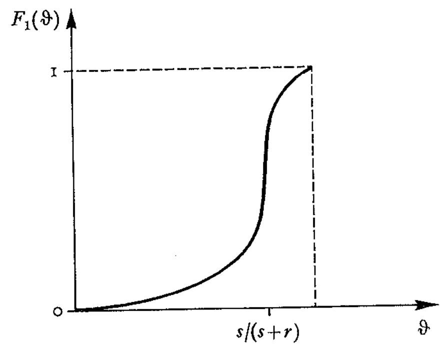


Figura 4.
Grafico della distribuzione finale dopo un numero finito di osservazioni: i valori di ϑ prossimi alla frequenza osservata sono i piú probabili.

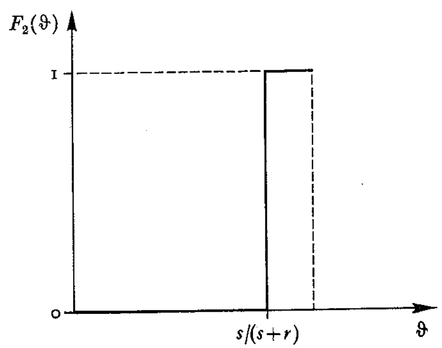


Figura 5.
Grafico della stessa distribuzione finale quando il numero delle osservazioni tende all'infinito: un unico valore di ϑ è possibile, quello coincidente con la frequenza osservata.

l'intuizione ha salvato lo statistico [oggettivista] dagli errori. La mia tesi è che il metodo bayesiano giustifica ciò che egli ha sempre fatto (reinterpretandolo e correggendolo; agg. d. trad.) e che sviluppa nuovi metodi che mancano nell'approccio "ortodosso"» [Lindley, citato e glossato in De Finetti 1970, p. 621].

Questo programma di reinterpretazione, correzione e sviluppo del paradigma «oggettivistico» entro quello bayesiano è in una certa misura già realizzato, grazie soprattutto ai lavori di De Finetti, Savage, Jeffreys, Good, Lindley, Jaynes e di un numero sempre crescente di ricercatori «convertiti» al paradigma bayesiano. Si vedrà negli ultimi paragrafi qualche semplice esempio di tali risultati. Naturalmente, molto resta ancora da fare; ma i risultati fin qui ottenuti sono già sufficienti a stabilire *anche sul piano strettamente operativo* la superiorità del nuovo paradigma. Come ha notato Jaynes, «i metodi ortodossi, se raffinati al massimo (usando test unilaterali, riportando livelli di significatività critici, usando riassunti sufficienti, o condizionando rispetto a tutta l'informazione ancillare), diventano matematicamente equivalenti ai metodi bayesiani basati su distribuzioni iniziali non-informative [le tanto criticate distribuzioni uniformi!] purché non vi siano parametri di disturbo, ed esistano riassunti sufficienti o un insieme completo di riassunti ancillari. Altrimenti, l'equivalenza non può essere ottenuta e il risultato bayesiano si dimostra superiore» [1976, p. 231].

Questo non significa naturalmente negare che, almeno in parte, fu lo stesso successo dei metodi «oggettivisti», soprattutto in biologia, a stimolare originariamente la ripresa critica dei metodi bayesiani.

A conclusione di questo paragrafo, val la pena di accennare alla possibilità di generalizzare la nozione di scambiabilità in varie direzioni. Non deve infatti sfuggire che non sono molti i casi che soddisfano esattamente lo schema di scambiabilità. Esso costituisce perciò un'idealizzazione a cui la varietà dei casi incontrata nella pratica è riducibile solo mediante ulteriori complicazioni. Si tratta però di complicazioni di carattere essenzialmente «tecnico» che lasciano invariato il «panorama concettuale».

Una prima importante generalizzazione è costituita dalla nozione di scambiabilità *parziale*; essa si ha supponendo che la classe di eventi considerata sia ripartita in piú sottoclassi entro ciascuna delle quali vale la scambiabilità usuale. Le catene di Markov sono uno dei processi piú interessanti concettualizzabili nei termini di questa nozione. [Per altri esempi, oltre che per uno sviluppo tecnico della nozione di scambiabilità parziale, si confronti De Finetti 1959, pp. 92-100].

Una seconda generalizzazione è costituita dalla nozione di scambiabilità *rispetto a classi di variabili aleatorie* (per la nozione di variabile aleatoria si veda il § 5.1). [Per una generalizzazione del teorema di De Finetti lungo queste linee cfr. ad esempio Hewitt e Savage 1955; Good 1965, pp. 21-23].

5. *Applicazioni statistiche.*

5.1. La stima puntuale.

Le distinzioni che caratterizzano i §§ 5.1-5.3 sono in realtà del tutto irrilevanti dal punto di vista bayesiano. Esse sussistono entro l'impostazione «oggettivistica» proprio perché, in mancanza di un criterio unitario per affrontare il problema generale dell'induzione statistica, essa è stata costretta ad affrontare queste classi di casi con metodi diversi. Al contrario, entro l'impostazione bayesiana, il metodo per affrontare tutte queste classi di casi è *unico*: esso consiste – come si è già sottolineato – nel determinare una distribuzione finale a partire da una verosimiglianza e da una distribuzione iniziale, via il teorema di Bayes. Sono state conservate semplicemente per facilitare un confronto tra le due impostazioni. Naturalmente, non sarà dato piú di qualche esempio schematico per ciascuna classe di casi. D'altra parte, l'impostazione bayesiana è troppo recente (nella sua forma postlaplaceana) perché la maggior parte di tali esempi possa essere considerato piú che un abbozzo provvisorio di soluzione dei corrispondenti problemi dell'impostazione «oggettivistica». Infine, sempre allo scopo di facilitare il confronto, sarà sfruttata tutta la libertà concessa dal teorema di rappresentazione e si parlerà quindi liberamente di «prove indipendenti» e di probabilità «incognita».

Dato che in questa classe di problemi svolge un ruolo centrale la nozione di variabile aleatoria, è da questa che è opportuno iniziare. Si è già visto che il dominio delle funzioni di probabilità è costituito da una classe $\mathfrak{S}$ sufficientemente ricca di eventi E, dove gli E sono sottoinsiemi di un insieme fissato S, da intendersi come l'insieme di «stati di cose possibili» per il problema considerato. Piú in generale, si può pensare ad S come a un insieme astratto di «punti». La terna (S, $\mathfrak{S}$, Prob) viene usualmente detta uno spazio di probabilità; in particolare S viene detto spazio delle alternative, o dei campioni, ed $\mathfrak{S}$ spazio degli eventi. L'assunzione standard è che lo spazio degli eventi abbia la struttura di un'algebra (o di una σ-algebra) di sottoinsiemi di S. Dato uno spazio di probabilità, una variabile aleatoria è una funzione X a valori reali definita su S tale che l'insieme di tutti i punti s di S tali che $X(s) \leq a$, è un evento (appartiene cioè a $\mathfrak{S}$), per tutti i reali a. È allora chiaro che se s'identificano gli eventi E con le loro funzioni caratteristiche $c_E(s)$ (dove $c_E(s) = 1$, se s appartiene a E; $= 0$, altrimenti), gli eventi non sono che casi speciali di variabili aleatorie. È evidentemente immediata l'estensione della nozione d'indipendenza al caso di variabili aleatorie.

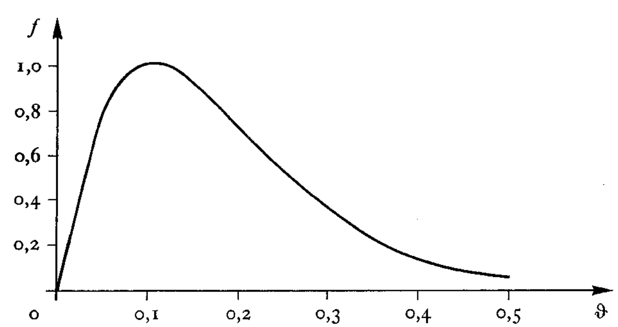


Figura 6.
La funzione di verosimiglianza determinata dalla distribuzione bernoulliana per $n = 10$ e $r = 9$.

Dato che ormai il lettore ne ha una certa familiarità, conviene iniziare dal problema di «stimare» la probabilità incognita che una moneta deformata dia testa. Entro l'approccio bayesiano, la soluzione di questo problema richiede – come si sa – la specificazione di due componenti:

a) un *modello dei dati* intesi come valori, $x_1, x_2, \ldots$ di variabili aleatorie $X_1, X_2, \ldots$, che consiste nella formulazione di una distribuzione comune per tali variabili, assunte come indipendenti subordinatamente al valore di uno o piú parametri, $\vartheta_1, \ldots, \vartheta_n$ caratteristici della distribuzione stessa. Restringendosi al caso di un solo parametro incognito, se la funzione di distribuzione $F(x|\vartheta)$ ammette una funzione di densità $f(x|\vartheta)$, il modello resta completamente specificato da quest'ultima. Essa non è che una generalizzazione al caso continuo della funzione di verosimiglianza. $f(x|\vartheta)$ va intesa come funzione di ϑ con x fissato. Si è già notato che la tendenza a concordare sulla funzione di verosimiglianza piú appropriata ad ogni dato problema dà a tale componente un'apparenza di «oggettività» del tutto illusoria. In realtà, è anch'essa solo l'espressione di un'opinione non meno aperta alla critica e alla discussione della componente costituita dalla distribuzione iniziale;

b) una distribuzione iniziale di probabilità sullo spazio dei parametri, che nel seguito si assumerà ammetta una funzione di densità, $\pi(\vartheta)$.

In questo problema, il parametro incognito è la probabilità p che la moneta in questione dia testa. L'assunzione di indipendenza subordinatamente al valore ϑ di p implica allora che il modello dei dati è quello della distribuzione bernoulliana

$$(13) \qquad f(x_1, \ldots, x_n|\vartheta) = \prod_i f(x_i|\vartheta) = \vartheta^{n-r}(1-\vartheta)^r,$$

dove le variabili casuali $X_1, X_2, \ldots$, di cui $x_1, x_2, \ldots$ sono i valori osservati, possono assumere due soli valori, 1, corrispondente all'evento «L'i-esimo lancio dà testa», e 0, corrispondente all'evento «L'i-esimo lancio dà croce», e $n-r=s$ è il numero degli 1 osservati, cosicché $s = \sum_i x_i$. Nella figura 6 è rappresentata tale funzione di verosimiglianza, per $s = 1$ e $r = 9$. Una caratteristica di questa funzione, condivisa dalla maggior parte di funzioni di verosimiglianza in uso, è che al crescere di n essa si «addensa» sempre piú in prossimità del valore di ϑ corrispondente a quello «osservato»; essa ha cioè un massimo sempre piú «appuntito» in prossimità di tale valore, per poi decrescere rapidamente sia a destra che a sinistra. Nel nostro caso, si tratta del valore $\vartheta = 1/10$, corrispondente alla frequenza osservata.

Un'altra importante caratteristica di questa funzione, pure condivisa dalla maggior parte delle funzioni in uso, è che essa dipende dal campione osservato soltanto per i numeri n e r. $f(\mathbf{x}|\vartheta)$ è quindi completamente determinata dalla funzione dei dati $t(\mathbf{x}) = \sum_{i=1}^{n} x_i$, già incontrata, dove $\mathbf{x} = (x_1, \ldots, x_n)$. $t(\mathbf{x})$ viene allora detto un *riassunto esaustivo* per $f(\mathbf{x}|\vartheta)$ o, piú semplicemente, per ϑ. Piú in generale, una funzione a valori reali dei dati $t(\mathbf{x})$ è un riassunto esaustivo per una data $f(\mathbf{x}|\vartheta)$ se e solo se $f(\mathbf{x}|\vartheta, t(\mathbf{x})) = f(\mathbf{x}|t(\mathbf{x}))$, se e solo se cioè $f(\mathbf{x}|\vartheta, t(\mathbf{x}))$ non dipende da ϑ. Nel nostro esempio, con $n=10$ e $r=9$, si ha da un lato $f(\mathbf{x}|\vartheta) = \vartheta(1-\vartheta)^9$, e dall'altro $f(t(\mathbf{x})|\vartheta) = \binom{10}{1}\vartheta(1-\vartheta)^9$, dato che vi sono $\binom{10}{1}$ sequenze binarie di lunghezza 10 caratterizzate da un solo «successo». Valendo infine in generale

$$f(\mathbf{x}|\vartheta, t(\mathbf{x})) = \frac{f(\mathbf{x}|\vartheta)}{f(t(\mathbf{x})|\vartheta)}$$

si ottiene immediatamente

$$f(\mathbf{x}|\vartheta, t(\mathbf{x})) = \frac{1}{\binom{10}{1}},$$

che non dipende da ϑ. La conoscenza di $t(\mathbf{x})$ ha reso superflua quella di ϑ ai fini della determinazione della densità per $\mathbf{x}$.

L'importanza delle funzioni di verosimiglianza che ammettono un riassunto esaustivo è determinata dal seguente teorema:

TEOREMA (principio di esaustività). *Se* $t(\mathbf{x})$ *è sufficiente per* $f(\mathbf{x}|\vartheta)$, *allora, data qualunque distribuzione iniziale,* $\pi(\vartheta|\mathbf{x}) = \pi(\vartheta|t(\mathbf{x}))$.

Per quanto riguarda la distribuzione iniziale, essa dovrà evidentemente dipendere dalla totalità delle Tue conoscenze, anteriormente all'esperimento, circa la natura del processo sotto osservazione. Dato tuttavia che tali conoscenze sono almeno in parte verbalizzabili, in questa misura esse possono essere corrette e criticate allo stesso modo di quelle da cui dipende la funzione di verosimiglianza. Mediante la discussione razionale, sarà perciò possibile nella maggior parte dei problemi mettere in comune le conoscenze di sfondo sopravvissute al vaglio della critica, e quindi assumere distribuzioni iniziali almeno approssimativamente uguali. La scelta di una distribuzione iniziale sarà dunque «soggettiva» solo nel senso di non essere determinata *univocamente* dalle conoscenze di sfondo circa la natura del processo considerato. Ma in questo senso non è meno «soggettiva» la scelta di una funzione di verosimiglianza o, piú in generale, di qualsiasi ipotesi scientifica. Anzi, Jaynes [1968; 1976] ha dimostrato che in taluni problemi sono sufficienti alcune assunzioni estremamente semplici e naturali per determinare univocamente la distribuzione iniziale, mediante l'applicazione del principio di massima entropia e di appropriati gruppi di trasformazione. Essere «soggettivisti» (o, come sarebbe meglio dire, per evitare le connotazioni usuali, e in particolare, la «carica emotiva» di questo termine, «bayesiani») invece che «oggettivisti» non significa allora sostenere che queste scelte sono arbitrarie, o che un'opinione vale un'altra, ma piuttosto che anche la scelta di un'opinione entro la classe di opinioni coerenti può essere discussa criticamente alla luce delle migliori teorie di cui si dispone circa la natura dei processi considerati. Se, dopo un confronto sufficientemente approfondito, non si riesce ad ottenere alcuna convergenza sulla distribuzione iniziale piú appropriata al problema in questione, questo inconveniente non sarà certo imputabile all'impostazione bayesiana, ma alle caratteristiche oggettive del problema. Si è visto tuttavia che – anche in casi del genere – tale impostazione ci dà una soluzione soddisfacente, eccetto in alcuni casi «patologici», pur di assumere la scambiabilità.

Nel caso del nostro problema, una scelta estremamente ragionevole di distribuzione iniziale sembra essere quella della cosiddetta *distribuzione* β:

$$\pi(\vartheta) = \frac{(a+b+1)!}{a!b!}\vartheta^a(1-\vartheta)^b \tag{14}$$

dove a e b sono due parametri maggiori di -1. Se ne discuterà la portata *intuitiva* dopo aver calcolato la distribuzione finale di ϑ. Da un punto di vista puramente *formale*, essa risulta particolarmente comoda dato che, come si vedrà, combinata alla funzione di verosimiglianza già specificata, determina, via il teorema di Bayes, una distribuzione finale dello stesso tipo di quella cui appartiene la funzione di verosimiglianza. Famiglie di distribuzioni iniziali che soddisfano questa condizione vengono dette *coniugate* rispetto alla famiglia data di funzioni di verosimiglianza.

Trattando il parametro incognito ϑ come una variabile aleatoria soggetta ad una distribuzione che ammette una densità di probabilità $\pi(\vartheta)$, il teorema di Bayes si generalizza nel modo seguente:

$$\pi(\vartheta|x) = \frac{\pi(\vartheta)f(x|\vartheta)}{\int_{-\infty}^{+\infty} f(x|\vartheta)\,\pi(\vartheta)\,d\vartheta};$$

in breve,

$$\pi(\vartheta|x) = K\,\pi(\vartheta)f(x|\vartheta),$$

dove K è una costante di proporzionalità il cui valore è determinato dalla condizione $\int_{-\infty}^{+\infty}\pi(\vartheta|x)\,d\vartheta = 1$. La distribuzione finale che si ottiene combinando in tal modo (13) e (14) è perciò la seguente:

$$\pi(\vartheta|x_1, \ldots, x_n) \propto \vartheta^{a+s}(1-\vartheta)^{b+r}. \tag{15}$$

Se l'integrale $\int_{-\infty}^{+\infty} |\vartheta|\, \pi(\vartheta|\mathbf{x})\, d\vartheta$ converge, allora $\int_{-\infty}^{\infty} \vartheta\pi(\vartheta|\mathbf{x})\, d\vartheta$ viene detto il valor medio o la previsione (condizionale) della distribuzione π e denotato da $E(\vartheta|x_1, ..., x_n)$. Piú in generale, $E(\vartheta^n)$ viene detto l'n-esimo momento della distribuzione, e $E((\vartheta - E(\vartheta))^n)$ l'n-esimo momento centrale, e in particolare per $n=2$, varianza della distribuzione. I simboli usuali per denotare il valor medio e la varianza di una distribuzione sono μ e σ^2. Infine, la radice quadrata della varianza, e cioè σ, viene detta scarto standard, e l'inverso della varianza precisione.

Non è difficile vedere che il valor medio della distribuzione finale (15) riportata alla pagina precedente è

$$E(\vartheta|x_1, ..., x_n) = \frac{a+s+1}{a+b+s+r+2}.$$

In vista del teorema di rappresentazione, esso coincide dunque con la probabilità che il prossimo lancio dia testa nel caso in cui la distribuzione iniziale $F(\vartheta)$ sia proporzionale a $\int_0^\vartheta Y^a(1-\vartheta)^b\, dY$. La distribuzione di Bayes-Laplace è dunque un caso speciale della distribuzione β per $a=b=0$. (15) consente di interpretare i parametri a e b della distribuzione β rispettivamente come il numero di «successi» e di «insuccessi» in un campione «immaginario» di $a+b$ osservazioni. In tal caso assegnare ad a e b particolari valori significherà considerare la propria conoscenza di sfondo altrettanto informativa di quella che si otterrebbe osservando a «successi» e b «insuccessi» su $a+b$ «osservazioni». Perciò, tanto piú è informativa la nostra conoscenza di sfondo, tanto maggiori a e b dovranno essere; in particolare, se essa è «simmetrica» rispetto ai due esiti possibili, dovrà essere $a=b$. Il caso Bayes-Laplace corrisponde dunque a una conoscenza di sfondo informativa al minimo, dato che $a=b=0$. Un caso interessante con a e b minori di 1 è il caso limite $a=b=-1$. Tale scelta determina la densità iniziale divergente proposta da Haldane [1945], $\vartheta^{-1}(1-\vartheta)^{-1}$. Il suo interesse consiste essenzialmente nel fatto che uno dei metodi «oggettivisti» di stima puntuale piú noti, il cosiddetto metodo di massima verosimiglianza di Fisher, dà come stima puntuale di ϑ precisamente il valor medio della distribuzione finale (15), e cioè s/n, assumendo come distribuzione iniziale la densità impropria di Haldane. Un modo equivalente per ottenere questo risultato consiste nel scegliere come stima puntuale di ϑ la *moda* della distribuzione finale (15), quel valore di ϑ cioè che rende massima la (15), assumendo come distribuzione iniziale quella uniforme.

È chiaro che la stima di massima verosimiglianza di ϑ sarà perfettamente ragionevole se il campione è grande. Le cose cambiano però nel caso di campioni piccoli. Si supponga ad esempio $s=0$. ϑ viene allora stimato uguale a 0. Ma un valore del genere determinerebbe decisioni ovviamente inaccettabili, ad esempio la decisione di dare a n contro 0, per qualunque n, l'uscita di croce al prossimo lancio, e cioè di pagare n lire, qualunque sia n, se esce testa al prossimo lancio, e di ricevere 0 lire se esce croce. Questo significa che la densità iniziale di Haldane va scartata e che la stima di massima verosimiglianza va

considerata una semplice approssimazione, per valori di $s+r$ sufficientemente grandi, alla «stima» bayesiana.

Un'alternativa interessante alla distribuzione di Bayes-Laplace è stata proposta da Jeffreys [1946]. Una caratteristica indesiderabile della distribuzione uniforme è che essa non è invariante rispetto a trasformazioni del parametro incognito. L'alternativa in questione doveva allora soddisfare la condizione che la probabilità di una data regione dello spazio dei parametri non variasse rispetto a cambiamenti di coordinate. Il risultato è una densità iniziale proporzionale a $\vartheta^{-(1/2)}(1-\vartheta)^{-(1/2)}$, e dunque una stima di ϑ uguale a $(s+(1/2))/(n+1)$.

Un metodo generale per determinare a e b, nel caso in cui la conoscenza di sfondo sia simmetrica, è stato proposto da Good [1965] ed è noto come «metodo dei risultati ipotetici». La simmetria implica naturalmente $a=b$. Dunque, la stima di ϑ è $(s+a+1)/(n+2(a+1))$. S'immagini allora di aver fatto n lanci, di cui $n-1$ hanno dato testa e 1 croce. Quanto grande deve essere n perché Tu sia disposto a dare al massimo a 2 contro 1 l'uscita di testa al prossimo lancio? In circostanze normali, la risposta di Bayes-Laplace, e cioè $n=4$, non sembra affatto ragionevole. Una risposta piú ragionevole come $n=40$ darebbe invece il seguente risultato. Si sa già che la stima di ϑ è $(s+a+1)/(n+2(a+1))$. Si sta perciò cercando il valore minimo di n tale che

$$\frac{n-1+a+1}{n+2(a+1)} = \frac{2}{3}.$$

Se la Tua risposta è 40,

$$\frac{40-1+a+1}{40+2(a+1)} = \frac{2}{3};$$

e dunque $a=36$. Può però ben darsi che Tu non sia in grado di specificare un unico valore, ma soltanto un intervallo di valori. In questo caso, il metodo dei risultati ipotetici darà semplicemente un intervallo in cui a è compreso. Come ha suggerito Good, si potrebbe allora scegliere una distribuzione iniziale per lo stesso a interamente concentrata su tale intervallo.

5.2. La stima zonale.

La stima zonale differisce da quella puntuale nella misura in cui il suo risultato non è un singolo valore del parametro incognito (il valor medio della distribuzione finale, nell'esempio del § 5.1), bensí un intervallo di valori cui esso appartiene con una certa probabilità. Tale intervallo viene naturalmente determinato in base alla distribuzione finale, che anche in questo caso ha un ruolo decisivo.

Piú precisamente, se $\pi(\vartheta|x_1, ..., x_n)$ è la distribuzione finale di ϑ, e $I_\beta(x_1, ..., x_n)$ è un qualunque intervallo di valori di ϑ, dipendente da $(x_1, ..., x_n)$ e β, dove $0 \leq \beta \leq 1$, tale che

$$\int_{I_\beta(\mathbf{x})} \pi(\vartheta|\mathbf{x})\, d\vartheta = \beta$$

si dice allora che $I_\beta(\mathbf{x})$ è un intervallo di confidenza per ϑ al 100β per cento (dato $\mathbf{x}$). $I_\beta(\mathbf{x})$ viene detto stima zonale di ϑ, e β livello di confidenza. Queste nozioni andrebbero qualificate dall'aggettivo «bayesiano» per distinguerle dalle corrispondenti nozioni «oggettiviste».

Naturalmente, gli intervalli di maggior interesse pratico sono di due tipi: I) quelli cui ϑ appartiene «molto probabilmente», dove questa espressione ha una interpretazione convenzionale in termini di $\beta=0{,}95$; e II) quelli per cui è equiprobabile che ϑ vi appartenga o meno; perciò in questo caso $\beta=0{,}50$.

Tuttavia, intervalli di questi tipi *non* sono unici: vi sono usualmente molti intervalli corrispondenti a un β prefissato. Si ottiene tuttavia un unico intervallo, imponendo la condizione addizionale che la densità di ogni punto entro l'intervallo sia maggiore di quella di qualunque punto esterno. Si può mostrare che l'intervallo cosí determinato è il piú «informativo», nel senso ovvio che è il piú breve tra tutti quelli con il livello di confidenza fissato. Esso non è tuttavia invariante rispetto a trasformazioni del parametro incognito.

Si supponga di essere interessati a una stima zonale della conduttività di un nuovo materiale isolante. Le prove, in questo caso, sono misurazioni della sua conduttività mediante uno strumento di misura, di cui è nota la precisione. L'assunzione usuale è anche qui che esse siano rappresentate da una sequenza di variabili aleatorie, $X_1, X_2, \ldots$, indipendenti, e soggette alla distribuzione normale o gaussiana, con valor medio μ incognito, che rappresenta il valore vero della grandezza misurata, e varianza σ^2 nota, il cui inverso $1/\sigma^2$ rappresenta appunto la precisione dello strumento di misura. Si denota una tale distribuzione con $N(\vartheta;\sigma^2)$ dove ϑ sta come al solito per il parametro incognito, e la si intende come funzione di ϑ.

Queste assunzioni determinano la seguente funzione di verosimiglianza:

$$f(x|\vartheta)=(2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(x-\vartheta)^2}{2\sigma^2}\right),$$

da cui segue immediatamente:

$$f(x_1,\ldots,x_n|\vartheta)=(2\pi\sigma^2)^{-n/2}\exp\left(-\sum_{i=1}^{n}\frac{(x_i-\vartheta)^2}{2\sigma^2}\right)$$

$$\propto\exp\left(-\frac{1}{2}\vartheta^2\frac{n}{\sigma^2}+\vartheta\bar{x}\frac{n}{\sigma^2}\right)$$

dove $\bar{x}=n^{-1}\sum_{i=1}^{n}x_i$;

$$\propto\exp\left(-\frac{1}{2}(\bar{x}-\vartheta)^2\frac{n}{\sigma^2}\right).$$

Si vede allora che fare n misurazioni «normali» equivale a farne una sola $\sqrt{n}$ volte piú precisa. Si vede inoltre che in questo caso $\bar{x}$ e n sono riassunti esaustivi.

Per dare un'immagine piú concreta, si consideri il seguente esempio numerico. Si può supporre – via una scelta appropriata delle unità di misura – che lo strumento di misura usato abbia uno scarto standard unitario. Fatte $n=10$ misurazioni, si ottiene la seguente sequenza di letture: 12,0; 15,2; 10,7; 12,4; 13,6; 13,6; 13,3; 13,9; 11,7; 11,9. La media del campione $\bar{x}$ è allora 12,83. La verosimiglianza di queste dieci misurazioni per ϑ è dunque proporzionale a

$$\exp(-5(12{,}83-\vartheta)^2). \tag{16}$$

Il suo grafico è in figura 7. Essa ha un massimo molto appuntito per $\vartheta=\bar{x}$, e poi decresce sia a destra che a sinistra molto rapidamente: $f(\mathbf{x}|\vartheta)$ è cioè molto «attiva» in un piccolo intervallo di valori di ϑ, e praticamente nulla altrove.

Resta – per la determinazione della distribuzione finale – da specificare la distribuzione iniziale $\pi(\vartheta)$. Una possibilità, suggerita soprattutto da ragioni di comodità matematica, consiste nell'assumere che anche la distribuzione iniziale è normale con parametri μ_0 e σ_0^2, è cioè $N(\mu_0;\sigma_0^2)$. Si dimostra in tal caso che la densità finale di ϑ è $N(\mu_n;\sigma_n^2)$, dove

$$\mu_n=\frac{\left(\frac{n\bar{x}}{\sigma^2}\right)+\left(\frac{\mu_0}{\sigma_0^2}\right)}{\left(\frac{n}{\sigma^2}\right)+\left(\frac{1}{\sigma_0^2}\right)} \tag{17}$$

$$\sigma_n=n\sigma^{-2}+\sigma_0^{-2}. \tag{18}$$

Dunque, la precisione finale è uguale a n volte la precisione dei dati piú la precisione iniziale. D'altra parte, il valor medio finale è la media ponderata del valor medio dei dati e del valor medio iniziale, dove i pesi sono le loro rispettive precisioni.

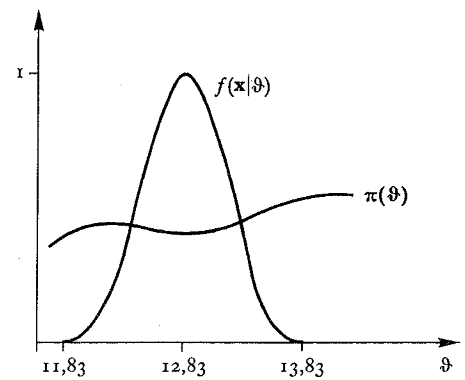


Figura 7.

Densità iniziale diffusa $\pi(\vartheta)$ e funzione di verosimiglianza $f(\mathbf{x}|\vartheta)$ nell'esempio della conduttività. Si noti che le due funzioni non sono tracciate sulla stessa scala verticale.

La scelta di $N(\mu_0; \sigma_0^2)$ come densità iniziale implica però opinioni iniziali molto ben definite circa la conduttività del materiale isolante sotto esame, e in particolare che l'intervallo $(\mu_0-2\sigma_0, \mu_0+2\sigma_0)$ è un intervallo di confidenza per ϑ al 95 per cento (prendendo per $\mathbf{x}$ la sequenza vuota). Naturalmente, tale opinione iniziale sarà tanto piú definita quanto maggiore sarà $1/\sigma_0^2$. Valori crescenti di σ_0 rappresenteranno perciò opinioni sempre meno definite circa ϑ. Al limite, come $\sigma_0 \rightarrow \infty$, la distribuzione finale $N(\mu_n; \sigma_0^2) \rightarrow N(\bar{x}; \sigma^2/n)$, che non dipende piú dalla distribuzione iniziale. Tale caso limite rappresenta una conoscenza di sfondo minimalmente informativa. Si vede in particolare che (17) e (18) si riducono ad affermare rispettivamente che la precisione finale è uguale a $\sqrt{n}$ volte quella dei dati, e il valor medio finale è uguale al valor medio del campione. Lindley ha dimostrato che questo risultato vale approssimativamente sotto assunzioni molto meno restrittive di $\sigma_0 \rightarrow \infty$; piú specificamente, esso vale per ogni distribuzione iniziale I) con densità quasi costante nell'intervallo in cui la funzione di verosimiglianza è «attiva», e cioè in un opportuno intorno di $\vartheta = \bar{x}$, e II) tale che i valori di ϑ fuori da questo intervallo non siano molto piú probabili. Deve trattarsi – nella terminologia di Savage [1959, pp. 42-43; 1964, pp. 20-23] – di una distribuzione iniziale *diffusa*. Come mostra la figura 7, nel caso dell'esempio della conduttività, $N(12{,}83; 1/10)$, intesa come funzione di ϑ, è quasi zero fuori dall'intervallo $12{,}83 \pm 1$; la funzione $\pi(\vartheta)$ varia molto poco entro questo intervallo, e non è mai molto piú grande fuori che dentro l'intervallo. In queste circostanze, il prodotto cui è proporzionale la distribuzione finale di ϑ, e cioè $N(\vartheta; \sigma^2)\,\pi(\vartheta)$, è ben approssimato per molti scopi da $N(12{,}83; 1/10)\,\pi(12{,}83)$; e dunque da $N(12{,}83; 1/10)$, assorbendo anche $\pi(12{,}83)$ nella costante di proporzionalità.

Da questi risultati segue infine immediatamente che ogni distribuzione iniziale diffusa può essere approssimata da una distribuzione iniziale uniforme, che ammetta una densità costante del parametro incognito sull'intervallo I dei suoi valori possibili. Nel caso in cui tale intervallo I sia illimitato, non esiste tuttavia alcun $\pi(\vartheta) = k$, tale che

$$(19) \qquad \int_I k \, d\vartheta = 1;$$

non esiste cioè alcuna funzione costante di ϑ normalizzabile; densità che violano (19) vengono usualmente dette *improprie*. Di esse si è fatto largo uso entro l'impostazione bayesiana. Ma non c'è alcuna ragione di scandalo. Il loro uso non sembra infatti *essenziale* per ottenere alcun risultato: se questo è vero, esse sono solo comode scorciatoie per ottenere risultati che potrebbero essere ottenuti ugualmente o mediante densità diffuse oppure mediante assunzioni piú realistiche circa il campo di valori del parametro incognito. Il primo punto è stato già implicitamente chiarito. Quanto al secondo, «applicare la distribuzione uniforme su un intervallo illimitato non rappresenta alcuno stato realistico d'informazione iniziale. Ad esempio se ϑ è la lunghezza di qualche oggetto materiale, sappiamo per certo che lo scarto standard non può essere minore della dimensione di un atomo, circa 10^{-8}, o maggiore della dimensione della Terra 10^9... Dunque, fuori da questo intervallo, la densità deve essere nulla» [Jaynes 1976, p. 249]. Questo è solo un esempio; ma sembra plausibile supporre che in ogni specifico problema la conoscenza di sfondo determinerà limiti inferiori e superiori finiti al parametro incognito considerato [Lindley 1973]. Usando l'uno o l'altro di questi metodi si dovrebbe perciò riuscire a reinterpretare qualunque uso di densità improprie come una semplice «finzione» matematica volta a semplificare il problema considerato. Tipici problemi usualmente risolti entro l'impostazione bayesiana mediante l'uso di densità improprie sono quelli della «stima» dello scarto standard di una distribuzione normale di cui è noto il valor medio, e quello della «stima» simultanea di valor medio e scarto standard di una distribuzione normale [cfr. Lindley 1965, pp. 26-46].

Questo paragrafo si concluderà specificando intervalli di confidenza nel caso dell'esempio della conduttività. Assumendo una densità diffusa, si ha in generale che l'intervallo $\bar{x} \pm 1{,}96(\sigma/\sqrt{n})$ è un intervallo di confidenza al livello 95 per cento; dunque, in particolare, la probabilità finale che la conduttività appartenga all'intervallo compreso tra 12,21 e 13,44 è 95/100. Se si aumenta il livello a 99/100, l'intervallo di confidenza naturalmente aumenta, e si ottiene $12{,}83 \pm 0{,}815$. Viceversa, se lo si diminuisce a 50/100, si ottiene un intervallo di confidenza compreso tra 12,618 e 13,041.

5.3. Test di significatività.

Entro l'impostazione bayesiana, la differenza tra «stima» e «test di significatività» è semplicemente una differenza tra due possibili punti di vista del ricercatore. I problemi di test sono quelli in cui il ricercatore è particolarmente interessato a uno specifico valore di un parametro incognito ed intende ottenere dati pertinenti alla valutazione di questa specifica ipotesi. Nel caso della stima invece – lo si è visto – non c'è alcun valore prefissato che svolga un ruolo privilegiato. Riprendendo l'esempio della conduttività, esso diventerebbe un problema di test, se vi fosse un qualche valore di ϑ, ad esempio ϑ_0, tale che si volesse sapere se la conduttività del nuovo materiale isolante è uguale (oppure maggiore oppure ancora minore) a ϑ_0. Una situazione tipica in cui si presenta un problema di test è quella in cui ϑ_0 è determinato da standard legali o economici. Ad esempio la legge potrebbe vietare di immettere sul mercato un materiale isolante a meno che la sua conduttività non sia minore di un valore ϑ_0 ben specificato. Oppure ancora, si sta sperimentando una nuova varietà di grano che s'intende sostituire alla varietà precedentemente coltivata se dà una media di raccolti maggiore di quest'ultima. Oppure ancora, si sta sperimentando un nuovo metodo di sigillare lampadine elettriche da sostituire al precedente se determina una vita media piú lunga.

In casi di questo tipo, si costruisce usualmente un'ipotesi, detta ipotesi *nulla*, che prevede un esito negativo alla sperimentazione, ad esempio che la conduttività del nuovo materiale sia maggiore o uguale a ϑ_0, che il raccolto medio della nuova varietà sia minore o uguale a quello della precedente, che la vita media delle lampadine sigillate con il nuovo metodo sia minore o uguale a quella delle precedenti. L'ipotesi nulla viene detta esatta (*sharp*) o semplice nel caso in cui determini completamente la funzione di verosimiglianza (specifichi cioè un particolare valore di ϑ), composita altrimenti. Le osservazioni vengono quindi svolte per valutare l'ipotesi nulla in rapporto a qualche altra ipotesi con essa incompatibile, detta ipotesi alternativa.

È precisamente sul modo d'interpretare questa nozione di valutazione che l'impostazione bayesiana si separa nettamente da quella «oggettivistica». Entro quest'ultima, si tratta essenzialmente di specificare una regola o una procedura per decidere se *respingere* l'ipotesi nulla, sulla base dei dati osservativi. Usando una terminologia bayesiana, e restringendosi alle ipotesi semplici, tale procedura ridotta all'osso consiste nella scelta di un livello di significatività β, e nella decisione di respingere l'ipotesi se il valore del parametro da essa specificato, ϑ_0, non appartiene all'intervallo $I_\beta(\mathbf{x})$ tale che:

$$\int_{I_\beta(\mathbf{x})} \pi(\vartheta|\mathbf{x})\, d\vartheta = \beta.$$

In queste circostanze, si dice anche che i dati sono significativi al livello $1-\beta=\alpha$.

Da un punto di vista bayesiano, non si tratta invece di respingere (o accettare) *ipotesi*, bensí di respingere (o accettare) *decisioni*, per cui l'unico fattore pertinente è l'utilità prevista, almeno se l'ordinamento di preferenza è coerente. Come è noto, si dovrà respingere una decisione se ne esiste un'altra con una utilità prevista maggiore. Nel caso qui in discussione con decisioni d con un insieme infinito di conseguenze, una per ciascun particolare valore di un dato parametro ϑ, la nozione di utilità prevista dalla decisione d, $\bar{U}(d)$, si generalizza in modo naturale ponendo

$$\bar{U}(d) = \int U(d, \vartheta)\, \pi(\vartheta)\, d\vartheta.$$

Ovviamente, nel caso in cui si venga a conoscenza di nuovi dati $\mathbf{x}$, le utilità pertinenti alla decisione saranno le utilità *finali*, $\bar{U}(d|\mathbf{x})$, dove

$$\bar{U}(d|\mathbf{x}) = \int U(d, \vartheta)\, \pi(\vartheta|\mathbf{x})\, d\vartheta.$$

Tuttavia, già *prima* di osservare i particolari dati $\mathbf{x}$, si è in grado di valutare con precisione la convenienza di eseguire l'esperimento di cui $\mathbf{x}$ è uno dei risultati possibili. Basterà infatti confrontare l'utilità della decisione, ad esempio d^* che massimizza $\bar{U}(d)$, e cioè $\bar{U}(d^*)$, con l'utilità della decisione, ad esempio d^{**}, che massimizza il *valor medio* (o previsione) di $U(d|\mathbf{x})$ rispetto a $\pi(\mathbf{x})$, e cioè alla densità dei vari possibili risultati dell'esperimento prima che esso sia eseguito, e cioè con

$$(20) \qquad \int \bar{U}(d^{**}|\mathbf{x})\, \pi(\mathbf{x})\, d\mathbf{x}.$$

Assumendo che il costo di eseguire l'esperimento in questione sia nullo, è ovvio che sarà razionale eseguirlo se e solo se $\int \bar{U}(d^{**}|\mathbf{x})\, \pi(\mathbf{x})\, d\mathbf{x} \geq \bar{U}(d^*)$. Non

è però difficile dimostrare che $\bar{U}(d^*)$ è necessariamente minore o uguale a (20); dunque, se il costo di un dato esperimento è nullo, sarà senz'altro razionale eseguirlo. Abbandonando quest'assunzione non realistica, la razionalità dipenderà evidentemente dal confronto tra la differenza tra $U(d^*)$ e (20), e il costo dell'esperimento. Queste conseguenze dell'impostazione bayesiana sono state esplorate in varie direzioni, soprattutto con riferimento alle decisioni economiche [cfr. Raiffa e Schlaifer 1961]. Su questo punto si veda anche l'articolo «Decisione» in questa stessa *Enciclopedia*.

Non solo, tuttavia, l'impostazione oggettivistica dei test di significatività è concettualmente confusa, ma inoltre buona parte delle regole formulate entro di essa viola uno dei principî piú naturali dell'induzione statistica, il cosiddetto principio di verosimiglianza:

Se due insiemi di dati x e y hanno le seguenti proprietà:

1) *la loro distribuzione dipende dallo stesso insieme di parametri,*
2) *le verosimiglianze per questi parametri di x e y sono uguali,*
3) *le densità iniziali dei parametri sono uguali in entrambi i casi,*

allora, x e y sono indistinguibili rispetto ad ogni inferenza circa i parametri.

Esso implica in particolare che se valgono le 1)-3), è irrilevante rispetto ad ogni inferenza circa i parametri il modo in cui i dati sono stati ottenuti. Ad esempio, nel caso di un campione binomiale bernoulliano la probabilità di osservare 10 «successi» su 100 prove, è la stessa sia che i dati siano stati ottenuti prefissando il numero di prove, sia che siano stati ottenuti continuando a sperimentare fino ad ottenere i 10 «successi». È questa conseguenza del principio che è contraddetta da molte procedure «oggettivistiche» basate su «distribuzioni di campionamento». Il principio risulta invece banalmente vero entro l'impostazione bayesiana, dato che sotto le condizioni 1)-3) il teorema di Bayes implica l'identità delle distribuzioni finali, e dunque l'indipendenza di ogni inferenza circa i parametri dal «disegno» dell'esperimento.

Data la complessità delle applicazioni piú interessanti (ad esempio al problema di Behrens-Fisher [cfr. Jaynes 1976, pp. 181-83; oppure Lindley 1965, II, pp. 76-95]), ci si limita a rielaborare a titolo illustrativo dei test di significatività l'esempio della conduttività. Assumendo una distribuzione uniforme ($\sigma_0 \to \infty$), la distribuzione finale è proporzionale alla funzione di verosimiglianza, e cioè con $\bar{x}=12{,}83$ e $\sigma=1$, a (16). Supponendo che standard di sicurezza impongano che $\vartheta < 13{,}3$, ϑ_0 è perciò in questo caso 13,3. Si è già visto che l'intervallo di confidenza al livello 95 per cento per ϑ è compreso tra 12,21 e 13,44. I dati non sono perciò significativi al livello 5 per cento, dato che ϑ_0 appartiene a questo intervallo. Non è però difficile vedere che lo diventano al livello 20 per cento, dato che l'intervallo di confidenza al livello 80 per cento è compreso tra 12,43 e 13,23. Ma nessuno di questi risultati è molto informativo, dipendendo entrambi da una scelta del tutto arbitraria del livello di significatività. Già piú informativa è la specificazione del livello *critico* di significatività, e cioè di quel valore di α al di sotto del quale i dati non sono significativi e al di sopra del quale essi lo diventano; nell'esempio si tratta di trovare un valore di α tale che

$$(21) \qquad 12{,}83+\lambda_\alpha\left(\frac{1}{\sqrt{10}}\right)=13{,}3,$$

dove

$$(22) \qquad \frac{1}{(\sqrt{2\pi})}\int_{-\infty}^{-\lambda_\alpha}\exp\left(-\frac{1}{2}y^2\right)dy=\frac{\alpha}{2}.$$

Ora (21) implica $\lambda_\alpha=1{,}486$; la sua sostituzione a λ_α in (22), che dà l'area sotto la funzione di densità normale standardizzata, implica $\alpha=0{,}138$. Dunque, il livello critico per la significatività dei dati è il 13,8 per cento. Pur restando ancora misteriose le ragioni per la scelta di un livello, in tal modo si sa almeno a quale livello i dati sarebbero *appena* significativi.

In un senso però questo risultato dice di piú di quanto non si volesse sapere all'inizio. Il problema era infatti quello di sapere se la conduttività non superava 13,3; non interessava invece conoscere un limite inferiore. Dunque, l'intervallo pertinente è compreso tra $-\infty$ e 13,3, e interessa conoscere il livello critico di significatività per *questo* intervallo. Ma tale livello critico coincide precisamente con la probabilità finale che ϑ superi 13,3. Dal momento che la distribuzione finale di ϑ è $N(12{,}83;\ 1/10)$, esso risulta uguale a

$$(23) \qquad \frac{1}{(\sqrt{2\pi})}\int_{-\infty}^{-k}\exp\left(-\frac{1}{2}y^2\right)dy,$$

dove $k=(13{,}3-12{,}83)\sqrt{10}=1{,}486$. Dunque (23) risulta uguale a 0,087, e cioè alla metà del livello critico precedentemente calcolato.

Se ne conclude che il modo migliore per impostare i test di significatività consiste nello specificare semplicemente la probabilità finale delle ipotesi considerate, e di lasciare poi alla regola di massimizzazione dell'utilità prevista l'utilizzazione di questi valori per scegliere la decisione ottimale.

Tuttavia, il tipo di test qui considerato è appropriato soltanto nel caso in cui la conoscenza di sfondo sia minimalmente informativa. Se si ha invece – prima di sperimentare – qualche ragione per supporre che l'ipotesi nulla esatta abbia una probabilità finita, si dovrà naturalmente usare una distribuzione mista. Questo tipo di situazione è stato trattato da Jeffreys [1961, pp. 245 sgg.; un esempio si trova anche in Cox e Hinkley 1974, pp. 394-95].

6. *Test di «ipotesi estreme nulle».*

I test considerati nel § 5.3 vengono detti *parametrici*, dato che in essi si assume *noto* il «modello dei dati», a meno di uno o piú parametri. Essi vengono usualmente contrapposti ai cosiddetti test non-parametrici, in cui il problema consiste essenzialmente nel determinare qual è il «modello dei dati» migliore.

[Per una trattazione bayesiana dei test non-parametrici, si vedano Ferguson 1973 e Antoniak 1974]. In questo paragrafo, ci si occuperà di un tipo di test diverso da entrambi, nella misura in cui il «modello dei dati» non ha carattere «statistico». Casi di questo tipo s'incontrano in medicina, giurisprudenza, e nella ricerca scientifica pura. Entro l'impostazione bayesiana, essi non differiscono tuttavia qualitativamente dai casi finora affrontati: prove di collaudo, controlli di qualità, diagnosi mediche, sentenze, e scelta di teorie scientifiche non sono processi qualitativamente diversi per il bayesiano. In ogni caso, una decisione dovrà essere presa condizionalizzando la distribuzione iniziale rispetto ai dati disponibili, ed utilizzando la risultante distribuzione finale per calcolare l'utilità prevista finale delle varie decisioni possibili.

Il caso su cui qui ci si soffermerà è quello della scelta fra teorie scientifiche, restringendolo ulteriormente al sottocaso in cui una delle teorie in competizione, combinata con la conoscenza di sfondo, implica logicamente i dati. Siano allora i dati $E_1, E_2, \ldots$, ed E una teoria tale che $E \cap B$, dove B è la conoscenza di sfondo, implica logicamente E_i, per ogni $i=1, 2, \ldots$ Si denoti con $E(n)$ l'intersezione di $E_1, \ldots, E_n$. La teoria alternativa è qui $-E$.

Il teorema di Bayes implica allora:

$$(24) \qquad \mathrm{Prob}(E|E_1\cap\ldots\cap E_n)=\frac{\mathrm{Prob}(E)\,\mathrm{Prob}(E_1\cap\ldots\cap E_n|E)}{\mathrm{Prob}(E_1\cap\ldots\cap E_n)}=\frac{\mathrm{Prob}(E)}{\mathrm{Prob}(E_1\cap\ldots\cap E_n)},$$

da cui è stato sistematicamente omesso B, dato che è un fattore costante. Ovviamente, se la probabilità iniziale di E è uguale a 0, nessuna evidenza potrà cambiarla, per cui sarà uguale a 0 anche la sua probabilità finale. Si supponga allora che $\mathrm{Prob}(E)\neq 0$. La formula (1) implica

$$(25) \qquad \mathrm{Prob}(E_1\cap\ldots\cap E_n)=\mathrm{Prob}(E_1)\,\mathrm{Prob}(E_2|E_1)\,\mathrm{Prob}(E_3|E(2))\ldots\ldots\mathrm{Prob}(E_n|E(n-1)).$$

Poste $\mathrm{Prob}(E|E(n))=p(n)$, $\mathrm{Prob}(E_n|E(n-1))=p_n$, si ha $p(n)/p(n-1)=1/p_n$, e

$$(26) \qquad p(n)=\frac{1}{p_1p_2\ldots p_n}p(0),$$

dove $p(0)=\mathrm{Prob}(E)$. Ne segue che $p(n)>p(n-1)$ se $p_n<1$. Dunque, se i dati non sono massimalmente probabili gli uni rispetto agli altri (e a B), ciascuno di essi accresce la probabilità finale di E, purché naturalmente la probabilità iniziale di E sia diversa da 0. È questa una versione precisa del famoso principio della concordanza (*consilience*) delle induzioni. Cosí è formulato in Leibniz: «Allorché l'ipotesi spiega facilmente fenomeni altrimenti di difficile interpretazione, e senza connessioni gli uni con gli altri, ... è molto probabile» [1703-1704, trad. it. II, p. 215]. E cosí in Kant: «Nondimeno, la probabilità di un'ipotesi può crescere e essere elevata al rango di un *analogo* della certezza, quando tutte le conseguenze *che abbiamo fino a allora verificato* si lasciano spiegare a partire dal principio supposto» [1800, ed. 1923 p. 85]. Nella nostra versione, tale principio è del tutto compatibile sia con la posizione di chi restringe la sua applicabilità al caso in cui gli E_i siano fatti *nuovi anticipati* da E, e solo successivamente verificati (come forse Leibniz, e sicuramente Whewell e i «falsificazionisti» [cfr. per tutti Lakatos 1970]), sia con la posizione di chi estende la sua portata a fatti già noti prima della formulazione di E (come sembra fare Kant, e come certamente gli «induttivisti» da Mill fino alla «scuola di Cambridge», Johnson, Broad, Keynes, Jeffreys) e oltre (Carnap, Hintikka).

Naturalmente, quanto piú improbabile è ciascun E_i alla luce della sola conoscenza di sfondo, tanto maggiore sarà il suo contributo alla «conferma» di E. Non segue tuttavia da (26) che, al crescere di n, la probabilità finale di E tende a 1 («l'analogo della certezza» di Kant).

Per specificare sotto quali condizioni questo vale, basta notare che

$$p(n) = \frac{p(0)}{p(0)\,\mathrm{Prob}(E(n)|E) + (1-p(0))\,\mathrm{Prob}(E(n)|-E)} = \frac{1}{1+\dfrac{(1-p(0))}{p(0)}\mathrm{Prob}(E(n)|-E)}.$$

Non è difficile vedere che questa espressione tende a 1 se e solo se esiste un $\varepsilon>0$, tale che per tutti gli n $\mathrm{Prob}(E_n|E(n-1)\cap -E)\leq 1-\varepsilon$. In tal caso infatti $\mathrm{Prob}(E(n)|-E)\leq(1-\varepsilon)^n$, e dunque, $\mathrm{Prob}(E(n)|-E)\to 0$, per $n\to\infty$.

Un'altra condizione equivalente a quella appena vista, ma piú semplice, può essere ottenuta a partire da (24)-(26). Si avrà infatti che $p(n)\to 1$, se e solo se $\mathrm{Prob}(E(n))\to p(0)$, per $n\to\infty$, e cioè se e solo se vale il seguente assioma del limite [Jeffrey 1965, p. 178]:

Per ogni $\varepsilon>0$, esiste un intero n tale che ogni membro della sequenza $p_1, p_2, \ldots$ successivo all'n-esimo è maggiore di $p(0)$ per meno di ε.

L'equivalenza tra l'assioma del limite e la condizione $\mathrm{Prob}(E(n))\to p(0)$, per $n\to\infty$, risulta immediata in quanto $\mathrm{Prob}(E(n))=p_1p_2\ldots p_n$, e $p_1p_2\ldots p_n\leq p(0)$.

Si è assunto che $p(0)\neq 0$. Dunque, perché valga l'assioma del limite, il prodotto infinito $p_1p_2p_3\ldots$ non deve valere 0. La teoria delle serie infinite dà una condizione necessaria e sufficiente perché questo non sia il caso:

Un prodotto della forma $\prod p_n$ è convergente e ha un valore diverso da 0 se e solo se:
1) nessuno dei suoi fattori $p_i=0$; 2) ciascuno dei suoi fattori $p_i<1$; e 3) $\Sigma(1-p_n)$ è convergente.

[Segue dai teoremi 1 e 5 di Knopp 1956, pp. 93 e 96]. Un esempio di prodotto infinito che ha un valore finito diverso da 0 è quello il cui n-esimo fattore è

$$p_n = \frac{n!-\frac{1}{2}}{n!}.$$

Si consideri infatti $\sum 1/2n!$ Applicando il test del quoziente di Cauchy, $\sum 1/2n!$ converge se $\lim(p_n/p_{n-1})$ esiste ed è minore di 1. Ma $p_n/p_{n-1}=1/n$, il cui limite esiste e vale 0.

Si è dunque mostrato che se una teoria spiega (nel senso che implica logicamente l'occorrenza di) una varietà di fenomeni ($E_1, E_2, \ldots$), l'osservazione di ciascun nuovo fenomeno accresce la probabilità finale della teoria, purché 1) la sua probabilità iniziale sia $\neq 0$, e 2) il nuovo fenomeno non possa essere previsto con certezza a partire dai precedenti combinati con la conoscenza di sfondo. Inoltre, esso l'accresce tanto piú quanto piú esso è *im*probabile alla luce della conoscenza di sfondo. Infine, tale probabilità finale tende a 1, al crescere di n, purché la probabilità iniziale sia $\neq 0$, e la probabilità di ciascun fenomeno relativamente agli altri e alla teoria alternativa non superi mai un valore fissato minore di 1.

In analogia con la precedente definizione, dati $E_1, E_2, \ldots$, una teoria E che li implica logicamente, e una funzione di probabilità Prob, si dice che Prob consente apprendimento *forte* dall'esperienza, relativamente a $E_1, E_2, \ldots$, ed E, se e solo se $\lim_{n\to\infty}\mathrm{Prob}(E|E(n))=1$. Dato che questo vale se e solo se $\lim_{n\to\infty}\mathrm{Prob}(E(n))=\mathrm{Prob}(E)\neq 0$, si avrà apprendimento forte quando il tasso di apprendimento nel senso del § 4 è sufficientemente veloce. In vista del teorema di rappresentazione, la circostanza se una data Prob consente o meno apprendimento forte dall'esperienza sarà dunque interamente determinata dalla forma della funzione di distribuzione F. Il teorema di De Finetti, come ha notato Hintikka [1971, pp. 338-39], chiarisce dunque in modo definitivo il significato di scommesse su teorie (universali): scommettere su una teoria (universale) significa semplicemente scommettere in un modo ben specificato su sequenze finite di eventi (singolari). Si noti che nessuna delle funzioni di distribuzione fin qui considerate determina una Prob che consente apprendimento forte dall'esperienza. Tali funzioni caratterizzano l'opinione di chi è disposto a dare a 0 contro n, per ogni n, qualunque teoria (universale), di chi cioè è praticamente certo che nell'estensione infinita dello spazio-tempo essa abbia almeno una eccezione. Nel caso speciale considerato nel § 5.1, questo dipende ovviamente dal fatto che tutte le funzioni ivi considerate determinano una densità «infinitesima» per i punti estremi dell'intervallo, e cioè 0 e 1, assegnano cioè una probabilità «infinitesima» alle due ipotesi generali «Tutti i lanci dànno testa» e «Tutti i lanci dànno croce». Jeffreys e Wrinch [cfr. Jeffreys 1961, pp. 117 sgg.] hanno perciò suggerito di assegnare una densità finita a tali punti estremi, di utilizzare cioè una funzione di distribuzione F *mista*; una possibilità è ad esempio la seguente:

$$F(\vartheta)=\begin{cases} 0 & \text{se } \vartheta<0 \\ \dfrac{(1+\vartheta)}{3} & \text{se } 0\leq\vartheta<1 \\ 1 & \text{se } \vartheta\geq 1. \end{cases}$$

Essa implica che $f(0)=f(1)=1/3$, e che $F(b)-F(a^-)=(b-a)/3$. Il grafico di F è rappresentato nella figura 8. Piú in generale, nel caso multinomiale (in cui cioè ciascuna variabile casuale ammette piú di due valori, ma comunque un numero finito) Hintikka [Hintikka e Niiniluoto 1974] ha mostrato che una data funzione Prob ammette apprendimento forte *solo se* viola il postulato di sufficienza di Johnson (o, equivalentemente, l'assioma di irrilevanza predittiva di Carnap) secondo cui la probabilità che una variabile casuale assuma un certo valore nella prova successiva dipende *solo* dal numero totale di prove precedenti e dal numero di queste ultime in cui è esemplificato quello stesso valore. Hintikka ha inoltre mostrato che se si assume che tale probabilità dipende *anche* dal numero di valori distinti esemplificati nelle prove precedenti, si ottengono allora funzioni di probabilità che consentono apprendimento forte. Un'altra interessante proposta in questo senso è stata avanzata da Good [1965, pp. 26-27, e cap. VIII].

Si supponga allora che sia data una funzione Prob che consente apprendimento forte. Che cosa implica l'impostazione bayesiana circa la scelta tra E e $-E$? Un'interpretazione corrente è che essa implichi la scelta della teoria che ha massima probabilità finale. Ma non è affatto questo il caso. Una scelta tra teorie scientifiche è una decisione come ogni altra, e dunque va operata in modo da massimizzare l'utilità prevista (finale). Il problema in questa classe di casi è che non è affatto chiaro quali siano le utilità coinvolte nell'accettare o rifiutare una teoria. Una possibilità naturale (suggerita da Hintikka [1968, trad. it. pp. 227-30] ed elaborata da Hilpinen [1968, capp. III, VIII e IX]) consiste nell'identificare l'utilità di accettare E, se è vera, con l'informazione $\mathrm{cont}(E) = 1 - \mathrm{Prob}(E)$ che in tal modo si ottiene, e l'utilità di accettare E, se è vera $-E$, con l'informazione in tal modo perduta, e cioè con $-\mathrm{cont}(-E)$. L'utilità prevista finale di accettare E è allora

$$\mathrm{Prob}(E|E(n))\,\mathrm{cont}(E) - \mathrm{Prob}(-E|E(n))\,\mathrm{cont}(-E).$$

Ovvie trasformazioni implicano allora che l'utilità prevista finale di E è uguale a $\mathrm{Prob}(E|E(n)) - \mathrm{Prob}(E)$. Dunque, la strategia bayesiana consiste nella scelta

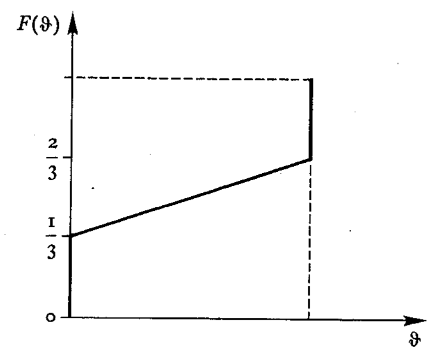


Figura 8.

Grafico di una funzione di distribuzione mista che consente apprendimento forte dall'esperienza.

della teoria che massimizza la differenza tra probabilità finale e iniziale. Come ha mostrato Jeffrey [1975, pp. 151-57] questa strategia è superiore alla strategia popperiana di scelta in termini di corroborazione o potere esplicativo.

A conclusione, due parole di cautela. È ovviamente *fanta*scientifico supporre che le probabilità pertinenti a queste scelte possano essere valutate in modo numericamente esatto. Il valore dell'impostazione bayesiana – in queste classi di casi – *non* consiste quindi nel dare una regola mediante cui tutte le controversie scientifiche possono essere risolte con un semplice «Calculemus». Consiste eventualmente nel dare un modello *qualitativamente* adeguato dei fattori che influenzano tali scelte, senza alcuna pretesa che possano venire in generale rappresentati numericamente. [Cfr. ad esempio De Finetti 1970, pp. 559-61].

Infine, è chiaro che questo metodo di confronto via una evidenza comune alle teorie in competizione va comunque generalizzato in modo da tener conto del fenomeno della «varianza di significato»: solo in questo modo esso risulterà applicabile a quelle coppie di teorie che – per uno slittamento di significato dei loro termini cruciali – non ammettono alcuna evidenza comune.

7. *Problemi aperti.*

In quest'ultimo paragrafo non verranno trattate le questioni connesse al problema «tecnico» di dare una ricostruzione adeguata entro l'impostazione bayesiana dei metodi «oggettivisti» validi. Non che problemi di questo genere manchino; ma sono troppo specifici perché sia opportuno affrontarli in questa sede. Si vuol qui accennare brevemente, piuttosto, ad alcuni problemi di carattere «fondazionale», anche se è indubbiamente dal successo nella soluzione dei primi che dipenderà la carriera scientifica futura del paradigma bayesiano.

7.1. L'«onniscienza» logica.

Nel § 5.1 sono stati identificati gli eventi con particolari sottoinsiemi di «punti» dello spazio S (oppure con la loro funzione caratteristica). Questo significa che ogni enunciato in una variabile libera che prende valori in S determinerà un unico evento, e dunque che enunciati logicamente equivalenti determineranno lo stesso evento. Perciò, è indifferente parlare di eventi oppure degli enunciati in una variabile libera che li determinano, pur di assumere in questo secondo caso che gli ordinamenti di preferenza sono invarianti rispetto a trasformazioni logicamente equivalenti degli enunciati considerati. La condizione di coerenza implica allora il seguente teorema:

TEOREMA. *Se l'enunciato $\varphi(w)$ implica logicamente l'enunciato $\psi(w)$, allora* $\mathrm{Prob}(\varphi(w)) \leq \mathrm{Prob}(\psi(w))$; *e in particolare: se l'enunciato $\varphi(w)$ è una verità logica, allora* $\mathrm{Prob}(\varphi(w)) = 1$.

È allora chiaro che per essere coerenti bisognerà conoscere tutte le conseguenze logiche di ogni enunciato che determina gli eventi dello spazio considerato, e in particolare sapere quali tra questi enunciati sono verità logiche. Perciò – notava Ramsey – «anche quando non vogliamo contraddirci, non riusciamo sempre a farlo: vi sono proposizioni matematiche la cui verità o falsità non può ancora essere decisa. Eppure, si potrebbe, umanamente parlando, aver ragione a credere in esse ad un certo grado per ragioni induttive, o d'altro genere», e ne concludeva che «una logica che si proponga di giustificare un grado di credenza del genere deve essere preparata a andare contro la logica formale, poiché a una verità formale la logica formale può solo assegnare una credenza di grado 1» [1926, trad. it. p. 208].

Se questa conclusione di Ramsey fosse vera, essa stabilirebbe l'inadeguatezza – nel senso specificato nel § 1 – del modello Ramsey - De Finetti - Savage a rappresentare la pratica induttiva. Il problema è ancora piú intricato dopo che i risultati di Gödel e Church hanno stabilito che – per classi di enunciati con una struttura sufficientemente ricca – la relazione di implicazione logica è *in linea di principio* indecidibile.

La miglior risposta al dubbio scettico di Savage «È possibile migliorare la teoria da questo punto di vista in modo che tenga conto del costo del pensare, o questo porterebbe a paradossi, come io sono propenso a credere, ma non sono in grado di dimostrare?» [1967, p. 308] è indubbiamente quella avanzata da Hintikka [1973, cap. X]. Si tratta però di una soluzione solo parziale. Infatti, la nozione generalizzata di coerenza proposta da Hintikka, pur «salvando» una vasta classe di comportamenti, esclusi come incoerenti dal modello Ramsey - De Finetti - Savage, continua ad escluderne alcuni a prima vista perfettamente razionali. Si tratta di alcuni di quei casi in cui, benché la relazione di implicazione logica sia decidibile, la sua complessità computazionale è tale da rendere razionale un comportamento incoerente anche nel senso generalizzato. Un esempio di questo tipo è quello costruito da Savage [1967, p. 308]. Si supponga che venga proposta una scommessa sull'evento: n (per $n = 0, 1, ..., 9$) è la milionesima cifra dello sviluppo decimale di π. La quota coerente (sia nel senso di Hintikka sia in quello di Ramsey - De Finetti - Savage) è 0 o 1. Ma per sapere se accettare o rifiutare alla quota proposta, dovremo impegnarci in un calcolo che con probabilità 9/10 potrebbe rivelarsi sprecato. Quindi, in conflitto con il modello, potrebbe essere razionale accettare la scommessa se la quota proposta è minore o uguale a 1/10.

7.2. L'evidenza «incerta».

Si tratta qui, piú che di una inadeguatezza del modello, di una sua limitazione. Infatti, esso restringe il processo di condizionalizzazione a quei casi in cui l'evento condizionante D è conosciuto con *certezza*. Dopo la scomparsa dei «dati puri», è però dubbio che si diano mai casi di questo genere. Il caso tipico è quello in cui – dopo una o piú osservazioni – si apprende la verità di D solo con un certo grado di probabilità. Un'estensione del modello in questa direzione è stata proposta da Jeffrey [1965, pp. 153-66].

7.3. L'a priori e l'esperienza.

Anche accettando l'estensione di Jeffrey, l'unico tipo di apprendimento dall'esperienza ammesso dal modello è quello in cui si calcola una distribuzione finale a partire da una distribuzione iniziale *fissata*, tenendo conto di uno o piú eventi osservati nel frattempo. Perché non considerare anche la possibilità che l'osservazione muti *anche* la distribuzione iniziale? Il problema è stato sollevato da Hintikka [1974, p. 6] ma non ha ancora ottenuto alcuna risposta soddisfacente.

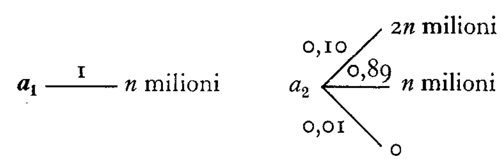
7.4. La vaghezza.

«Un caso particolarmente evidente in cui noi non conosciamo noi stessi completamente (e in cui perciò non possiamo comportarci in accordo con il modello) è rivelato dalla nostra incertezza, o vaghezza, circa le nostre preferenze tra scelte relativamente semplici quali fra 5000 lire e un biglietto a teatro. Alcuni hanno tentato di riflettere il fenomeno della vaghezza nel modello; altri credono invece che, benché si debbano certo fare i conti con tale fenomeno, esso eluda qualsiasi formalizzazione» [Savage 1967, p. 308]. Si è già visto nel § 5.2 che – indipendentemente dalla possibilità di incorporarlo esplicitamente nel modello – il fenomeno della vaghezza è *in pratica* trattabile. Suppes [1975] e Fishburn [1973] hanno però mostrato che esso è formalizzabile in modo soddisfacente entro una opportuna estensione del modello.

7.5. Il paradosso di Allais.

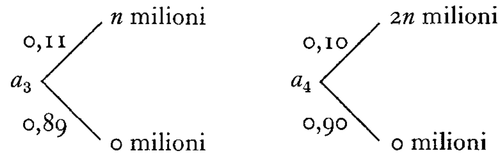
Il paradosso di Allais [1953] mette in discussione l'adeguatezza del modello, nel senso precisato nel § 1: esso pretende cioè di stabilire l'esistenza di una classe di casi in cui il nostro comportamento induttivo è sufficientemente uniforme, ma in conflitto con il modello. Si tratta dunque, *prima facie*, di una classe di casi paradigmatici incompatibili con il modello.

Si consideri allora un agente e guadagni maggiori o uguali a n milioni che siano sufficientemente grandi rispetto al suo patrimonio. (Ciascuno si scelga il valore di n adatto al proprio). Si supponga che venga offerta all'agente in questione una scelta tra i due seguenti atti, a_1 e a_2:



(la cifra a sinistra del guadagno indica la probabilità degli eventi il cui esito è il guadagno in questione). È un dato di fatto sperimentale – ma nessuno si sorprenderà – che la maggior parte degli agenti scelgano a_1. Ma – fin qui –

niente di male. Si supponga però che allo stesso agente venga offerta una scelta ulteriore tra i due seguenti atti, a_3 e a_4:



È di nuovo un dato di fatto sperimentale – e di nuovo nessuno si sorprenderà – che la maggior parte degli agenti scelgono a_4.

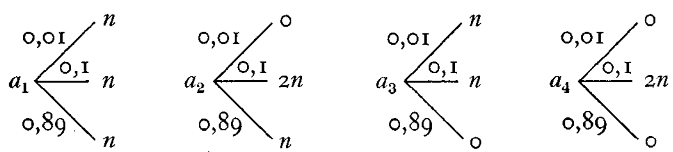
È noto però (cfr. p. 390) che – in base al modello in questione – a è preferito ad a' se e solo se $\sum U(c_i)\,\mathrm{Prob}(H_i) > \sum U(c_i')\,\mathrm{Prob}(H_i)$. Dunque la coppia di scelte precedenti implica le due seguenti disuguaglianze:

$$U(n) > 0{,}10\,U(2n) + 0{,}89\,U(n) + 0{,}01\,U(0)$$
$$0{,}11\,U(n) + 0{,}89\,U(0) > 0{,}10\,U(2n) + 0{,}90\,U(0).$$

Sommando membro a membro, si ottiene immediatamente la contraddizione:

$$U(n) + 0{,}90\,U(0) > U(n) + 0{,}90\,U(0).$$

Questo stabilisce che una struttura di preferenze come quella considerata non ammette alcuna rappresentazione in termini del modello, e cioè che non esiste alcuna funzione di utilità compatibile con essa. La ragione è semplice: esso viola il principio della cosa sicura. Per vederlo, basta rappresentare i quattro atti sotto la seguente forma ovviamente equivalente:



Dato che a_3 e a_4 sono ottenuti rispettivamente da a_1 e a_2 rimpiazzando una loro conseguenza *comune*, e cioè l'esito n con probabilità 0,89, con una stessa conseguenza, e cioè l'esito 0 con probabilità 0,89, e lasciando tutto il resto invariato, allora il principio della cosa sicura implica che se a_1 è preferito ad a_2, allora a_3 è preferito ad a_4. Quel che rende paradigmatica la classe di casi isolata dal paradosso di Allais è che l'uniformità di comportamento persiste anche dopo che gli agenti scoprono la violazione del principio in questione. In effetti, in questa classe di casi, la violazione del principio è collegata ad un argomento a prima vista valido. Da un lato, scegliendo a_1 si è certi di ottenere un consistente miglioramento del nostro livello di vita, che non si è disposti a mettere a repentaglio per un mutamento ulteriore, ma incerto; dall'altro, essendovi poche probabilità di vincere sia scegliendo a_3 che a_4, si considera trascurabile una differenza di probabilità dell'1 per cento rispetto a una differenza di guadagno del 100 per cento.

È opportuno fare un passo indietro. Il paradosso di Pietroburgo aveva messo in luce il ruolo dell'*avversione al rischio* in certe classi di casi ben specificate. La concavità della funzione di utilità proposta da Daniele Bernoulli era volta a tener conto precisamente di questo fattore nella formazione di decisioni razionali. La moderna generalizzazione del modello classico ha quindi consentito di render conto anche di certi tipi di atteggiamento verso il rischio completamente diversi, ammettendo come funzioni di utilità e probabilità una classe di funzioni molto piú estesa. Ma quel che sembra mostrare il paradosso di Allais è proprio la falsità della congettura che tutti i tipi di atteggiamento razionale verso il rischio possano essere rappresentati entro il modello Ramsey - De Finetti - Savage. Esso mette in luce il ruolo decisivo in alcuni casi di formazione di decisioni *prima facie* razionali di un *effetto di certezza o sicurezza* di cui il modello in questione non riesce a tener conto.

Il paradosso non ha avuto fino ad oggi alcuna soluzione soddisfacente. È probabile però, parafrasando Descartes su Galileo, che questo «modo di filosofare sia tanto piú vicino alla verità quanto piú facilmente si possono scoprire i suoi errori» [Descartes 1638, p. 1025].

Può dunque ben darsi che questo paradosso segnali una inadeguatezza del modello superabile soltanto da una teoria interamente nuova. Ben difficilmente tuttavia, in queste circostanze, un candidato a questo ruolo potrà essere considerato soddisfacente, a meno che non implichi la sua verità approssimata. Questo è il modo delle rivoluzioni scientifiche in una scienza matura; e «nessuna teoria potrebbe avere in sorte un destino migliore che quello di indicare la strada per la costruzione di una teoria piú comprensiva, in cui essa continua a vivere come caso limite» [Einstein 1917, trad. it. pp. 102-3]. [M. M.].

Allais, M.
1953 *Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et des Axiomes de l'Ecole Americaine*, in «Econometrica», XXI, pp. 503-46.
Antoniak, C. E.
1974 *Mixtures of Dirichlet Processes with Applications to Bayesian Non-parametric Problems*, in «Annals of Statistics», II, pp. 1152-74.
Bernoulli, D.
1738 *Specimen theoriae novae de mensura sortis*, in *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, tomo V, pp. 175-92.
Cox, D. R., e Hinkley, D. V.
1974 *Theoretical Statistics*, Chapman and Hall, London.
De Finetti, B.
1937 *La Prévision; ses Lois Logiques, ses Sources Subjectives*, in «Annales de l'Institut Poincaré», VII.
1959 *La Probabilità e la Statistica nei rapporti con l'Induzione, secondo i Diversi Punti di Vista*, CIME, Varenna.
1970 *Teoria delle probabilità. Sintesi introduttiva con appendice critica*, Einaudi, Torino.
Descartes, R.
[1638] Lettera a Mersenne dell'11 ottobre, in *Œuvres et Lettres*, Gallimard, Paris 1953, pp. 1024-39.
Domotor, Z.
1978 *Axiomatization of Jeffrey's Utilities*, in «Synthese», XXXIX, pp. 165-210.

Einstein, A.
1917 *Über die spezielle und allgemeine Relativitätstheorie (gemeinverständlich)*, Vieweg, Braunschweig (trad. it. in *Relatività. Esposizione divulgativa*, Boringhieri, Torino 1967, pp. 41-140).
Ferguson, T. S.
1973 *A Bayesian Analysis of Some Non-parametric Problems*, in «Annals of Statistics», I, pp. 209-30.
Fishburn, P. C.
1973 *Utility Theory with Inexact Preferences and Degrees of Preference*, in J. Leach, R. Butts e G. Pearce (a cura di), *Science, Decision and Value*, Reidel, Dordrecht, pp. 98-114.
Good, I. J.
1965 *The Estimation of Probabilities*, Mit Press, Cambridge Mass.
Goodman, N.
1955 *Fact, Fiction and Forecast*, Bobbs-Merrill, Indianapolis.
Haldane, J. B. S.
1945 *On a method of estimating frequencies*, in «Biometrika», XXXIII, pp. 222-25.
Hardy, G. H.
1949 *Divergent Series*, Oxford University Press, London.
Hewitt, E., e Savage, L. J.
1955 *Symmetric Measures on Cartesian Products*, in «Transactions of the American Mathematical Society», LXXX, pp. 470-501.
Hilpinen, R.
1968 *Rules of Acceptance and Inductive Logic*, North-Holland, Amsterdam.
Hintikka, J.
1968 *The Varieties of Information and Scientific Explanation*, in B. van Rootselaar e J. F. Staal, *Logic, Methodology and Philosophy of Science. Proceedings of the III International Congress*, North-Holland, Amsterdam, pp. 311-31 (trad. it. in *Induzione, accettazione, informazione*, Il Mulino, Bologna 1974, pp. 211-34).
1971 *Unknown Probabilities, Bayesianism, and de Finetti's Representation Theorem*, in R. C. Buck e R. S. Cohen (a cura di), *Boston Studies in the Philosophy of Science*, VIII. *PSA 1970, In memory of R. Carnap*, Reidel, Dordrecht, pp. 325-41.
1973 *Logic, Language-games and Information*, Oxford University Press, London (trad. it. Il Saggiatore, Milano 1975).
1974 Prefazione a *Induzione, accettazione, informazione*, Il Mulino, Bologna, pp. 3-7.
Hintikka, J., e Niiniluoto, I.
1974 *An Axiomatic Foundation for the Logic of Inductive Generalization*, in M. Przelecki, K. Szaniawski e R. Wojcicki (a cura di), *Formal Methods in the Methodology of Empirical Sciences*, Reidel, Dordrecht 1976, pp. 57-81.
Humburg, J.
1971 *The Principle of Instantial Relevance*, in R. Carnap e R. Jeffrey (a cura di), *Studies in Inductive Logic and Probability*, I, University of California Press, Berkeley Cal., pp. 225-34.
Hume, D.
1739 *A Treatise of Human Nature*, Noon, London 1739-40 (trad. it. in *Opere*, vol. I, Laterza, Bari 1971, pp. 5-665).
1740 *An Abstract of a Treatise of Human Nature*, Borbet, London (trad. it. *ibid.*, pp. 669-88).
Jaynes, E. T.
1968 *Prior Probabilities*, in «IEEE Transactions on System Science and Cybernetics», SSC-4, pp. 227-41.
1976 *Confidence Intervals vs Bayesian Intervals*, in W. L. Harper e C. A. Hooker (a cura di), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, II, Reidel, Dordrecht, pp. 175-213, 229-57.
Jeffrey, R.
1965 *The Logic of Decision*, McGraw-Hill, New York.
1975 *Replies*, in «Synthese», XXV, pp. 149-58.
1978 *Axiomatizing the Logic of Decision*, in C. A. Hooker e altri, *Foundations and Applications of Decision Theory*, I, Reidel, Dordrecht.

Jeffreys, H.
1946 *An Invariant Form for the Prior Probability in Estimation Problems*, in «Proceedings of the Royal Society», A. 186, pp. 453-61.
1961 *Theory of Probability*, Oxford University Press, London 1961³.
Kant, I.
1800 *Logik. Ein Handbuch zu Vorlesungen...*, Nicolovius, Königsberg; ora in *Werke*, vol. IX, De Gruyter, Berlin und Leipzig 1923.
Keynes, J. M.
1933 *Essays in Biography*, Macmillan, London (trad. it. Einaudi, Torino 1974).
Knopp, A.
1956 *Infinite Sequences and Series*, Dover, New York.
Lakatos, I.
1970 *Falsification and the Methodology of Scientific Research Programs*, in I. Lakatos e A. M. Musgrave (a cura di), *Criticism and the Growth of Knowledge*, Cambridge University Press, London, pp. 91-196 (trad. it. Feltrinelli, Milano 1976, pp. 164-276).
Leibniz, G. W.
[1703-704] *Nouveaux essais sur l'entendement humain*, in *Œuvres philosophiques latines et françaises*, Raspe, Amsterdam und Leipzig 1765 (trad. it. Laterza, Bari 1926).
Lindley, D. V.
1965 *Introduction to Probability and Statistics, from a Bayesian Point of View*, Cambridge University Press, London.
1971 *Bayesian Statistics, a Review*, SIAM, Philadelphia.
1973 *Discussion*, in W. L. Harper e C. A. Hooker (a cura di), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, II, Reidel, Dordrecht, pp. 59-61.
Quine, W. van Orman
1969 *Natural Kinds*, in *Ontological Relativity*, Columbia University Press, New York.
Raiffa, H., e Schlaifer, R.
1961 *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, Boston.
Ramsey, F. P.
[1926] *Truth and Probability*, in R. B. Braithwaite (a cura di), *The Foundations of Mathematics, and Other Logical Essays*, Routledge and Kegan Paul, London 1931 (trad. it. Feltrinelli, Milano 1964).
Savage, L. J.
1954 *The Foundations of Statistics*, Wiley, New York.
1959 *La Probabilità Soggettiva nei Problemi Pratici della Statistica*, CIME, Varenna.
1964 *Subjective Probability and Statistical Practice*, in L. J. Savage e altri, *The Foundations of Statistical Inference*, Methuen, London.
1967 *Difficulties in the Theory of Personal Probability*, in «Philosophy of Science», XXXIV, pp. 305-10.
1972 *The Foundations of Statistics*, Dover, New York 1972².
1973 *Probability in Science: a Personalistic Account*, in P. Suppes, L. Henkin, A. Joja e G. C. Moisil (a cura di), *Logic, Methodology and Philosophy of Science, Proceedings of the IV International Congress*, North-Holland, Amsterdam, pp. 417-28.
Suppes, P.
1975 *Approximate Probability and Expectation of Gambles*, in «Erkenntnis», IX, pp. 153-61.

Nell'ambito generale delle relazioni fra causalità e determinismo (cfr. **causa/effetto, determinato/indeterminato, caso/probabilità**) il problema dell'induzione è quello della validità di un ragionamento basato sull'esperienza di casi già verificatisi, contrapposto alla deduzione operata dal ragionamento nei sistemi formali (cfr. **empiria/esperienza, induzione/deduzione**).

Nel contesto induttivo la pratica è primaria (cfr. quindi **teoria/pratica** e anche **esperimento**) e i metodi sono quelli delle **probabilità** e della statistica (cfr. **rappresentazione statistica**). Pertanto le giustificazioni del ragionamento induttivo hanno il carattere delle usuali spiegazioni scientifiche che non forniscono certezze ma consentono una migliore organizzazione dei dati in strutture che evidenziano le leggi generali agenti (cfr. **spiegazione**, ma anche **legge**, **dato**, **struttura** e **certezza/dubbio**).

La maggior parte delle applicazioni è relativa a specifici modelli decisionali (cfr. **modello**, **decisione**) e, sotto la condizione di **coerenza**, corrisponde al caso di una sequenza di prove (cfr. **deduzione/prova**) indipendenti soggette a una stessa legge di carattere statistico nella quale si assumono come incogniti alcuni parametri (cfr. **distribuzione statistica**). Il ragionamento risulterà giustificato nella misura in cui contribuisce a decisioni ottimali (cfr. **evento**, **comportamento e condizionamento**, **previsione**).

La condizione di scambiabilità inoltre fornisce una giustificazione locale all'**apprendimento** dall'esperienza, nel quale la probabilità che si realizzi un **evento futuro** è tanto maggiore quanto maggiore è la frequenza osservata di eventi analoghi.

Probabilità

1. *Considerazioni introduttive.*

1.1. La probabilità: chi è costei?

Prima di rispondere a tale domanda è certamente opportuno chiedersi: ma davvero «esiste» la probabilità? e cosa mai sarebbe? Io risponderei di no, che non esiste. Qualcuno, cui diedi questa risposta (ribadita, col motto in tutte maiuscole – PROBABILITY DOES NOT EXIST – nella prefazione all'edizione inglese di *Teoria delle probabilità* [1970]), mi chiese ironicamente perché mai, allora, me ne occupo.

Mah! potrei anche dire, viceversa e senza contraddizione, che la probabilità regna ovunque, che è, o almeno dovrebbe essere, la nostra «guida nel pensare e nell'agire», e che per ciò mi interessa. Soltanto, mi sembra improprio, e perciò mi urta, vederla *concretizzata* in un sostantivo, 'probabilità', mentre riterrei meglio accettabile e piú appropriato che si usasse soltanto l'aggettivo, 'probabile', o, meglio ancora, soltanto l'avverbio, 'probabilmente'.

Dire che la probabilità di una certa asserzione vale 40 per cento appare – purtroppo! – come espressione concreta di una verità apodittica. Non pretendo né desidero che tale modo di esprimersi vada bandito, ma certo è che l'asserzione apparirebbe assai piú appropriatamente formulata se la si ammorbidisse dicendo, invece, che quel fatto lo si giudica «probabile al 40 per cento», o, meglio ancora (a parte che suona male), che ci si attende «al 40 per cento – probabilmente» che sia o che risulti vero.

Il guaio è che il realismo (come acutamente osservò Jeffreys) ha il vantaggio che «il linguaggio è stato creato da realisti, e per di piú da realisti molto primitivi», ed è per ciò che «noi abbiamo larghissime possibilità di descrivere le proprietà attribuite agli oggetti, ma scarsissime di descrivere quelle direttamente conosciute come sensazioni» [1939, p. 394].

Da ciò la mania (che forse per altri è invece indizio di saggezza, serietà, acutezza) di assolutizzare, di concretizzare, di oggettivizzare perfino quelle che sono soltanto proprietà dei nostri atteggiamenti soggettivi. Non altrimenti si spiegherebbe lo sforzo di fare della Probabilità qualcosa di *nobler than it is* (sempre parole di Jeffreys), nascondendone la natura soggettiva e gabellandola per oggettiva. Secondo la spiritosa fantasia di Hans Freudenthal si tratterebbe di uno strano pudore per impedire di farci vedere la Probabilità «come Dio l'ha fatta»: occorre «una foglia di fico», e spesso la si riveste tutta di foglie di fico rendendola addirittura invisibile o irriconoscibile.

1.2. Le probabilità: pretesamente oggettive.

Vi sono molti casi (spesso banali, ma anche no) nei quali le valutazioni di probabilità dei vari individui coincidono o tendono a coincidere. È abbastanza «na-

turale» (anche per un profano) attribuire, ad esempio, probabilità 40 per cento all'estrazione di pallina bianca da un'urna che ne contiene 100 di cui 40 bianche, o anche (piú o meno) se, anziché conoscerne la composizione, si sa che in 100 estrazioni con reimbussolamento (e rimescolamento) le estrazioni di palline bianche sono state 40.

Si è detto «naturale» (e tra virgolette) perché si tratta pur sempre di un giudizio probabilistico soggettivo (anche se, di solito, appare naturale e viene accettato da tutti). Non si tratta di sottigliezza sofistica: si tratta del fatto che un'opinione, in quanto tale, è sempre soggettiva, personale; è, cioè, tutt'altra cosa che un dato oggettivo (quali, ad esempio, la vera composizione dell'urna o l'effettiva frequenza osservata). A prescindere poi dal fatto che, per quanto riguarda la frequenza, essa è solo «probabilmente» vicina alla composizione dell'urna, e varia da un gruppo di «prove» a un altro.

È tuttavia un fatto che, in casi siffattamente schematici, piú o meno tutti giungono a valutazioni piú o meno concordanti, considerate per ciò dalla piú parte degli autori come espressione di «probabilità *oggettive*». Ma sarebbe piú appropriato, in tali casi, e verrà qui (ove occorra) seguito, l'uso del termine neutro 'probabilità ***pubbliche***', suggerito da Leonard Jimmie Savage (acutissimo pensatore e impareggiabile amico, scomparso, purtroppo, anzi tempo), oppure, come mi è sembrato (ripensandoci) ancor piú appropriato, 'probabilità *consuete*' (conformi a consuetudine): è infatti inutile, ingiustificato e fuorviante attribuir loro qualifiche piú ambiziose.

Vero è, come dato di fatto, che il consenso su certe valutazioni di probabilità è spesso piú o meno generale. E ciò costituisce un fatto concreto, una circostanza che può avere interesse in sé (ed essere utile in quanto favorisce mutua comprensione e consenso). Ma – attenzione! – essa non avrebbe alcun valore, avrebbe anzi un valore *fallace e negativo*, se venisse fraintesa come velleitaria e pretestuosa giustificazione di credenze di tipo superstizioso: la credenza, anzitutto, nella «esistenza» di una fantomatica «probabilità oggettiva», magari camuffata di volta in volta sotto le tradizionali spoglie della dea Fortuna e della strega Scalogna, cui attribuire tutto quel poco o tanto di bene e di male che a ciascuno viene largito.

Queste considerazioni introduttive non pretendono, né potrebbero, fornire fin d'ora indicazioni positive sul senso in cui occorre intendere la nozione di probabilità, precisando e perfezionando l'idea intuitiva che tutti ne abbiamo. Al contrario, sono intese a sgombrare il terreno da troppe idee preconcette, sia grossolane o sofisticate, che tuttora imperversano.

1.3. La presente occasione.

L'occasione che scaturisce dall'iniziativa di questa *Enciclopedia* sembra suscettibile di favorire un costruttivo chiarimento, un sostanziale passo in avanti nell'auspicata direzione.

I due magistrali articoli di Stefan Amsterdamski, «Caso/probabilità» e «Causa/effetto» (vol. II, pp. 668-87 e 823-45), aprono infatti la visuale su di una vasta e complessa tematica, assai analoga a quella prevista – sia pure come «sottofondo» – per la presente trattazione: una trattazione di carattere piú tecnico ma anche concettuale, la quale risulterà pertanto arricchita e meglio precisata in un puntuale confronto.

Confronto e *non* contrapposizione, direi, in quanto si tratta di proporre e cercar di giustificare una scelta univoca e precisa entro il largo ventaglio delle opzioni prospettate, o almeno non escluse, nei due già citati articoli.

In forma schematica, e approfittando della possibilità di far riferimento all'ampia e approfondita panoramica di Amsterdamski, posso precisare fin d'ora la mia posizione in poche parole dicendo che, delle due interpretazioni della probabilità ivi prospettate (pp. 674-75), escluderei senz'altro la prima secondo la quale «le asserzioni probabilistiche riguarderebbero gli eventi e sarebbero analitiche», mentre potrei accettare – in una versione invero molto radicalizzata – la seconda, riformulandola come segue: «La probabilità, pur essendo sempre una caratteristica dei giudizi, non è mai un concetto logico; le asserzioni contenenti valutazioni probabilistiche non sono mai analitiche in quanto esprimono sempre e soltanto il grado di credenza che, nel suo presente stato d'informazione, il soggetto che giudica attribuisce all'oggetto dell'asserzione. Sinteticamente, essa caratterizza, cioè, l'atteggiamento del soggetto conoscente nei riguardi di una data asserzione».

Per chiarire la situazione in forma piú esplicita basta chiedersi quali risposte può dare una persona interrogata riguardo a un *evento*, cioè a *una data affermazione* (dotata di senso univoco e per lei comprensibile). Evidentemente, le risposte possibili, tra cui ciascuno può scegliere quella che corrisponde allo stato delle sue attuali conoscenze al riguardo, sono, in senso *oggettivo*, tre: «Sí», «No», «Non so». La differenza essenziale fra le tre risposte sta nel fatto che (in qualunque versione) le due estreme: «Sí» (o «Vero», o «Certo») e «No» (o «Falso», o «Impossibile») sono dotate di un senso univoco, di un carattere definitivo e categorico, mentre quella intermedia «Non so» (o «Dubbio», o «Incerto») non ha invece che un carattere provvisorio in quanto esprime solamente il perdurare di una attuale ignoranza o indecisione tra il «Sí» e il «No», che sono le sole due risposte definitivamente concludenti.

In tale situazione di incertezza, ciascuno potrà propendere piú o meno sensibilmente per il «Sí» o per il «No», ed esprimere tale sua propensione dicendo che l'affermazione gli appare piú o meno probabile. Ma frasi del genere sono sempre vaghe, non impegnative, di dubbia interpretazione, magari a volte anche volutamente equivoche, come o quasi come i famigerati responsi della Sibilla, del tipo: «Ibis, redibis / non / morieris in bello», con libertà di immaginare la virgola prima o dopo del «non».

1.4. Come eliminare tale vaghezza?

Chi si limita ad esprimere la propria opinione dicendo che qualcosa è «molto» o «poco» probabile, che la sua probabilità è piú o meno «piccola» o «grande», dice ben poco; comunque, niente di preciso. È però sempre possibile (e, quando

l'indicazione vaga non basta, necessario) tradurre il proprio convincimento, il proprio grado di fiducia, in un'indicazione numerica, come 10 per cento, 40 per cento, 75 per cento di probabilità. E non c'è dubbio che ogni persona, anche poco o affatto istruita, sappia esprimere correttamente in tale forma le proprie opinioni e, analogamente, comprendere il significato di quelle esposte da altri. Si vedranno, ad esempio, nei §§ 1.7-1.8, i cenni esplicativi riguardanti i pronostici probabilistici sul calcio.

Semmai, il rischio è quello di esser stati deviati, allontanati, dalla concezione naturale causa una certa moda incomprensibilmente imperversante, favorevole a certe disgraziate concezioni della probabilità, banali, artificiose, e, per sovrappiú, fuorvianti e limitative delle capacità d'intendere di quanti vi si assuefanno.

La sola concezione che (come si spiegherà in seguito) abbia senso, l'unica che comporti una vera comprensione del significato e della validità del ragionamento probabilistico, è quella genuina di un qualunque «uomo della strada»: quella che ci guida in ogni attimo ed azione della nostra vita, anche se inconsciamente, con elaborazioni mentali e sintesi istintive piú rapide di quelle di un qualunque calcolatore elettronico.

Il «calcolo delle probabilità» (in quanto *calcolo*) può servire in casi artificialmente complessi, ma sempre considerandolo come un sussidiario dell'intuizione e non come sostituto (o come possibile sostituto) di essa.

In questo senso, si dovrebbe insistere soprattutto per far considerare il ragionamento probabilistico *non come un sostituto* bensí come *uno strumento integrativo* delle capacità intuitive che tutti (uomini ed altri animali) possediamo. Tali capacità, secondo una felice espressione (di cui mi spiace non ricordare chi ne sia l'autore) costituiscono una *built in machinery* nel nostro cervello (un macchinario innato). Ed è quindi da ciò che si deve partire. Si tratta (si ponga ben attenzione!) di comprensione effettiva, anche se un po' rozza, ed occorrerà soltanto *approfondirla e affinarla*; sarebbe invece un *regresso* sostituire questa comprensione intuitiva e pratica con delle *pseudodefinizioni* (!) della *probabilità*: pseudodefinizioni – purtroppo di *moda*! – che si autodefiniscono *oggettivistiche*. Se ne riparlerà a suo tempo e luogo.

Ma occorre anzitutto indicare (come verrà fatto nei prossimi §§ 1.5-1.6) un procedimento *operativo* atto a *misurare* la probabilità di un evento E: ripetiamo (meglio ripeterci fino alla noia pur di evitare il rischio di fraintendimenti) che 'evento' significa «caso unico ben determinato».

Attenzione: abitualmente il termine 'evento' viene invece usato in senso generico, per indicare tutti gli eventi di un certo tipo, detti «prove» di quell'«evento». Ciò comporta molti inconvenienti ed inestricabili confusioni; per evitarli si potrebbe dire che certi eventi (analoghi) sono «prove di un medesimo *fenomeno*» (ma senza intendere con ciò che siano ugualmente probabili o indipendenti od altro salvo che non sia esplicitamente detto). La terminologia attuale è, oltretutto, ambigua, perché a volte si considera anche il caso in cui la probabilità di un «evento» *varia* di prova in prova (ma allora vuol dire che la probabilità si riferisce, anche per gli oggettivisti, non all'«evento» secondo l'accezione oggettivistica, bensí all'evento (caso singolo) secondo la terminologia conforme alla concezione soggettivistica. Ciò dovrebbe bastare per far riconoscere a chiunque – chissà come cosí non è? – che la concezione (e la stessa terminologia) degli *oggettivisti* altro non è che una vuotaggine confusionaria).

1.5. «Previsione» e «scarto» (quadratico medio).

Finora si sono considerati soltanto eventi e loro probabilità, ma si tratta solo di un caso particolare di quello piú generale dei numeri aleatori e della loro previsione. Come caso particolare (già noto) è un numero aleatorio ogni evento, se, con convenzione di cui si vedrà sempre meglio l'appropriatezza, si identifica l'evento E col numero aleatorio che vale 1 se E è vero e 0 se è falso. (Spesso lo si chiama «indicatore di E», ma senza alcun costrutto: una distinzione senza differenza non crea che apparenti ed inutili doppioni di parole, oltre a contravvenire una norma sacrosanta: «Entia non sunt multiplicanda sine necessitate»).

Un numero aleatorio, X, può assumere un numero finito di valori (come i punti da 1 a 6 con un dado o da 2 a 12 con due dadi, o da 1 a 90 alla tombola) o tutti i valori reali entro un intervallo «verosimile» se si tratta ad esempio della «temperatura di domani mattina» a un dato osservatorio meteorologico.

Nel caso *discreto* (valori possibili in numero finito: $x_1, x_2, ..., x_n$) basterà indicare le probabilità $p_1, p_2, ..., p_n$ attribuite a ciascuno di essi, e si avrà una distribuzione *discreta*; altrimenti una distribuzione *continua*, nel caso piú regolare con una densità, $f(x)$ (probabilità $f(x)\,dx$ che X cada tra x e $x+dx$, per dirla in termini comprensibili anche se criticabili). C'è anche un caso intermedio (in certo senso «patologico»): vedansi la figura 1 e relativa didascalia. (Per maggiori informazioni cfr. l'articolo «Distribuzione statistica» in questa stessa *Enciclopedia*).

Il nostro attuale obiettivo è molto limitato ed elementare; tuttavia, il modo di considerarlo è inteso a preparare il terreno per una discussione semplice ma

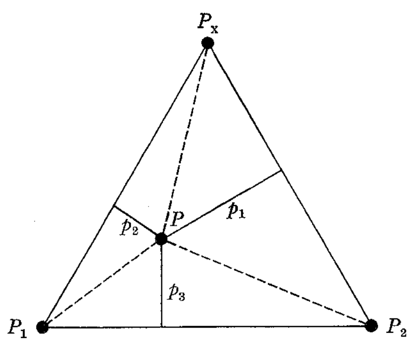


Figura 1.

Triangolo equilatero: per ogni punto interno P la somma delle tre distanze dai lati è costante (e precisamente è uguale all'altezza). Perciò, nel caso di tre eventi incompatibili ed esaustivi (come i risultati «1», «x», «2» in una partita di calcio), ogni opinione sull'esito di una data partita (probabilità p_1, p_x, p_2) è rappresentabile come un punto del triangolo.

critica sul modo appropriato di valutare la previsione, $m = \mathbf{P}(X)$, nonché lo scarto (quadratico medio) di X, che si indica con sigma di X, $\boldsymbol{\sigma}(X)$. Il suo quadrato $\boldsymbol{\sigma}^2(X) = \mathbf{P}(X-m)^2$ (con $m = \mathbf{P}(X)$) si chiama «varianza» di X. In parole: σ è la radice della media di $(X-m)^2$, cioè del quadrato degli scostamenti di X dalla media m. Come dice il nome stesso, σ fornisce una misura (inversa) dell'addensamento della distribuzione attorno al valor medio (o, in termini meccanici, baricentro). Si può anche considerare lo scarto quadratico medio da un punto (o valore) x diverso dalla media m; lo si indichi σ_x. È facile vedere che $\sigma_x^2 = \sigma^2 + (x-m)^2$; in termini geometrici, σ_x è l'ipotenusa del triangolo rettangolo di lati σ_x ed $(m-x)$, e quindi il baricentro è il punto rispetto al quale il momento è minimo. (Ed è, del resto, intuitivo che, se l'asse di rotazione passa per il baricentro, la massa vi è ravvicinata e il momento diminuisce). Ed è questa la conclusione che serve: il baricentro, oltre che come punto di equilibrio, è anche caratterizzato dall'essere il punto rispetto al quale il momento è minimo; si dispone pertanto di due metodi per trovare il baricentro di un solido (nel caso che ci interessa: una sbarra): 1) è il punto per cui si deve sospendere la sbarra affinché rimanga in equilibrio; 2) è il punto della sbarra che occorre tener fisso affinché, facendo ruotare la sbarra (beninteso, a parità di velocità angolare), l'energia o (come forse è piú familiare) la «forza viva», sia minima.

Immagino e comprendo lo stupore del lettore: chi mai farebbe tanti tentativi per misurare l'energia per rotazioni con assi diversi fino a individuare il minimo di «forza viva» e quindi il baricentro? Ha ragione, anzi ragioni da vendere...; ma, nel caso che ci interessa, l'aspetto meccanico scompare e rimane per analogia la questione di convenienza tra gli analoghi metodi nel contesto probabilistico.

E qui sta il punto: nel caso della probabilità la misura diretta, anziché essere la piú appropriata come nel caso meccanico, si riduce a profferire una cifra per la probabilità «ad occhio», senza alcun ausilio di controlli o correttivi; invece la procedura indiretta – cioè una «stima», ma collegata ad una «penalizzazione» (appropriata, nel senso di *proper scoring rule*) serve ad affinare la sensibilità degli «stimatori» e a vagliarne l'abilità tenendo conto (mediante i punteggi, *scores*, di ogni stima) dell'abilità dimostrata nel complesso della loro attività in tale campo. Naturalmente, se da una parte occorre buona compenetrazione con lo spirito del procedimento, occorre – e del resto è questa la motivazione del farne uso – una buona competenza e informazione nel campo specifico (nel nostro esempio, (cfr. §§ 1.7 e 1.8), valore e situazione delle squadre del campionato di calcio). Quanto alla regola di penalizzazione quadratica, che già avevo applicata in concorsi probabilistici sul calcio, appresi poi che era già nota (*Brier's rule* 'regola di Brier') ed applicata in America per dare indicazioni probabilistiche per la pioggia nei bollettini meteorologici diffusi ogni mattina da radio, Tv e giornali.

Beninteso, non avrebbe senso pensare che una previsione basata su questo tipo di procedure ed informazioni sia di per sé migliore (sarebbe miracolismo!); sta di fatto, però, che il metodo fornisce un autocontrollo, nonché un controllo comparativo se gli addetti o partecipanti a tali pronostici sono parecchi e possono (a posteriori) confrontare pronostici e risultati di tutti e diagnosticare il perché taluno va piú bene che male e talaltro piú male che bene.

	x										
	00	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
$(1-x) =$	100	90	80	70	60	50	40	30	20	10	00
$y=x(1-x) =$	00	09	16	21	24	25	24	21	16	09	00
Ordinate delle tangenti nel punto 0 =	00	01	04	09	16	25	36	49	64	81	100
Ordinate delle tangenti nel punto 100 =	100	81	64	49	36	25	16	09	04	01	00
Dislivello =	+100	+80	+60	+40	+20	00	−20	−40	−60	−80	−100
Retta a =	09	13	17	21	25	29	33	37	41	45	49
Retta b =	64	58	52	46	40	34	28	22	16	10	04

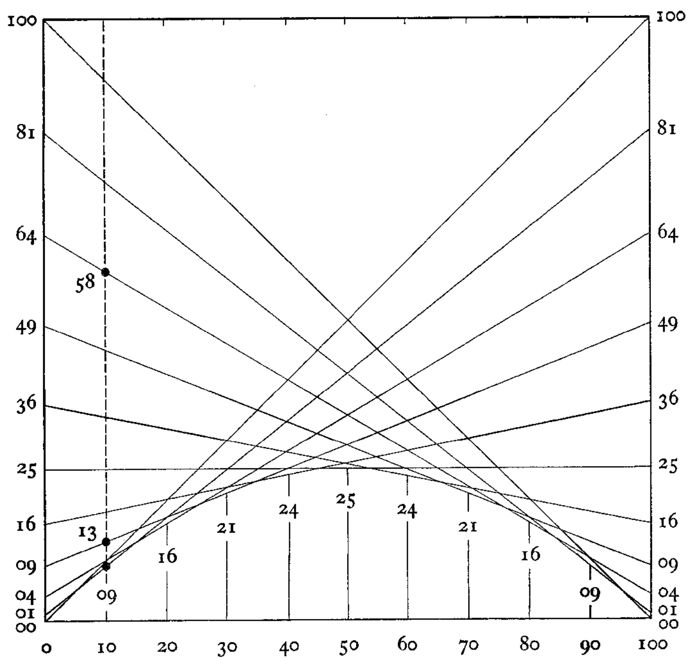


Figura 2.

Le rette corrispondono alle combinazioni di penalizzazione tra cui il metodo consente di scegliere (si può ridurre la penalizzazione in uno dei due casi a spese di un aumento nell'altro: per abbassare l'ordinata in un estremo si alza nell'altro). L'ordinata di una retta nel punto p è la previsione di penalizzazione per chi sceglie quella retta e attribuisce all'evento in questione la probabilità p. In tal caso il minimo ottenibile è dato dall'ordinata della parabola (nessuna retta vi passa al di sotto!) e la scelta ottima è quella della retta tangente alla parabola in corrispondenza all'ascissa p.

1.6. Una presentazione alternativa.

Può riuscire istruttiva e appropriata, sotto vari punti di vista, un'illustrazione anche in forma grafica del senso e del funzionamento delle valutazioni basate sulla minimizzazione del quadrato dell'errore (che, nel gergo statistico, si chiama «regola di Brier»).

La figura 2 mette in evidenza, visivamente, come e *perché* una regola di penalizzazione appropriata *obblighi* ciascuno, *nel suo proprio interesse*, a comportarsi in accordo con quanto segue dalla sua valutazione di probabilità e ad esprimere sinceramente tale sua valutazione.

La figura rappresenta un quadrato di lato unitario con l'arco di parabola $y = x(1-x)(0 \leq x \leq 1)$ e le tangenti ad essa per ogni decimo dell'ascissa.

(Per comodità tutti i valori sono indicati moltiplicati per 100, cioè, ad esempio, 100 anziché 1 e 24 anziché 0,24).

1.7. Pronostici probabilistici.

È importante intrattenersi sull'argomento dei pronostici probabilistici per vari motivi.

Il motivo teorico consiste nel mostrare come il concetto informatore della «regola di Brier» si trasporti, con le stesse utili proprietà, dal caso di due sole eventualità (eventi) a quello di tre (o piú).

Il motivo esemplificativo-psicologico consiste nell'illustrare la validità educativa di esercizi sistematici di valutazioni di probabilità, riferendo su concorsi di pronostici probabilistici riguardanti le partite del campionato di calcio.

Ed infine, incidentalmente ma appropriatamente, verrà messa in luce l'antitesi di mentalità di educatività e di moralità (in senso lato) tra i giochi-scommesse in cui si stimola la sciocca «furbizia» del «tirare a indovinare» e quelli in cui si tratta di dare una valutazione quanto piú «obiettiva» e spassionata possibile.

Quanto all'educatività e all'importanza pratica, si vedranno subito dopo (§ 1.9) le analoghe esperienze nel campo (nientemeno!) delle prospezioni petrolifere!

Nel caso del calcio (come per molti altri giochi) i risultati possibili per ogni partita sono tre: «1» = vittoria, «x» = pareggio, «2» = sconfitta (sempre con riferimento alla squadra ospitante). Ogni pronostico probabilistico consiste pertanto nell'indicare le tre probabilità, p_1, p_x, p_2 (di somma = 1, ossia 100 se le indicazioni sono fatte, come è usuale, in percento), ed è opportuno pensarle come masse (o «pesi») collocate nei vertici P_1, P_x, P_2 di un triangolo equilatero. La penalizzazione è il quadrato della distanza tra il punto-pronostico, P, e il punto-risultato: P_1 o P_x o P_2: ovvia estensione della «regola di Brier» a due (e anche, volendo, piú) dimensioni.

Statisticamente, la proporzione dei tre risultati «1» - «x» - «2» è in media 50 : 30 : 20 (50 per cento per vittoria in casa, 30 per cento per pareggio, 20 per cento per vittoria esterna), ma, è chiaro, queste sono indicazioni statistiche globali, mentre ogni caso singolo differisce per molte circostanze da tener presenti e vagliare attentamente: rapporto di valori tra le due squadre che si affrontano, fattore campo, condizioni del tempo, assenze e sostituzioni di giocatori per malattie, incidenti, squalifiche, importanza per la classifica dell'una e/o dell'altra squadra, ecc.: tutte cose che ciascuno dovrà tener presenti, nella misura in cui ne è informato, e vagliare con cura.

È bene sottolineare e ricordare insistentemente, per non cadere in distorsioni di visuale di tipo superstizioso o fatalistico o metafisicheggiante, che non si tratta di «scoprire» un preteso e fantomatico «valore vero» di ascose «probabilità oggettive», bensí di indicare il valore che *ciascuno a suo modo* (come nella commedia di Pirandello) vi attribuisce. Sperabilmente, lo farà previa attenta riflessione sui pro e sui contro, in conformità alla misura in cui ciascuno propende per l'una o l'altra delle tre possibilità. È *questa* la probabilità nell'*unico senso che appare valido*, universalmente valido.

(Esistono però, come si vedrà (§ 2.3), altre sedicenti «definizioni» che non possono venir considerate e accettate come tali, ma soltanto valide come criteri ausiliari per la valutazione – sempre, beninteso, soggettiva – delle probabilità).

1.8. Pronostici e concorsi pronostici.

Concorsi pronostici sulle partite del campionato di calcio, nell'illustrata forma *significativa*, sono stati ripetuti per parecchi anni (presso l'Università di Roma, con partecipazione anche di colleghi e di studenti di altre sedi). In forma «significativa» significa «nel modo già indicato»: significa cioè che si tratta dell'opposto del diseducativo criterio del «tirare a indovinare», del «tentare la fortuna», come al Totocalcio ove si tratta di «predire» il risultato secco (o «1» o «x» o «2»), o come al famigerato gioco del Lotto. Il quale – sia detto per inciso – concorre anche, indirettamente, a perpetuare incorreggibili diffuse idiozie, quali l'attesa con crescente fiducia di numeri «arretrati» o suggeriti da sogni o da astrologi o da «maghi» o dalla «cabala» o da calcoletti cervellotici... e chissà cos'altro!

Nulla vi è in comune, nei pronostici probabilistici, con tali disgustose forme di «predizioni secche», che appena di poco appaiono meno peggiori del sullodato Lotto (in quanto, nel calcio, la scelta fra «1» - «x» - «2» implica almeno un po' di riflessione). Meno male che tali superstiziose fole e scimunitaggini giovano allo Stato, e che molti cittadini, magari evasori fiscali o riluttanti e dispiaciuti nel pagare le debite tasse ed imposte, dimostrano un immenso anche se involontario (e pertanto non meritorio) zelo nel versare abbondantemente denaro per tale «tassa sulla imbecillità».

Il concorso pronostici probabilistico richiede invece ad ogni partecipante di indicare per ogni partita le probabilità che egli attribuisce ai tre risultati possibili, e, trattandosi di una «regola di penalizzazione appropriata», ha convenienza ad esprimerle sinceramente ed esattamente, dopo aver vagliato il valore delle squadre, preso nota di assenze per malattie o squalifiche, dello stato di forma dei giocatori, delle condizioni meteorologiche previste e dell'influenza di tutto ciò

(e di quant'altro ritiene rilevante) sul rendimento dell'una e dell'altra squadra.

Tale attività si è dovuta interrompere da qualche anno causa difficoltà create da scioperi e disservizi nel servizio postale e da analoghe complicazioni entro l'Università. È allo studio la possibilità di riprenderla in altra sede e con partecipazione di pronosticatori piú qualificati (giornalisti sportivi ed altre persone legate all'ambiente calcistico e sportivo).

Comunque, sarebbe auspicabile che tale capacità di esprimere in termini di probabilità il grado di attendibilità o di fiducia che uno attribuisce a risultati possibili di una qualsiasi azione o iniziativa venisse apprezzata e incoraggiata, al fine di venire effettivamente sfruttata, con consapevolezza e coerenza, per vagliare accuratamente il pro e il contro di ogni elemento che influisce sul risultato di ogni possibile decisione.

Particolarmente significativo e istruttivo a tale riguardo risulterà l'esempio che verrà illustrato nel § 1.9, considerate anche le necessarie nozioni di «numeri aleatori» e loro «previsione» introdotte nel § 1.10.

1.9. Ruolo della previsione in decisioni importanti.

Gli esempi finora introdotti riguardavano situazioni piú o meno di carattere ludico, in particolare risultati sportivi, e ciò sembrava utile per avviare e far entrare nello spirito della trattazione senza dover superare – sperabilmente – eccessive riluttanze. Forse, dopo aver appreso e meditato il modo in cui le valutazioni probabilistiche hanno un ruolo essenziale in situazioni di gioco, risulterà però ora – sempre «sperabilmente» – accettabile l'affermazione che i medesimi criteri e procedimenti sono applicabili, come sono stati effettivamente applicati, con risultati significativamente validi in campi ove è altissima l'*importanza pratica* di una attenta e accurata valutazione (da parte di esperti dei diversi rami) dei fattori e delle circostanze che, con le loro probabilità, incidono sulla probabilità da attribuire ad ipotesi di risultati globali piú o meno favorevoli.

Molte di tali questioni sono trattate sotto l'etichetta di «Ricerca operativa» (*Operation Research*), e alcuni esempi semplici, ma utili a scopo illustrativo, si possono vedere nell'articolo «Decisione» di questa *Enciclopedia* (vol. IV, pp. 421-84).

Ma l'esempio piú significativo, e in cui meglio appare la connessione fra tante valutazioni fatte da esperti diversi, è certamente quello relativo alla decisione di intraprendere, e poi di proseguire, e in quale modo, le ricerche petrolifere in una data zona, oppure di abbandonarle.

Presupposto per tale decisione è l'acquisizione di elementi di giudizio (geologici, ecc.) da parte di esperti, di una attenta e accurata valutazione da parte loro delle prospettive di successo o insuccesso – in termini di utile o perdita – di una tale costosissima impresa.

Riguardo ad esperienze su questo particolare ma assai rilevante e istruttivo problema, e al tipo di argomentazioni interconnesse cui conduce, vale la pena di segnalare soprattutto il libro *Decisions under Uncertainty: Drilling Decisions by Oil and Gas Operators* di Grayson jr [1958]. Egli descrive come sia riuscito ad ottenere dagli esperti (geologi, ingegneri, ecc.) di esprimere in *valutazioni probabilistiche* (*numeriche*) i loro giudizi sulle prospettive di successo di ricerche in una data località, anziché usare (come in precedenza era abituale) frasi studiatamente vaghe e ragionamenti sofisticati con comode riserve a titolo cautelativo... quasi ad imitazione della già menzionata Sibilla.

In base ad informazioni probabilistiche dettagliate (cioè, concernenti varie sottoipotesi sulla natura e ricchezza dei presunti giacimenti) diviene possibile anche stabilire, mediante un'analisi delle previsioni probabilistiche relative a diverse circostanze, la convenienza o meno (speranza di risparmio o timore di perdita) per ogni ulteriore esperimento di questo o quel tipo (ad esempio, perforazione di un pozzo di sondaggio o prospezione sismica) atto a consigliare o sconsigliare, a seconda dell'esito, la decisione finale (o, eventualmente, quella di rinviare la decisione procedendo, prima, ad ulteriori indagini).

Probabilmente molti saranno perplessi e troveranno ridicolo fare dei calcoli «campati in aria» (assimilando ad «aria», magari ad «aria fritta», le probabilità soggettive sia pure stimate da esperti); certamente, esse non possono dare alcuna certezza, ma l'indicazione di un grado di probabilità presentato come tale è il massimo grado ottenibile di informazione oggettiva: un'indicazione comunque molto piú dotata di senso di responsabilità, e quindi di attendibilità, che non una «certezza» fasulla, asserita con leggerezza, o un responso «oggettivo» ma ambiguo. Confucio, del resto, non aveva già detto che la parola 'certezza' era una di quelle che si sarebbero dovute abolire?

Tutto ciò appare naturale per chiunque, libero da preconcetti assolutistici, tenga conto del fatto che tutto è incerto, ma che per decidere occorre e basta basarsi su ciò che si sa (con certezza) e su ciò che si ritiene probabile, piú o meno probabile, sulla base di ciò che si sa e di ciò che non si sa. Ed ogni informazione arricchisce questo sfondo sempre incompleto, ma soltanto l'onniscienza potrebbe completarlo: guai a chi, rinunziando ad avvalersi dell'informazione possibile, decide a vanvera o rinunzia a decidere (o decide secondo pregiudizi generici, ritenendo di dovervisi attenere alla cieca, senza vagliarne l'appropriatezza e l'opportunità che variano caso per caso).

1.10. Probabilità, previsione, prezzo.

Ma – ci si potrà obiettare – non è un'inutile complicazione il riferimento a «regole di penalizzazione» dal momento che ciò (come si è visto) equivale all'affermazione banale, chiara per chiunque, che $\mathbf{P}(E)$ (sia ad esempio $\mathbf{P}(E)=0{,}40$) significa che 0,40 Lire è il *prezzo equo* per ricevere una Lira se E si verifica? Si usa 'una Lira' come termine generico: chi sentisse il bisogno di riferirsi a una scala piú attuale potrebbe intendere per 'Lira' una Kilolira (mille Lire) od altro importo a suo piacimento. Meglio però non troppo piccolo da rendere insignificante il risultato né troppo grande per evitare il divario tra valore monetario e *utilità* (cfr. il già citato articolo «Decisione»).

Una siffatta brutale identificazione della probabilità a prezzo avrebbe però il difetto di condurre in una deprecabile situazione di «gioco», nel senso magistralmente esposto da John von Neumann e Oskar Morgenstern nella loro famosa opera *Theory of Games and Economic Behavior* (1947) (cfr. l'articolo «Giochi» in questa stessa *Enciclopedia*).

Una tale situazione di «gioco» dà spesso adito, infatti, ad astuzie, a mercanteggiamenti o tentativi di mercanteggiamento, sicché il prezzo non sarebbe un dato certo e significativo su cui ci si possa basare. Se non si ponesse attenzione a tali inconvenienti la stessa probabilità verrebbe a confondersi con un frutto di patteggiamenti, di un labile compromesso tra chi vorrebbe spendere meno e chi vorrebbe incassare piú di quanto potrebbe venire ragionevolmente stabilito.

Questa critica non inficia tuttavia l'idea di considerare la probabilità come un prezzo: è soltanto necessario ricorrere ad uno «strumento di misura» insensibile ai menzionati fattori di distorsione. E tali strumenti – le «regole di penalizzazione appropriate» – si conoscono già, pur non avendone finora rilevato la proprietà che qui interessa.

A questo punto (per non ripetere due volte lo stesso discorso) conviene introdurre, oltre agli eventi, anche i «numeri aleatori», ad esempio $X=x_1E_1+x_2E_2+ +\ldots+x_nE_n$ (dove $E_1, E_2, \ldots, E_n$ formano una *partizione*: sono cioè incompatibili ed esaustivi, nel senso che se ne verifica certamente uno e uno solo): X è pertanto (come mostra la scrittura) il numero aleatorio che assume il valore x_1 se si verifica E_1 (e cosí via: $x_2, \ldots, x_n$ se si verificano, rispettivamente, $E_2, \ldots, E_n$).

Naturalmente, si possono considerare anche numeri aleatori con un'infinità (discreta o continua) di valori possibili: ad esempio, pensando ad un numero X (qualunque, o soltanto razionale) scelto «a caso» – cioè con densità uniforme di probabilità – tra 0 e 100, e quindi con probabilità $(x''-x')/100$ di trovarsi in qualunque intervallino (x', x'') contenuto in $(0, 100)$. Ma, per il momento, ci si limita al caso elementare di valori possibili in numero finito per non dover parlare di derivate e integrali.

Per sviluppare l'argomento in termini matematici (pur cercando di evitare discorsi in forma astrusa per non spaventare i profani) è necessario introdurre alcuni concetti e simboli (del resto già usati in casi particolari).

Anzitutto il simbolo $\mathbf{P}$, comodo per indicare indifferentemente sia *probabilità* (nel caso di eventi, ad esempio $\mathbf{P}(E)$), e sia *previsione* (nel caso di numeri aleatori, per esempio $\mathbf{P}(X)$). Però, con una interpretazione unitaria e banale, riferentesi a una scommessa unitaria, $\mathbf{P}(X)$ si può anche dire «*prezzo* di X» (prezzo da pagare per ricevere l'importo incognito X quando sarà noto), e cosí $\mathbf{P}(E)$, prezzo di E (di «una Lira» se si verifica E).

1.11. Probabilità (e previsione): sempre subordinate.

Parlare (come è stato fatto finora, «sic et simpliciter») di eventi e di numeri aleatori, come di enti cui riferire probabilità e rispettivamente previsione, è però un nonsenso. Per giustificare tale colpa occorre dire che era tuttavia utile far cosí per evitare troppe complicazioni tutte d'un colpo e per attirare maggiormente l'attenzione su di esse ora, facendo notare e correggere la provvisoria (e «calcolata») dimenticanza.

Dire che la probabilità di un dato evento, E, vale, ad esempio, $\mathbf{P}(E)=0{,}40$, non è un'affermazione avente un senso compiuto, a meno che non si pensi sottinteso il secondo *essenziale* fattore: il nostro attuale stato di conoscenza. Lo si indichi con H_0. Allora, a rigore, si dovrebbe indicare la scrittura completa, cioè $\mathbf{P}(E|H_0)$. Se quella che si vuole considerare è la probabilità subordinata all'ulteriore conoscenza o «ipotesi» H, quindi ad HH_0, si avrà $\mathbf{P}(E|HH_0)$ ove H_0 serve per rammentare lo stato di conoscenza attuale, mentre H è l'ipotesi aggiuntiva sotto la quale ci interessa stimare la probabilità di E.

Anziché $\mathbf{P}(E)$ e $\mathbf{P}(E|H)$ *dovremmo* pertanto scrivere sempre $\mathbf{P}(E|H_0)$ oppure $\mathbf{P}(E|HH_0)$, rispettivamente per ricordare e indicare quale sia il nostro stato di conoscenza, oppure, inoltre, quale sia l'ulteriore *ipotetica* circostanza H da aggiungervi, interessando conoscere quale sarebbe detta probabilità condizionandola a tale ampliata conoscenza (o informazione).

Ho detto «dovremmo», e non «dovremo», perché la continua indicazione e ripetizione di H_0 risulterebbe inutilmente ingombrante. Tuttavia, andrà *sempre* tenuto presente che questo «H_0» dovrà *sempre* intendersi *sottinteso*, mai soppresso come cosa superflua. E può essere sottinteso soltanto se dal contesto risulta in modo non dubbio quale sia la situazione (per quanto riguarda le circostanze rilevanti al riguardo).

La probabilità di un evento E dato un H si esprime, in base al «teorema delle probabilità composte»: $\mathbf{P}(EH)=\mathbf{P}(E)\cdot\mathbf{P}(H|E)$ od anche (è ovvia la simmetria) $\mathbf{P}(EH)=\mathbf{P}(H)\cdot\mathbf{P}(E|H)$.

Tenendo conto di tale identità è possibile ricavare, per $\mathbf{P}(E|H)$, l'espressione seguente:

$$\mathbf{P}(E|H)=\frac{\mathbf{P}(EH)}{\mathbf{P}(H)}=\mathbf{P}(E)\,\frac{\mathbf{P}(H|E)}{\mathbf{P}(H)}.$$

A parole: la probabilità di E, subordinandola ad H, si modifica nel medesimo rapporto in cui si modifica la probabilità di H subordinandola ad E.

È questo il fondamentale *teorema di Bayes*, base del ragionamento induttivo, e in particolare della statistica matematica (quando non venga ridotta a ricettari empirici piú o meno grossolani).

La principale fonte di errori e malintesi, nel campo probabilistico-statistico, consiste proprio nel considerare certi dati come se fossero dotati di senso assoluto, non pensando che esso è sempre relativo a un certo stato di conoscenze.

Eppure ciò sembra difficile da far capire (o «inghiottire»: a molti ripugna)! Quanti non insistono nel sostenere che esistano «probabilità oggettive» (e perché no, allora, anche quadrati circolari?!)

A chiunque parli di probabilità oggettive si dovrebbe dare una risposta drastica: la sola probabilità oggettiva, per un qualunque evento E, è $\mathbf{P}(E|E)=1$ nell'ipotesi che E si verifichi e $\mathbf{P}(E|\tilde{E})=0$ nell'ipotesi che E non si verifichi. (Su questo punto si veda anche il § 3.9).

1.12. Il «**punto**», dopo le considerazioni introduttive.

Gli argomenti e le considerazioni finora svolti hanno (come già espresso nel titolo) carattere e scopo introduttivo, ma sotto una duplice visuale: l'una di chiarire alcuni aspetti generali del ragionamento probabilistico e del suo significato effettivo, e l'altra di precisarli (quanto piú elementarmente possibile, ma in modo concettualmente preciso) come preparazione alla trattazione matematica (e, necessariamente, piú organica) da svilupparsi a suo tempo.

È quindi opportuno, in questo momento, fare il «punto» della situazione cui si è giunti, riflettendo sinteticamente su ciò che è stato detto e delineando un abbozzo panoramico degli argomenti ed aspetti che andranno sviluppati in seguito. Naturalmente, gli sviluppi comporteranno in genere una trattazione in forma matematica, senza però appesantirla con tecnicismi; fatta – si potrebbe dire – per aiutare a comprendere il «succo», in forma matematica, anche a coloro che sono o si sentono «digiuni» in matematica ma non cadono nell'errore di rifiutare ogni aiuto per capire una spiegazione in forma idonea per chiunque abbia interesse ad afferrare il «succo» usualmente nascosto «sotto il velame delli *sgorbi* strani»: quegli «sgorbi» che sono, per lui, le formule e i simboli che vi compaiono.

In chiusura di questo primo paragrafo è necessario indicare alcuni simbolismi e forme di scrittura che occorreranno in seguito: aiutano alla concisione e al risparmio di spazio, e quindi alla chiarezza. (Purché uno si degni di abituarvisi: è un po' faticoso – specie per coloro che si sentono «profani» o «refrattari» alla matematica – ma vorrei dire loro, per incoraggiarli (ma *con convinzione*, non per illuderli o per ingraziarmeli), che non si tratta né di loro inettitudine né di indigeribilità della matematica, bensí di indigeribilità dell'insegnamento matematico formalistico-mnemonico-astratto nelle scuole; salvo, beninteso, parecchie lodevoli eccezioni).

Oggetto della teoria delle probabilità sono gli *eventi* e i numeri aleatori (potrebbero considerarsi anche punti aleatori, funzioni aleatorie, passeggiate aleatorie, processi aleatori, ecc.). Gli eventi aleatori si indicano in genere con E e indici (E_0, E_1, E_2, ...) oppure altre maiuscole (A, B, C, ...); i numeri aleatori con maiuscole a fine alfabeto (X, Y, Z, ... oppure X_0, X_1, X_2, ...)

Il simbolo $\mathbf{P}$ significa sia *probabilità* sia *previsione*: probabilità se riferito a un evento, ad esempio $\mathbf{P}(E)$; previsione se riferito ad un numero aleatorio, ad esempio $\mathbf{P}(X)$; il simbolo $\boldsymbol{\sigma}$ indica lo scarto quadratico medio (o «scarto standard»): $[\boldsymbol{\sigma}(X)]^2=\mathbf{P}(X-m)^2$ dove $m=\mathbf{P}(X)$.

Ogni evento, E, si identifica col numero aleatorio che vale 1 se E si verifica e 0 se non si verifica. Le operazioni aritmetiche hanno (naturalmente) il medesimo significato per numeri aleatori che nel caso abituale; interessa però aggiungerne altre: col segno $\sim$ («tilde»: il segno che in spagnolo si sovrappone alla lettera n (ñ) per farla pronunziare come «gn» in italiano; ad esempio «giugno») si indica il complemento ad 1: $\sim x=1-x$ (sovrapponendolo: $\tilde{x}$, quando si tratta di una sola lettera); in particolare, per un evento, E, $\sim E$ (o $\tilde{E}$) significa «negazione di E» (infatti, il segno «tilde» scambia vero con falso e viceversa).

Altre operazioni logiche (su eventi ma anche su numeri, aleatori o no) sono quelle di «sup» e «inf», indicate con $\vee$ e $\wedge$: $A \vee B$ indica il maggiore (e, analogamente, $A \wedge B$ il minore) tra i numeri A e B (e lo stesso per piú termini: ad esempio $A \wedge B \wedge C$ ed $A \vee B \vee C$ significano rispettivamente che *tutti* i tre eventi sono veri, o che lo è almeno uno). Per dare un esempio un pochino piú complesso, $(A \vee B) \wedge (C \vee D)$ significa che c'è almeno un evento vero in entrambe le parentesi.

Per dare un esempio relativo a numeri aleatori, basta pensare che gli A, B, C, D del caso precedente siano numeri qualunque (non piú solo 0 o 1): in tal caso il significato è «il minore tra i massimi di ciascuna coppia».

A seconda delle sue preferenze, il lettore potrà cercar di assimilare fin dall'inizio tali concetti e simbolismi, oppure ricordare che può ricorrere a queste pagine ogni qual volta abbia bisogno di decifrare un caso singolo o di rinfrescarsi le idee.

2. *Molteplicità anche di concezioni.*

2.1. Un preambolo pirandelliano.

Parafrasando un brano di Pirandello nel romanzo *Uno, nessuno, centomila* («parafrasandolo» col sostituire «probabilità» a «realtà» e «sento» a «mi do»), il discorso potrebbe iniziare cosí: «Ci fosse fuori di noi, per voi e per me, ci fosse una signora probabilità mia e una signora probabilità vostra, dico per se stesse, e uguali, immutabili. Non c'è. C'è in me e per me una probabilità mia: quella che io sento, e una probabilità vostra in voi: quella che voi sentite; le quali non saranno mai le stesse, né per voi né per me».

Sarebbe stato impossibile, senza l'aiuto di Pirandello, esprimere questo concetto (e, in nuce, l'essenza della nostra tesi) in un modo cosí preciso, completo, efficace; rimane però da chiarire la specifica interpretazione – anzi, le due *opposte* interpretazioni – in cui potrebbe sembrare appropriato intenderlo nel presente contesto.

Questa citazione pirandelliana si presta infatti – nel tentativo qui presentato di suo adattamento in campo probabilistico – a due diverse interpretazioni, esprimenti i due aspetti complementari delle tesi qui contrapposte; quella *soggettivista* dove si ha unicità d'interpretazione e molteplicità di valutazioni, e quella *oggettivista* dove si ha una molteplicità d'interpretazioni ciascuna delle quali si traduce nell'unicità (o pretesa unicità?) della corrispondente valutazione.

Per chiarire un po' meglio fin d'ora le posizioni contrapposte dei soggettivisti e degli oggettivisti si aggiungono le precisazioni che seguono:

Nel campo dei *soggettivisti* si ha un'*unica* concezione ammissibile basata soltanto sul requisito della *coerenza*, e dove la definizione in senso operativo della probabilità si traduce nella «regola di Brier» (o simili; cfr. § 1.6).

Ma è proprio nell'ambito di tale concezione che l'illimitata molteplicità delle valutazioni di probabilità ammissibili (conformi all'opinione di ciascuno: «Ciascuno a suo modo») si presenta come cosa naturale e *necessaria*. Necessaria in dipendenza del fatto stesso che ogni valutazione (ciascuna delle «una-nessuna-centomila») è, per definizione: *soggettiva*, nel senso che riflette *non* circostanze oggettive, oggettive di per sé, bensí l'opinione che se n'è fatto, sia pure *in base ad esse*, l'individuo che le valuta, e *specifica*, nel senso di riferirsi specificamente alla probabilità di un «evento», inteso sempre come «caso singolo univocamente individuato» nelle date circostanze (e *non* al modo degli oggettivisti che usano «evento» in senso generico e chiamano «prove di tale evento» tutti gli eventi di quel certo tipo).

Nel campo degli *oggettivisti* si ha invece – secondo il gusto di ciascuno: «Ciascuno a suo modo» – una fungaia di (una? nessuna? centomila?) «definizioni» (piú o meno cervellotiche, e che piú appropriatamente, come si vedrà, dovrebbero dirsi «pseudodefinizioni»), le quali – almeno nelle pie intenzioni dei loro fautori – dovrebbero conferire, *motu proprio eorum*, alla probabilità di ogni evento («evento» da interpretarsi – quel che è peggio – come un ammasso incomprensibile e stravagantemente «collettivistico»?!) il diritto a fregiarsi del titolo onorifico di «oggettive». È inutile dire quale assurda confusione ciò possa ingenerare; la miglior prova è data dagli stessi oggettivisti che distinguono – palesemente contraddicendosi! – il caso in cui «tutte le prove» siano «ugualmente probabili» e il caso in cui la probabilità «varia di prova in prova». Accettando sul serio tale formulazione, sarebbe naturale concludere che ogni evento ha probabilità o uno o zero a seconda che si verifichi oppure non si verifichi!

Poiché tentativi prematuri di spiegazioni e chiarimenti riguardo alle molteplici diatribe sul significato delle probabilità (diatribe che si riducono di regola a «dialoghi tra sordi») riuscirebbero oscuri e finirebbero per confondere ancor piú le idee anziché facilitarne la comprensione, sembra consigliabile seguire una via di mezzo: *dapprima* (nel seguito di questo § 2) prospettare il senso delle diverse concezioni e discutere l'appropriatezza di diverse terminologie e notazioni, aggiungere qualche cenno storico al riguardo, discuterne la validità (se esiste, ed entro quali limiti), ma sempre a scopo di preliminare orientamento in vista della trattazione piú approfondita e precisa, che *poi* (nel § 3, ed ultimo) sarà sviluppata un po' piú col necessario rigore (anche matematico). La lettura e comprensione dovrebbe tuttavia risultare facilitata anche ai lettori non troppo agguerriti in fatto di conoscenze matematiche, dato che gli sviluppi di formule e i risultati matematici, nella maggior parte, non saranno che la traduzione in termini precisi di quanto sarà già stato fatto intravedere da varie osservazioni critiche che verranno sviluppate nel seguito del presente secondo paragrafo. E, naturalmente, sarà considerato acquisito quanto premesso nel § 1.

In particolare, e soprattutto, si tenga sempre presente la «regola di Brier» (cfr. § 1.4), che sarà sempre considerata come lo strumento-base per la misura (e, sostanzialmente, per la definizione operativa) della probabilità. Si rammenti, tuttavia, che essa è equivalente a quella banale ($\mathbf{P}(E)$ è il valore di «uno» – una Lira, oppure un Dollaro, se si vuole dare un nome all'unità – da ricevere se E è vero), salvo la situazione di «gioco» («io credo ch'ei credesse ch'io credessi») che potrebbe falsare la decisione e che grazie alle «regole di penalizzazione appropriate» viene eliminata.

2.2. «Eventi»: ambiguità da eliminare.

Per evitare di discutere di probabilità *nel vuoto* o nell'ambiguo (come purtroppo può capitare e spesso capita) è certo opportuno – e direi addirittura (a mio avviso) necessario – introdurre subito alcune precisazioni terminologiche (almeno in parte nuove). Si tratta anzitutto di stabilire in senso univoco il significato di 'evento' e di introdurre i due termini 'fatto' e 'fenomeno' da sostituirsi ad 'evento' nei sensi in cui, per evitare equivoci, non dovrebbe mai piú essere usato.

Il termine 'evento' dovrebbe, a tal fine, venire riservato al senso di «caso *unico* perfettamente *specificato* (in anticipo)»: ad esempio, il pareggio in una ben precisata partita di calcio; l'aumento della percentuale di voti per una data lista nelle prossime elezioni in confronto alle precedenti; la cattura (entro un preciso limite di tempo) di un dato criminale ora evaso, ecc.

Al contrario, indicazioni *generiche* come «la cattura di *un* evaso», «il pareggio in una (non specificata) partita di calcio», «un forte acquazzone», mancano dei requisiti necessari per consentire una risposta univoca e certa, «Sí» o «No», e pertanto *non* costituiscono «eventi» nel senso precisato, ma soltanto «fatti».

Per dare un chiarimento concreto e completo su di un esempio: il fatto che cada (o che sia caduto) un fulmine è un *fatto* (che può «accadere»); il fatto che esso colpisca o abbia colpito un dato edificio causando danni coperti da assicurazione (secondo le condizioni di polizza vigenti e le clausole convenute) è un *evento* (che può «verificarsi»); la «caduta di un fulmine» (in senso generico: dove e quando che sia) è un *fenomeno* (che può «ripetersi», sempre e dovunque).

Piú radicalmente ancora, va scartato l'uso del termine 'evento' in senso generico, conformemente alla locuzione confusionaria disgraziatamente invalsa di «prove di un evento» per dire «eventi» (in genere, piú o meno analoghi) e, quel che è peggio, «probabilità di un evento» (!) per probabilità di ciascuno di tali eventi, detti (in virtú di tale loro intruppamento) «prove di quello stesso evento». Gli stessi oggettivisti, però, si smentiscono, in quanto parlano anche del caso in cui «la probabilità... varia di prova in prova», in contrasto coll'«assioma» che la probabilità debba riguardare, per aver senso, senso *collettivo*, un gran numero (o, secondo i piú raffinati, una successione *infinita* (!)) di «prove».

È penoso, ma doveroso, segnalare queste palesi assurdità (non si potrebbe spiegare, salvo per l'assuefazione, come esse possano non apparire palesi anche al piú sprovveduto mortale che vi ponga il minimo di attenzione!)

L'opportunità di queste precisazioni terminologiche è scaturita da discussioni critiche su tali problemi durante un corso di lezioni all'Istituto nazionale di alta matematica (Roma, marzo-maggio 1979).

2.3. Certe «definizioni».

Anziché di «definizioni» della probabilità sarebbe piú appropriato parlare di «pseudodefinizioni»; esse non sono in genere che dei conati di definizione: conati somiglianti a quelli di chi volesse sollevarsi da terra tirando verso l'alto i lacci delle proprie scarpe. Un'altra immagine – altrettanto bella e del tutto diversa, dovuta all'indimenticabile Leonard Jimmie Savage – ribadisce e arricchisce il medesimo concetto dicendo che «è impossibile fare una omelette probabilistica senza spezzare uova probabilistiche». Fuori di metafora (e si potrà notare e apprezzare sempre piú quanto dette metafore siano appropriate!) è insensato cercar di foggiare una definizione usando il medesimo termine che si vuol definire, o altri che lo presuppongono. Ed è proprio questo, invece, il sistema cui tentano di aggrapparsi gli aspiranti ideatori e gli illusi scopritori della «probabilità oggettiva».

Le «definizioni» correnti si basano poi entrambe sul circolo vizioso di supporre già noto il significato di probabilità, almeno nel senso di saper distinguere se certi dati eventi *sono* o *non sono* «ugualmente probabili». È necessario infatti – per la «definizione classica» – «avere una partizione in n risultati *ugualmente probabili* e sapere che m sono favorevoli a un dato evento E» per dire che $\mathbf{P}(E) = m/n$, oppure – per la «definizione frequentista» – che «su n "prove" se ne sono verificate m» per dire che la frequenza è stata m/n. (Ma in quel singolo caso, e quindi «per caso». Di per sé, il fatto di stimare la loro probabilità in «circa m/n» è frutto di un'illazione infondata, a meno che non si tratti di molti eventi giudicati «scambiabili»: cfr. cenno nel § 2.7 e sviluppi nel § 3).

Come conclusione: in entrambi i casi la frazione m/n può essere una scelta piú o meno ragionevole ma non obbligatoria; è bene riflettere caso per caso senza elevare a teoremi o a dogmi delle semplici norme di buon senso affinate con la familiarità a ragionare sull'incertezza.

Non è detto, naturalmente, che parlando di scommesse si debba pensare ad esempi nel senso piú deplorevole e dannoso del termine (lotto, lotterie, giochi d'azzardo con carte o dadi o roulette, ecc.); fortunatamente rientrano nello schema anche operazioni formalmente analoghe ma di motivazione e direzione opposta, come il risparmio che dà una protezione generale contro ogni rischio imprevisto o piú o meno genericamente prevedibile, e come, piú specificamente, le assicurazioni che coprono ogni genere di rischi, incoraggiando la preveggenza anziché l'incoscienza, l'estraneità anziché la soggezione a certe manie sciocche e difficilmente curabili per chi ne è vittima come il fumo, l'alcool, la droga. Tutte cose, tra l'altro, che, sia pure indirettamente, danneggiano purtroppo anche chi ne è immune ed estraneo.

E conviene rammentare, qui, che ogni valutazione di probabilità è sempre *subordinata*, o condizionata. Anziché $\mathbf{P}(E)$ si dovrebbe a rigore scrivere sempre $\mathbf{P}(E|H_0)$ o $\mathbf{P}(E|H_0H)$ per indicare che la valutazione è fatta nello stato di conoscenza H_0 e, rispettivamente, condizionatamente *anche* all'ipotesi H.

2.4. Certezza, incertezza, probabilità.

Sembra un dannato destino quello di molte scienze che vedono sopravvivere credenze da esse ridicolizzate e smentite dai fatti, e che trovano in tal modo inquinato il loro campo dal tentativo di intrusioni da parte di squallidi residui superstiziosi e cabalistici. Si pensi all'astronomia travisata in supporto di sproloqui astrologici! Eppure, è strano: non siamo forse ben lontani dal medioevo e prossimi al 2000?

Una delle credenze superstiziose piú pericolose e aberranti è quella che ammette ed afferma che «esistono» delle «probabilità oggettive». Tale binomio, oltre che a «quadrato circolare» come già fatto, potrebbe essere abbinato, per la sua contraddittorietà, a «ghiaccio bollente», o «luce nera», o «pioggia asciutta». È chiaro che la probabilità oggettiva (volendo dar senso, sia pure artificiosamente e provvisoriamente, a tale locuzione) non potrebbe essere se non il valore (logico) di *verità*, e cioè:

«1» («certezza», o «vero», o «certo») se l'evento si è verificato o si verificherà; ossia, sinteticamente, «Sí»;
«0» («impossibilità» o «falso») o, sinteticamente, «No» nel caso opposto;

al quale (nella logica dell'incertezza, propria della non-onniscienza umana) è necessario aggiungere

«?» («incertezza», o «dubbio», o «incerto») o, sinteticamente, «Non so», od anche, recitando il ben noto detto dannunziano, «Forse che Sí, Forse che No» (cfr. fig. 3).

La situazione cui si è pervenuti nel momento attuale – con tre livelli di conoscenza: «Sí», «No», «Non so» – è la situazione della «logica dell'incerto», con tre «valori di verità». Quello intermedio è quello dell'incertezza, ma essa va qui considerata come una situazione unica senza differenziazioni tra «il *piú* e il *meno* probabile».

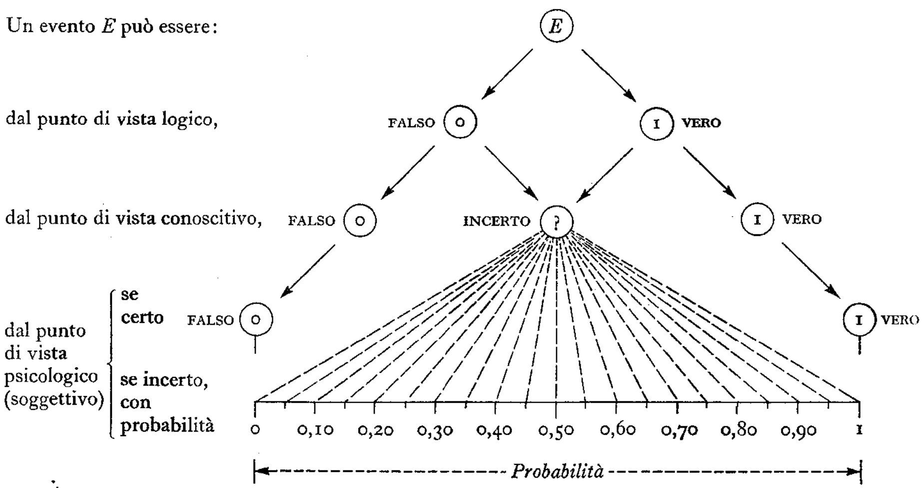


Figura 3.
I tre livelli di conoscenza di un evento.

Questa «logica a tre valori» («1», «0», «?»), che non mi consta sia mai stata teorizzata o utilizzata, può effettivamente risultare significativa per sviluppare espressioni logiche, o matematiche, o logiche e matematiche insieme. Tutto è analogo al «calcolo letterale» dell'algebra elementare, e si possono poi ottenere anche le corrispondenti espressioni in termini di *probabilità* e di *previsione* quando se ne facciano valutazioni ed elaborazioni probabilistiche. (Si tratterebbe del banale procedimento di sostituzione di indicazioni in lettere con le corrispondenti valutazioni numeriche come probabilità e previsioni). Questo non è che un cenno; ma basti darne qui una esemplificazione per rendere chiaro tutto il significato della distinzione tra livello «incertezza» e livello «probabilità».

Si sa che il numero $X = x_1E_1 + x_2E_2 + \ldots + x_nE_n$, supponendo che gli eventi E_h siano incompatibili ed esaustivi (in parole povere: che si debba verificare certamente uno e uno solo di essi), ci darà un guadagno x_1 se si verificherà l'evento E_1 (e cosí per tutti gli altri). Si tratta di un numero (qui, in particolare, di un guadagno) incerto. E non lo si dice «aleatorio»: è utile infatti (anzi, per la presente finalità, essenziale) distinguere nettamente «incertezza» da «probabilità»: si sarebbe detto «aleatorio» se si conoscesse (cioè si fossero stimate, ecc., non importa come) le probabilità degli eventi E_h: allora (ma solo allora) si potrebbe parlare della previsione, $\mathbf{P}(X)$, del guadagno aleatorio X (beninteso, soggettiva, come è soggettiva ogni probabilità).

Il caso piú generale, di n eventi E_h (non disgiunti; $h = 1, 2, \ldots, n$) si riconduce subito al caso precedente di una partizione in N ($N \leq 2^n$) eventi disgiunti, C_i, detti «costituenti». Essi sono ottenibili dal prodotto logico $E_1E_2 \ldots E_n$ cambiando in tutti i 2^n modi possibili alcuni degli E_i nella loro negazione $\tilde{E}_i$; non è detto però, naturalmente, che tutti i 2^n prodotti siano non vuoti; perciò essi sono *al piú* 2^n, non 2^n senz'altro.

2.5. Dall'incertezza alla probabilità.

Le considerazioni che precedono avevano uno scopo assai semplice e meramente preparatorio: intendevano mostrare fin dove si poteva portare avanti il discorso e la trattazione matematica restando nel campo dell'incerto, per poi passare direttamente dal campo della semplice *incertezza* a quello in cui l'incertezza, venendo tradotta in *probabilità* mediante stime dirette o indirette o mediante calcoli piú o meno complessi su di esse basati, fornisce gli elementi necessari e desiderati come base per prendere le decisioni nel modo piú ragionevole e vantaggioso.

La conclusione è ora semplicissima. Si tratta soltanto di sostituire a tutti gli eventi (siano ad esempio $E_1, E_2, E_3, \ldots, E_n$) le rispettive probabilità, ossia applicare l'operatore lineare $\mathbf{P}$ («probabilità» di) ottenendo $\mathbf{P}(E_1)$, $\mathbf{P}(E_2)$, $\mathbf{P}(E_3)$, ..., $\mathbf{P}(E_n)$. Analogamente, applicando $\mathbf{P}$ al numero aleatorio $X = x_1E_1 + x_2E_2 + \ldots + x_nE_n$, si ha per la *previsione* $\mathbf{P}(X) = \mathbf{P}(x_1E_1) + \ldots + \mathbf{P}(x_nE_n) = x_1\mathbf{P}(E_1) + \ldots + x_n\mathbf{P}(E_n) = p_1x_1 + \ldots + p_nx_n$ (ove si pone $\mathbf{P}(E_i) = p_i$).

Quanto al modo di stimare tali probabilità nulla c'è da aggiungere a quanto detto in generale, salvo far presente una circostanza utile come controprova. La somma delle n probabilità deve ovviamente dare 1 ma in pratica la somma delle probabilità stimate se ne discosterà piú o meno. Se la differenza è piccola basta alterare proporzionalmente (moltiplicandole per il necessario coefficiente, poco inferiore o poco superiore ad 1); altrimenti conviene ripensare alle singole probabilità nel dubbio che una (o piú) presenti una deviazione macroscopica.

2.6. Le «pseudodefinizioni» tuttora in voga.

È necessario parlare anche delle «pseudodefinizioni» della probabilità, sia perché sono quelle usualmente (purtroppo!) presentate tuttora come autentiche «definizioni (!)», e sia perché – considerandole non nella mentita veste di definizioni bensí come criteri ausiliari per la valutazione di probabilità in certi tipi di circostanze – possono costituire spesso un valido punto d'appoggio.

Beninteso, come definizione *vera* si considera sempre quella diretta: $\mathbf{P}(E)$ è il «prezzo equo» per una scommessa che faccia vincere l'importo «uno» (una Lira, un Dollaro, quel che altro si voglia) se l'evento E si verifica. Beninteso, «equo» secondo la valutazione dell'interessato. Tuttavia, per eliminare il carattere di «gioco», di «rischio», di «azzardo», conviene fare una di quelle «scommesse col morto» considerate fin dal § 1.5 e chiamate «regole di penalizzazione appropriate» (*proper scoring rules*). Si rammenta che, all'opposto delle usuali scommesse, questa specie di «scommessa col morto» tende a rendere minimo il rischio anziché a crearlo. Si tenga presente, senza ripeterne qui la spiegazione, la «regola di Brier», particolarmente elementare e pertanto chiarificatrice (anche grazie alla sua interpretazione meccanica).

Le altre cosiddette «definizioni» tuttora imperversanti non dovrebbero in alcun modo venir chiamate «definizioni»; escludendo di considerarle tali possono però talvolta, se intese ed usate con discernimento, costituire criteri utili per agevolarci o guidarci, in particolari circostanze, alla valutazione soggettiva delle probabilità.

Quella che pretende di «definire» la probabilità come rapporto, $p = m/n$, tra il numero dei «casi favorevoli» e dei «casi possibili» *supposti... ugualmente probabili* esprime una proprietà esatta ma pressoché tautologica; non è comunque una definizione perché presuppone di saper già cosa significhi «ugualmente probabili». Altrettanto poco «definizione» sarebbe quella di «peso», «volume» «carica elettrica», ecc., che riconducesse ad una altrettanto non definita nozione di «uguaglianza di peso», «di volume», ecc.; ciò fisserebbe la *scala* (lineare, non logaritmica, non ..., ecc.) ma non permetterebbe di distinguere quale sia la grandezza che viene chiamata «peso» e quale venga chiamata «volume», ecc.: eppure è proprio su questa distinzione sostanziale che si sorvola (come se «tutto ciò che l'Autore tace non sapendolo spiegare» potesse, per chissà quale miracolo, riuscire chiaro al Lettore!)

Ma al peggio non si è ancora arrivati: non si è ancora arrivati al gradino peggiore di confusionismo che è quello in cui si cerca di identificare, o almeno assimilare tra loro, due nozioni che richiedono assolutamente, per essere comprese in modo corretto, di venir considerate in certo senso antitetiche e tuttavia legate

da molteplici e reciproci influssi e rapporti: la probabilità e la frequenza. Il peggio è la concezione *frequentista*, che, come l'idra dalle sette teste, presenta sempre nuove varianti di conati di «definizioni frequentistiche»: a mano a mano che quelli precedenti vengono rintuzzati, ecco pullularne versioni sempre piú artificiose e infelici.

Sembrerebbe di dover essere giunti, in questo modo, ad un limite invalicabile, ma l'esperienza in tal campo e un detto popolare romano inducono a prudenza. Secondo tale detto, «il peggio non è mai morto»! E infatti, per liberarsi dall'indeterminatezza della frequenza, la si sostituisce con l'irraggiungibile «frequenza-limite» (conoscibile... dopo la fine dell'eternità!) Comunque, il discorso si disperde, inevitabilmente, in mille rivoli.

2.7. Fraintendimenti: guardarsene!

Tentar di passare sistematicamente in rassegna questi «mille rivoli» sarebbe fatica improba e inutile. Basti soffermarsi a titolo esemplificativo su qualche fraintendimento in cui è facile cadere o che può lasciarci confusi.

I malintesi si possono ricondurre, sempre o quasi sempre, alla tendenza ad interpretare in senso oggettivistico delle considerazioni che sono valide solo in senso soggettivo, o a travisare in senso assoluto dei ragionamenti che sono validi solo in senso relativo.

Esempio tipico del primo malinteso è il confondere «probabilità zero» con «impossibilità» (sarebbe come dire che un insieme di misura nulla, ad esempio un punto o una linea su un piano, *è* l'insieme vuoto!)

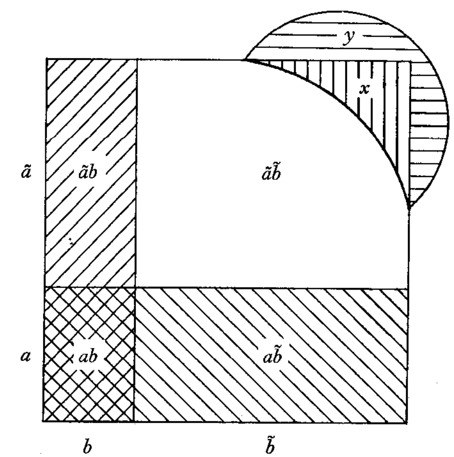


Figura 4.

La figura (quadrato di lato 1 diviso in quattro rettangoli) indica le probabilità dei quattro casi, AB, $A\tilde{B}$, $\tilde{A}B$, $\tilde{A}\tilde{B}$, essendo A e B indipendenti. Se però, fermi restando i rettangoli tratteggiati, si altera (allargandolo o restringendolo) il rettangolo bianco $\tilde{a}\tilde{b}$ (che diverrebbe $\tilde{a}\tilde{b}-x$ risp. $\tilde{a}\tilde{b}+y$) l'indipendenza stocastica non sarebbe piú rispettata.

Il secondo malinteso ha luogo, in molti casi, per il fatto di non tener conto della dipendenza di ogni conclusione dallo stato di conoscenza in cui ci si trova. Per dirlo con la notazione già usata, si tratta di pensare all'evento E «nel vuoto» anziché in rapporto al nostro presente stato d'informazione, H_0; di per sé – lo si rammenti – $\mathbf{P}(E)$ non ha senso a meno che si sottintenda H_0, pensando che s'intenda scritto $\mathbf{P}(E|H_0)$.

Ed è chiaro che se, al posto di H_0, si avranno altre «ipotesi» H_0', H_0'', ... diverse, non solo cambieranno valore le probabilità $\mathbf{P}(E)$ ma anche, in genere, conseguentemente, le relazioni tra esse. Per illustrare tale fatto si veda l'esempio rappresentato nella figura 4, osservando dapprima il quadrato (di lato $=1$) diviso nei quattro rettangoli ab, $a\tilde{b}$, $\tilde{a}b$, $\tilde{a}\tilde{b}$ (dove, si ricordi, la tilde indica negazione, o, aritmeticamente, complemento ad 1 : $\tilde{a}=1-a$, $\tilde{b}=1-b$), e poi lo stesso quadrato privato del pezzetto x, o invece accresciuto del pezzetto y, in seno al quale a e b risultano *correlati* (risp. positivamente e negativamente).

Ciò mostra che, a rigore, si dovrebbe sempre specificare rispetto a quale *stato di conoscenza* (H_0; H_0H; ...) l'indipendenza stocastica viene affermata: non esiste l'«indipendenza in sé» (come la «cosa in sé» di certi filosofi).

Analogamente, l'indipendenza stocastica può anche sussistere, anziché *tout court*, soltanto «subordinatamente a una data ipotesi» oppure «subordinatamente a ciascuna di certe ipotesi incompatibili»: è il caso che dà luogo, ad esempio, alla «scambiabilità» che verrà presentata – data la necessità di piú ampie premesse e di strumenti e ragionamenti alquanto piú delicati e complessi – nel § 3.

Il peggiore fraintendimento (e, sembra, il piú radicato, tanto che chi ne è immune rischia di venir considerato un idiota o uno squilibrato,... e di ricevere lettere di insulti!) è però quello che induce molti «competenti» o «intenditori» a ritenere molto probabile, al Lotto, l'estrazione di un numero ritardato (cioè che non è uscito da molte settimane su una data ruota, o, occasione ancor piú ghiotta per gli «intenditori», su nessuna ruota!) È inutile dire a tali «intenditori» che i numeri «non hanno memoria» e che non c'è quindi alcun motivo di pensare che l'essere stati estratti poche o molte volte piú o meno recentemente non sia, come lo è, un fatto passato privo di qualsiasi influenza sulle circostanze in cui l'estrazione avrà luogo. E «acqua passata non macina piú».

2.8. Probabilità e frequenza.

Fra i molti equivoci e le molte distorsioni d'interpretazione che riducono spesso le discussioni sulla probabilità a «dialoghi tra sordi», primeggiano indubbiamente quelle che concernono le relazioni tra *probabilità* e *frequenza*. Vi sono addirittura delle scuole che pretenderebbero di identificarle (!), di considerare i due termini come sinonimi, come inutili doppioni l'uno dell'altro.

In tal senso sono orientati specialmente molti statistici, ma anche, sia pure con svariate sfumature e «abbellimenti», molti autori di estrazione filosofica o propensi a filosofanteggiare. Lo scopo, ambito a prezzo di qualsiasi distorsione, è quello di *negare* (in sostanza) l'autentica nozione di probabilità – quella *soggettiva*, quella «naturale» del non abbastanza fuorviato e catechizzato «uomo

della strada» – per sostituirla con incredibili frutti di macchinose elucubrazioni.

'Probabilità' e 'frequenza' – secondo certe vedute tuttora abbastanza in voga, specie negli ambienti degli statistici – sono sinonimi o quasi, piú o meno intercambiabili, come gemelli identici o forse come un individuo e la sua ombra (per ricollegarsi a una fantasia di Bontempelli).

Solo quest'ultima immagine, però, risulta concettualmente appropriata, nel senso che la statistica rileva oggettivamente i fatti (li enumera, li classifica, li elabora,...) traendone indicazioni significative e certe; mentre la probabilità (*ars conjectandi* 'arte di congetturare', come la chiamò il suo stesso fondatore, Giacomo Bernoulli) fornisce a ciascuno il modo di esprimere il proprio grado di fiducia nelle varie ipotesi di cui si interessa.

Le nostre piú o meno istintive valutazioni di probabilità dipendono da una sintesi di esperienze favorevoli e sfavorevoli, vissute o sentite raccontare, di tipo piú o meno affine ai casi di cui attualmente ci si preoccupa. In casi schematici e ripetitivi è naturale pensare che le cose andranno in futuro piú o meno conformemente alle esperienze del passato, a meno di non ritenere che ci siano motivi di prevedere piú o meno sensibili miglioramenti o peggioramenti. Ciò significherebbe rispettivamente aumento o diminuzione della probabilità per fatti desiderabili o viceversa; ecc.

È bene, comunque, cercare di schematizzare un po' le varie situazioni.

Si pensi anzitutto a un certo numero di eventi qualunque, E_1, E_2, ..., E_n; beninteso, eventi aleatori. La loro somma, $S_n=E_1+E_2$, ..., E_n, è il numero dei successi; un numero per noi incognito, quindi aleatorio. Si può però darne la *previsione*, $\mathbf{P}(S_n)=\mathbf{P}(E_1)+\mathbf{P}(E_2)+...+\mathbf{P}(E_n)$, come somma delle probabilità degli E_h (l'additività vale comunque, siano gli E_h compatibili o incompatibili, logicamente e/o stocasticamente dipendenti o indipendenti tra loro).

Piú in generale, ciò vale anche per ogni numero aleatorio $X=x_1E_1+x_2E_2+...+x_nE_n$ (ci si riferisce qui al caso piú semplice di numeri aleatori con un numero finito di valori possibili: x_1, x_2, ..., x_n con probabilità $p_1=\mathbf{P}(E_1)$, $p_2=\mathbf{P}(E_2)$, ..., $p_n=\mathbf{P}(E_n)$, naturalmente, di somma $=1$). Concettualmente ciò vale anche se la probabilità (e, se ciò giova a dare un'immagine piú intuitiva, la si pensi come massa) è distribuita, anziché su un numero finito di punti, su un'infinità o addirittura con continuità su tutto l'asse da $-\infty$ a $+\infty$ o su una parte qualsiasi di esso.

Ma, per ora, interessa soltanto la previsione della frequenza, $\mathbf{P}(S_n)$, facendo notare anche che $\mathbf{P}(S_n/n)=\mathbf{P}(S_n)/n$ è la probabilità media degli n eventi considerati. D'altra parte, se si indica con F_0, F_1, F_2, ..., F_n gli eventi consistenti nell'aver nessun successo, o uno, o due, ecc., fino a tutti n, si potrebbe esprimere $\mathbf{P}(S_n)$ con una diversa espressione, e cioè $\mathbf{P}(S_n)=\mathbf{P}(F_1)+2\mathbf{P}(F_2)+3\mathbf{P}(F_3)+...+n\mathbf{P}(F_n)$. Nel caso (o nell'ipotesi) che gli n eventi E_h ($h=1, 2, ..., n$) siano ugualmente probabili e indipendenti, la probabilità $\omega_h^{(n)}$ ($h=0, 1, ..., n$) di h successi su n prove vale $\binom{n}{h}p^h q^{n-h}$ (ove $q=1-p$ è la probabilità di insuccesso (in una qualunque prova); si tratta (come è chiaro) dei termini dello sviluppo della potenza n-esima del binomio $(p+q)$; la loro somma è ovviamente $=1$).

Questo, però, non è che un caso molto particolare. Un caso abbastanza piú generale (benché, in certo senso, differisca poco dal precedente) è quello della *scambiabilità* (che sarà sviluppato nel § 3 con l'attenzione che richiede: si tratta del caso che, con locuzione impropria e contraddittoria, viene comunemente descritto come di «indipendenza con probabilità costante ma incognita»).

Nel presente contesto è però necessario insistere, invece, su ciò che ha validità generale, indipendentemente dalle ipotesi piú comuni che sono indubbiamente idonee in molti casi, ma che sarebbe un errore (o un travisamento? o una superstizione?) considerare come norme valide *in generale* salvo quelle «eccezioni che (con detto comicamente ineffabile) confermano la regola»!

Ma non basta. Occorre sottolineare come sia ingannevole l'idea grossolana, conforme però a certe vedute tuttora abbastanza diffuse (specie nel campo degli statistici), secondo le quali probabilità e frequenza sarebbero da considerarsi sinonimi, o quasi.

È importante, data tale situazione, chiarire e precisare rigorosamente fino a che punto il significato di 'probabilità' e quello di 'frequenza' concordino, e indicare da quale punto e in quale senso le due nozioni risultino contrapposte.

Se si considerano n eventi $E_1, E_2, ..., E_n$, la loro somma $X=E_1+E_2+...+E_n$ è il numero m dei successi (degli E_h che si verificano, cioè che sono $=1$), ed X/n ne è la *frequenza* (o m/n, come è piú usuale, dato che si pensa già noto il valore m di X). Ma, prima di conoscerne l'esito, cosa è possibile dire di X? Non certo il valore effettivo; se ne può però indicare la *previsione*, $\mathbf{P}(X)=\mathbf{P}(E_1+E_2+ +...+E_n)=$ somma delle probabilità.

Per esprimersi in forma ragioneristica, rendendo piú «palpabile» il significato, basterà dire, concludendo, che la differenza tra probabilità e frequenza, o tra previsione e realizzazione, consiste nella necessaria distinzione tra valutazione *preventiva* (necessariamente piú o meno incerta e soggettiva) e valutazione *consuntiva* (ovviamente certa e oggettiva).

2.9. Valutazioni condizionate.

Vari accorgimenti possono essere spesso d'aiuto per valutare accuratamente la probabilità da attribuire a un dato evento E. Come primo esempio, può a volte aiutare il fare piú valutazioni condizionate a diverse ipotesi, siano $H_1, H_2, ..., H_s$, incompatibili ed esaustive (cioè tra le quali una e una sola risulterà essere quella vera). Se uno attribuisce alle s ipotesi le probabilità $q_i=\mathbf{P}(H_i)$, risulterà che, conseguentemente, egli dovrà valutare $p=\mathbf{P}(E)=\sum_i p_i q_i$. In particolare, se si tratta di distinguere solo due ipotesi, H e $\tilde{H}$, si avrà $p=p'\,q+p''\tilde{q}$.

Analogamente, nel valutare la probabilità di un evento E che sia il prodotto di due o piú altri eventi $E=E_1E_2$ (o $E=E_1E_2...E_n$), conviene confrontare la valutazione soggettiva diretta di $\mathbf{P}(E)$ con quelle indirette come $\mathbf{P}(E)=\mathbf{P}(E_1)\mathbf{P}(E_2|E_1)$ (o, rispettivamente $\mathbf{P}(E)=\mathbf{P}(E_1)\cdot\mathbf{P}(E_2|E_1)\cdot\mathbf{P}(E_3|E_1E_2)...\mathbf{P}(E_n|E_1E_2...E_{n-1})$, anche cambiando comunque l'ordine degli n eventi). Ci saranno delle discordanze (la coerenza in casi complessi non è visibile di primo acchito) e si dovrà vagliare quali, fra i ritocchi atti a ristabilire la coerenza, dànno luogo al complesso di valutazioni globalmente soddisfacente come espressione delle proprie opinioni.

Un caso analogo, ma molto piú semplice, è quello in cui si abbiano da valutare le probabilità degli n eventi $E_1, E_2, ..., E_n$ costituenti una *partizione* ($E_1+ +E_2+...+E_n\equiv 1$; cioè, se ne deve verificare uno e uno solo). Evidentemente, anche la somma delle probabilità deve dare 1; ma sembra dia maggiore affidamento il procedere alla valutazione delle n probabilità $p_1=\mathbf{P}(E_1)$, $p_2=\mathbf{P}(E_2)$, ..., $p_n=\mathbf{P}(E_n)$ singolarmente (senza le probabili tentazioni di aggiustare via via gli addendi per arrivare ad 1); verificare poi quanto la loro somma si scosti in piú o in meno da 1, e in base a ciò ripensare quali valori sembri ragionevole ritoccare per eliminare tale differenza.

Tutto ciò, beninteso, non può rientrare in forma troppo «ufficiale» nei precetti della teoria delle probabilità secondo il concetto di coloro che ne fanno un'astrazione perfetta, immutabile, apodittica: ma è bene che sia cosí, altrimenti probabilità e teoria delle probabilità cesserebbero di essere creature vive e vitali riducendosi a spoglia imbalsamata o addirittura a nudo scheletro.

Oltre che come immagine descrittiva, il termine 'scheletro' è appropriato per sottolineare l'impostazione astratta e meramente «assiomatica» che molte scuole impongono alla teoria delle probabilità, rinunziando ad ogni scelta (buona o cattiva che sia) di una interpretazione da dare al termine 'probabilità'. Potrebbero abolirlo, e dire 'teoria della misura' (con un qualsiasi aggettivo di loro gradimento), e cesserebbe ogni rischio di confusione ed ogni motivo di recriminazione.

La teoria delle probabilità *si serve* della matematica, ma *non* in astratto, bensí per applicazioni concrete nei problemi di previsione, e non si basa su assiomi artificiosi bensí è essa stessa – come ben disse (salvo errore) Henri Poincaré – «il buon senso ridotto a calcolo».

2.10. Le «certezze» col «quasi».

Per finire (e proprio, purtroppo, anche nel senso dei «per finire» umoristici) occorre segnalare (ma – beninteso! – soltanto per guardarsene) le affermazioni in cui si parla di «certo» e di «impossibile»... col «quasi» (un «quasi» talora espresso ma spesso addirittura sottinteso).

È chiaro che questa voluta imprecisione, intesa a «dire e disdire» nonostante «la contradizion che no l'consente», è particolarmente esiziale se inquina fin dall'inizio il discorso dal quale si pretenderebbe di estrarre lo spunto per le *definizioni*: qualunque cosa si dica in seguito risulterà allora fatalmente ambigua e, a rigore, priva di senso. Eppure le piú tipiche (pseudo)-definizioni «oggettivistiche» della probabilità si sforzano per l'appunto di ridurre il senso di «probabilità» all'esistenza di certi comportamenti pretesamente obbligati dalle «leggi del caso» in (lunghe) successioni di «*prove*».

Il caso piú estremo è quello della cosiddetta «concezione statistica» in cui la probabilità viene addirittura «definita» (!) come la frequenza (cioè la percentuale di successi) su «un grande numero di prove».

La critica alla definizione frequentista (e la spiegazione del perché la si dovrebbe dire semmai «pseudodefinizione») sta nel fatto che tra probabilità e frequenza sussistono parecchi legami in entrambi i sensi che non consentono però di confondere le parti in certo senso *opposte* che giocano nei ragionamenti e nei fatti. In forma di scherzoso apologo si può paragonare il caso di probabilità e statistica a quello dei due gemelli identici di Plauto, che, col loro entrare e uscire dalla scena, provocavano continui equivoci.

Occorre ribadire (cfr. § 2.8) che le probabilità riguardano il «preventivo» di ciò che ci si aspetta accada (o sia accaduto ma non ci sia ancora noto), mentre le frequenze indicano ciò che realmente è accaduto. Nei casi in cui i dati non noti relativi (in genere) al futuro riguardano frequenze, la valutazione di probabilità riguarderà i valori piú o meno ritenuti ragionevoli da attendersi per esse; si tratterà di *previsione* di frequenze.

A tale riguardo, accade spesso che la previsione della frequenza sia ritenuta abbastanza «certa» o «buona», nel senso di attribuire piccola probabilità a scarti sensibili dal previsto.

Per limitarsi al caso piú banale, di Testa e Croce, se la moneta appare non deformata, è naturale pensare che lo scarto fra i risultati Testa o Croce sia piccolo (dell'ordine di grandezza della radice del numero dei colpi); ma non si deve attribuire ciò ad un meccanismo o magia che tende a correggere le deviazioni favorendo la faccia che è in minoranza. Il procedere a sempre piú numerosi colpi non ha alcuna tendenza alla compensazione: l'avvicinamento alla situazione «equa» avviene, ma non per compensazione bensí soltanto per «sommersione»: la differenza si *diluisce* e scompare per il prevalere dei risultati successivi.

La «tendenza» alla compensazione non ha nulla di intenzionale o guidato: su 10 colpi a Testa o Croce si può ottenere una qualsiasi delle $2^{10} = 1024$ successioni; non si deve pensare che una successione *data* con 5T e 5C abbia probabilità maggiore di quelle con tutte T o con tutte C; ma la probabilità di una successione *qualunque* con 5T e 5C ha probabilità $1/4$ (esattamente $252/1024$) perché 5T e 5C si possono ordinare in 252 modi differenti; la probabilità di 6T e 4C (o viceversa) è $210/1024$; quindi la probabilità di ottenere parità con al piú uno scarto di 1 è $672/1024$ (praticamente, $2/3$).

A parte l'utilità di spiegazioni esplicative su esempi semplici, è opportuno (come conclusione del presente § 2) sottolineare ancora – e magari «enfasizzare» (le «voci» riprese dai pavidi puristi sono spesso le piú efficaci) – la distinzione fra i termini che hanno significato oggettivo e quelli che hanno significato soggettivo.

La verità o falsità di un'affermazione, o evento, è un fatto oggettivo, la sua probabilità è un fatto soggettivo; lo stesso vale per il valore (vero) di un numero (numero aleatorio per chi non ne conosce il valore effettivo), e per la sua previsione che è soggettiva. Per due eventi, il fatto che siano logicamente compatibili (in base alle informazioni certe che qualcuno ne ha) è un fatto oggettivo; per chi non ne ha è soggettivo. La frequenza (in un certo gruppo di eventi) è un fatto oggettivo se essa è conosciuta con certezza; nel caso opposto, la sua previsione è soggettiva.

E, per finire, si aggiunga ancora il caso-limite (sia in senso matematico, sia in senso interpretativo): quello che – a volerlo prendere sul serio e non come una insensatezza – permetterebbe di conoscere le probabilità solo dopo la fine dei tempi. È il caso della cosiddetta «concezione statistica» in cui la probabilità viene addirittura «definita» (!) come la frequenza (cioè la percentuale di successi) su «un gran numero di prove», ed anzi, in una sua versione piú spinta, su... un'infinità di «prove» (probabilmente richiedenti di continuarle fino alla fine dell'eternità). Fra i piú impegnati sostenitori di tale concezione si possono segnalare il matematico Richard von Mises e il filosofo Hans Reichenbach; da menzionare, del primo, *Wahrscheinlichkeitsrechnung und ihre Anwendung* (1931), e, del secondo, *Wahrscheinlichkeitslehre* (1935).

3. *«Ab omni naevo vindicata»?*

3.1. Perché «complicare le cose semplici»?

Il proposito di presentare la probabilità e la teoria delle probabilità in forma «ab omni naevo vindicata» apparirà indubbiamente come uno dei piú ardui, tale è l'inestricabile connessione di vedute e concezioni e terminologie radicalmente disparate e spesso anche intrinsecamente inconsistenti che contraddistinguono il variegato campo dei cultori specifici o collaterali della teoria delle probabilità.

Non si tratta affatto, tuttavia, di difficoltà inerenti alla nozione di probabilità e alla teoria delle probabilità, bensí semplicemente dell'effetto di storture d'interpretazione, di ambiguità di concetto e di linguaggio che sono in voga fra i cultori di versioni artefatte della teoria, e, peggio ancora, delle molteplici superfetazioni che ne derivano. Mai come in questo campo riesce indispensabile l'assillo di cui parla Giovanni Papini quando dice, riferendosi all'amico Mario Calderoni, che «a lui premeva insegnare con quali cautele e quali accorgimenti si possa giungere a ottenere delle proposizioni che abbiano un senso» (*Stroncature*, n. 14). Nel campo della probabilità – finché perdurerà il confusionismo imperversante – sembrerebbe quasi utopistico riuscire a tanto.

Non però perché il ragionare in termini di probabilità sia qualcosa di difficile o astruso, bensí perché tale lo si rende sovrapponendo al significato intuitivo, limpido e genuino, di probabilità, delle deformazioni che lo rendono oscuro e vuoto. Accettando invece, secondo le spiegazioni e indicazioni già date, il *naturale* significato *soggettivo* della probabilità – liberata da contraffazioni e da insulsi camuffamenti oggettivistici – tutto diventa assolutamente chiaro, sia per chi lo voglia accettare e sia per chi (anche senza volerlo accettare) non disdegni di apprendere e comprendere cosa ciò comporti.

È necessario, a tal fine, porre attenzione alle precisazioni terminologiche occorrenti per eliminare e sostituire e correggere locuzioni improprie, confuse, fuorvianti, che risentono delle deviazioni «oggettivistiche» o pretesamente tali. Già gran parte delle osservazioni critiche presentate nei due precedenti paragrafi avevano espressamente l'intento di segnalare e far riconoscere le manchevolezze di fraseologie ambigue e devianti, inquinate di oggettivismo: oggettivismo che inevitabilmente dà luogo ad errori e nonsensi impossibili da correggere con semplici ritocchi.

Tali nonsensi, infatti, come scrisse molto efficacemente e spiritosamente Bernard O. Koopman, «al contrario della Guardia di Napoleone si possono sempre far arretrare, ma mai arrendersi e scomparire».

3.2. Delle impostazioni assiomatiche.

In realtà, di impostazioni «assiomatiche» ne esistono molte e molto diverse, ma la distinzione preliminare e radicale è la fondamentale dicotomia fra le due concezioni in cui la probabilità si riferisce *a*) a un *evento* (nell'accezione qui fissata di «caso singolo univocamente specificato»), e si potrebbe dirla concezione *chiara*; o invece *b*) a una *collettività* di eventi *in un qualche senso* «analoghi» (che, nel gergo oggettivistico, si dicono «prove di uno stesso evento» e che, per evitare ambiguità, si potrebbero chiamare – come qui si propone – «prove di uno stesso fenomeno»). A volte si pensa a collettività numerose ma finite, ma a volte qualcuno pensa addirittura a successioni infinite; comunque tutte queste sovrastrutture non giovano che a «complicare le cose semplici», a recarsi dalla località A alla vicina B, non direttamente, bensí percorrendo tutto un cerchio massimo intorno alla Terra tranne il tratto AB. Non sembra eccessiva cattiveria battezzare tale concezione come *confusionaria*. D'altronde, a chiunque s'interessi a un qualche fatto, o *evento*, premerà valutare la probabilità di *quell'evento* (*Hic Rhodus, hic salta!*) e non avrà scopo, in genere, curarsi di altri eventi, piú o meno analoghi, magari (secondo una fraseologia corrente) «prove dello stesso *evento*» (che andrebbe semmai corretta – lo si ribadisce! – in «eventi» che sono «prove di uno

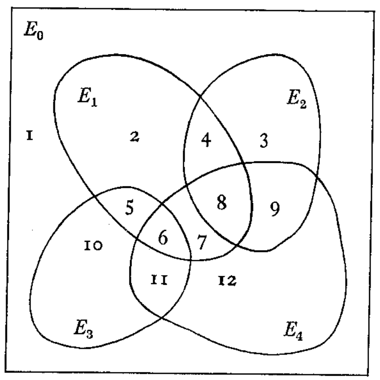


Figura 5.

Il quadrato rappresenta schematicamente tutti i «casi possibili» (punti); le quattro «patate» delimitano gli eventi E_1, E_2, E_3, E_4 e le loro dodici intersezioni (a due a due e a tre a tre): E_0 è la parte del quadrato esterna a tutte le «patate», e indica l'evento «Nessuno degli E_h ($h = 1, 2, 3, 4$) si verifica».

stesso fenomeno»). Può darsi, naturalmente, che questi eventi (o «prove»), specie se in qualche senso «analoghi», vengano considerati ugualmente probabili, o anche (stocasticamente) indipendenti, ecc. (Ma ciò va detto: non è da considerarsi sempre tacitamente affermato se non è esplicitamente negato).

La dicotomia fra concezioni *chiare* e concezioni *confusionarie*, come quella fra *soggettivisti* e *oggettivisti*, non esauriscono la sia pur sommaria rassegna che occorre per dare una efficace anche se pallida idea dell'insieme.

Si può anzi cominciare dagli *astrattisti*: coloro che si occupano sostanzialmente di «teoria della misura» in spazi astratti qualunque, e chiamano «eventi» dei sottoinsiemi e «probabilità» una loro «misura» (additiva, magari completamente additiva) con misura $= 1$ per tutto lo spazio. È questo l'esempio piú spinto di quel modo di vedere dei formalisti che vantano la matematica come quella scienza in cui «non si sa di cosa si parla né se ciò che si dice è vero o falso». Si veda la figura 5 con quattro eventi rappresentati da «patate» e che, con le loro intersezioni, dànno luogo a una partizione in dodici eventi.

Tutto bene; ma se si cura l'aspetto formale senza badare soprattutto al significato pratico, concreto (e in questo caso «concreto» significa *soggettivo autentico* e non *oggettivo fasullo*!), non è lecito pretendere che le conclusioni valide per convenzione in *quella* teoria astratta debbano valere anche nei problemi concreti e pratici concernenti la probabilità. Per fare un solo esempio, non appare lecito pretendere che valga l'additività completa: nel caso (sia pure considerato inaccettabile da molti autori) di «un intero scelto a caso» ogni intero ha probabilità nulla ma tutti insieme (un'infinità numerabile) hanno probabilità $= 1$.

Quanto agli oggettivisti del tipo «classico», che si basano sulle suddivisioni in «casi ugualmente probabili», si può riconoscere che c'è modo, spesso comodo, di ricondursi ad esemplificazioni di quel tipo; però il giudizio di «uguale probabilità» è non definibile salvo in senso soggettivo, oppure... con varianti verbali o perifrasi: anziché «ugualmente probabili» dire «ugualmente possibili» (che è peggio!), o addirittura «uguali» (peggio che peggio!) In conclusione, il metodo pretesamente oggettivo od oggettivistico ha senso ed è accettabile ed applicabile se e soltanto se lo si *concreta* in senso *soggettivistico* anziché evocare presunti fantasmi oggettivistici.

Non è un gioco di parole: sembra giusto asserire che è piú oggettiva una cosa soggettiva considerata come tale anziché una cosa che viene considerata come oggettiva mentre tale qualifica non può essere avallata senza riserve.

Passando agli oggettivisti di formazione statistica, si giunge talora a veder addirittura non solo confondere, ma perfino identificare (!) probabilità e frequenza. Ciò significa, in sostanza, scambiare l'attesa di un fatto con la sua realizzazione e constatazione, il «preventivo» col «consuntivo». Questa distorsione è terribilmente grande non solo perché oscura entrambi i concetti, bensí, peggio ancora, perché, identificandoli, li trasforma in un ibrido mostro bicipite. Guardando piú a fondo, l'argomentazione è ancor piú inconcludente; per esprimersi in modo sensato, corretto, si dovrà parlare di «scambiabilità» (cfr. § 3.5) anziché di indipendenza stocastica.

C'è qualche conclusione che si può trarre da tutto ciò?

Probabilmente sí, e precisamente nel senso di riconoscere che nel *fondo* (spesso trascurato o negato o svisato), qualcosa, spesso anche molto, si può salvare e utilizzare di quanto dicono le diverse teorie, ma con un «purché»: purché nell'interpretare tutti i termini e tutte le affermazioni o definizioni o nozioni o conclusioni si abbia sempre cura di vivificarle facendovi scorrere la linfa salutare del soggettivismo.

Senza di ciò – lo si può ben affermare senza esitazione – tutto si ridurrebbe a un vaniloquio: proprio come affermava la *boutade* di Bertrand Russell citata sopra.

3.3. Qualcosa che si dice «eccezionale».

Vi sono molte specie, piú o meno fondamentalmente analoghe, di fraintendimenti che fanno giudicare «eccezionale» il verificarsi di qualche fatto, o circostanza, o situazione, e ritenere «accettabili» come «normali» altri fatti o circostanze o situazioni del tutto analoghi. Il fatto piú tipico a questo riguardo è quello che fa ritenere necessario che in una lunga successione di colpi a Testa e Croce, oppure di lanci di un dado o di due dadi, le 2 facce della moneta (e, rispettivamente, le 6 facce del dado o le 36 coppie di facce dei due dadi, *debbano* presentarsi (prolungando le «prove») circa nella proporzione prevista (1/2, o 1/6, o 1/36), ed inoltre trovarsi *in disordine* (non ad esempio sempre Testa, né sempre alternatamente Testa e Croce, né Testa nella prima metà e Croce nella seconda, né una Testa e poi sempre Croce, né in altre modalità «regolari»)). Tant'è vero che tali risultati si scarterebbero come «non ammissibili», «non regolari» (nel senso di non abbastanza irregolari); eppure di per sé non presentano nulla di anomalo. Essi hanno probabilità $(1/2)^n$, $(1/6)^n$, $(1/36)^n$; è molto piccola se n è grande ma è esattamente la stessa di qualunque altra successione, non importa se piú o meno regolare o irregolare (quale che sia il senso – molto arbitrario! – in cui uno potrebbe interpretare tali distinzioni!)

E allora perché meravigliarsi? Lo stupore può essere giustificato dalla sorpresa, ma non dal fatto che la probabilità sia piccola. Ogni fatto, se lo si precisa con tutti i dettagli, ha probabilità piccolissime, ed anche nulla se la precisione è assoluta (non sbagliare la posizione di un millimetro, non sbagliare l'istante di un microsecondo, ecc.).

Si può notare, d'altra parte, l'utilizzazione che viene fatta dagli statistici sperimentatori di tabelle di «numeri casuali» (*random numbers*) allo scopo di eseguire «scelte a caso» di individui o oggetti od altro onde approfondire certi studi (nell'impossibilità di esaminare tutti gli individui, o oggetti, o avvenimenti, od altro) limitandosi ad esaminare un «campione rappresentativo»; la scelta «a caso» dovrebbe eliminare (o almeno ridurre di molto) il rischio di scelte distorte (ad esempio con sproporzionata rappresentanza di persone del tipo piú abbordabile). La cura di scegliere un campione cercando che risulti rappresentativo è il principale requisito per rendere attendibili le previsioni basate su di esso (ad esempio nei sondaggi).

Guardando nel verso opposto, uno potrebbe dire che si sente sicuro perché

i rischi cui si espone sono minimi: ma se sono parecchi o si ripetono frequentemente la risultante può facilmente essere fatale. Tutte queste ovvie riflessioni, a cosa possono giovare? Dovrebbero giovare a non prendere mai troppo sul serio le impressioni di tranquillità, come è giustificato in parte da una scritta profondamente significativa che campeggia in una trattoria di Trastevere: «Il caso ci protegge – piú che qualsiasi legge». È vero, ma vale anche il viceversa. E perciò «non fidarsene è meglio».

3.4. Indipendenza (stocastica) e correlazione.

È bene premettere che esistono diverse proprietà che si chiamano «indipendenza» (tra eventi, tra numeri aleatori). Si accenna dapprima alla indipendenza (o invece dipendenza) *lineare*: se X e Y sono numeri aleatori (qualunque), $Z=aX+bY+c$ è combinazione lineare di X e Y (la costante c è inessenziale); ovviamente la **P** (previsione, o, in particolare, probabilità) è lineare (additiva) cosicché, nell'esempio, sarà $\mathbf{P}(Z)=\mathbf{P}(aX+bY+c)=a\mathbf{P}(X)+b\mathbf{P}(Y)+c$. È questa la piú semplice forma di dipendenza funzionale. Un esempio relativo ad eventi: se A e B sono eventi incompatibili la loro unione (o «evento somma») $A\vee B$ coincide con la somma (aritmetica) $A+B$; se non sono incompatibili (se cioè la loro intersezione non è vuota, e sia $C=AB$), la loro unione $A\vee B$ non è piú $A+B$ bensí $A+B-C$. Bastino questi cenni a titolo informativo.

Piú interessante forse è notare una circostanza che, dopo aver visto un esempio, è ovvia, ma di primo acchito può sembrare incredibile: l'indipendenza stocastica a due a due tra n eventi qualsiansi $E_1, E_2, \ldots, E_n$ non implica la loro indipendenza; cioè, il fatto che per ogni coppia di eventi sia $\mathbf{P}(E_hE_k)=\mathbf{P}(E_h)\mathbf{P}(E_k)$ non implica che debba essere anche $\mathbf{P}(E_1E_2\ldots E_n)=\mathbf{P}(E_1)\mathbf{P}(E_2)\ldots\mathbf{P}(E_n)$. Si consideri il piú semplice controesempio, illustrato nella figura 6. I quattro eventi sono rappresentati dai tre rombi E_0+E_1, E_0+E_2, E_0+E_3 di probabilità (= area) 1/2; l'intersezione è il triangolo centrale E_0 di probabilità (= area) 1/4, cioè $1/2\times 1/2$, come si voleva dimostrare, e non $1/8=(1/2)^3$ come se sussistesse l'indipendenza fra tutti e tre i rombi e non solo a due a due.

Nel caso di numeri aleatori il significato di indipendenza (stocastica) è sostanzialmente il medesimo; limitandosi, per rimanere al livello elementare, al

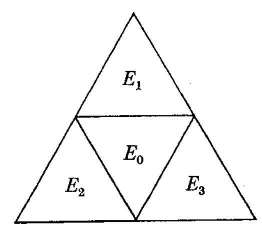


Figura 6.
L'indipendenza stocastica a due a due tra n eventi non implica la loro indipendenza.

caso di numeri aleatori con un numero finito di valori possibili (e siano X ed Y, di valori rispettivamente $x_1, x_2, \dots, x_n$ con probabilità $p_1, p_2, \dots, p_n$ ed $y_1, y_2, \dots, y_m$ con probabilità $q_1, q_2, \dots, q_m$), esso significa che $\mathbf{P}(XY)=\mathbf{P}(X)\mathbf{P}(Y)=$ somma dei termini $p_h q_k x_h y_k$.

Si può considerare, in certo senso, come una forma piú «debole» di indipendenza un'altra proprietà di cui si farà cenno piú avanti: la «scambiabilità»; si può in certo modo anticiparne il senso dicendo che si tratta di «indipendenza *condizionata* alla conoscenza di dati che al momento *non si conoscono*». E si può approfittare dell'occasione per rilevare anche su questo esempio la necessità logica di certe correzioni a terminologie inveterate ma disgraziatamente fuorvianti causa inconsistenze od ambiguità o facilità di sottintendere precisazioni che non possono essere sottintese.

3.5. La scambiabilità.

Il termine 'scambiabilità' è stato introdotto (dallo scrivente) per sostituire una precedente denominazione inaccettabile poiché di per sé contraddittoria: quella di «eventi indipendenti ed ugualmente probabili con probabilità incognita».

Dicendo «eventi scambiabili» s'intende correggere una delle peggiori assurdità terminologiche: nelle condizioni cui allude la precedente descrizione l'uguale probabilità e l'indipendenza non possono coesistere a meno che la probabilità non sia conosciuta; allora si è nel caso usuale (detto in genere delle «prove ripetute»). Se invece la probabilità (ad esempio il numero di palline bianche e nere) non si conoscesse (si sapesse, ad esempio, che l'urna è stata «scelta a caso» (con uguale probabilità) fra due, di cui una contiene 6 palline bianche e 4 nere, e l'altra viceversa 4 bianche e 6 nere), le estrazioni *non sarebbero indipendenti*. Infatti, a mano a mano che si ripetono delle estrazioni, si sarà giustamente indotti a ritenere che l'urna prescelta sia quella che contiene il maggior numero di palline del colore uscito piú spesso. (Per esercizio – se si trattasse di un testo scolastico – si potrebbero fare esempi numerici, e indicare, per ogni momento (ad esempio dopo 10, o 15, o 20 estrazioni), quali probabilità dovrebbero darsi al fatto che l'urna da cui vengono fatte le estrazioni sia quella con prevalenza di palline dell'uno o dell'altro colore. Qui non interessa fare esercizi, ma basti far notare che, evidentemente, si propenderà per ritenere che l'urna prescelta sia quella con maggior numero di palline del colore che è stato estratto piú spesso).

Quindi, non c'è indipendenza, bensí (in questo caso) spostamento dell'attesa verso il colore presentatosi piú spesso. Sussiste però la scambiabilità, nel senso che tutte le successioni (che variano solo per l'ordine, per esempio con 8 estrazioni di pallina bianca e 10 di pallina nera) hanno la medesima probabilità.

È cosa, in fondo, banale; ma le confusioni che possono derivare dall'uso non sufficientemente inequivocabile del linguaggio sono terribili. Sembrerà forse una affermazione stravagante e assurda sostenere l'importanza di cose che potrebbero sembrare minuzie, ma è proprio dal fatto di considerarle tali e di esprimersi in modi ambigui o sconclusionati che si arriva nel modo piú diretto a crearsi intorno un viluppo inestricabile di concetti confusi e di parole usate a casaccio.

Comunque, la locuzione sbagliata di «indipendenza con probabilità uguale ma incognita» merita di essere ricordata come «memento» per riflettere su tutte le terminologie e accertarsi di capire quali hanno senso e quali sono vuote, o, peggio ancora, contraddittorie.

3.6. Previsione e scarto (quadratico medio).

Il discorso finora svolto aveva carattere e scopo piuttosto orientativo, come sembra particolarmente necessario in un campo dove spesso i veri significati vengono soffocati e sacrificati a favore di astrazioni e formalismi nonché (talvolta) perfino di elucubrazioni piú o meno sconnesse. Si spera che le precedenti considerazioni critiche possano avere per lo meno aiutato a sgomberare un po' il terreno da parecchie delle peggiori insidie che vi allignano.

Ed ora, che fare? Fare una sintesi di tutto ciò che potrebbe costituire la materia di un trattato (o magari trattatello) sarebbe cosa insieme troppo pesante e troppo poco efficace. Nel tentativo di fare un discorso piú appetibile per chi desideri acquisire un po' di familiarità e sicurezza, nonché un po' di competenza, nel ragionare sensatamente in termini di probabilità, dovrebbe giovare maggiormente una sia pur piccola serie di esemplificazioni interessanti e sperabilmente (con appena un po' di sforzo) intuitivamente accessibili.

Non sembri troppo banale (o addirittura offensivo per il lettore) se si inizia dal processo di Testa e Croce; sono molte (e spesso anche elevate e complesse) le considerazioni che scaturiscono dallo studio di questo classico argomento. Il riferimento a tale caso è comunque qui appropriato per considerazioni su *previsione* e *scarto* (scarto quadratico medio: $\boldsymbol{\sigma}(X)=\sqrt{\mathbf{P}(X-m)^2}$, ove $m=\mathbf{P}(X)$: previsione di X); considerazioni che varranno, sostanzialmente, anche per tutti i casi analoghi piú o meno semplici.

Si indichino con $E_1, E_2, \dots, E_h, \dots$ gli eventi «ottenere Testa all'h-esimo colpo a Testa e Croce» (e quindi $\tilde{E}_h$, o $1-E_h$, l'ottenere Croce); piú semplicemente, basta dire che E_h vale 1 oppure 0 a seconda che l'h-esimo colpo dà Testa oppure Croce. Si suppone al solito di ritenere che le probabilità dei due risultati siano uguali (1/2 e 1/2) e che le prove siano indipendenti (cioè che i risultati precedenti non modifichino l'opinione circa le probabilità dei casi successivi).

Ciò equivale a dire, in altri termini, che tutte le diverse possibili successioni di n risultati (Testa o Croce; oppure 0 od 1) vengono giudicate ugualmente probabili; precisamente, ciascuna di probabilità $1/2^n$. (Ad esempio per $n=10$ si avrebbe $2^{10}=1024$, quindi probabilità circa 1/1000, e per $n=20$ circa un milionesimo). Ed è il caso di rammentare, in questa occasione, per evitare di cadervi, quei facili fraintendimenti per cui una probabilità molto *piccola* (o molto *grande*) si *confonde* con *impossibilità* (rispettivamente con *certezza*).

Come già detto (e come è ovvio) il numero di successi su n colpi fissati (il successo sia per esempio Testa) è $X=E_1+E_2+\dots+E_n$; finché di tali eventi non si conosce l'esito, X è sconosciuto ed è possibile soltanto dirne qualcosa in termini probabilistici. Si potrà esprimere la *previsione* di X, $\mathbf{P}(X)$, come somma delle probabilità $\mathbf{P}(E_h)$; per saper restringere un po' l'indeterminatezza di tale conoscenza sarebbe utile ogni indicazione sull'ordine di grandezza dello scostamento del valore vero di X da quello della detta stima. Il dato consueto (e perciò appunto è chiamato spesso «scarto standard») lo si indica con $\boldsymbol{\sigma}(X)$ ed è definito come *radice della varianza* $\boldsymbol{\sigma}^2(X) = \mathbf{P}(X-m)^2$. Da notare che se si considerassero le distanze da un punto m' diverso dal baricentro m si otterrebbe $\mathbf{P}(X-m')^2$ *maggiore* (e precisamente aumentato di un termine proporzionale al quadrato della distanza $m-m'$ fra il baricentro e il punto scelto in luogo di esso come riferimento). Veniamo allora al nostro caso (di n eventi ugualmente probabili e indipendenti); si è supposto che abbiano probabilità $p = 1/2$ (ma si potrebbe considerare il caso di p qualunque, od anche di p_h diverso da evento ad evento).

Comunque, per $p = 1/2$, la previsione del numero di successi in n colpi è $n/2$, e lo scarto standard è $\sqrt{n/2}$; si può indicare convenzionalmente (come d'uso) tale fatto dicendo che il numero di successi in n colpi (di probabilità $1/2$, e indipendenti) è $n/2 \pm \sqrt{n/2}$. (Per fare un esempio: su 200 colpi (con probabilità di «successo» $1/2$ ad ogni colpo, e indipendentemente dal risultato degli altri) il numero di successi dev'essere (nel senso detto), $100 \pm \sqrt{100} = 100 \pm 10$, ossia tra 90 e 110). Si badi che tali modi di esprimersi sono utili come indicazione grossolana; non c'è però troppo da preoccuparsene perché nei casi pratici serve piú una *mentalità* intuitivamente allenata a vagliare i rischi e ponderare le decisioni in base a una visione panoramica del pro e del contro. Questo non deve però significare «decidere a vanvera» e senza ponderazione, ma tener presente quanto ci sia di troppo o di troppo poco in una schematizzazione matematica prima di affidarsi ciecamente ai consigli che se ne traggono.

3.7. La tendenza alla distribuzione normale.

Le presenti considerazioni sul caso di Testa e Croce vanno approfondite, sia per l'interesse dell'argomento in sé, sia per i legami con molte delle piú consuete applicazioni. Anziché limitarsi a indicare previsione e scarto (risp. $n/2$ e $\sqrt{n/2}$) si possono indicare le probabilità di ogni singolo valore possibile per la frequenza su un certo numero n di colpi. Esse sono date dalla tabellina in figura 10 dell'articolo «Distribuzione statistica» di questa stessa *Enciclopedia* (vol. IV, p. 1208) fino ad $n = 6$: su 64 casi possibili ce n'è 1 con sempre Testa (o sempre Croce), 6 con 5 volte Testa e 1 Croce (o viceversa), 15 con 4 volte Testa e 2 Croce (o viceversa), e infine 20 con 3 volte Testa e altrettante Croce.

Su $n = 10$ colpi, tra le $2^{10} = 1024$ successioni possibili di 10 tra T e C (e per semplificare si arrotonderà a 1000) ce n'è 1 di tutti T (o tutti C), 10 con 1 C e 9 T, 45 con 2 C e 8 T, 120 con 3 C e 7 T, 210 con 4 C e 6 T, e 252 con 6 C e 6 T (e poi simmetricamente, scambiando T e C). È chiaro che, disegnando l'istogramma, si avrebbe la forma di una collina simmetrica, e si può dire subito (tralasciando la dimostrazione) che la sua forma si avvicinerebbe sempre piú (e, al limite col crescere di n all'infinito, coinciderebbe) col diagramma della *distribuzione normale* (o gaussiana, dal nome di Gauss); (cfr. il già citato «Distribuzione statistica», p. 1210). Può interessare una proprietà geometrica del diagramma di tale distribuzione: la collina che si ottiene facendo ruotare tale curva intorno all'asse centrale ha tutte le sezioni verticali simili (uguali alla sezione centrale, salvo il sempre maggiore appiattimento man mano che ci se ne allontana).

Un processo analogo a quello di Testa e Croce può essere immaginato in due, tre (od anche piú) dimensioni, sia pensando che ogni passo possa essere non solo di *avanti o indietro* su una retta, ma in una delle quattro direzioni (sempre «avanti» o «indietro») nel piano, o delle otto nello spazio a tre dimensioni, o in direzioni qualsiansi (indipendentemente dalla direzione degli assi). Si hanno cosí schemi di «passeggiate aleatorie», che, tra l'altro, possono rappresentare schematizzazioni di moti del tipo browniano (moto disordinato di molte particelle che si urtano procedendo cosí a zigzag). Ed anzi (in una schematizzazione un po' semplificata), varrebbe, nel processo di diffusione, una formula (per la densità) identica a quella di Testa e Croce ($N\omega \pm \sqrt{N\omega}$).

Tornando allo schema di Testa e Croce, va segnalata la semplicità ed efficacia con cui il semplice ma potente «ragionamento di Desiré André» facilita e rende intuitiva la soluzione dei problemi tipo «rovina dei giocatori» (cfr. fig. 7).

Esso si basa sul «principio di rovesciamento»; si ha la medesima probabilità per la traiettoria segnata con linea continua e per quella che (dopo l'intersezione

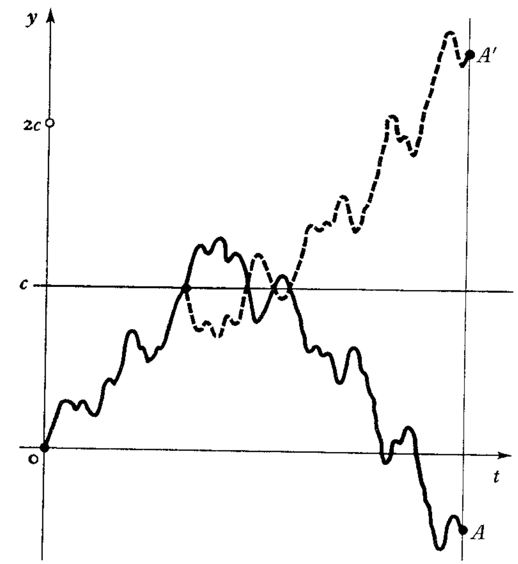


Figura 7.

Ragionamento di Desiré André nel caso di una barriera.

Le traiettorie che, dopo toccato il livello c, ne sono al di sotto alla fine dell'intervallo che interessa (punto A) corrispondono biunivocamente per simmetria a quelle che terminano in A' (simmetrico rispetto alla barriera $y = c$). Di qui (in un processo simmetrico) l'ugual probabilità di terminare in A' o terminare in A dopo toccato il livello c, ed anche, di terminare a un livello $> c$, oppure ad un livello inferiore ma avendo toccato il livello c. Il punto $2c$ sull'asse y è indicato in quanto «sorgente fredda» nel metodo di Lord Kelvin.

col livello c) ne è immagine rovesciata rispetto a detta linea di livello. Nel caso di scommesse eque (probabilità $1/2$ e $1/2$) c'è la stessa probabilità di trovarsi, nell'istante finale, nel punto A o nella sua immagine speculare A'. Ciò significa che c'è ugual probabilità di terminare in A' oppure di *terminare in A dopo aver* toccato il livello c. Problemi del genere sono spesso suscettibili di soluzione elegante con ragionamenti intuitivi di grande efficacia; tale è ad esempio la determinazione della probabilità che un «laccio» (il tratto di traiettoria zigzagante tra due successivi *zeri*) consti di 2, 4, 6, ... (necessariamente numero pari) passi unitari.

Un esempio interessante che fa vedere in modo particolarmente intuitivo la conclusione e l'efficacia del principio di rovesciamento è il cosiddetto «problema dello scrutinio»; si tratta di chiedersi quale sia la probabilità che durante uno scrutinio (supposto che su n voti quelli favorevoli siano in maggioranza in numero di $h > n/2$), essi si siano trovati talvolta in minoranza nel corso dello spoglio. Come la figura 8 mette senz'altro in evidenza, ogni traiettoria che inizi con un passo verso il basso dà luogo ad un'altra che inizi con un passo verso l'alto. Ma il primo passo (come ogni altro) ha probabilità $(n-h)/n = 1 - h/n$ di essere uno degli $n-h$ passi discendenti. La probabilità di annullamento è doppia, $2h/n$, e quella di non-annullamento è quindi $1-2h/n$.

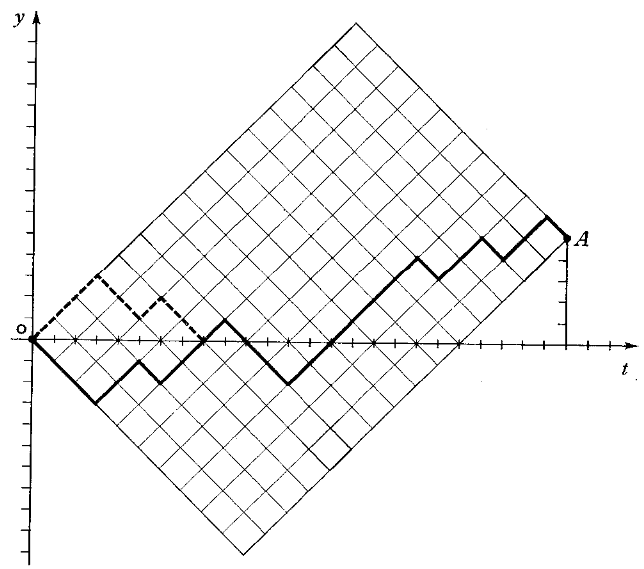


Figura 8.

Ragionamento di Desiré André: problema dello scrutinio (ossia: distribuzione ipergeometrica).

Le traiettorie da O ad A con primo passo discendente corrispondono biunivocamente (per simmetria del tratto fino al primo raggiungimento dell'asse t) a quelle con primo passo ascendente che però toccano l'asse t.

Si accenna ancora ad una conclusione che sembra incredibile: si tratta di chiedere quale sia la previsione della lunghezza di un laccio, cioè del numero di «prove» a Testa e Croce fino al primo ritorno all'equilibrio, cioè a zero. Evidentemente ciò può avvenire dopo 2 colpi (con probabilità $1/2$) o dopo 4, 6, un numero pari qualsiasi,... o anche mai (sebbene tale eternità abbia probabilità zero); ma la previsione è tuttavia infinita (è data dalla somma di una serie che diverge come quella di termini che tendono a zero soltanto come $n^{-1/2}$ (cioè $1/\sqrt{n}$)).

3.8. Riflessioni su presunti paradossi.

I presunti «paradossi» derivano dall'impressione che il calcolo delle probabilità conduca *sempre* ad avvalorare le ipotesi e le previsioni piú appiattite, quelle conformi alla media e alla vetusta massima *in medio stat virtus*. Ma cosí non è: la teoria delle probabilità considera tutti i casi, svariatissimi, rispetto ai quali siamo in condizioni di incertezza, e ciascuno (coll'ausilio di tale teoria o seguendo la sua intuizione, corretta o distorta che sia, nel seguire sia pur istintivamente e grossolanamente certe sue concezioni) dovrà scegliere la decisione e la via che riterrà preferibile, caso per caso.

I casi piú frequenti (o almeno quelli per abitudine considerati «normali») saranno quelli in cui tutti si attendono che permanga piú o meno costante la frequenza osservata nel passato recente (e meglio se anche nel passato meno recente), col «disordine» dovuto al «caso». E probabilmente con eccessiva faciloneria: una sequenza di 10 Teste consecutive ha probabilità assai piccola ($1/2^{10} = 1/1024$, diciamo un millesimo), ma ciò non significa che non si otterrà mai (salvo «miracoli» o... «eccezioni che confermano la regola» secondo un modo di dire stravagante); è da attendersi che il fatto si ripeta in media circa una volta ogni 1000 colpi.

Piú illuminante è l'indicazione data da Willy Feller circa la durata di «permanenza in vantaggio» fra due giocatori che puntano rispettivamente l'uno su Testa e l'altro su Croce ininterrottamente per un anno (un colpo ogni ora, od ogni minuto, od ogni secondo: la numerosità non conta perché sia tanto grande da condurre alle stesse conclusioni del caso limite di infiniti colpi). Per dare un'idea concreta di ciò che è «naturale» prevedere in dette circostanze, si riporta un esempio (di Feller): «Si pensi di giocare continuamente a Testa e Croce per un anno (un colpo ogni ora, o minuto, o secondo: praticamente valgono sempre le conclusioni limite per infiniti colpi); sembrerebbe che, per ragioni di simmetria, i due contendenti dovrebbero trovarsi in vantaggio (complessivamente) ciascuno circa per metà del tempo. Invece: c'è appena probabilità del 30 per cento che entrambi stiano in vantaggio per piú di 100 giorni (circa 28 per cento del tempo totale), mentre c'è probabilità del 50 per cento che uno dei due vi rimanga meno di 54 giorni (15 per cento del tempo), del 20 per cento che vi stia meno di 9 giorni (2,4 per cento del tempo), del 10 per cento che vi stia meno di 2 giorni e mezzo (ossia meno dello 0,6 per cento, e l'avversario piú del 99,4 per cento)». (Da notare poi che l'esser stato piú a lungo in vantaggio non implica maggior probabilità di aver vinto! Infatti tutti i lacci finiscono (per definizio-

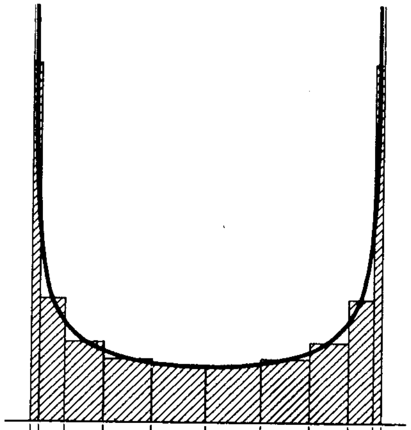


Figura 9.

Densità della distribuzione arcoseno. L'istogramma indica la densità media in ogni intervallo fra i decili. La curva è il diagramma della densità. L'equazione (se l'intervallo-base si assume come (0,1)) e $f(x) = k/\sqrt{x(1-x)}$. La densità è infinita negli estremi.

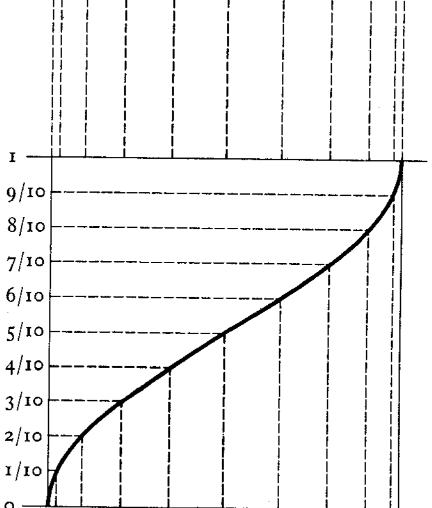


Figura 10.

Funzione di ripartizione della distribuzione arcoseno (ottenibile col modello in fig. 11). Le ascisse segnate sono quelle dei decili come risulta dalle corrispondenti ordinate. I dieci intervalli fra i decili sono ugualmente probabili (probabilità 1/10); notare quanto piú si addensi la probabilità verso gli estremi.

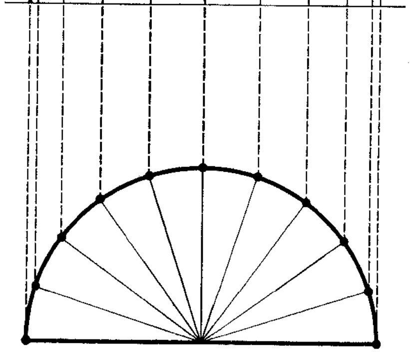


Figura 11.

Si considera la distribuzione di probabilità della proiezione (sul diametro) di un punto «scelto a caso» (densità costante) su una semicirconferenza (o circonferenza). Tale distribuzione si presenta pertanto, ad esempio, se si misura in un istante «a caso» la posizione (o la velocità) di un punto che effettua oscillazioni armoniche. La divisione della semicirconferenza in 10 parti uguali (18°) dà i decili.

ne) in parità, ed è soltanto l'ultimo laccio (anzi frazione di laccio) a decidere la vittoria (o il pareggio)).

Il carattere «squilibrato» di tale processo (probabilità maggiore per squilibri alti anziché modesti) ne fa qualcosa di molto diverso (sotto questo aspetto) dai giochi consueti (tipo Testa e Croce) dove le probabilità maggiori si addensano verso il centro (parità) e svaniscono allontanandosene. Non si tratta però di un caso «eccezionale», ché anzi ne va notata la presenza in fenomeni naturali. L'esempio piú chiaro è quello dei fenomeni periodici (andamento sinusoidale) come ad esempio la proiezione di un punto che percorre una circonferenza a velocità costante, o (approssimativamente, in natura) l'alternarsi dell'alta e bassa marea. È chiaro che l'andamento non è una spezzata a zigzag, in cui si alternerebbero tratti in salita e in discesa rettilinei con passaggio istantaneo e brusco dall'uno all'altro caso, bensí c'è un lento passaggio dalla fase crescente a quella decrescente attraverso un intervallo di quasi stazionarietà. Le figure 9-11 e le relative didascalie bastano a completare la spiegazione.

3.9. Bayes, o del ragionamento induttivo.

È giusto porre ora in testa al titolo il nome di Bayes perché intorno ad esso si è scatenata e tuttora perdura la contrapposizione tra bayesiani e antibayesiani: il suo nome può considerarsi come «segno di contraddizione», come vessillo in una battaglia tra fazioni contrapposte e inconciliabili.

Oggetto della contesa è il fondamento del ragionamento induttivo, del ragionamento che precisa il senso e il modo in cui si fanno, ed è giustificato fare, delle *previsioni*, in termini di *probabilità*, basandosi sull'esperienza, e precisamente – in particolare – sull'osservazione della *frequenza* dei successi in un numero (possibilmente grande) di casi «analoghi» a quello (o quelli) di cui c'interessa prevedere il risultato. In termini piú generali si tratta di vedere il modo in cui «la» (o «le») probabilità in questione vengono modificate in seguito all'acquisizione di ulteriori informazioni, H, che – aggiunte a quelle che costituivano il precedente stato di conoscenza, H_0 – dànno come nuovo stato di conoscenza $H_0' = H \cdot H_0$.

Il teorema di Bayes dice che (come si è già accennato) la probabilità $\mathbf{P}(EH)$ del prodotto di due eventi E ed H è data da $\mathbf{P}(H) \cdot \mathbf{P}(E|H)$, o, simmetricamente, da $\mathbf{P}(E) \cdot \mathbf{P}(H|E)$ da cui risulta che è

$$\mathbf{P}(E|H) = \mathbf{P}(E) \frac{\mathbf{P}(H|E)}{\mathbf{P}(H)} \qquad \mathbf{P}(H|E) = \mathbf{P}(H) \frac{\mathbf{P}(E|H)}{\mathbf{P}(E)};$$

a parole: la probabilità di E, subordinandola ad H, si modifica nello stesso rapporto in cui si modifica la probabilità di H subordinandola ad E.

Le precedenti formulazioni e conclusioni, considerate valide ed anzi ovvie nell'impostazione soggettivistica, sono al contrario ferocemente avversate ed esecrate dagli oggettivisti. Tale contrasto è esploso ripetutamente sia in congressi sia in scambi polemici su riviste statistiche o matematiche, e in particolare al congresso dell'Istituto Internazionale di Statistica a Vienna (1973) dove la difesa del bayesianismo fu affidata a una *contributed paper* (del presente autore). Il titolo rivela di per sé chiaramente l'impostazione: *Bayesianism: its unifying role for both the foundations and the applications of statistics.*

Da parte dei soggettivisti fu fatto notare che la loro posizione è «naturale», e il teorema di Bayes ne è parte in modo ovvio; anzi – come appropriatamente soggiunse Cornfield – è talmente ovvio che è *overly solemn to call it a theorem at all.* I contraddittori non seppero dire nulla che denotasse idee contrastanti ma rispettabili; si limitarono perciò a sfogarsi strappandosi (metaforicamente) le vesti per il sacrilego rifiuto ad adorare le presunte «Probabilità oggettive» come un novello vitello d'oro.

3.10. L'atteggiamento bayesiano.

Sembrerebbe logico, per chiunque abbia una mentalità immunizzata rispetto ad ogni superfetazione oggettivistica, che tutti i suoi ragionamenti e pensieri fossero ispirati (anche inconsapevolmente rispetto a tecnicismi) ad un'attenzione per ogni nuova scoperta o conoscenza tale da farla incorporare in senso bayesiano nel precedente complesso delle sue conoscenze, con tutti gli eventuali conseguenti mutamenti e arricchimenti di esse.

Probabilmente tutti già fanno cosí: fanno un continuo aggiornamento del loro orizzonte globale, riordinandone piú o meno automaticamente tutto il contenuto; dimenticano o mettono in disparte (in una «memoria esterna», per dirlo in termini di informatica) ciò che preme di meno, e collocano quelle per loro piú importanti in posizioni di piú rapido accesso e in piú stretto collegamento con altre zone della memoria.

L'analogia dovrebbe giovare in entrambi i sensi, incoraggiando ad usare le proprie facoltà mentali nel modo migliore per utilizzare (selezionandolo) tutto l'input che ci proviene dall'esterno per aggiornare il complesso delle cose piú o meno ricordate, richiamandole per fondersi coi nuovi apporti. È certo che a tale opera collabora – anche forse a nostra insaputa, ai limiti dell'inconscio – la nostra mente per perfezionarla, la nostra attenzione per precisarla, la nostra fantasia per precorrere le possibilità di farne uso nel modo piú soddisfacente.

Ed è questo, al di là delle applicazioni piú specialistiche e tecniche, il ruolo che ha la probabilità, e in particolar modo il ragionamento bayesiano, per contribuire congiuntamente – nel corso del cammin di nostra vita – a norme oggettive di razionalità e lungimiranza. [B. D. F.].

De Finetti, B.
1970 *Teoria delle probabilità. Sintesi introduttiva con appendice critica*, Einaudi, Torino.
Grayson, C. J. jr
1958 *Decisions under Uncertainty. Drilling Decisions by Oil and Gas Operators*, Harvard University Division of Research, Boston.
Jeffreys, H.
1939 *Theory of Probability*, Clarendon Press, Oxford.

La probabilità si presenta, a prima vista, come un'entità paradossale, perché di essa altrettanto si potrebbe negare l'**esistenza** quanto affermarla come ovunque presente (cfr. anche **essere, idea**). Ma ciò forse inerisce piú che al **concetto** alla sua concretizzazione (cfr. **astratto/concreto**) o piuttosto ancora al suo uso indiscusso e a-problematico, che riversa dalla parte dell'oggetto ciò che invece piú propriamente appartiene al soggetto (cfr. **dato, soggetto/oggetto**). Le necessità scientifiche o pratiche (cfr. **scienza, teoria/pratica**) ove interviene la probabilità impongono naturalmente che in determinati contesti si debbano usare leggi (cfr. **legge**) piú o meno empiriche (cfr. **empiria/esperienza, esperimento**; e anche **convenzione, operatività**) ricorrendo a metodi (cfr. **metodo**) piú o meno attendibili (cfr. **anticipazione, invenzione, ipotesi, modello**); che venga usato un **linguaggio** determinato e specifico (cfr. anche **formalizzazione, logica**) il quale certo impone l'uso di segni determinati o di particolari metafore (cfr. **codice, metafora, segno, significato**); ma è un'istanza **metafisica** ingiustificabile far sí che questo vasto spettro di azioni e di pensieri debba raccogliersi in definizioni pseudo-oggettive e contraddittorie (cfr. anche **opposizione/contraddizione, errore, dicibile/indicibile**). Liberato il concetto di probabilità dai vincoli di una falsa oggettività è possibile suggerirne un uso pertinente nei vari contesti ove esso si dà (cfr. **caso/probabilità, causa/effetto**), legandolo appropriatamente ai concetti e ai procedimenti del **calcolo** dell'**induzione statistica**, alle valutazioni quantitative (cfr. **qualità/quantità**) e in generale collegandolo in modo piú adeguato alle varie esigenze del comportamento umano (cfr. **comportamento e condizionamento, decisione, giochi**).

Rappresentazione statistica

1. *Classificazioni e paradossi.*

«Fece dunque il Signore Iddio dal suolo ogni sorta di animali terrestri e tutti i volatili del cielo, li condusse all'uomo, per vedere come costui li avrebbe chiamati: qualunque nome infatti avesse posto l'uomo a ciascun animale, quello sarebbe stato il suo nome. E l'uomo impose nomi a tutti gli animali domestici e ai volatili del cielo e a tutte le fiere della terra» [*Genesi*, 2, 19-20]. Nella tradizione biblica a un'opera divina (dunque oggettiva), la creazione degli animali, corrisponde un'opera umana (dunque soggettiva), la formazione dei nomi delle categorie in cui vanno classificati. Il primo gradino di quest'opera umana è il nome comune: «Per dirla in breve, l'espressione linguistica con la quale si riconosce la uguaglianza convenzionale di dati gruppi di eventi e la diseguaglianza di essi rispetto ad altri gruppi, qualificati con nomi diversi. Chi dice cane, sa di affermare l'uguaglianza di entità assai varie, che vanno dal minuscolo *chihuahua* al gigantesco *san bernardo*, ma che pure differiscono assai piú radicalmente che non fra di loro dal bue, dalla farfalla, da un cristallo, da una scintilla elettrica» [Boldrini 1965, p. 92]. Ma il nome comune è solo un primo gradino di una tipica costruzione gerarchica: il testo biblico consente al primo uomo non solo la scelta dei nomi per gli esseri viventi, ma anche la loro graduazione in speci e generi. «Nasce da ciò una ulteriore conquista che merita un particolare esame. Allo stesso modo che i nomi comuni compendiano le concettualizzazioni di esperienze riconosciute come analoghe (cane, falco, ...) i generi riuniscono i nomi fra i quali l'analisi oggettiva riconosce... relazioni di analogia (mammiferi, uccelli, ...) Si tratta dunque di una procedura di uniformizzazione dal basso verso l'alto, da classi piú ristrette e meglio specificate ad altre via via piú comprensive, che può venire spinta tanto innanzi quanto si vuole» [*ibid.*]. Sul nodo della classificazione si intrecciano dunque cultura e natura, convenzione ed essenza (un tema che il lettore ritrova in altri articoli della presente *Enciclopedia*).

Dalla διαίρεσις platonica alle «classificazioni naturali» di Linneo, dalla tavola di Mendeleev degli elementi chimici alla sistematica delle particelle «elementari», scienza e metafisica ritrovano la partizione in classi come paradigma ricorrente. Un primo schema che occorre arricchire: «Contare le cose, i fenomeni, gli eventi classificati, cioè i casi, è operazione normale della vita e della scienza: ricorre ad essa la lavandaia, che scrive la lista del bucato; se ne serve il sociologo, che organizza censimenti della popolazione; ne fa uso il filologo, che esegue l'inventario delle voci, degli stilemi, dei suoni, dei segni, delle varianti, delle interpolazioni, degli scolii nelle opere letterarie; e l'astronomo si affatica in essa, cercando di reperire ed enumerare le stelle, i pianeti, gli asteroidi, tutti i corpi vaganti nello spazio; mentre l'archeologo enumera i vasi e lo storico ricerca i committenti delle opere pubbliche romane» [*ibid.*, p. 227].

È questo lo sfondo concettuale della statistica descrittiva: essa sottende larga

porzione della modellistica scientifica come dello stesso linguaggio quotidiano: «Le sei facce di un dado, le 90 palline della tombola, le 52 carte da poker sono statistiche, come lo sono la "popolazione" delle stelle e quando vengono contati o misurati, sono ancora statistiche gli elettroni, i fotoni, le molecole nella goccia d'acqua, le dimensioni delle camomille e dei girasoli e ogni altro numero o misura di cose» [*ibid.*, p. 228]. Questa caratterizzazione informale di statistica «competerebbe – a voler sottilizzare – persino allo stesso nome comune come tale... Ci sono nomi comuni, come quelli di talune malattie rarissime, di taluni pezzi archeologici, di certi preziosi incunabula, dei francobolli piú apprezzati dai collezionisti, di speciali varietà di orchidee, che rimarrebbero inespressivi se nei trattati, nei cataloghi, nelle citazioni orali non li accompagnasse regolarmente il numero delle unità note» [*ibid.*].

Ora, questo rendere espressivo ciò che «inespressivo» resterebbe se non venisse specificato il «numero delle unità note», è tutto fuorché non problematico. Ogni problema interessante di rappresentazione statistica sottende – implicitamente o esplicitamente – un problema di inferenza o induzione statistica, dunque uno sfondo di conoscenza teorica. Nel racconto biblico Adamo impone nomi prima di aver gustato il frutto «desiderabile per avere la conoscenza» [*Genesi*, 3, 6]. Ma questa è una semplice illusione: nella stessa statistica descrittiva il processo di rilevazione dei dati non è mai innocente. Lo sfondo teorico s'insinua infatti fin dalla strutturazione del campo sperimentale in un sistema di classificazione. Ad esempio all'anagrafe i nuovi nati vengono classificati (tra l'altro) come maschi o femmine. Che cosa c'è di piú ovvio, si dirà? ovvero di meno carico di teoria – in una parola, di innocente? Si prenda allora quell'ufficio dell'anagrafe di goodmaniana memoria [Goodman 1955, pp. 59 sgg.] che classificava i neonati in *f*aschio e *m*emmina. In quell'ufficio *faschi* venivano considerati i neonati di sesso femminile (secondo la nostra classificazione: questa è una traduzione!) nati prima del 1984 oppure quelli di sesso maschile nati nel o dopo il 1984; *memmine* invece quelli di sesso maschile nati prima del 1984 oppure quelli di sesso femminile nati nel o dopo il 1984. È chiaro che, fino al 1984, qualunque seriazione statistica (per cui si veda piú oltre) basata sulla prima classificazione sarà quantitativamente indistinguibile da qualunque seriazione statistica basata sulla seconda classificazione. Tanti faschi quante femmine, tanti maschi quante memmine. Ma le due classificazioni fanno tutta la differenza del mondo dal punto di vista dell'inferenza statistica. Si assuma infatti che una stima della probabilità che nasca un maschio basata sui dati degli usuali uffici anagrafici sia pari a p, diciamo 0,51. Questa stima varrà in particolare per i maschi nati nel o dopo il 1984. Ma una stima basata sui dati del secondo ufficio anagrafico darà invece una probabilità pari a 0,49 per i maschi nati nel o dopo il 1984. Stime incompatibili, dunque, basate sugli stessi dati quantitativi.

La nostra preferenza intuitiva va naturalmente alla stima basata sulla classificazione standard. Certo, si è portati a dire, la classificazione non-standard introduce nel problema una componente temporale del tutto arbitraria. Questione di punti di vista, replicherebbe chi assumesse la classificazione non-standard come «naturale». Un maschio non è che una memmina nata prima del 1984 oppure un faschio nato nel o dopo il 1984... La situazione – è facile vedere – è completamente simmetrica. E allora?

È solo mettendo in gioco la totalità delle nostre conoscenze teoriche, almeno in campo biologico, che si è in grado di privilegiare la classificazione «standard» rispetto a quella «deviante». Senza insistere ulteriormente sui dettagli di questa linea di soluzione [per cui si vedano ad esempio Hesse 1969; Quine 1969], questo «enigma» basta a mostrare il carattere fortemente problematico di ogni processo di rilevazione dei dati, e la sua connessione con l'inferenza statistica.

2. *Protocolli, seriazioni (e paradossi).*

Questa connessione è già esplicita nelle *Natural and Political Observations on the Bills of Mortality* di John Graunt [1662] in cui la scoperta dei «dati statistici» è finalizzata alla soluzione di uno specifico problema d'inferenza: la stima della popolazione di Dublino, Londra e Parigi. «Questo rende Graunt in sostanza il fondatore della pratica di registrare sistematicamente nascite, matrimoni, e *cause* di morte non per ragioni ecclesiastiche, ma a uso dello stato» [Pearson 1921-23, ed. 1978 p. 33]. Lo stesso Graunt rispondeva a conclusione delle sue *Observations* a chi si chiedeva la ragione di «tante fatiche e affanni (all this laborious bustling and groping)» che «è molto soddisfacente dedurre inferenze cosí astruse e inattese da questi poveri *Bills of Mortality* tanto disprezzati» [1662, ed. 1665 pp. 143 sgg.]. Un secolo dopo il prussiano Johann Peter Süssmilch, nel 1741, constatava che il materiale per la determinazione di quello che chiamava «l'ordine divino» (cioè la stabilità dei rapporti statistici) esisteva nei registri parrocchiali almeno dal tempo della Riforma. «Ma chi, – aggiungeva, – mai ne fece uso a questo scopo prima di Graunt? La scoperta era facile come quella dell'America, ma Colombo mancava» [citato e discusso in Merz 1903-904, ed. 1965 II, p. 564, nota].

Concediamo a questo punto allo statistico le classificazioni usuali e mettiamo tra parentesi la connessione tra rilevazione dei dati e inferenza per considerare piú in dettaglio il processo il cui esito finale è la rappresentazione statistica.

Per «successione statistica» o «protocollo statistico» s'intende usualmente la registrazione di un insieme finito di osservazioni entro un dato sistema di classificazione. Ad esempio, la rilevazione della statura di un campione di cento individui potrebbe essere rappresentata dalla seguente tabella:

Massimo Galuzzi	[170, 175)
Renato Betti	[180, 185)
Giulio Giorello	[185, 190)
Luciano Lovera	[165, 170)
Giuseppe Papagno	[185, 190)
Giampaolo Caprettini	[190, 195)
...	...
...	...

In questa tabella si nota immediatamente che Giorello e Papagno hanno la medesima statura, rispetto al sistema di classificazione usato – basato su intervalli uniformi semiaperti, chiusi inferiormente e aperti superiormente. Questo fatto li rende indistinguibili per chi disponesse solo dell'informazione codificata dalla tabella. Sembra perciò un passo del tutto ovvio omettere dalla tabella le specificazioni individuali e dare semplicemente la *numerosità* di ciascuna delle classi prese in considerazione (per questa terminologia, cfr. anche l'articolo «Distribuzione statistica» in questa stessa *Enciclopedia*). Si ottiene in tal modo una «seriazione statistica», cioè una successione di coppie ordinate il cui primo elemento è una delle classi del sistema usato e il cui secondo elemento è il numero degli individui osservati che appartengono a quella classe.

Un'altra volta, tuttavia, il passo non è affatto ovvio, ma incorpora un'assunzione molto forte. Per drammatizzare, si consideri la seguente eventualità. Si sono estratte da un'urna cento palline che si distinguono semplicemente per il fatto di essere bianche o nere. Il protocollo statistico pertinente registra alle estrazioni pari il colore nero e alle estrazioni dispari il colore bianco. Ma la seriazione statistica corrispondente registra semplicemente cinquanta palline nere e cinquanta palline bianche. Che differenza fa, ci si chiederà? La differenza emerge una volta ancora se ci si pone dal punto di vista dell'inferenza statistica. Sulla base del protocollo statistico, è ragionevole stimare la probabilità che la prossima (la centounesima) pallina sia bianca molto prossima a 1. Sulla base della seriazione statistica invece, è ragionevole stimare la stessa probabilità molto prossima a 1/2. La preferenza intuitiva – di nuovo – va alla prima stima. In questo caso, la ragione è abbastanza ovvia. Il campione osservato induce a supporre di aver a che fare con un fenomeno di carattere periodico. Dunque l'informazione che si omette passando da un protocollo a una seriazione statistici non è a priori irrilevante. La sua rilevanza dipende essenzialmente dalla natura *fisica* del fenomeno considerato. Coinvolge dunque una volta di piú la nostra conoscenza di sfondo.

Proprio per questo è opportuno rendere esplicita l'ipotesi che giustifica il passaggio dai protocolli statistici alle seriazioni statistiche, e cioè l'ipotesi secondo cui individui indistinguibili rispetto alla classificazione considerata vanno trattati allo stesso modo, nota come ipotesi di scambiabilità (per cui si vedano gli articoli «Distribuzione statistica» e «Induzione statistica» in questa stessa *Enciclopedia*). Sulla sua base, il protocollo statistico precedente potrebbe dare luogo alla seguente seriazione statistica:

Stature	*Numerosità*	*Stature*	*Numerosità*
[140, 145)	1	[175, 180)	10
[145, 150)	2	[180, 185)	5
[150, 155)	4	[185, 190)	2
[155, 160)	10	[190, 195)	1
[160, 165)	20	[195, 200)	1
[165, 170)	25	[200, 205)	1
[170, 175)	18		

3. *Rappresentazioni e convenzioni.*

Le rappresentazioni finora esaminate possono essere intese come rappresentazioni numeriche di rilevazioni statistiche. Queste rilevazioni ammettono naturalmente anche una rappresentazione geometrica. Nonostante i meriti di questo tipo di «traduzione» dei dati nel rendere piú immediatamente evidenti certe caratteristiche teoricamente significative di un protocollo o di una seriazione, non tutte le traduzioni sono ugualmente buone. Un solo esempio varrà a mettere in guardia da traduzioni «perverse». Si consideri il seguente protocollo che descrive l'andamento della spesa pubblica (in miliardi di lire) nel corso di un semestre specificato, per esempio da maggio a novembre del 1947:

Maggio	19,50
Giugno	19,45
Luglio	19,40
Agosto	19,55
Settembre	19,35
Ottobre	19,65
Novembre	20,15

E si considerino ora, seguendo Huff [1954], le due rappresentazioni grafiche dello stesso protocollo (fig. 1). Il grafico a sinistra è volto a suscitare l'impressione di un drammatico aumento della spesa pubblica, mentre l'altro tende a incoraggiare l'opinione che nulla d'importante stia succedendo. Tuttavia, entrambe le rappresentazioni codificano esattamente la stessa informazione. L'unica differenza consiste nella scelta di una diversa unità di misura sull'asse delle ordinate. Qual è piú «fedele ai fatti»? A priori, nessuna delle due. Tutto dipende in effetti

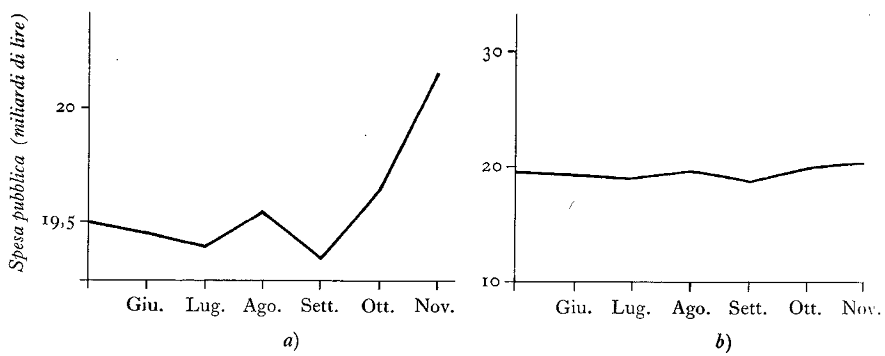


Figura 1.
Uso di un grafico «truccato» che amplifica le differenze fra le ordinate. In *a*) la spesa pubblica risulta in aumento, in *b*) risulta costante.

dall'andamento della spesa pubblica nel paese in questione in un periodo di tempo, precedente a quello considerato, sufficientemente lungo. Si ponga ad esempio che le variazioni massime siano contenute entro un margine dello 0,01 per cento. In tal caso, sarà la prima rappresentazione piú «fedele ai fatti». Se il margine fosse invece significativamente piú ampio, lo sarebbe la seconda. Perciò, convenzioni per la rappresentazione grafica di «distribuzioni di frequenza» (per questa nozione si veda in questa stessa *Enciclopedia* l'articolo «Distribuzione statistica»), come ad esempio la cosiddetta «regola aurea dei tre quarti» [per cui si veda Runyon e Haber 1976, trad. it. p. 44], hanno solo un valore pragmatico. È semplicemente il modo in cui vanno usualmente le cose a determinare la maggior adeguatezza di una convenzione rispetto all'altra.

Non solo tuttavia la scelta delle unità di misura dipende dalla natura del fenomeno che s'intende rappresentare, ma anche il tipo «qualitativo» della rappresentazione. Si prenda la seguente seriazione statistica che specifica il numero di studenti immatricolati nel 1980-81 in Italia, distribuiti per facoltà (dati approssimati):

Scienze matematiche, fisiche e naturali	23 237
Farmacia	5 257
Medicina e chirurgia	21 318
Ingegneria	16 907
Architettura	10 422
Agraria	6 297
Economia e commercio	30 744
Scienze politiche	6 960
Giurisprudenza	33 724
Lettere	21 113
Magistero	19 496
Lingue	3 855
Altre	5 540
Totale	204 870

Il fatto che le «celle» del sistema di classificazione utilizzato non ammettano nessun ordinamento «naturale», rende del tutto arbitraria la costruzione di uno dei due assi. Una soluzione usuale è quella rappresentata in figura 2 in cui risultano arbitrari 1) l'ordinamento delle celle e 2) la lunghezza della loro base. Invece, la scelta dell'asse delle ascisse per rappresentare le materie di studio, e delle ordinate per il numero degli studenti, e non viceversa, dipende dal fatto che l'interpretazione piú naturale dell'espressione «il numero degli immatricolati» è nei termini della funzione «il numero degli immatricolati (in Italia nel 1980-81) iscritti alla facoltà x», dove x può assumere i valori specificati nella colonna di sinistra della seriazione. A ciascuno di questi valori, il nostro protocollo associa un *unico* valore numerico (cfr. fig. 2). Questa condizione di «univocità in seconda sede» – nota come condizione di Dirichlet – consente d'interpretarlo come

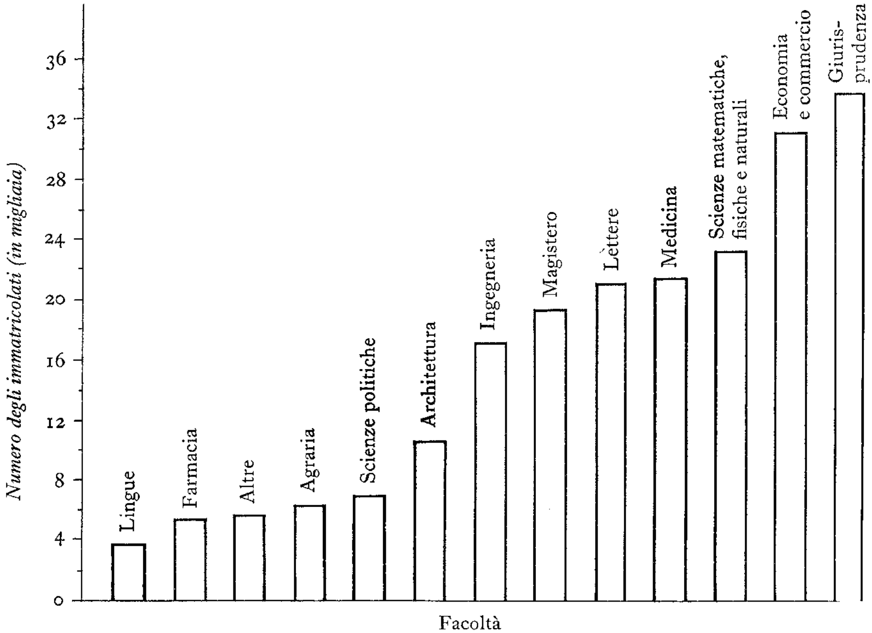


Figura 2.
Numero degli studenti immatricolati in Italia nel 1980-81 distribuiti per facoltà.

grafico di una funzione (si veda per questo l'articolo «Funzioni» in questa stessa *Enciclopedia*). Non è invece difficile vedere che in generale il converso non dà luogo a funzioni, poiché in linea di principio consente violazioni della condizione di univocità (si pensi, per esempio, a una seriazione uguale a quella di p. 589 tranne che per agraria cui si assegna ora il valore 6960, lo stesso assegnato a scienze politiche; in questo esempio fittizio a 6960 corrispondono *due* distinte facoltà). Funzioni di questo genere sono note nella letteratura statistica come «variabili casuali (o aleatorie)» [per cui si veda ad esempio l'ormai classico Kendall e Stuart 1969, pp. 186 sgg.]. Ma si veda anche il commento di De Finetti: «Non sembra tuttavia una locuzione felice, e tanto meno necessaria» [1970, p. 26].

4. *Dal discreto al continuo.*

A rappresentazioni qualitativamente simili a quelle della figura 2 dànno luogo anche quei protocolli o seriazioni basati su sistemi di classificazione le cui celle, pur ammettendo un ordinamento «naturale» non ammettono nessuna metrica «naturale». Ad esempio, se la nostra popolazione fosse costituita dai minerali di una certa regione geografica e fossimo interessati alla loro distribuzione di

frequenza, la scala di Mohs costituirebbe certamente un criterio naturale per assegnare un ordine, dal meno duro al piú duro, ai vari minerali lungo l'asse delle ascisse. Tuttavia, dato che in questo caso le uniche rappresentazioni numeriche possibili sono di tipo ordinale (per cui si veda l'articolo «Numero» in questa stessa *Enciclopedia*), la «distanza» tra le varie celle risulterebbe del tutto arbitraria. La scelta usuale è ancora quella di una distanza uniforme.

Un caso che s'incontra piú di frequente è quello in cui le celle del sistema di classificazione ammettono – oltre a un ordinamento naturale – anche una metrica «naturale». Sarà allora quest'ultima a determinare le distanze pertinenti. Le rappresentazioni geometriche cosí indotte vengono dette «istogrammi a canne d'organo giustapposte». Si supponga ad esempio di essere interessati alla distribuzione di una certa popolazione per classi d'età. Si hanno qui infinite possibilità di ritagliare le celle dello «spazio delle età», poiché quest'ultimo è un intervallo finito di razionali positivi. Ovviamente piú fine è la partizione, maggiore è l'informazione che si ottiene; ma, ai fini di ogni data applicazione larga parte di questa informazione risulta ridondante. La finezza della partizione va perciò bilanciata rispetto alle sue applicazioni. Sono queste ultime a determinare la scelta della partizione piú appropriata entro l'infinità di partizioni possibili.

Si supponga allora di dover pianificare l'edilizia scolastica per una certa comunità: quanti asili-nido, quanti asili, quante elementari, quante medie inferiori, ecc. In tal caso sarà pertinente un istogramma basato su celle la cui grandezza è determinata dagli intervalli d'età che caratterizzano le popolazioni scolastiche corrispondenti. Una seriazione statistica potrebbe quindi dare come risultato:

Classi d'età	*Abitanti*
(0, 3]	850
(3, 6]	753
(6, 11]	2000
(11, 15]	1350
(15, 19]	1951
(19, 25]	2256

Si otterrebbe perciò l'istogramma della figura 3, in cui la lunghezza della base dei rettangoli è proporzionale alla lunghezza dell'intervallo d'età corrispondente.

Il tipo di rappresentazione grafica fin qui utilizzato è evidentemente basato su un opportuno sistema di coordinate cartesiane. Come è noto, non è questo l'unico sistema di rappresentazione. Almeno a partire da Newton [cfr. Kline 1972, p. 319], è di utilizzazione corrente anche il sistema delle coordinate polari. A priori, i due sistemi sono «equivalenti», nel senso che esiste una traduzione canonica di ogni grafico in coordinate cartesiane in un grafico in coordinate polari e viceversa. In particolare, la traduzione nella seconda direzione è data dalla seguente coppia di relazioni

$$(1) \qquad \begin{aligned} x &= r \cos \vartheta \\ y &= r \sin \vartheta \end{aligned}$$

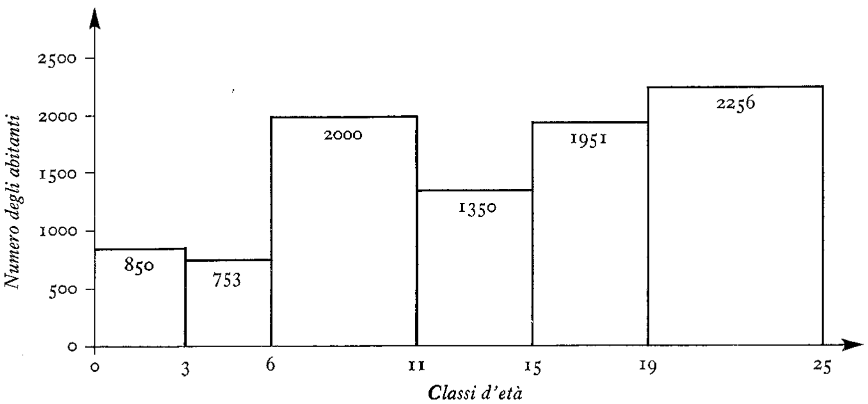


Figura 3.

Istogramma a canne d'organo giustapposte che rappresenta la distribuzione per età degli abitanti di un'ipotetica comunità.

dove $r > 0$ è la lunghezza del segmento di retta che unisce il punto considerato, P, all'origine e ϑ è la misura in radianti dell'angolo formato da tale segmento con l'asse positivo delle x. r e ϑ sono detti rispettivamente raggio vettore e anomalia. Una volta di piú quindi la scelta tra i due sistemi dovrà essere fatta in base a considerazioni di carattere pragmatico.

In generale, le rappresentazioni in coordinate polari sono appropriate a quei fenomeni statistici che hanno un andamento ciclico, «per mettere in rilievo le fluttuazioni delle cosiddette *serie cicliche*», caratterizzate cioè da variazioni ricorrenti [cfr. Livi 1968, p. 73]. S'identifica a tale scopo il periodo di tempo pertinente con l'intero angolo di 2π radianti, che viene quindi suddiviso in tanti «spicchi» quanti sono gli intervalli in cui è suddiviso tale periodo. Dato che questi intervalli di tempo svolgono il ruolo di celle del sistema di classificazione sottostante, l'anomalia degli spicchi dovrà essere proporzionale alla lunghezza degli intervalli. Infine la lunghezza dei raggi vettori dovrà essere proporzionale al valore della funzione che caratterizza il fenomeno nell'intervallo di tempo corrispondente (o a uno degli estremi dell'intervallo). La figura 4 dà una rappresentazione dell'andamento della produzione annua di un'industria per gli anni 1974 e 1975 secondo le quantità prodotte mese per mese.

Naturalmente, si sarebbero potuti suddividere gli anni considerati, invece che in 12 mesi, in 48 settimane, in 302 giorni (supponendo che ogni festivo sia stato soppresso), ecc. Al crescere della finezza della partizione (e dunque del numero delle osservazioni), il grafico si approssimerà a un grafico continuo. Tuttavia, dato che ogni protocollo o seriazione particolare è necessariamente finito, i grafici continui vengono usualmente ottenuti mediante i cosiddetti metodi d'interpolazione (per cui si veda l'articolo «Distribuzione statistica» in questa stessa *Enciclopedia*). Non si tratta in ogni caso di semplici accorgimenti matematici:

essi implicano sempre ipotesi empiriche molto forti, nella misura in cui il loro risultato è quello di assegnare un valore alla funzione considerata rispetto a una infinità di «punti» non osservati. Questo significa che per ogni data rappresentazione di un fenomeno osservato, per quanto fini siano la partizione su cui è basato il protocollo o la seriazione corrispondente, esisteranno infiniti modi di «completarla», che implicano previsioni (o retrodizioni) *incompatibili* rispetto a una *infinità* di punti. Il problema della scelta tra tali «completamenti» non è che un caso particolare del problema dell'induzione (per cui si veda l'articolo

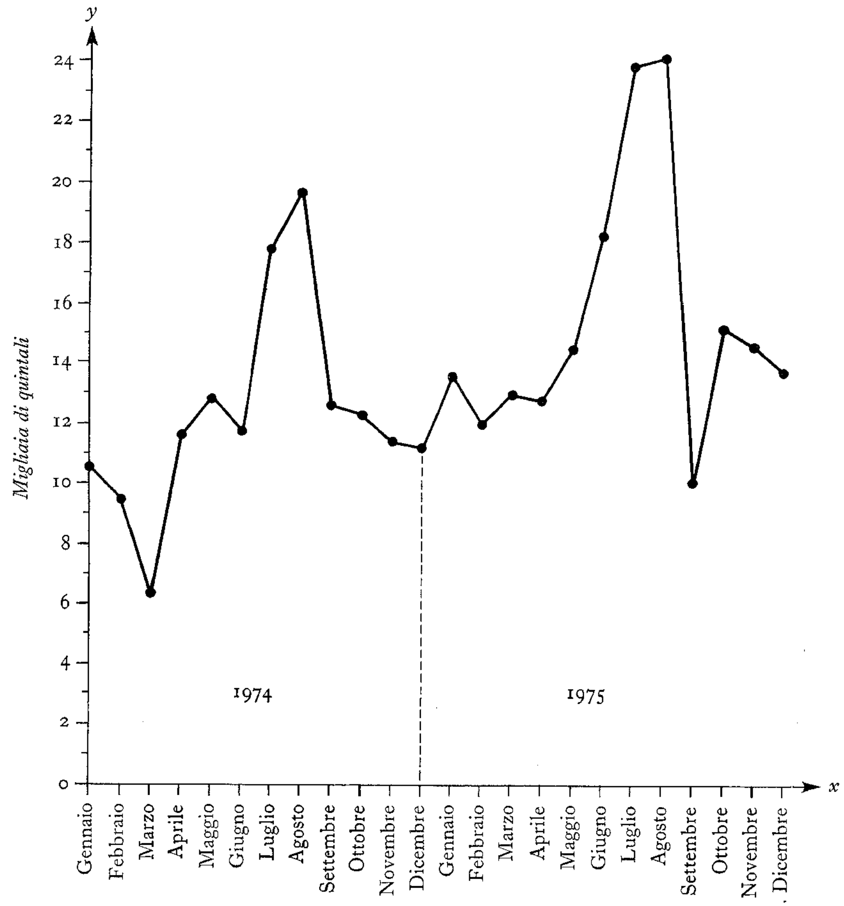


Figura 4.

Rappresentazione in coordinate cartesiane dell'andamento della produzione annua di un'industria per gli anni 1974-75, secondo le quantità (migliaia di quintali) prodotte mese per mese.

«Induzione statistica» in questa stessa *Enciclopedia*) [cfr. ad esempio Popper 1959, trad. it. pp. 126 sgg.].

Nell'esempio della figura 4, l'ipotesi piú semplice (uno dei tanti criteri di scelta sul mercato) è quella lineare che consiste nel congiungere ogni coppia di punti immediatamente successivi con un segmento di retta. Dato l'andamento ciclico del fenomeno, una rappresentazione in coordinate polari sarebbe forse stata piú appropriata. Tuttavia, in questo caso l'espediente di norma utilizzato (si veda del resto la figura 5), consistente anche qui nel congiungere ogni coppia di punti immediatamente successivi con un segmento di retta, pur visualizzando meglio l'andamento ciclico del fenomeno (il lettore confronti le rappresentazioni nelle figure 4 e 5 dello *stesso* protocollo) non corrisponde piú a un'ipotesi di linearità: come ha notato Gini [citato in Livi 1968, p. 75 nota 1], per rappresentare la linearità in coordinate polari occorre congiungere punti immediatamente successivi con segmenti curvilinei in modo che le variazioni intermedie dei raggi vettori siano proporzionali alle variazioni dell'angolo polare. Si ottiene cosí un grafico che è un «incollamento» di segmenti della spirale di Archimede (analiticamente: $r = a\vartheta$, per $a > 0$). L'ipotesi lineare è comunque, con ogni probabilità, empiricamente falsa (il criterio della semplicità non è infallibile!) In realtà, nel caso del fenomeno considerato, *ogni* ipotesi – non solo quella lineare – risulte-

	Produzione 1974	*Produzione 1975*
Gennaio	10,6	13,6
Febbraio	9,5	12,0
Marzo	6,3	13,1
Aprile	11,6	12,9
Maggio	12,8	14,5
Giugno	11,8	18,3
Luglio	17,8	23,9
Agosto	19,7	24,1
Settembre	12,6	10,1
Ottobre	12,3	15,2
Novembre	11,4	14,6
Dicembre	11,2	13,8

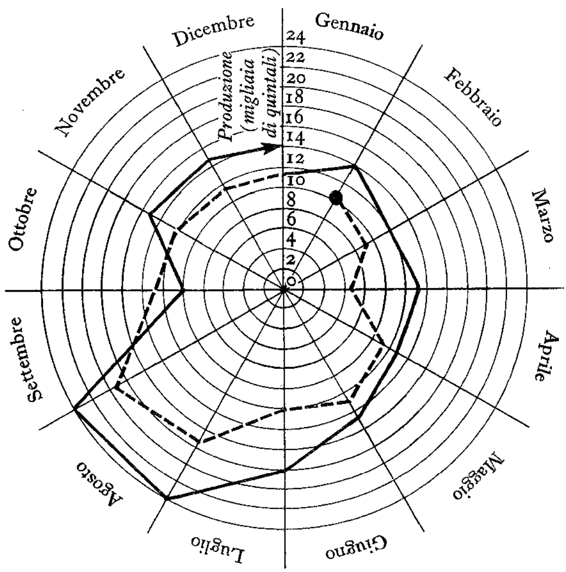


Figura 5.
Rappresentazione in coordinate polari dell'andamento della produzione annua di un'industria per gli anni 1974 (linea tratteggiata) e 1975 (linea continua), secondo le quantità (migliaia di quintali) prodotte mese per mese. (Da Brugnoli e Messori 1980).

rebbe non solo falsa, ma inoltre arbitraria, per lo meno rispetto a un intervallo di tempo come quello considerato. Perché in un caso come questo possano emergere regolarità statistiche significative, si dovrebbero sicuramente considerare intervalli di tempo piú lunghi. Tali regolarità ridurrebbero allora l'arbitrarietà della scelta e l'ipotesi risultante sarebbe – anche se eventualmente falsa – per lo meno significativa.

5. *Un caso storico: l'«uomo medio» di Quételet.*

Da questo punto di vista, la ricerca di una funzione continua – anzi analitica! – per interpolare (ed estrapolare!) un dato insieme di punti osservati risulta interessante solo nei casi in cui si possa supporre che essa costituisce un'ipotesi profonda sulla «natura» del fenomeno esaminato, e non una semplice «finzione» matematica, per fornire un'agile rappresentazione visiva dello stesso. A questo riguardo, uno degli esempi piú significativi è costituito dall'uso che fece Quételet della curva normale (per cui si vedano gli articoli «Distribuzione statistica», IV, p. 1210, e «Probabilità», X, p. 1180, in questa stessa *Enciclopedia*), rappresentata in figura 6, per interpretare la variabilità dei caratteri antropometrici lineari o piú in generale, come fece successivamente Galton, la variabilità di fenomeni biologici. Il nucleo metafisico del programma di Quételet consisteva nell'analogia tra il suo «uomo medio» e il centro di gravità di un corpo in meccanica: «L'uomo che considero qui è, nella società, l'analogo del centro di gravità dei corpi; è la media intorno a cui oscillano gli elementi sociali: sarà, se si vuole, un ente fittizio per il quale tutte le cose si svolgeranno conformemente ai risultati medi ottenuti per la società» [1835, I, p. 21]. In tal modo, «saremo in grado di fissare le leggi a cui egli [l'uomo medio] è stato soggetto nelle diverse nazioni fin dalla nascita» [*ibid.*, p. 23]. L'insistente riferimento alla meccanica (specialmente alla meccanica celeste, sulla scia di Newton e Laplace [cfr. a questo proposito le osservazioni di Merz 1903-904, II, pp. 579 sgg.]) è peraltro spia della consapevolezza dell'urgenza epistemologica della questione (piú sopra richiamata) delle regolarità statistiche. «Le cause regolari e *periodiche*, che dipendono o dal perio-

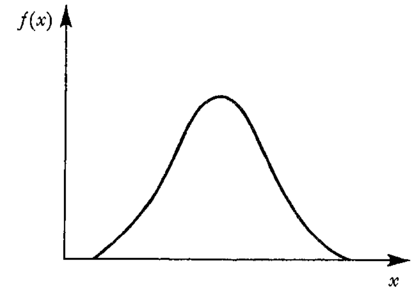


Figura 6.
Esempio di curva normale.

do annuale o dal periodo diurno, esercitano sulla società degli effetti piú pronunziati e che variano entro limiti maggiori degli effetti combinati *non periodici*, prodotti annualmente grazie al concorso di tutte le altre cause che agiscono sulla società; in altri termini, il sistema sociale, nel suo modo di essere, sembra essere piú dissimile da se stesso nel corso di un anno o anche nel semplice spazio di un giorno che durante due anni consecutivi, se si bada alla crescita della popolazione» [Quételet 1835, II, p. 323].

Il carattere «fittizio» dell'«uomo medio» di Quételet – proprio in quanto consapevole prodotto dell'astrazione soggiacente alla modellizzazione tipica dell'«aritmetica politica» di un Petty e della sua stessa «fisica sociale» – consente – e non ostacola – la comprensione del «gioco» della ripetizione e della differenza. Nella stessa «clinica»: «Poiché nella stragrande maggioranza dei casi il malato non può presentare alcuna osservazione soddisfacente fatta sulla propria persona, né alcuno degli elementi che gli sono peculiari, il medico si trova costretto a ricondurlo alla scala comune e ad assimilarlo all'uomo medio. Ciò, in fondo, sembra presentare il minimo delle difficoltà e di inconvenienti; ma può causare anche gravi fraintendimenti in qualche circostanza; è infatti il caso di fare osservare qui che le leggi generali relative alle masse sono essenzialmente false quando vengono applicate a individui: il che non significa, pertanto, che non le si possa consultare con successo: e gli scarti vanno sempre tenuti in considerazione» [*ibid.*, pp. 268-69].

L'enigma di fondo – non semplice questione di teoria, ma elemento costitutivo di pratiche (per esempio della pratica clinica, ove l'excisione dei tratti specifici, gli «éléments particuliers» di Quételet, caratterizza il consolidarsi di metodi nuovi di diagnosi e terapia [cfr. Foucault 1963]) – è risolto da Quételet avanzando la congettura secondo cui «le leggi di sviluppo dell'uomo medio restano approssimativamente le stesse nei vari secoli e variano soltanto per la grandezza dei massimi» [1835, II, p. 271].

6. *Asteroidi e toraci: come dominare «l'irrazionale».*

Si era in tal modo delineata un'interpretazione della curva normale che dava un nuovo senso empirico alle ipotesi fisico-matematiche da cui era stata originariamente derivata da Moivre e Laplace da un lato, da Gauss dall'altro. Tali derivazioni implicavano però un sottile, ma cruciale slittamento dalla nozione di frequenza (fin qui centrale in quest'articolo) a quella di probabilità. (Sul rapporto tra di esse si veda ancora «Distribuzione statistica» (IV, §§ 3.1-3.4), in questa stessa *Enciclopedia*).

Moivre e Laplace l'avevano infatti ottenuta come funzione limite (fig. 7*c*) di una successione di funzioni a scala del tipo illustrato nella figura 7*a*, *b*, analizzando la distribuzione dei risultati di un gioco di cui sono note le probabilità dei risultati «elementari».

«Il leit-motiv matematico del calcolo delle probabilità, la cui orchestrazione prosegue ancora ai nostri giorni, si basa sullo *schema di Bernoulli*, che non è altro che una idealizzazione del gioco di testa e croce» [Loève 1978, p. 283; si vedano anche gli articoli «Induzione statistica» e «Probabilità» in questa stessa *Enciclopedia*]. Tale schema (cosí battezzato in onore di Jakob Bernoulli) può venir compendiato dalla seguente formula, ove si indica con S_n il numero delle realizzazioni di un evento «elementare» A (per esempio l'uscita di testa in un particolare lancio della moneta) di probabilità p in n prove «indipendenti» e si pone $q=1-p$:

$$(2) \qquad P(S_n=k)=\binom{n}{k}p^k q^{n-k}$$

ove, al solito, $P(S_n=k)$ per $k=1, 2, \ldots$ si legge «probabilità che S_n sia uguale al numero k». La legge dei grandi numeri, dovuta a Jakob Bernoulli, afferma allora che, per ogni $\varepsilon>0$, quando $n\rightarrow+\infty$, si ha:

$$(3) \qquad P\left(\left|\frac{S_n}{n}-p\right|<\varepsilon\right)\rightarrow 1.$$

D'altra parte, il teorema del limite centrale afferma che, se $p\cdot q\neq 0$, per ogni valore reale x, quando $n\rightarrow+\infty$

$$(4) \qquad P\left(\frac{S_n-np}{\sqrt{npq}}<x\right)\rightarrow\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy.$$

La (4) è nota anche come teorema di Moivre-Laplace. Era stato in particolare Moivre, nel corso delle varie edizioni della sua *Doctrine of Chances* (1718, 1740, 1756: la storia è ricostruita da Todhunter [1865, cap. IX; per la (4) cfr. in particolare pp. 143 sgg.]) a ottenerla nel caso $p=1/2$, aggiungendo che il suo procedimento era esemplare anche per il caso generale. Si deve però a Laplace la dimostrazione esplicita. Nella (4), al secondo membro, sotto il segno di integrale, compare la funzione $(1/\sqrt{2\pi})e^{-y^2/2}$, il cui grafico è detto curva normale.

Diverso era stato il cammino di Carl Friedrich Gauss. La situazione problematica iniziale venne rappresentata da un «piccolo pianeta Cerere, [che] impri-

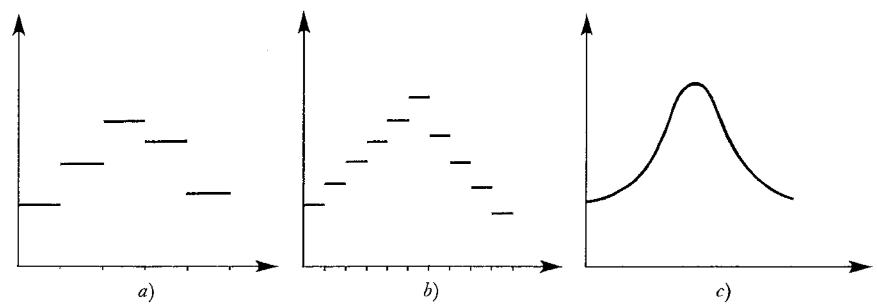


Figura 7.
La curva normale come limite di una successione di funzioni a scala.

gionò... il suo impareggiabile spirito quando egli aveva ventiquattro anni, e si inoltrava a grandi passi in quelle regioni ancora vergini che dovevano diventare l'impero della matematica moderna» [Bell 1937, trad. it. p. 245]. Proprio nel primo giorno del secolo XIX era stato scoperto un nuovo pianeta, l'asteroide Cerere; poche settimane piú tardi il piccolo corpo celeste era stato perduto di vista. Ora, «calcolare un'orbita per mezzo degli scarsi dati dei quali si disponeva era un compito da preoccupare perfino Laplace; Newton aveva dichiarato che tali problemi sono i piú difficili in astronomia matematica. I calcoli necessari per determinare l'orbita di Cerere con esattezza sufficiente perché essa non sfuggisse ai telescopi nel suo giro intorno al sole, scoraggerebbero forse molte macchine calcolatrici ai nostri giorni» [*ibid.*]. Ma non Gauss. In questa sede non importa tanto il successo conseguito da Gauss nel problema originario, quanto lo schema intellettuale che sottende i risultati della sua *Theoria combinationis observationum erroribus minimis obnoxiae*: «La stima di una grandezza a partire da osservazioni, con un errore piú o meno grande, può essere paragonata a un gioco d'azzardo in cui non si può che perdere e in cui ogni errore corrisponde a una perdita. Il rischio, in un gioco del genere, è misurato dalla perdita probabile. Pertanto la questione di sapere quale perdita deve essere assegnata a un errore dato non è affatto chiara. In realtà, la scelta di questa perdita dipende, almeno in parte, dalla nostra valutazione» [1821, p. 7]. Gauss ricorreva qui al metodo che fu detto dei minimi quadrati per fornire una legge degli errori di osservazione, metodo delineato tre anni prima da Legendre e consistente nel scegliere il quadrato dell'errore come valutazione della perdita: in questo modo ottenne la «legge normale», cioè «il suggerimento rappresentativo della curva normale, o curva a campana, che ancor oggi porta il suo nome, in cui anche l'occhio vede che, in linea teorica, la frequenza va da un massimo per l'errore zero a un minimo per l'errore massimo» [Maros Dell'Oro 1965, p. 408]. Laplace, pochi anni dopo, scelse invece il valore assoluto dell'errore come valutazione della perdita probabile, ottenendo non la curva normale, ma quella che venne poi detta «prima legge di Laplace».

È comunque interessante rivedere l'approccio di Gauss alla luce dell'atteggiamento bayesiano (per cui si vedano in questa stessa *Enciclopedia* gli articoli «Induzione statistica», VII, in particolare pp. 390 sgg., e «Probabilità», X, in particolare pp. 1158 e 1185-86). I dati del problema originario della *Theoria combinationis* erano costituiti da n misurazioni della grandezza considerata $x_1, \ldots, x_n$ e si chiedeva di determinare la probabilità che un valore $x \in \mathbf{R}$ fosse il suo valore «vero»; in termini bayesiani si tratta di un problema di determinazione di una probabilità finale. Il teorema di Bayes garantisce la verità della

$$P(x|x_1, \ldots, x_n) = \frac{P(x)P(x_1, \ldots, x_n \mid x)}{\int_{-\infty}^{+\infty} P(x_1, \ldots, x_n|x)P(x)\,dx} \tag{5}$$

che consente una soluzione del problema quando si sia in grado di determinare la probabilità iniziale, $P(x)$, e la verosimiglianza $P(x_1, \ldots, x_n \mid x)$. Gauss suppose $P(x)$ distribuita uniformemente, e cioè che qualunque valore della grandezza

incognita fosse ugualmente probabile, sulla base della considerazione che non si disponeva di alcuna ragione per supporre piú probabile un qualunque particolare intervallo di valori della grandezza. Si trattava evidentemente di un'applicazione del principio d'indifferenza o di ragion non sufficiente. Questa ipotesi consente di eliminare dalla (5) il fattore $P(x)$ riducendo la determinazione della probabilità finale a quella della verosimiglianza. A questo scopo Gauss suppose che *a*) le misurazioni sono indipendenti; *b*) la probabilità di commettere un errore in ciascuna di esse è funzione solo di quest'ultimo; *c*) tale funzione è simmetrica, continua e derivabile; *d*) essa assume il suo massimo in corrispondenza della media aritmetica delle misurazioni effettuate. Da queste ipotesi, si deriva la

$$P(x) = (\sigma\sqrt{2\pi})^{-1} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \tag{6}$$

che costituisce la forma analitica generalizzata della curva normale, dove $\mu \in \mathbf{R}$ è il valor medio di x e $\sigma \in \mathbf{R}^+$ è lo scarto (quadratico medio) di x. Ponendo $\mu = 0$ e $\sigma = 1$, si ottiene il caso speciale della curva normale considerato a proposito di Moivre (cfr. la (4) a p. 597).

Il merito di aver introdotto la curva normale spetta dunque a Gauss, che vi era pervenuto muovendo dal problema degli errori di misurazione *via* il teorema di Bayes oppure di Moivre e quindi di Laplace, che l'avevano ricavata come espressione limite del modo di disporsi dei risultati di un gioco? Discussioni circa la priorità di una scoperta (si tratti della curva normale come, poniamo, della legge di conservazione dell'energia o della rappresentazione geometrica di numeri complessi, ecc.), come già in altri articoli della presente *Enciclopedia* si è via via rilevato, sono significative nella misura in cui permettono di cogliere le differenze tra i contesti della scoperta in cui si inscrivono i risultati dei singoli ricercatori (altrimenti scadono al rango di curiosità da bottega per eruditi e collezionisti maniaci di aneddoti stravaganti). Nel caso qui trattato, la derivazione di Moivre e quindi di Laplace risolve un problema di probabilità diretta mentre la derivazione di Gauss prende le mosse dal tentativo di risolvere un tipico problema di probabilità inversa. A parere di chi scrive, inoltre, la differenza contestuale fa inclinare la bilancia al punto di vista di Gauss: sotto il profilo di una «logica della scoperta scientifica» (per riprendere una locuzione cara a Karl Popper) che valuti la crescita scientifica in termine di accresciuta capacità di modellizzazione dei processi reali, è l'approccio gaussiano che pare maggiormente fertile sul piano euristico. Com'è noto, «la curva di Gauss per la distribuzione degli errori delle misure risultava [per esempio]... adatta anche per la velocità delle molecole di un gas (in base alla formula $p = mNV^2/3$, dove p è la pressione, m la massa molecolare, N il numero delle molecole e V^2 il valore medio del quadrato della velocità), per la distribuzione delle pallottole intorno al centro di un bersaglio, per l'altezza o il perimetro toracico dei chiamati alla visita di leva, per il quoziente di intelligenza, ecc.» [Maros Dell'Oro 1965, p. 357]. Se Gauss si limitò solo ad applicazioni sul campo dell'astromonia, Quételet (cfr. quanto si è già detto alle pp. 595-96) faceva della curva normale uno strumento di modellizzazione indispensabile in altri campi.

Se le ricerche di Quételet mostravano che la legge di Gauss governava, poniamo, anche la distribuzione dei toraci dei giovani scozzesi o quella della statura dei coscritti francesi, mezzo secolo dopo Francis Galton, nel suo *Natural Inheritance* (1889), riconosceva in essa «la suprema legge dell'irrazionale», qualcosa che tipicamente mancava alla razionalità del pensiero della Grecia classica (a fortiori, dunque, qualcosa di tipicamente *moderno*), ma che «sarebbe stata personificata dagli antichi Greci e deificata se l'avessero conosciuta» [citato e discusso in Scardovi 1978, p. 12]. Con Galton il programma di Quételet (e Gauss) conobbe un ulteriore slittamento creativo: fondatore dell'eugenica, ideatore del sistema delle impronte digitali per individuare i delinquenti, studioso di meteorologia, Galton mirava a ritrovare la curva degli errori in «collettivi» di grandezze omogenee variabili, compiendo cosí un'ulteriore generalizzazione.

Importanti contributi successivi a una chiarificazione concettuale e tecnica del ruolo della curva normale sono non solo i lavori di Pearson (per cui si veda l'articolo «Distribuzione statistica» in questa stessa *Enciclopedia* (IV, p. 1222)), che generalizzò la (2) a un'ampia classe di funzioni appropriate alla rappresentazione analitica di seriazioni statistiche, ma anche quelli meno noti di Bortkiewicz sui cosiddetti fenomeni rari. Questi scopre [1898] che un'altra legge probabilistica, quella che a suo tempo Poisson aveva dedotto come caso limite della distribuzione binomiale quando le probabilità erano molto piccole, trovava conferma nell'osservazione sperimentale: per esempio, nella distribuzione dei morti per calcio di cavallo nell'esercito prussiano rispetto ai corpi d'armata e agli anni.

Questi risultati di Pearson e Bortkiewicz sono rappresentativi della linea di sviluppo della statistica dall'inizio di questo secolo, caratterizzata dall'abbandono della «metafisica influente» di Quételet, giocata sul ruolo privilegiato della curva normale, e dalla costruzione di una molteplicità di modelli statistici piú adatta a caratterizzare tutta la varietà dei fenomeni casuali. [S. M. e B. M.].

Bell, E. T.
1937 *Men of Mathematics*, Simon and Schuster, New York (trad. it. Sansoni, Firenze 1966).

Boldrini, M.
1965 *Teoria della statistica*, in M. Boldrini (a cura di), *Teoria e metodi della statistica*, vol. I, Giuffrè, Milano, pp. 3-341.

Bortkiewicz, L. von
1898 *Das Gesetz der kleinen Zahlen*, Teubner, Leipzig.

Brugnoli, A., e Messori, F.
1980 *Appunti di statistica descrittiva*, Edagricole, Bologna.

De Finetti, B.
1970 *Teoria delle probabilità. Sintesi introduttiva con appendice critica*, Einaudi, Torino.

Foucault, M.
1963 *Naissance de la clinique. Une archéologie du regard médical*, Presses Universitaires de France, Paris (trad. it. Einaudi, Torino 1977[3]).

Gauss, C. F.
[1821] *Theoria combinationis observationum erroribus minimis obnoxiae. Pars prior*, in «Commentationes societatis regiae scientiarum Gottingensis recentiores», V, 1823; ora in *Werke*, vol. IV, Königliche Gesellschaft der Wissenschaften zu Göttingen, Göttingen 1873, pp. 1-26.

Goodman, N.
1955 *Fact, Fiction and Forecast*, Bobbs-Merrill, Indianapolis.

Graunt, J.
1662 *Natural and Political Observations... Made upon the Bills of Mortality*, Martyn and Allestry, London 1665[3].

Hesse, M. B.
1969 *Ramifications of "Grue"*, in «British Journal for the Philosophy of Science», XX, pp. 13-25.

Huff, D.
1954 *How to Lie with Statistics*, Norton, New York.

Kendall, M. G., e Stuart, A.
1969 *The Advanced Theory of Statistics*, Griffin, London.

Kline, M.
1972 *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York.

Livi, L.
1968 *Elementi di statistica*, Cedam, Padova.

Loève, M.
1978 *Calcul des probabilités*, in J. Dieudonné e altri, *Abrégé d'histoire des mathématiques (1700-1900)*, vol. II, Hermann, Paris, pp. 277-313.

Maros Dell'Oro, A.
1965 *Storia della statistica*, in M. Boldrini (a cura di), *Teoria e metodi della statistica*, vol. I, Giuffrè, Milano, pp. 345-532.

Merz, J. Th.
1903-904 *A History of European Thought in the Nineteenth Century*, Blackwood, Edinburgh 1903-904[2]; ed. Dover, New York 1965.

Pearson, K.
[1921-23] *The History of Statistics in the 17th and 18th Centuries*, Griffin, London 1978.

Popper, K. R.
1959 *The Logic of Scientific Discovery*, Hutchinson, London (trad. it. Einaudi, Torino 1970).

Quételet, A.-J.-L.
1835 *Sur l'homme et le développement de ses facultés; ou essai de physique sociale*, Bachelier, Paris.

Quine, W. van Orman
1969 *Natural Kinds*, in *Ontological Relativity and Other Essays*, Columbia University Press, New York.

Runyon, R. P., e Haber, A.
1976 *Fundamentals of Behavioral Statistics*, Addison-Wesley, Reading Mass. (trad. it. Inter European Editions, Amsterdam 1977).

Scardovi, I.
1978 *Adolphe Quételet tra determinismo e accidentalismo*, in «Atti dell'Accademia delle Scienze dell'Istituto di Bologna. Classe di scienze morali», LXVI, 2, pp. 5-44.

Todhunter, I.
1865 *A History of the Mathematical Theory of Probability*, Macmillan, Cambridge-London.

L'**osservazione** di una vasta classe di fenomeni (cfr. **fenomeno**) – quelli cosiddetti casuali (cfr. **caso/probabilità, determinato/indeterminato**) – dà luogo a una varietà di rappresentazioni (cfr. **rappresentazione**) grafiche a carattere necessariamente finito e discreto, dati i vincoli (cfr. **vincolo**) naturali sulle possibilità di rilevazione dei dati (cfr. **dato**) da parte di osservatori – siano essi umani (cfr. **uomo**; e anche **soggetto/oggetto**) o macchine (cfr. **macchina**) – in un tempo (cfr. **tempo/temporalità**) finito. Questa limitazione si applica tanto nel caso di osservazioni di carattere qualitativo (cfr. **qualità/quantità**) quanto in quello di misurazioni (cfr. **misura**). Ma accrescendo la grandezza del campione, tali rappresentazioni si approssimano (cfr. **approssimazione**) a un profilo continuo (cfr. **continuo/discreto**). Di qui il tentativo d'individuare **funzioni** analitiche (cfr. **locale/globale**) corrispondenti non solo per ragioni di semplicità (cfr. **semplice/complesso**) matematica, cioè come strumenti di **calcolo**, ma anche come **ipotesi** sulla natura profonda dei fenomeni (cfr. anche **reale**) modellizzati (cfr. **modello**, **teoria/modello**). Caso tipico è la celebre distribuzione (cfr. **distribuzione statistica**) di **probabilità** nota come «curva normale», di così larga applicazione nella **induzione statistica**.

Teoria/pratica

1. *Le «scuole dell'irragionevolezza».*

Τοῦ γὰρ εἶναι δοκοῦντος ἀγαθοῦ χάριν πάντα πράττουσι πάντες 'Proprio in grazia di quel che pare bene tutti compiono tutto'. Cosí Aristotele nella *Politica* [1252a, 2-3] e Samuel Butler parafrasa all'inizio di *Erewhon*: «Non si dà azione che non si fondi su un equilibrio di considerazioni». Negli uomini – si legge nell'*Etica Nicomachea* [1113a, 24-25] – «oggetto della volontà è ciò che sembra bene: per chi è virtuoso ciò che è veramente bene, per chi è vizioso quello che capita». Oggetto della volontà è il fine, mentre gli oggetti «della deliberazione e del proposito sono i mezzi che riguardano il fine»; questi sono tipicamente gli ambiti della virtú [*ibid.*, 1113b, 4-5]; ché la virtú (ἀρετή) è rettitudine del proposito, la saggezza (φρόνησις) è rettitudine della deliberazione [cfr. in particolare *ibid.*, 1142b]. Né la distinzione concettuale significa reale separazione: la saggezza senza la virtú morale sarebbe mera «accortezza», capacità che, come già avvertiva Platone nella *Repubblica* [519a] degenera in malizia quando è messa al servizio dell'«animuccia propria dei malvagi sapienti»; la virtú senza saggezza non è nemmeno piena ἀρετή (nel senso anche di 'eccellenza'), semmai è semplice «disposizione alla virtú» che opportunamente condotta può evolvere in «eccellenza».

Ma «si delibera su ogni cosa, e ogni cosa può essere oggetto di deliberazione, oppure intorno ad alcune cose non vi può essere deliberazione?» [*Etica Nicomachea*, 1112a, 18-20]. La risposta di Aristotele («Ciascun uomo delibera intorno alle cose che egli stesso può realizzare. Quanto poi alle scienze esatte e autosufficienti non si può deliberare: ad esempio riguardo ai segni grafici, non v'è dubbio su come si debbano scrivere le lettere» [*ibid.*, 1112a, 35 - b, 1-2]) ricompare oggi autorevolmente [Rawls 1971, per esempio p. 49]: se, ad esempio, si disponesse di un'accurata spiegazione dei moti celesti che pure non viene trovata «attraente» non per questo si sarebbe in grado di mutare i movimenti dei corpi celesti per renderli conformi a una teoria piú attraente. Che si disponga di una meccanica celeste «elegante» è, forse, solo un caso fortunato.

Ma una volta situati nell'opportuno quadro teorico i principî della grammatica di una lingua o strutturata in un dato paradigma la meccanica celeste, non sono inconcepibili revisioni o cambiamenti radicali: la storia delle «rivoluzioni scientifiche» [cfr. Kuhn 1962, e l'articolo «Paradigma» in questa stessa *Enciclopedia*] lo mostra abbondantemente. E revisioni e rivoluzioni sono promosse – nell'impresa scientifica come in altre attività – in nome di valori che si strutturano in «ideali» che possono anch'essi cambiare nel tempo e nello spazio (cfr. l'articolo «Scienza» in questa stessa *Enciclopedia*). «Una teoria, per quanto elegante e economica, deve essere rifiutata o rivista, se non è vera; allo stesso modo leggi e istituzioni, non importa quanto efficienti e ben congegnate, devono essere riformate o abolite, se non sono giuste» [Rawls 1971, p. 3]. E questo è già un ideale nel senso precisato, anzi uno schema di ideale che andrà «riempito» indicando cosa s'intende per «vero» o per «giusto», ecc.

L'equilibrio di considerazioni che nei termini piú generali sembra sottendere una gamma assai vasta di «forme di vita» – dai paradigmi consolidati al disegno delle istituzioni – è dunque una sorta di «equilibrio riflessivo» (riflessivo proprio perché, almeno parzialmente, sono noti i principî cui si conformano i nostri giudizi e le premesse da cui essi derivano [cfr. *ibid.*, pp. 20, 48-51]) non immune dal gioco delle innovazioni che producono a vari livelli il mutamento (cfr. gli articoli «Equilibrio/squilibrio», «Innovazione/scoperta», «Semplice/complesso», «Sistema» in questa stessa *Enciclopedia*). La determinazione delle condizioni in cui questo equilibrio va conservato o invece abbandonato è il problema che Goethe enunciava nei termini negativi della scelta tra ingiustizia e disordine (cfr. del resto le osservazioni conclusive del già citato articolo «Equilibrio/squilibrio»): la questione s'intreccia dunque a quella di definire cosa intendere come «comportamento razionale».

Il riferimento a quelle «rotture dell'equilibrio» che sono le rivoluzioni scientifiche permetterà di mettere in luce qualche aspetto della questione. Per esempio: a fronte delle innovazioni promosse da Galilei o da Descartes già un Vico – a detta almeno di Habermas [1971, trad. it. p. 81] – sarebbe stato consapevole di una «perdita di forza ermeneutica nella penetrazione teorica di situazioni da dominare praticamente». E nel vichiano *De nostri temporis studiorum ratione* [1708] si legge: «Non operano saggiamente coloro che negli usi pratici della prudenza civile s'avvalgono dei medesimi criteri di giudizio adoperati dalla scienza» (trad. it. p. 194). Piú rigidi sono i dettami del Metodo piú si accresce l'incertezza nell'agire; per dirla ancora con Habermas, l'«oggettivazione scientifica» è cosí intenzionalmente lontana dalla «prassi della vita» che la stessa applicazione delle cognizioni mediante essa acquistate rimane incontrollata [1971, trad. it. pp. 83-84].

L'annunzio baconiano di una filosofia pratica come scienza e il progetto caro a Hobbes di dare corpo a questa speranza in nome di una «Ragione» che è soprattutto «Calcolo» (cfr. del resto la citazione all'inizio dell'articolo «Calcolo» in questa stessa *Enciclopedia*, II, p. 373) non finiscono per tramutarsi nell'utopia negativa degli abitanti di Erewhon descritta nel XXI capitolo dell'opera di Butler? «La vita, sostengono, sarebbe intollerabile se gli uomini in tutti i loro atti fossero guidati dalla ragione e soltanto dalla ragione. Questa ci travia inducendoci a tracciare linee troppo rigide e precise, e a definire ogni cosa attraverso il linguaggio – il quale linguaggio, come il sole, prima dà vita e poi inaridisce. La logica conduce alle idee estreme, ma le idee estreme sono sempre assurde, mentre il giusto mezzo è illogico». Proprio adducendo, del resto, le vicende dell'impresa scientifica da Galileo ad Einstein o a Bohr, questo punto di vista è portato alle estreme conseguenze da quel maestro delle erewhoniane «scuole dell'Irragionevolezza» che è Paul Feyerabend. «I razionalisti, – scrive quest'ultimo [1980], – vogliono che ci si comporti sempre in modo razionale; ossia che si prendano decisioni secondo regole e criteri che essi e i loro amici considerano importanti e fondamentali. L'esempio della scienza

indica che un tale comportamento non conduce ad alcun risultato: il mondo fisico è troppo complesso per poter essere dominato e compreso con l'ausilio di metodi 'razionali'. Ma il mondo sociale, il mondo del pensiero e del sentimento umano, della fantasia umana, il mondo della filosofia, della poesia, delle scienze, il mondo della convivenza politica è ancora piú complicato. Ci si deve forse attendere che i razionalisti abbiano successo in questo mondo, dopo aver fallito nel mondo fisico? E non è meglio perciò fondare il comportamento sociale su decisioni concrete di esseri umani che conoscono con precisione il loro ambiente, cosí come i desideri, le attese, le speranze, le fantasie dei loro simili, piuttosto che affidarsi alle regole di dotti che si sono trovati di fronte a questo ambiente al piú nei libri dei loro colleghi, nei quali esso appare inoltre gravemente deformato?» (trad. it. pp. 31-32).

2. *«Scienza» e «saggezza».*

Poiché per Vico come per Feyerabend «Aristotele non è un cane morto» [Feyerabend 1978, pp. 53-65], sarà interessante muovere proprio dalle aristoteliche determinazioni della differenza tra scienza e saggezza, tra ἐπιστήμη e φρόνησις: mentre la prima mira a «verità eterne», la seconda ha a che fare solo con ciò che è «verisimile». (Dunque qui è la motivazione della ostilità di Vico alla nuova – e antiaristotelica – scienza matematizzante: la saggezza avendo minori pretese dal punto di vista teorico porta nella pratica a una maggior sicurezza. Qui è pure la radice della feyerabendiana preferenza per «le decisioni concrete di esseri umani che conoscono con precisione il loro ambiente»: tramontato l'ideale della *scientia aeterna*, la rigidità delle norme e dei canoni è ostacolo alla stessa crescita scientifica).

Ma la caratterizzazione aristotelica del ragionamento del φρόνιμος 'saggio' o 'prudente' come di colui che «deve sapere entrambe le cose, l'universale e il particolare, ma soprattutto questo» [*Etica Nicomachea*, 1141b, 22-23], attraverso il cosiddetto sillogismo pratico, la cui conclusione è la προαίρησις 'decisione', è una semplice «mostruosità dal punto di vista della teoria della conoscenza di Aristotele..., una *contradictio in adiecto* come un 'quadrato rotondo'. Ma resta... un contrassegno per una difficoltà» [Joachim 1902-17, p. 209]. E nel libro VI dell'*Etica Nicomachea* la difficoltà si scinde in due distinte ἀπορίαι, vere aporie fondatrici [per questa terminologia, cfr. Thom 1980, in particolare pp. 149-50].

La prima concerne l'utilità e della scienza e della saggezza. La «scienza» (nel quadro concettuale e linguistico di questo articolo: la «teoria») «non riguarda nessuna delle cose che si generano», dunque non considera alcuna delle cose per cui l'uomo può essere felice [*Etica Nicomachea*, 1143b, 20]; la «saggezza» (cioè la «pratica») «si occupa bensí di ciò, ma tuttavia che bisogno vi è di essa?» [*ibid.*, 20-21]. Anche se la saggezza riguarda ciò che per l'uomo è bello, giusto, buono, cioè le azioni che è compito dell'uomo retto compiere, non è per il fatto di essere consapevole di ciò che l'uomo diventa piú capace di agire, proprio come non diventa piú capace in quel che riguarda salute o robustezza per il fatto che possiede la medicina o la ginnastica. E anche se si ammette che la saggezza contribuisce alla rettificazione del carattere, «per quelli che sono già virtuosi essa non sarebbe affatto utile; ma non lo sarebbe neppure per quelli che non lo sono: infatti non importa per nulla che abbiano la saggezza essi stessi oppure che diano ascolto a coloro che la posseggono» [*ibid.*, 30-33], proprio come, del resto, è il malato e non il sano che ricorre ai servigi della medicina; ma non studia medicina egli stesso, si limita a chiamare il medico.

Tuttavia [si veda in particolare *ibid.*, 1144a, 3-6] scienza e saggezza sono un bene in sé, semplicemente in quanto «virtú»: la scienza, se anche non prende in considerazione «nessuna delle cose per cui l'uomo può essere felice» [*ibid.*, 1143b, 19], produce la felicità, «non però come la medicina produce la salute, ma come la salute produce la salute» [*ibid.*, 1144a, 4], dunque per il fatto stesso di essere posseduta. La saggezza, infine, costituisce felicità perché l'opera dell'uomo – in quanto animale politico e buon cittadino – è semplicemente incompatibile senza di essa.

La seconda aporia concerne conseguentemente le relazioni di «scienza» e «saggezza» con la politica. «Si può poi ritenere una forma di conoscenza il conoscere ciò che è utile a sé; ma occorre far molta distinzione. Cosí sembra che sia saggio chi conosce ciò che lo riguarda e si dedica solo a ciò, mentre invece gli uomini politici si occupano di molte cose» [*ibid.*, 1141b, 35 - 1142a, 3]. I politici debbono dunque realizzare una sorta di ideale di vita garantendo l'autonomia dei filosofi proprio come i medici eliminano gli ostacoli al libero dispiegamento della salute: esercitando dunque un controllo sul malato nell'interesse della salute, non un controllo sulla salute medesima.

E il saggio, in quanto anch'egli esercita una «virtú architettonica», pare collocarsi su un piano superiore a quello del politico. Questi mediante decreti si limita a mettere in atto (prontamente) quello che il saggio delibera (con la necessaria lentezza) [cfr. *ibid.*, 1141b, 25-30].

Ma questa «fabbrica dell'universo» etico e politico insieme (poiché in Aristotele la politica, come dottrina della vita associata buona e giusta, è naturale prosecuzione dell'etica: cfr. del resto l'articolo «Politica» in questa stessa *Enciclopedia*, X, in particolare pp. 855-57) è proprio in un equilibrio che l'innovazione scientifica prodotta dai Galilei, dai Keplero, dai Newton, ecc. è destinata col tempo e in un processo tutt'altro che lineare a rompere. Com'è noto, Aristotele distingueva tra «saggezza» e «arte», tra φρόνησις e τέχνη proprio in quanto l'una era intesa alla prassi in senso stretto (πρᾶξις) e l'altra alla creazione (ποίησις): etica e politica, in ultima istanza, si rivolgevano sempre alla formazione del carattere e poco avevano in comune con «l'abilità» dell'artigiano e/o dell'artista nel produrre opere. Nel quadro di un Hobbes in cui la politica è, *in primis*, «artificio» (cfr. il citato articolo «Politica», pp. 857 sgg.) la nuova scienza – a un tempo matematizzante ed efficace – è un elemento influente. «Per Hobbes... la massima di Bacone *scientia propter potentiam* è già ovvia; la specie umana deve il suo progresso massimamente alla tecnica», che include anche la tecnica politica del corretto ordinamento dello Stato [Habermas 1971, trad. it. p. 78]. Emerge qui il ruolo della tecnica come «capacità di affrontare compiti oggettivati» – dalla realizzazione delle macchine a quella dello «animale artificiale» che costituisce lo Stato.

Per coloro che individuano un tratto tipico della «modernità» nella cancellazione della distinzione aristotelica [cfr. Arendt 1958; Gadamer 1960; Habermas 1971, ecc.] Kant è forse l'esito piú coerente: qui il comportamento etico dell'individuo, libero solo internamente, è chiaramente distinto dalla legalità delle sue azioni esterne. E, analogamente, è separato dalla politica: quest'ultima – come l'abilità di chi costruisce o manipola le macchine piú diverse – rientra nella «competenza tecnica propria di una dottrina utilitaristica della saggezza» [Habermas 1971, trad. it. p. 78].

3. *Casi e regole.*

Lo slittamento del problema da Hume a Kant costituirà alla luce di quanto detto una pietra di paragone. Nel suo *Treatise of Human Nature* Hume aveva sottoposto a critica la «chimera» tipica delle «disquisizioni metafisiche» e anche di quelle «popolari» secondo cui «ogni creatura razionale ha l'obbligo di regolare le proprie azioni secondo i dettami della ragione» [1739, trad. it. p. 433] introducendo una distinzione che positivisti logici e filosofi analitici ampiamente riprenderanno (cfr. l'articolo «Etica» in questa stessa *Enciclopedia*, V, pp. 917 sgg.): «La ragione è la scoperta della verità o della falsità. La verità e la falsità consistono in un accordo o in un disaccordo o con le *reali* relazioni delle idee, o con l'esistenza e i dati di fatto *reali*. Perciò qualsiasi cosa non sia suscettibile di questo accordo o disaccordo non può essere né vera né falsa, e non può mai essere oggetto della nostra ragione. È ora evidente che le nostre passioni, volizioni e azioni non sono suscettibili di un simile accordo o disaccordo, poiché sono dei fatti e delle realtà originari, completi in se stessi, e che non implicano alcun riferimento ad altre passioni, volizioni e azioni. Perciò è impossibile dichiararle vere o false, contrarie o conformi alla ragione» [*ibid.*, p. 484; cfr. anche pp. 433-39]. Ma non si cade per questa via in un altro celebre luogo comune nelle «declamazioni popolari», l'irrilevanza della «comprensione» per l'«azione»?

Da un quesito del genere prende le mosse Kant: l'intelletto dell'uomo, lo *human understanding* di Hume, si limita a prender atto della legalità naturale (cfr. del resto sopra, p. 168); le azioni umane nei loro «effetti fenomenici» saranno via via oggetto delle discipline pertinenti (dalla fisica alla fisiologia, ecc.). L'«idea cosmologica di libertà» e la sua relazione con «la necessità universale della natura» [1787, trad. it. pp. 446 sgg.] va invece spiegata attribuendo alla ragione pratica e non semplicemente speculativa un tipo di *causalità* che si dispiega proprio negli *imperativi* «che nell'intero dominio pratico assegniamo come regole alle nostre attività» [*ibid.*, p. 449]. Kant realizza il nesso tra «teoria» e «pratica» non solo come semplice correlazione mezzi-fini ma anche – e soprattutto – come riconduzione di casi a regole. Se il nome di 'teoria' andrà allora riservato alle regole «pensate come principî generali», facendo astrazione inevitabilmente da una quantità di condizioni che pure hanno influenza sulla loro applicazione, a sua volta 'pratica' non denoterà qualsiasi atto ma solo «quello che attua uno scopo ed è pensato in rapporto a certi principî della condotta rappresentati nella loro generalità» [1793, trad. it. p. 237]. E si ritrova sottostante il problema del «termine medio» (cioè dello schematismo): «Che tra la teoria e la pratica vi debba anche essere un termine medio di congiunzione e di passaggio dall'una all'altra, sia pure la teoria perfetta quanto si vuole, è evidente. Infatti al concetto intellettivo che contiene la regola deve aggiungersi un atto del giudizio, per il quale l'uomo pratico distingue se il caso cade o no sotto la regola» [*ibid.*]. La necessità di evitare il regresso all'infinito (un hegeliano direbbe «la cattiva infinità») spiega anche perché il nesso teoria-pratica possa slittare in una contrapposizione teoria/pratica: «Siccome per il giudizio non si possono dar sempre dinuovo regole a cui rivolgersi nella sussunzione (poiché ciò andrebbe all'infinito) cosí può avvenire che vi siano teorici che nella loro vita non possono mai diventar pratici, perché ad essi manca la capacità di giudicare. Cosí, ad esempio, si hanno medici e giureconsulti, che hanno compiuto ottimamente i loro studi, ma che, dovendo dare un consiglio, non sanno come fare» [*ibid.*].

Ma non è lecito «tenere in dispregio la teoria». Infatti «del meccanico empirico, nel caso della meccanica in generale, o dell'artigliere, nel caso della dottrina matematica della balistica, che volessero affermare che la teoria relativa è certamente costruita sottilmente, ma nella pratica non ha valore, perché nell'applicazione l'esperienza dà tutt'altri risultati della teoria, non si può che ridere (infatti, se nel primo caso si aggiunge anche la teoria dell'attrito, nel secondo quella della resistenza dell'aria, cioè se si fa in generale ancor piú teoria, questa si accorderà perfettamente con l'esperienza)» [*ibid.*, p. 238]. Quindi comportarsi razionalmente vuol dire prendere decisioni sulla base delle *migliori* teorie che la scienza dell'epoca mette a disposizione.

4. *La teoria dell'utilità.*

4.1. Problematica generale.

Da quanto detto nel § 3 deriva che «la conoscenza delle regolarità della natura e delle condizioni in cui occorrono le trasformazioni del mondo fenomenico rendono possibile all'uomo manipolare queste condizioni stesse e produrre nuove possibilità di azione... Cosí, la funzione della scienza consiste nell'espandere l'insieme delle possibilità a disposizione degli uomini; la funzione del piano razionale [per la pratica] è quello di istituire un ordine per le azioni entro tali insiemi di possibilità in espansione» [Ostrom 1964, pp. 88-89].

Il modello dell'azione razionale come coordinazione di mezzi e fini e quello della sussunzione di casi sotto regole si possono entrambi riportare, dunque, a uno schema piú generale. Tanto per fissare le idee, si tratta, dato un insieme di

«possibilità a disposizione» (cioè di possibili azioni realizzabili) di sceglierne una che massimizzi (o minimizzi) un certo indice. Simbolicamente: sia $\mathbf{x}$ un'azione di un dato insieme F di azioni realizzabili e sia $f(\mathbf{x})$ un indice (una funzione) che in un qualche senso «valuta» $\mathbf{x}$: si tratta allora di trovare (almeno) un $\mathbf{x}^*$ in F tale che $f(\mathbf{x}^*) \geq f(\mathbf{x})$ per ogni $\mathbf{x}$ di F. Ora, per dirla di nuovo con Aristotele, «se si chiamasse sapienza quella che riguarda il nostro utile, vi sarebbero molte sapienze» [*Etica Nicomachea*, 1141a, 30-31]. Il primo punto problematico è qui la scelta appropriata dell'indice f. Quella particolare «sapienza» che è la teoria economica ha a lungo insistito che in molti contesti profitti e/o perdita rappresentano indici accettabili: e tuttavia in altri contesti non sono nemmeno pertinenti. Infine: «Il concetto di "utilità" è stato inteso, per secoli, come una sorta di qualità inerente a dei beni. Ci è voluto del tempo perché ci si accorgesse che non esiste niente del genere; si può parlare solo di preferenze e si può risolvere tutto il problema di base della misura dell'utilità solo adottando questo nuovo e differente modo di considerare il vecchio fenomeno del "valore" economico» [Morgenstern 1966, p. 133].

Si consideri un semplicissimo caso: un certo individuo, per esempio Clemente, si reca in una libreria per comprarsi un libro. Il problema è allora: come Clemente riduce, mediante la selezione di un indice, la scelta di un libro all'acquisto di uno di indice massimo? «Da un punto di vista strettamente operazionistico, potremmo addirittura sopprimere il problema: l'unica cosa da fare sarebbe quella di osservare che cosa di fatto acquista» [Luce e Raiffa 1957, p. 15]. Ma questo *wait and see* non è molto soddisfacente. Perché non osservare invece il comportamento di Clemente in situazioni piú ristrette – per esempio sottoponendogli qualche catalogo di novità librarie o qualche recensione e cercando di capire i suoi gusti dalle sue reazioni – e da questo predire quale sarà il suo acquisto? (Naturalmente alcuni – marxisti e freudiani per esempio – potrebbero obiettare che Clemente non sa esprimere o non conosce le sue «vere» preferenze: questo non è però un argomento contro il tipo di approccio qui delineato. Al piú indica che il modello dovrà essere sofisticato in modo da tener conto di eventuali «blocchi» o «condizionamenti» di Clemente (cfr. del resto in questa stessa *Enciclopedia* gli articoli «Apprendimento» e «Comportamento e condizionamento»). Comunque il comportamento di Clemente resta «razionale» nella misura in cui soddisfa i postulati di razionalità di cui si tratta poche righe piú oltre. È anche ovvio che la nozione di razionalità – o di comportamento razionale – enucleata dalla teoria dell'utilità non è quella di razionalità «perfetta» ma quella di razionalità «minima»: cfr. la successiva discussione alle pp. 180-81). Infine «se siamo in grado di ordinare le alternative e assegnare un indice numerico [cioè un f tale che $f(A) \geq f(B)$ se e solo se Clemente preferisce il libro A al libro B], allora siamo in grado anche di asserire in modo del tutto *tautologico* che l'individuo in questione ha scelto un'alternativa di indice massimo» [*ibid.*, p. 16]. La natura ordinale dell'indice va sottolineata; per fissare le idee, si supponga che i libri della libreria – cioè le alternative praticabili – siano soltanto tre, A, B, C. Si può compendiare la scelta di Clemente dicendo, per esempio, che se A è il libro che Clemente mette in cima alle sue preferenze e quindi vengono, nell'ordine, B e C, ai tre libri A, B, C toccano rispettivamente le «utilità» 3, 2, 1. Ma perché non 30 per A, 20,18 per B e 3,14 per C? In realtà qualsiasi tripla a, b, c con $a > b > c$ andrebbe bene. Il problema concettuale sottostante è stato affrontato nell'articolo «Numero» in questa stessa *Enciclopedia*, in particolare nel § 7: qui basta aggiungere che «ampi settori del pensiero economico possono venir conservati postulando la sola scala ordinale delle preferenze... per le alternative senza dover puntellare il tutto con delle latenti "utilità" [cardinali...] Si può sostenere però che introdurre dei numeri non porta nessun svantaggio, anzi consente di compendiare in modo compatto i dati ordinali» [*ibid.*].

Infine un cenno al nucleo matematico di quest'approccio alla decisione razionale. Esso risiede sostanzialmente nella dimostrazione del fatto che «se le preferenze di un dato individuo soddisfano certi assiomi di coerenza e di continuità, queste preferenze ammetteranno una rappresentazione in termini di una funzione di utilità ben definita (addirittura continua) [per la dimostrazione si veda ad esempio Debreu 1959, pp. 55-59]. Di conseguenza, per un tale individuo il comportamento razionale – che si è definito in questo modello formulato in termini di preferenze e alternative realizzabili – sarà equivalente alla massimizzazione dell'utilità (teorema di massimizzazione dell'utilità)» [Harsanyi 1976, p. 94]. Sui postulati di coerenza e continuità si tornerà, in un contesto piú generale, alla p. 180.

4.2. Un esempio: programmazione lineare e teoria dei giochi a due persone a somma zero.

Il tipo di situazioni piú sopra esaminate è sostanzialmente quello detto delle decisioni individuali in condizioni di certezza: svariati casi analizzati dalla teoria economica, dalla psicologia e dalle cosiddette scienze manageriali vi rientrano a pieno titolo. L'archetipo è, per molte situazioni, il modello che Walras ricalcò sulla grande tradizione della meccanica «quando cominciò a formulare il suo sistema di equazioni simultanee allo scopo di descrivere le interrelazioni complesse tra il prezzo e la produzione» [Morgenstern 1966, p. 131]. Sarà quindi interessante trattare in questa sede un esempio che richiede una genuina estensione di tale approccio, dato che nel caso gli usuali strumenti di calcolo – ricerca dei massimi e dei minimi di una funzione in analisi e princípî variazionali per ricercare le funzioni – si rivelano insufficienti.

Esempio: una dieta si può schematizzare nel modo seguente: 1) Si denotano n cibi con $C_1, \dots, C_n$. Una dieta allora non è altro che un insieme di prescrizioni circa la quantità di cibo che va consumata al giorno: x_1 unità di $C_1, \dots, x_n$ unità di C_n. A ciascuna dieta $\mathbf{x} = (x_1, \dots, x_n)$ si può associare quindi il suo potere nutritivo in relazione a tutte quelle sostanze che le conoscenze biologiche classificano come nutritive: ferro, calcio, vitamina C, ecc. Tale potere nutritivo sarà dato da un'espressione lineare della forma:

$$(1) \qquad a_1x_1 + a_2x_2 + \dots + a_nx_n$$

ove a_1 rappresenta la quantità poniamo di ferro per una unità di C_1, ecc. e analogamente per le altre sostanze nutritive. Naturalmente certi a_i (e x_i) possono essere 0, ma nessuno sarà negativo. 2) Gli esperti hanno poi stabilito certe esigenze minimali per una buona nutrizione. La dieta $\mathbf{x}$ è dunque sottoposta alla disuguaglianza lineare:

$$(2) \qquad a_1x_1 + a_2x_2 + \dots + a_nx_n \geq a$$

per il ferro, e ad analoghe disuguaglianze per le altre sostanze nutritive. 3) Vincoli come (2) non rappresentano ovviamente alcun problema se si possono considerare gli x_i, cioè le quantità giornaliere dei vari cibi C_i, prescindendo dai costi. Ma in molti casi – per esempio in un ospedale – si devono spesso scegliere «diete» che minimizzino i costi. Per ogni dieta $\mathbf{x} = (x_1, \dots, x_n)$ il costo verrà definito ovviamente dalla

$$(3) \qquad p_1x_1 + p_2x_2 + \dots + p_nx_n$$

ove p_i sono i prezzi delle unità dei cibi C_i.

Si tratta allora di scegliere una dieta che soddisfi le esigenze della nutrizione (i vincoli (2)) e minimizzi i costi (cioè la (3)) nei vincoli considerati. Si tratta di un caso esemplare dello schema della programmazione lineare in cui compaiono 1) delle azioni, ciascuna delle quali è rappresentata da un vettore a n componenti reali (le «diete»); 2) condizioni di realizzabilità, cioè disuguaglianze o uguaglianze che impongono dei vincoli sugli atti possibili (le «esigenze minimali della nutrizione»); 3) un indice associato a ciascuna azione che è la media ponderata degli n numeri che rappresentano l'azione (cioè la funzione che assegna a ciascuna azione il suo «costo»). Il problema è quello di trovare un'azione (cfr. 1)) che soddisfi i vincoli (cfr. 2)) e minimizzi il costo (cfr. 3)).

Ora, come si è visto nell'articolo «Giochi» in questa stessa *Enciclopedia*, in particolare nel § 5, il paradigma della programmazione lineare può essere soddisfacentemente «tradotto» in quello della teoria dei giochi, che ha il suo archetipo, notoriamente, nel fondamentale trattato di Neumann e Morgenstern [1947] e viceversa: piú precisamente, a ogni problema di programmazione lineare può venir associato un gioco a due persone a somma zero e viceversa; sicché, ogniqualvolta un problema di programmazione lineare è risolubile, la soluzione del problema può venir interpretata come soluzione del gioco associato e viceversa. La costituzione di un «manuale di traduzione» all'interno della teoria matematica della convessità [per un'esposizione dei risultati fondamentali cfr. ad esempio Luce e Raiffa 1957, pp. 408-23; e, per una sintesi, il citato articolo «Giochi», VI, p. 507 in particolare] è qui doppiamente interessante. Sotto il profilo euristico esso fornisce una linea di ricerca per non pochi problemi legati da una sorta di «aria di famiglia» a quelli tipicamente di programmazione lineare [per rilevanti esempi cfr. Luce e Raiffa 1957, pp. 18-19. Cfr. anche l'articolo «Decisione» in questa stessa *Enciclopedia*, IV, pp. 447-50]. Sotto il profilo epistemologico esso sottolinea la rilevanza dello slittamento, operato da Neumann e Morgenstern [1947], consistente nel sostituire alla nozione di guadagno quella di guadagno sperato nella teoria dei giochi a due persone a somma zero coll'ammettere strategie miste, cioè combinazioni aleatorie di strategie pure (cfr. ancora l'articolo «Giochi», pp. 804-5 in particolare). Un giocatore che ha a disposizione m strategie pure $s_1, ..., s_m$ può costruire una strategia mista $\mathbf{x} = x_1 s_1, ..., x_m s_m)$ ove gli x_k – con $0 \leq x_k \leq 1$ e $\sum_{k=1}^{m} x_k = 1$ – rappresentano le probabilità che quel giocatore associa alle strategie s_k. Per fissare le idee: si supponga $m = 2$. In tal caso la strategia $x = (s_1/2, s_2/2)$ deve essere rappresentata dall'esperimento consistente nel lanciare una moneta non truccata; e ogni altra strategia dal lancio di una moneta opportunamente truccata. (Allora, le strategie pure saranno determinate dal lancio di una moneta con due teste o due croci!) Ma perché legare il proprio destino all'esito di tali esperimenti? Si ponga che a uno dei due giocatori sia consentita una di queste due opzioni: la prima che non consente d'ignorare i dettami della moneta e la seconda che non obbliga invece ad attenervisi. «Dal momento che la seconda opzione include tutte le possibilità disponibili per il giocatore sotto la prima opzione, piú altre, sembrerebbe indubbio che sia questa che vada preferita... Sono le strategie pure che vanno tra loro raffrontate sulla base dei loro meriti specifici. Di conseguenza il concetto di strategia mista è un utile espediente matematico [come Neumann e Morgenstern mostrano] ma è completamente irrealistico» [Luce e Raiffa 1957, p. 75]. L'argomento piú comune in difesa della prima opzione – cioè delle strategie miste – è, notoriamente, che il carattere aleatorio di un gioco protegge ogni giocatore contro l'astuzia dell'avversario che, indovinando una scelta sicura, potrebbe allora adattarvi il proprio gioco. Ma questa difesa non è pertinente nel caso in cui «l'avversario» non è in alcun modo interessato al comportamento del giocatore, per esempio quando non è a conoscenza della matrice di pagamento del gioco o è una qualche entità impersonale («la natura», per esempio). In tal caso è meglio sostenere che «sotto il profilo psicologico la prima opzione va preferita alla seconda, contrariamente a quanto detto precedentemente, proprio perché non ci permette di cadere preda dell'umana fragilità» [*ibid.*]. È un po' come per un individuo che voglia seguire una dieta: rende pubblico il suo proposito e accetta la scommessa che non infrangerà la dieta, in modo che *dopo* non sarà libero di cambiar parere e di ottimizzare le sue azioni secondo i gusti di *quel* momento, per esempio divorando una moltitudine di pasticcini.

La «morale» della storia è comunque chiara: anche in condizioni di decisioni o in condizioni di certezza può essere «razionale» affidarsi all'alea di una moneta, di un dado, ecc.

5. *Decisioni in condizioni d'incertezza.*

5.1. Karl Popper e i «fantasmi della notte».

Un buon punto di partenza per una teoria delle decisioni individuali in condizioni d'incertezza (di cui quanto trattato nel § 4 risulterà un caso limite) sarà offerto dallo stesso brano di Kant [1793] citato alla fine del § 3 (cfr. p. 172).

Senza compromettersi a proposito di «teorie perfette», Kant prospetta la crescita teorica come una sequenza di teorie perfettibili che garantisce, in ultima analisi, la razionalità della pratica, intesa almeno come ποίησις. (A parte è l'autonomia relativa del comportamento morale: «Una vita indipendente dalle animalità e anche da tutto il mondo sensibile» [Kant 1788, trad. it. p. 313]. Autonoma, ma idealmente affiancata all'ordine cosmologico, in quanto dispiegamento dell'ordine interiore: «Il cielo stellato sopra di me e la legge morale dentro di me» [*ibid.*]).

Ma fino a che punto sono realistiche le scene della crescita della scienza – la ἐπίδοσις aristotelica di teorie perfette a teorie perfette e il continuo raffinamento teorico, immaginato da Kant, di un corpo di dottrine sostanzialmente incentrate sulla meccanica newtoniana? Nel 1899 il fisico Ludwig Boltzmann se paragonava lo sviluppo delle scienze della natura ancora fino agl'inizi dell'Ottocento «alla [ordinata] crescita di una vecchia città, che gradualmente si sviluppa grazie alle nuove costruzioni intraprese dai suoi industriosi cittadini», assimilava invece lo sviluppo successivo a quello di «una moderna metropoli americana», ove le vecchie costruzioni vengono spietatamente rase al suolo per far posto alle nuove [1899, ed. 1974 p. 77]. Si noti che Boltzmann diceva ciò *prima* delle due grandi rivoluzioni – relatività e quanti – che hanno cambiato nel Novecento l'immagine fisica del mondo.

Ora – ha scritto Imre Lakatos [1970] – ciò che pare caratterizzare una delle piú articolate epistemologie del nostro tempo, quella di Karl Popper, è che essa ha saputo cogliere «tutte le implicazioni del crollo della teoria scientifica meglio corroborata di tutti i tempi: la meccanica newtoniana e la teoria della gravitazione di Newton. Dal suo punto di vista, l'atteggiamento corretto non sta nella cautela nell'evitare errori, ma nella spietatezza nell'eliminarli. Audacia nelle congetture da un lato e severità nelle confutazioni dall'altro: questa è la ricetta di Popper. L'onestà intellettuale non consiste nel cercare di consolidare o stabilire la propria posizione dimostrandola (o "probabilificandola") – consiste piuttosto nello specificare con precisione le condizioni alle quali si accetta di rinunciare alla propria posizione» (trad. it. p. 165).

Non si entrerà nel merito di questa epistemologia (che è trattata in altri articoli di questa stessa *Enciclopedia*). Ma varrà la pena di prendere le mosse dalla soluzione negativa offerta da Popper al problema humeano dell'induzione. Popper riformula il problema logico dell'induzione nel modo seguente: «Siamo giustificati razionalmente nel ragionare da esempi o da controesempi di cui abbiamo avuto esperienza alla verità o alla falsità delle leggi corrispondenti, o a esempi dei quali non abbiamo avuto esperienza?» [1974*b*, p. 1020]. La risposta di Popper è che non siamo giustificati nell'inferire da un esempio la verità della legge corrispondente. Ma siamo giustificati nell'inferire da un controesempio la falsità di qualsiasi legge di cui esso sia appunto controesempio. La conoscenza teorica è tutta congetturale; ma «ci possono essere preferenze razionali per alcune delle congetture rivali: alcune possono essere migliori di altre, almeno in due sensi: possono essere piú informative e cosí piú interessanti, piú audaci; e possono resistere meglio a controlli piú severi» [*ibid.*, p. 1023].

Come molte «chimere» circa il comportamento razionale degli uomini smascherate dall'indagine spassionata delle reali «faccende della vita comune» e dalla retta filosofia, cosí l'induzione come «logica della scoperta scientifica» si dissolverebbe «come i fantasmi della notte all'apparire del mattino» (l'immagine è di Hume [1739, trad. it. p. 481]). Il problema logico dell'induzione è cosí liquidato.

Resta però un problema pragmatico dell'induzione. Mentre il teorico può anche non avere alcuna preferenza, «l'uomo di azione pratica» non può permettersi lussi del genere: egli «deve sempre *scegliere* tra alternative piú o meno definite, dal momento che anche *l'inazione è un tipo di azione*» [Popper 1974*b*, p. 1025]. Ma poiché ogni azione presuppone un insieme di aspettative, cioè di teorie sul mondo, a quale teoria si affiderà l'uomo d'azione? Ed esisterà qualcosa come una «scelta razionale»? Il problema pragmatico si scinde in due:

a) A quale teoria dovremmo affidarci per l'azione pratica, da un punto di vista razionale?

b) Quale teoria dovremmo preferire per l'azione pratica, da un punto di vista razionale?

a) Per Popper – e per altri della sua scuola [cfr. Watkins 1965], la corroborazione delle teorie è tipicamente analitica, ovvero dall'attribuzione a una teoria di un alto grado di corroborazione *non* segue alcuna predizione circa l'eventuale prosecuzione di tale successo in futuro. Dunque, conclude lo stesso Popper, «da un punto di vista razionale, non dovremmo fare "affidamento" su alcuna teoria, perché nessuna teoria si è dimostrata vera o di essa si può mostrare che sia vera ("affidabile")» [Popper 1974*b*, p. 1025]. In questo modo, come ha osservato Lakatos [1965], Popper lascia completamente senza risposta – nella sua ricostruzione della pratica scientifica – il problema della razionalità degli «uomini d'azione» che si affidano a tutte quelle teorie scientifiche che vengono di fatto applicate nella tecnologia.

b) Rispondendo a varie obiezioni del genere (Salmon, Lakatos, ecc.) Popper concede tuttavia che, se pur non si dà affidabilità nel senso di *a*), «dal momento che *dobbiamo* scegliere, sarà "razionale" scegliere la teoria meglio controllata» [Popper 1974*b*, p. 1025]. Riemerge qui la tematica già aristotelica della προαίρησις: «Dimentichiamo per ora le teorie che "usiamo" o "scegliamo" o su cui "basiamo le nostre azioni pratiche" e consideriamo solo il *proposito* o la *decisione* che ne risulta (fare X; non fare X; non fare nulla; e cosí via). Tale proposta può, speriamo, essere criticata razionalmente; e se siamo degli agenti razionali vorremo che essa sopravviva, se possibile, al maggior numero di controlli critici che possiamo passare in rassegna. *Ma questa critica userà liberamente le teorie scientifiche meglio controllate in nostro possesso.* Conseguentemente ogni proposta che ignora queste teorie (ovviamente, dove sono rilevanti) cadrà sotto i colpi della critica. Nel caso rimanga qualche proposta, sarà razionale adottarla. [Ma] perché, si potrebbe chiedere, la critica razionale usa teorie ben controllate, però poco affidabili? Tuttavia, la risposta è la stessa di prima. Decidere di criticare una proposta pratica dal punto di vista della medicina moderna (anziché, poniamo, in termini di frenologia) è in sé un tipo di decisione "pratica" (in ogni caso può avere conseguenze pratiche). Quindi la decisione razionale è sempre: adottare metodi critici che hanno essi stessi resistito a critiche severe» [*ibid.*, pp. 1025-26].

Come in Kant [1793] (cfr. p. 172), anche in questa risposta di Popper è implicito un «regresso all'infinito»: tuttavia esso, a detta di Popper, «è innocuo» [1974*b*, p. 1026]. Anzi esso giustifica lo scarto che nella pratica scientifica c'è sempre da attendersi tra teorie e realizzazioni tecnologiche – contro ogni versione della *adaequatio rei et intellectus*, ogni «filosofia dell'identità», ogni forma di storicismo. La scena, rispetto a Kant, è cambiata in tanto in quanto – tramontate le teorie «perfette» e spuntata la conoscenza «fallibile e congetturale» – la stessa perfettibilità kantiana è sostituita dalle valutazioni popperiane, tipicamente analitiche, circa l'aumento del grado di corroborazione. Ma la tesi è ancora quella secondo cui le nostre decisioni pratiche dipendono dalle teorie «migliori» (migliori, ora, nel senso di Popper). Tuttavia questa *non* è una risposta al problema pragmatico dell'induzione.

5.2. Il paradigma (neo)bayesiano.

Piú precisamente: la «discussione critica» cui Popper fa appello può benissimo lasciar sopravvivere non una, ma due o *piú* teorie tra loro incompatibili, tutte ugualmente ben controllate. D'altro canto, il problema pragmatico dell'induzione sorge proprio in quei casi in cui le migliori teorie che si hanno a disposizione *non* determinano in modo *univoco* quali eventi si realizzeranno. «Se cosí non fosse, la nostra incertezza sarebbe *totalmente* eliminata, e il problema della scelta di una linea di azione ottimale, *in condizioni di incertezza*, non si porrebbe nemmeno. Non è perciò una soluzione di *questo* problema la proposta di basare l'azione pratica sulla teoria meglio controllata» [Mondadori 1979, p. 28]. Certo le teorie meglio controllate strutturano, per cosí dire, lo spazio della possibilità sottostante a qualunque scelta di linea ottimale, ma non determinano direttamente quel che in questa sede si è convenuto di chiamare «il comportamento razionale».

L'idea viva nel razionalismo classico – dalla ragione come calcolo di Hobbes alla *Mathesis Universalis* di Leibniz – di rendere *calcolabile* la pratica umana, «non sempre utilizzata per un certo periodo nel pensiero economico dopo Walras» per non pochi motivi (tra cui, se pure tale calcolabilità fosse in linea di principio assicurata, tutti i problemi legati alla complessità computazionale), «è oggi superata dalla convinzione che una concezione realistica della pratica sia inconciliabile con un ideale di razionalità di questo tipo, che non corrisponde alla fallibilità di principio delle decisioni umane» [Albert 1978, p. 26]. Ma se la formulazione attuale di una «prassi razionale» deve tener conto «della critica dei modelli proposti dal razionalismo classico» in nome del fallibilismo e ancora non disperdersi in una miriade di soluzioni ad hoc, «confinate alle situazioni specifiche di determinati settori» [*ibid.*, p. 22], essa dovrà vertere proprio sulle decisioni in condizioni d'incertezza, utilizzando i modelli di decisione in condizioni di rischio calcolabile o di certezza come soluzioni approssimative, semplici casi limite alla luce di modellizzazioni piú sofisticate. Ma dalle conclusioni tratte dalla discussione svolta nel § 5.1 si potrebbe avere l'impressione che in questo contesto si è ancora in quella che Kuhn chiamerebbe una situazione preparadigmatica.

Invece un paradigma c'è già. Non proviene né da Hume, né da Kant, né da Popper. Nasce frammentariamente – nel secolo XVII – attraverso i primi tentativi di modellizzazione di «comportamenti induttivi» adoperando come «esemplari» i giochi di azzardo; si consolida con l'emergere del cosiddetto «approccio bayesiano», fino ai primi decenni dell'Ottocento; conosce quindi una lunga eclisse fino agli anni '20 del nostro secolo, quando un nuovo slittamento creativo (dovuto a Ramsey, De Finetti, Savage) consente una ricostruzione unitaria della pratica induttiva in un quadro «neobayesiano» (si vedano del resto gli articoli «Decisione» e «Induzione statistica» in questa stessa *Enciclopedia*). L'idea è che se un agente fa le proprie scelte in modo da soddisfare alcuni assiomi di base, necessariamente si comporta in modo da scegliere la linea d'azione con il massimo indice previsto. Inoltre, sotto la stessa ipotesi, necessariamente rivede le proprie opinioni alla luce di nuove informazioni in accordo con il teorema di Bayes (per cui si veda l'articolo «Probabilità» in questa stessa *Enciclopedia*, X, p. 1185).

Poiché il paradigma è ampiamente illustrato in altri articoli della presente *Enciclopedia*, ci si limiterà qui a un cenno sugli assiomi di base. Detto in breve, un primo assioma, quello detto di coerenza, si limita a imporre la transitività dell'ordinamento delle preferenze (se A è preferito o indifferente a B, e B è preferito o indifferente a C, allora A è preferito o indifferente a C); un assioma di continuità afferma che, nel caso che l'agente preferisca A a B e B a C, esiste una probabilità p tale che resterà indifferente tra la certezza di ottenere B e una linea di azione (o «lotteria») che dà A o C rispettivamente con probabilità p e $p-1$. Infine il cosiddetto «principio della cosa sicura» può venir cosí formulato: sia L_0 una lotteria che dà il premio A_1 se occorre l'evento E_1 e il premio A_2 se occorre l'evento E_2; sia L_0' una lotteria uguale a L_0 tranne che per il fatto che sostituisce ad A_1 un altro premio A_1' che l'agente preferisce ad A_1; l'agente preferirà allora L_0' a L_0 o almeno resterà indifferente tra i due (quest'ultima eventualità è contemplata per il caso in cui l'agente assegni probabilità zero ad E_1).

Ora, se il comportamento dell'agente soddisfa tali assiomi, esistono un'unica funzione di probabilità e una funzione di utilità (unica a meno di trasformazioni lineari positive – e dunque cardinale) tale che l'agente si comporta in modo da massimizzare la sua utilità attesa (cioè la somma dei prodotti delle utilità per le probabilità presa sull'insieme degli stati di cose possibili: per questa terminologia cfr. il citato articolo «Induzione statistica», VII, pp. 385, 390). È questo il teorema della massimizzazione dell'utilità attesa [per un'agile dimostrazione cfr. Anscombe e Aumann 1963].

Com'è noto – cfr. del resto quanto osservato al § 7 di «Induzione statistica» – non mancano nel paradigma (neo)bayesiano problemi aperti e anomalie. Ma di fronte alle molte obiezioni al «bayesianesimo [per un campionario delle quali cfr. Watkins 1977], una buona linea di difesa pare quella che insiste sul fatto che «per confutarlo... non basta argomentare contro il principio della massimizzazione dell'utilità attesa in sé. Esso, invece, può venir confutato solo mostrando che alcuni o tutti i suoi assiomi di base mancano di forza logica» [Harsanyi 1977*a*, p. 382]. Dunque tali assiomi sono «princípî di razionalità»: ora l'assioma della coerenza può essere certo violato in non poche occasioni e tuttavia pare difficile rinunziare a un principio che, essenzialmente, afferma che l'agente sa quel che vuole e sa anche che quel che vuole è coerente; il principio della cosa sicura, d'altro canto, non è che una riformulazione – nel contesto delle «lotterie» – del cosiddetto «principio di dominanza» (se in qualunque stato di cose possibile l'utilità di eseguire una certa azione A non è minore di quella che consegue dall'eseguire un'altra azione B, allora si deve eseguire l'azione B: un principio che opera in svariatissimi contesti e all'interno, tra l'altro, della stessa pratica scientifica [cfr. Giorello e Mondadori 1978, pp. 141 sgg.]); né difficoltà sembrano sussistere in particolare per l'assioma di continuità [cfr. Harsanyi 1977*a*, pp. 383-84]. Dunque questi assiomi sono princípî di razionalità *minimali* (cfr. quanto detto a p. 173. Harsanyi ha per altro ottimo gioco sui critici nel mostrare che gli approcci alternativi al bayesianesimo che essi propongono finiscono per accettare spesso delle varianti degli assiomi in questione).

Ciò non significa che le cose vadano in modo del tutto pacifico per il paradigma (neo)bayesiano: basterebbe ricordare «la critica agli assiomi e ai postulati della scuola americana» [Allais 1953]. E, a proposito dell'«effetto di certezza» che il paradosso di Allais mette in luce (cfr. ancora «Induzione statistica», VII, pp. 425-27), forse è vero che solo «un'attempata zitella schizofrenica» potrebbe conformare il suo comportamento ai dettami bayesiani [Watkins 1977, pp. 375-76]; ma questa non è ancora una ragione per scartare il paradigma. In assenza di un programma che superi tale inadeguatezza, il conformarsi agli assiomi di razionalità indicati va, seppur provvisoriamente, ancora «etichettato» come *razionale* (anche al prezzo di includere tra gli agenti «razionali» zitelle piú o meno schizofreniche). Del resto *in dubio pro theoria*: in conformità con le «norme» di non poca buona epistemologia [cfr. per esempio Lakatos 1970; Stegmüller 1973; ecc.]. Ma su questo punto si tornerà alle pp. 193-95.

5.3. Verso una teoria generale del comportamento razionale.

Di nuovo la teoria dei giochi. «Tra i diversi modi di teorizzare i conflitti – che corrispondono ai diversi significati della parola 'conflitto' – la principale linea di divisione corre tra coloro che, trattato il conflitto come una situazione patologica, ne cercano le cause e i modi per sanarlo e quelli che prendono il conflitto come un dato naturale e studiano il comportamento che ad esso è associato. Per questi ultimi sussiste un'ulteriore linea di divisione tra quelli che studiano i partecipanti a un conflitto in tutta la loro complessità – interessandosi al comportamento "razionale" e "irrazionale", conscio e inconscio, e alle motivazioni profonde tanto quanto ai calcoli – e quelli che fissano la loro attenzione soprattutto su un tipo di comportamento abile, consapevole, razionale» [Schelling 1960, p. 3].

Coloro che costruiscono modelli di situazioni di conflitto servendosi della teoria dei giochi – piú precisamente della teoria dei giochi a *n* persone – rientrano a pieno titolo in quest'ultima categoria, come del resto si evince da altri articoli di questa stessa *Enciclopedia* (cfr. in particolare «Giochi», «Conflitto» e «Tattica/strategia»). Qui si tratteranno quegli aspetti che possono venir prospettati come un'estensione del punto di vista (neo)bayesiano abbozzato nel § 5.2. La teoria dei giochi apparirà allora come «la teoria del comportamento razionale di *due* o *piú* individui razionali che interagiscono fra di loro, ciascuno dei quali è determinato a massimizzare il proprio interesse – sia o non sia puramente egoistico – come è specificato dalla sua funzione di utilità (funzione di pagamento). Va osservato che benché certi giocatori – o anche tutti – possano assegnare alte utilità a obiettivi chiaramente altruistici, ciò di per sé non previene un conflitto di interessi tra di loro, in quanto non va esclusa l'eventualità che essi assegnino alte utilità ad obiettivi altruistici assai *differenti* e magari fortemente conflittuali» [Harsanyi 1976, p. 97]. D'altra parte situazioni tipiche della teoria dei giochi «possono venir considerate come un caso speciale di incertezza, dal momento che in generale nessuno dei giocatori è in grado di predire l'esito o di conoscere le probabilità associate ai diversi esiti possibili» [*ibid.*, p. 96]. Una sintesi è dunque auspicabile.

I due rematori di Hume. Da una situazione di coordinazione pura a una di conflitto parziale. La stessa teoria dei giochi ha richiamato l'interesse [Schelling 1960; Lewis 1969; ecc.] sulle situazioni di «pura coordinazione». Il modello va cercato nella discussione humeana della *convenzione* come fondamento della proprietà e della giustizia [Hume 1739, trad. it. pp. 512 sgg.]. La convenzione rappresenta in tale contesto «una consapevolezza generale per l'interesse comune, consapevolezza che tutti i membri della società esprimono l'un l'altro, e che li induce a regolare la loro condotta in base a certe regole» [*ibid.*, p. 517]. Si consideri, dice ancora Hume, il caso di due uomini che, sospingendo una barca a forza di remi, cooperano «in virtú di un accordo o di una convenzione» [*ibid.*, p. 518] per attraversare un fiume. Ora, se si suppone che i due uomini in barca abbiano entrambi l'intenzione di remare di concerto in modo da raggiungere la stessa destinazione; anzi, che siano disposti a modificare ciascuno la propria remata in modo da sincronizzarla con quella dell'altro, la loro coordinazione è ovvia.

Ma si supponga invece che, anche se entrambi vogliono attraversare il fiume, ciascuno ritenga vantaggioso per sé attraversarlo dividendo equamente la fatica coll'altro, senza essere cosí ansioso di arrivare alla meta da ritenere vantaggioso di sobbarcarsi la fatica di remare da solo. E si supponga ancora che, pur avendo ciascuno una lieve preferenza per attraversare il fiume in tempo breve, tale preferenza non sia abbastanza forte da indurre uno dei due ad aggiungere i suoi sforzi a quelli dell'altro, se quest'ultimo sta già remando. Anzi ciascuno può dire tra sé e sé: «Se è l'altro che sta remando, perché dovrei remare anch'io? E se

l'altro non sta remando, perché dovrei cominciare io per primo?» La situazione è in questo caso [cfr. Mackie 1980, pp. 88-90] slittata in un conflitto parziale: l'esito per cui entrambi non remano o smettono di remare – e dunque non giungeranno mai a destinazione – è chiaramente subottimale. Non sarebbe stato preferibile per entrambi attraversare il fiume dividendo a metà la fatica?

Tosca/Scarpia: un gioco non a somma zero. Come ben sanno gli amanti dell'opera, Scarpia, capo della polizia, tiene in prigione il pittore Cavaradossi, amante di Tosca, condannato a morte. Ma Scarpia aspira ai favori di Tosca. Nell'atto II della *Tosca* si assiste alla contrattazione fra i due: «SCARPIA ...Via, mia bella signora, | sedete qui. Volete che cerchiamo | insieme il modo di salvarlo? E allora sedete... e favelliamo... TOSCA Quanto? SCARPIA Quanto? TOSCA Il prezzo!» Infine è concluso «l'orribile mercato». Tosca accondiscenderà ai desideri di Scarpia se questi libererà Cavaradossi inscenando una finta fucilazione. («SCARPIA Cedo. A misero prezzo | tu, a me una vita, io, a te chieggo un istante!»)

Si costruisca allora la matrice di pagamento di Tosca/Scarpia: le loro preferenze sono indicate – al solito – con unità arbitrarie (fig. 1). Ora, di fronte a tale matrice, Tosca ragiona nel modo seguente: «Supponiamo che Scarpia mantenga la parola. In tal caso mi conviene ingannarlo, in modo da salvare Cavaradossi senza cedere a Scarpia. Ma se Scarpia mi inganna e ordina al plotone di usare pallottole vere, anziché finte, allora mi conviene sicuramente ingannarlo. Quindi la migliore strategia consiste nell'ingannarlo». Ma anche Scarpia arriva a un'analoga conclusione. Si ingannano entrambi. Tosca pugnala Scarpia e nell'atto III troverà il suo amante crivellato di colpi. Gli specialisti di teoria dei giochi e gli amanti dell'opera possono discutere all'infinito se si

	B_1	B_2
A_1	2, 2	0, 3
A_2	3, 0	1, 1

Figura 1.

La matrice dei pagamenti del gioco Tosca/Scarpia (o «dilemma del prigioniero»). In questo gioco la coppia di strategie (A_2, B_2) è un punto di equilibrio, dal momento che la risposta migliore del giocatore 1 (ad esempio: Tosca) a B_2 è A_2 mentre la risposta migliore del giocatore 2 (ad esempio: Scarpia) ad A_2 è B_2. Il gioco non ha altri punti di equilibrio. Se i due giocatori usano le loro strategie di equilibrio A_2 e B_2 ottengono i pagamenti (1, 1) (è il finale della *Tosca*). Ovviamente i due giocatori si troverebbero entrambi meglio impiegando le strategie A_1 e B_1 che potrebbero dare loro i pagamenti (2, 2). Ma queste due strategie *non* formano un punto di equilibrio, come il lettore può agevolmente verificare. Il gioco è noto nella letteratura specializzata come «dilemma del prigioniero»: due prigionieri, interrogati separatamente, possono confessare un delitto di lieve entità commesso in comune oppure accusarsi reciprocamente di un delitto grave, alla condizione che chi accusa viene liberato se non viene a sua volta accusato e chi è accusato riceve una dura condanna.

tratti di un doppio «doppio gioco», di un «quadruplo» gioco o di una tragedia... Resta il fatto che Tosca e Scarpia si sono accordati nella speranza di trarne entrambi vantaggio. Ma poiché non avevano fiducia l'uno nell'altro hanno perso entrambi, proprio come nel caso dei rematori nella interpretazione «perversa». Quindi, affinché un accordo funzioni, è essenziale che entrambe le parti abbiano fiducia e che ognuna creda che l'altra rispetterà l'accordo, in caso contrario, i giocatori agiranno unicamente nel loro interesse. Non c'è niente di dannoso nell'agire in vista del proprio interesse e anzi questa è notoriamente una scelta oculata in ogni gioco a due persone a somma zero. Ma in un gioco non a somma zero – come quello di Tosca/Scarpia – valutare la strategia unicamente in base al proprio interesse porta a un esito manifestamente subottimale (cfr. del resto le osservazioni nell'articolo «Modello», IX, pp. 414-15).

Si riprenda in esame il caso dei due uomini in barca nell'interpretazione «perversa». Il dilemma sarebbe davvero sciolto se una sorta di autorità esterna, una specie di sovrano hobbesiano, costringesse i due a remare? Il problema cosí è solo spostato: perché infatti sottomettersi a tale autorità esterna? Una soluzione contrattualistica sarebbe ancora basata su una convenzione: non sarebbe nemmeno una soluzione, dal momento che anche qualora fosse stato stipulato, ciascuno avrebbe ancora gli stessi motivi per rompere l'accordo che prima aveva per non remare. «Non possiamo a questo punto assumere che ci sia un qualsiasi sentimento morale o, ovviamente, qualche sovrano hobbesiano» [Mackie 1980, p. 89], pena un regresso all'infinito. La situazione è dunque senza sbocco? Nel caso dei due uomini in barca, «sorprendentemente uno sbocco c'è» [*ibid.*]. Ciascuno dei due può dire tra sé e sé: «*Forse* l'altro remerà a patto che lo faccia anch'io. Che male c'è a provare?» Cosí uno dei due comincia a remare, senza sforzarsi troppo e guarda cosa farà l'altro. L'altro, «con lo stesso atteggiamento sperimentale» [*ibid.*], si chiede se il fatto che anche lui rema incoraggerà l'altro e comincia anche lui a remare moderatamente; se allora il primo gli risponde remando un po' di piú, ben presto entrambi remeranno a pieno ritmo. Ma, forse, uno comincia a rallentare quando l'altro aumenta il suo impegno: questi se ne accorge e minaccia di smettere; per pure ragioni egoistiche quello che aveva rallentato riprende a remare con piú impegno, ecc.

C'è quindi una sostanziale differenza tra questo caso dei rematori e quello di Scarpia e Tosca: l'uno è iterabile, l'altro no. Dunque «l'atteggiamento sperimentale» può rappresentare una via d'uscita nei giochi tipo «dilemma del prigioniero» (per questa terminologia cfr. la didascalia della figura 1) se l'iterazione è consentita. Ma – e non solo nelle opere liriche – molti casi reali sono tipici «dilemma del prigioniero» *senza* iterazione.

Un altro aspetto cruciale è rappresentato dalla sostanziale differenza tra giochi cooperativi – ove i giocatori prendono impegni che vengono fatti valere (come promesse vincolanti, accordi e minacce che devono essere rese effettive nelle condizioni previste) – e giochi non cooperativi in cui questo non si verifica (si segue qui la definizione di Harsanyi [1976] che modifica quella originale di Nash [1950*a*; 1951]). Se un gioco come quello della figura 1 è concepito come non cooperativo la soluzione è quella subottimale rappresentata dall'unico punto di equilibrio di Nash (A_2, B_2) – il finale della *Tosca*. Se il gioco fosse stato invece giocato come un gioco cooperativo, allora l'esito sarebbe stato la soluzione «cooperativa» (A_1, B_1) che *non* è un punto di equilibrio. Tutta una letteratura ha recentemente insistito sugli aspetti negativi della non-cooperazione: ora «quel che si ha da fare in tal caso, se è possibile farlo, è *mutare* un gioco non cooperativo in uno cooperativo, ove è possibile, rendendo gli accordi vincolanti, piuttosto che pretendere di vivere in un mondo fittizio, dove lasciamo i giochi non cooperativi cosí come sono, ma poi li analizziamo come se fossero cooperativi, se cosí ci aggrada» [Harsanyi 1976, p. 104].

La modificazione della matrice di un gioco (per un esempio, cfr. fig. 2) che a prima vista pare un «dilemma del prigioniero» resta dunque un difficile problema aperto: ma esso è un problema antecedente, per cosí dire, alla modellizzazione via teoria dei giochi, il cui scopo è definire la soluzione dei giochi stessi, una volta specificata correttamente la matrice dei pagamenti. E, proprio dal punto di vista che qui interessa – quello di una teoria del comportamento razionale – occorre aggiungere che la soluzione di un gioco non cooperativo deve essere sempre un punto di equilibrio (nel senso di Nash; cfr. ancora «Modello», p. 414). Infatti «se la soluzione comprendesse una strategia di un dato giocatore che *non* è la sua miglior risposta alle strategie degli altri giocatori in quella soluzione, allora proprio la previsione che gli altri giocatori useranno le loro strategie in quella soluzione renderà razionale per quel giocatore *non* usare la sua strategia di quella soluzione (ma usare piuttosto una strategia che è la risposta migliore alle strategie che egli si aspetta che gli altri giocatori usino). Quindi questa pretesa "soluzione" non soddisfa l'idea intuitiva di soluzione» [*ibid.*, p. 202].

Il problema della contrattazione. «Le situazioni economiche di monopolio contro monopsonio, del commercio di stato tra due nazioni e di negoziazione tra datore di lavoro e sindacato possono essere considerati tutte come problemi

	B_1	B_2
A_1	2, 2	0, 1
A_2	1, 0	1, 1

Figura 2.

La matrice di cui alla figura 1 («dilemma del prigioniero») viene modificata assumendo che i due giocatori attribuiscano una considerevole disutilità a usare una strategia non cooperativa come A_2 o B_2 quando l'altro giocatore usa una strategia cooperativa come A_1 o B_1: piú precisamente si è supposto che entrambi i giocatori, assegnino una disutilità di due unità a un esito del genere. Si ha cosí $(A_2, B_1)=(3-2, 0)=(1, 0)$ e $(A_1, B_2)=(0, 1)$. Di conseguenza il gioco non è piú un dilemma del prigioniero, in quanto si hanno ora due punti di equilibrio (A_1, B_1) e (A_2, B_2) [cfr. Harsanyi 1976, p. 103]. Naturalmente (A_1, B_1) non è un equilibrio in termini di strategie dominanti, ma solo in termini di migliori risposte [cfr. la critica di Watkins 1977, p. 357 che difende il punto di vista «prudente»; per una risposta, cfr. Harsanyi 1977*a*, pp. 389-91].

di contrattazione [*bargaining problems*]» [Nash 1950*b*, p. 155]. La modellizzazione proposta da Nash della «contrattazione» non verrà discussa qui [per una esposizione che tiene conto di critiche e approcci alternativi, cfr. Luce e Raiffa 1957, in particolare pp. 124-37; di notevole interesse è inoltre Harsanyi 1977*b*, cap. VIII in particolare] se non nelle linee molto generali, seguendo uno spunto dello stesso Nash [1951, p. 295] per cui l'analisi di un gioco cooperativo deve cominciare con la costruzione di un modello di una situazione di contrattazione o «gioco di contrattazione» che rappresenta appunto la preliminare contrattazione dei giocatori. «Il suggerimento di Nash si basa sull'assunto che una stretta cooperazione tra i giocatori in un gioco cooperativo richiede usualmente un accordo precedente sui pagamenti che, in moltissimi casi, può venir ottenuto solo grazie a una contrattazione tra i giocatori. Ma questa contrattazione in sé ha la natura di un gioco non cooperativo, a meno che non si voglia assumere che i giocatori si sono accordati in un gioco di contrattazione sussidiario ancora precedente su come agire nel gioco di contrattazione principale: un'assunzione piuttosto implausibile, che comporta inoltre un regresso all'infinito» [Harsanyi 1976, p. 111]. È in questo modo che si rende possibile sia una non banale unificazione concettuale dei giochi cooperativi sia un punto di vista piú profondo cui riportare unitariamente giochi cooperativi e non cooperativi, rappresentato dal modello della contrattazione che nel programma di Nash va strutturato come un gioco non cooperativo che richiede un attento studio dei suoi punti di equilibrio.

Punti di equilibrio perfetti. Con l'introduzione da parte di Nash del concetto di punto d'equilibrio la stragrande maggioranza dei ricercatori in teoria dei giochi ha per un certo lasso di tempo creduto che l'unica richiesta di razionalità in un gioco non cooperativo consistesse nel formare con le strategie dei giocatori un punto di equilibrio.

Si consideri allora il seguente gioco non cooperativo a due persone dato in forma estesa (fig. 3). La prima mossa spetta al giocatore 1. Questi può scegliere tra le mosse a_1 e a_2. Se sceglie a_1 il gioco termina con il pagamento (1, 3) ai due giocatori, senza che il giocatore 2 abbia ancora fatto una mossa. Ma se il

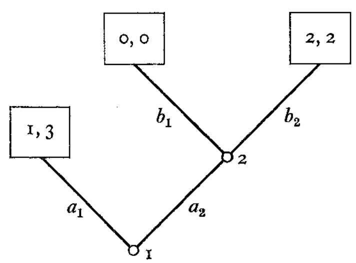


Figura 3.

Un gioco rappresentato in forma estesa, cioè come un albero (per questa terminologia si veda l'articolo «Grafo» in questa stessa *Enciclopedia*). L'esempio è tratto da Harsanyi 1976, p. 105.

giocatore 1 sceglie la mossa a_2 il giocatore 2 può scegliere tra le mosse b_1 e b_2. Se sceglie b_1, il gioco terminerà coi pagamenti (0, 0); se sceglie b_2, il gioco terminerà coi pagamenti (2, 2). Se ci si limita a considerare lo stesso gioco in forma normale (fig. 4), si vede che esso ha due punti di equilibrio $E_1 = (A_1, B_1)$ e $E_2 = (A_2, B_2)$. Ora, E_2 non desta preoccupazioni. Ma E_1?

Si osservi anzitutto che il giocatore 1 userà la strategia A_1 *solo se* si aspetta che il giocatore 2 usi la strategia B_1 (se il giocatore 2 usa B_2, il giocatore 1 si trova meglio con la strategia A_2); da parte sua il giocatore 2 preferirà a prima vista l'esito (A_1, B_1) che gli dà 3 unità di utilità all'esito (A_2, B_2) che gliene dà solo 2. Il giocatore 2 dovrebbe dunque cercare di indurre il giocatore 1 a usare la strategia A_1, cioè a fare la mossa a_1, per esempio minacciandolo di usare altrimenti la strategia B_1, cioè di punirlo facendo la mossa b_1. Ma questa minaccia è una sorta di tigre di carta: facendo la mossa b_1 il giocatore 2 non solo punirebbe il giocatore 1, ma anche se stesso (in quanto b_1 riduce a zero i pagamenti di entrambi, mentre b_2 dà ad entrambi pagamenti di 2 unità). «In conclusione $E_1 = (A_1, B_1)$ è un punto di equilibrio irrazionale per il fatto che si basa sulla irragionevole assunzione che il giocatore 2 dovrebbe punire il giocatore 1 se quest'ultimo fa la mossa a_1, anche se questa mossa punitiva finisce col ridurre non solo il pagamento del giocatore 1 ma anche quello dello stesso giocatore 2» [Harsanyi 1976, p. 107]. Con Selten [1965] – che per primo ha indicato «patologie» di questo tipo – si chiameranno punti di equilibrio imperfetto i punti come E_1, riservando a quelli come E_2 l'appellativo di punti di equilibrio perfetto.

Ma, non diversamente che nel caso delle anomalie piú resistenti alle nostre teorie del mondo fisico o dei controesempi a una dimostrazione matematica o di un animale che non rientra in una data classificazione, sono i tentativi di spiegare come si generino tali «patologie» quelli che si rivelano piú fertili e interessanti. Ci si domanda, nella fattispecie, come una strategia irrazionale tipo B_1 possa entrare in un punto di equilibrio. La risposta è che finché i giocatori seguono le strategie A_1 e B_1, il giocatore 2 non si troverà mai nella posizione di dover compiere la mossa «irrazionale» b_1 prescritta da B_1: questa strategia impone al giocatore 2 di fare la mossa b_2 solo se il giocatore 1 ha fatto la mossa a_2 (come piú sopra si è detto) e ciò non succederà se il giocatore 1 si attiene alla strategia A_1 che impone a_1 al posto di a_2.

Ma si supponga ora che il giocatore 1 che ha adottato la strategia A_1 sia in

	B_1	B_2
A_1	1, 3	1, 3
A_2	0, 0	2, 2

Figura 4.

Lo stesso gioco della figura 3, ma in forma normale. [Per un bilancio dei vantaggi e svantaggi della rappresentazione in forma di albero e in forma normale cfr. per esempio Luce e Raiffa 1957, cap. III].

grado di compiere la mossa a_1 solo con una certa probabilità $1-\varepsilon$ con $0<\varepsilon<1$ e sia quindi costretto alla mossa a_2 con probabilità $\varepsilon \neq 0$. Non sarà dunque priva di costo *ora* per il giocatore 2 la strategia B_1 quando il giocatore 1 adotta A_1, in quanto ora il giocatore 1 può fare la mossa a_2 con probabilità ε *diversa da zero* e di conseguenza il giocatore 2 risponderà con la mossa b_1 anche qui con probabilità *diversa da zero*. In questo modo la strategia B_1 non sarà la migliore risposta ad A_1 e $E_1 = (A_1, B_1)$ non rappresenterà piú un punto di equilibrio.

Questo ragionamento ha una portata generale [cfr. Harsanyi 1976, pp. 107-108]: si tratta di assumere nel modello che ogniqualvolta un giocatore cerca di fare una data mossa, ci sia una probabilità *arbitrariamente piccola ma non nulla* ε di compiere un «errore» che lo costringa a ripiegare su una mossa diversa da quella che intendeva fare, in modo che *ogni* possibile mossa occorre con qualche probabilità *non nulla*. Come in piú di un articolo della presente *Enciclopedia* si è osservato, una buona modellizzazione dei processi reali che pretende di essere qualcosa di piú di un esercizio di matematica pura deve tener conto del fatto che abbiamo a che fare sempre con situazioni perturbate [cfr. del resto Duhem 1906, trad. it. pp. 154 sgg.; Thom 1980, pp. 4-5, 153-54 nota 4]. L'assunzione del gioco perturbato non fa che rendere piú realistico il modello (per esempio: si potrebbe interpretare il comportamento degli agenti in un gioco perturbato come una concessione alla «razionalità limitata» di Simon [1960]; di fatto gli agenti non paiono in grado di attenersi a un modello di razionalità «assoluta» – cioè «non perturbata» – poiché le alternative sono praticabili solo entro certi margini e inoltre gli agenti decidono non sinotticamente ma sequenzialmente).

Come conseguenza nell'albero che rappresenta il gioco in forma estesa (cfr. ancora la figura 3) ogni nodo sarà raggiungibile con una certa probabilità (non nulla). Si ha poi un premio addizionale: si può infatti mostrare che se nel gioco perturbato si prendono i punti di equilibrio e si passa al limite per ε tendente a zero si ottengono proprio i punti di equilibrio perfetto del gioco originario. Essi possono quindi venir caratterizzati come quei punti di equilibrio del gioco originario che restano punti di equilibrio anche nel gioco perturbato. Essi godono, per cosí dire, di una sorta di stabilità.

Un concetto «bayesiano» di soluzione per i giochi non cooperativi. Un punto debole del programma di Nash che, come si è visto, pone l'enfasi sui giochi di contrattazione, è che molti giochi interessanti hanno troppi punti di equilibrio. Si consideri per esempio la seguente contrattazione a due persone. Giuseppe e Clemente devono spartirsi una somma di centomila lire e se non si accordano su come ripartirla, entrambi ricevono zero come pagamento. La situazione può venir modellizzata ricorrendo a una contrattazione siffatta: si indicano con x_1 e x_2 le richieste rispettive di Giuseppe e di Clemente, sottoposte al vincolo $0 \leq x_1, x_2 \leq 100\,000$. Se $x_1 + x_2 \leq 100\,000$, Giuseppe ottiene lire x_1 e Clemente lire x_2; altrimenti per $x_1 + x_2 > 100\,000$, sia Clemente sia Giuseppe ottengono entrambi lire 0. Intendendo x_1 e x_2 come numeri razionali (per non dire reali!), il gioco ha un numero infinito di punti di equilibrio: ogni coppia (x_1, x_2) ove $x_1 + x_2 = 100\,000$. Ma ci si limiterà a far richieste di denaro corrispondenti a numeri interi di lire! Anche in questo caso ci sono sempre troppi punti di equilibrio: 100 001 per la precisione. E, inoltre, 99 999 sono punti di equilibrio perfetto (ovviamente non lo sono gli «estremi»: (0, 100 000) e (100 000, 0)).

La tecnica, schizzata piú sopra, del gioco perturbato non è dunque sufficiente qui per selezionare un solo punto particolare di equilibrio come soluzione del gioco. Nuova difficoltà, nuovo slittamento creativo: Harsanyi e Selten, a partire dal 1974, hanno proposto «un'estensione del punto di vista bayesiano, che tanto successo ha avuto nell'analisi delle situazioni in cui si tratta delle decisioni del singolo, all'analisi dei giochi non cooperativi» [Harsanyi 1976, p. 113]. In breve (e limitandosi al caso di due giocatori) il procedimento consiste nell'immaginare che ciascun giocatore costruisca una distribuzione di probabilità a priori sulle strategie dell'avversario. Alla luce di questa iniziale assegnazione di probabilità, ciascun giocatore calcola che assegnazione di probabilità alle sue strategie pure fornirebbe la risposta migliore alle strategie dell'avversario. Fatto questo, rivede le sue distribuzioni di probabilità sulle strategie dell'avversario in modo da far sí che esse tengano conto della *sua* risposta migliore, com'è stimata fino a quel momento. Fatto questo, riesamina la sua replica migliore, ecc. Questo procedimento (tecnicamente noto come *tracing procedure* [cfr. Harsanyi 1975]) permette infine di definire un'unica soluzione del gioco, cui convergono le distribuzioni di probabilità dei giocatori man mano che vengono «riviste».

Anche in questo caso non si tratta di un mero espediente tecnico: il procedimento matematico «intende modellare il processo psicologico, che potremmo chiamare processo di soluzione, attraverso il quale le attese dei giocatori convergono a uno specifico punto di equilibrio che costituisce la soluzione del gioco» [Harsanyi 1976, pp. 114-15].

6. *Teoria «della» pratica e «pratica teorica».*

6.1. Contro la tesi dell'«a-teoricità».

Anche chi è «in pieno disaccordo» con il bayesianesimo riconosce che la ripresa dei punti di vista di «pionieri» come Nash o Zeuthen entro il programma di Harsanyi di una teoria unificata delle decisioni razionali è «una vittoria del punto di vista bayesiano» [valga per tutti Watkins 1977, p. 351]: senza tener conto, tra l'altro, della proposta di estendere tale punto di vista all'etica stessa, reinterpretando l'utilitarismo come massimizzazione del livello dell'utilità media di tutti gli individui nella società [cfr. per esempio Harsanyi 1976, capp. II-V; 1977*b*, cap. IV] e cercando di sciogliere alcune tipiche riserve relative all'utilitarismo tradizionale [cfr. per esempio Harsanyi 1977*b*, pp. 62-64; 1977*c*] in un tentativo di conciliazione di ποίησις e πρᾶξις, di «precetti dell'abilità» e linee d'azione disinteressate.

Ma, anche a prescindere da quest'ultima problematica (per cui si veda però anche il successivo § 6.3), è interessante che l'approccio bayesiano non solo sembra costituire una teoria della pratica almeno parzialmente soddisfacente (cfr. le osservazioni a p. 181) per tutti quei casi in cui si può supporre che il nostro «avversario» sia indifferente alla nostra scelta di una strategia, in breve nei cosiddetti «giochi contro la natura», ma ammette estensioni interessanti (di cui la *tracing procedure* illustrata sopra è un esempio) anche a un ambito che a prima vista potrebbe sembrargli estraneo, quello della teoria dei giochi propriamente detta.

Inoltre, non si tratta affatto di una teoria a-teorica della pratica, come è stato sostenuto da Lakatos. Il punto è delicato e merita qualche delucidazione. Nella sua ricostruzione razionale dello sviluppo della logica induttiva, Lakatos ha battezzato «postulato di Jeffreys e Keynes» l'assunto secondo cui la probabilità a priori $p(G)$ di una legge è diversa da zero [1965, pp. 330-34]. Nel sistema di logica induttiva sviluppato da Carnap risulta però $p(G)=0$ per qualsiasi genuina asserzione universale [1950, in particolare pp. 570-71]. Quanto a Popper [1974*a*], questi si spinge a proclamare che a tali asserzioni «si *dovrebbe* attribuire la "probabilità" zero... quantunque il loro grado di corroborazione possa essere piú grande di zero» (trad. it. p. 151). Ora, $p(G)=0$ implica $p(G|E)=0$ per ogni resoconto sperimentale E: dunque non è possibile, negando il postulato di Jeffreys e Keynes, discutere e confrontare in termini probabilistici quanto l'esperienza «sostenga» (*supports*) le varie asserzioni universali. «Si può quindi concludere che l'assunto $p(G)=0$ per asserzioni universali G ha rappresentato un fattore decisivo che ha spinto Carnap a respingere una logica induttiva delle *teorie* e Popper a respingere una logica *induttiva* delle teorie» [Niiniluoto e Tuomela 1973, p. 215]. Ma il sistema presentato da Niiniluoto e Tuomela riprendendo Hintikka fornisce una misura di probabilità p che attribuisce probabilità non nulle a genuine asserzioni universali, pur facendo proprie alcune richieste di Popper (nel non far coincidere probabilità e grado di corroborazione) e di Lakatos (circa l'eccesso di corroborazione) [cfr. *ibid.*, pp. 218, 136].

Non solo, ma come ha mostrato Hintikka [1971], dal teorema di rappresentazione di De Finetti (per cui si veda ancora «Induzione statistica», VII, pp. 396-405) segue la possibilità di reinterpretare soggettivisticamente (cioè in termini di quozienti di scommessa) tali probabilità non nulle assegnate ad asserzioni universali. In queste circostanze, dove H è un'ipotesi, E un resoconto sperimentale e T una teoria, i risultati citati stabiliscono dunque che ha perfettamente senso parlare (da un punto di vista bayesiano) non solo di $p(H)$ e di $p(T)$, ma anche di $p(H \mid E \& T)$. Dunque, entro un approccio bayesiano le nostre stesse decisioni dovranno dipendere dalle migliori teorie disponibili. Proprio come suggeriva Popper (cfr. p. 179). Tuttavia, qui si sa in piú quale dev'essere la forma di questa dipendenza per un soggetto razionale. Perciò, da un punto di vista (neo)bayesiano le teorie influenzeranno la pratica in tanto in quanto influenzano la valutazione delle probabilità pertinenti alla formazione della decisione in questione via la regola della massimizzazione dell'utilità attesa (cfr. p. 181).

6.2. La «pratica teorica».

È ancor piú sorprendente che questo stesso paradigma possa venir applicato in modo del tutto naturale anche a un tipo di pratica, a prima vista esterna al suo orizzonte concettuale, fatto di linee di azione, perdite e guadagni, strategie pure e miste, ecc. Si fa qui riferimento alla pratica *teorica* e cioè alle regole che governano la scelta dei programmi di ricerca entro la comunità scientifica. Questa tematica ha suscitato una controversia di grande interesse che ha visto come protagonisti popperiani e carnapiani: scegliere il programma di ricerca che massimizza il contenuto, oppure il grado di conferma, oppure il grado di corroborazione, oppure ancora un'opportuna funzione di tali quantità? Il limite della maggior parte delle soluzioni proposte è indubbiamente il loro carattere ad hoc. Ed è proprio questo limite che il paradigma (neo)bayesiano promette di superare.

«In dibattiti recenti, è stata spesso espressa la speranza che le idee fondamentali della moderna teoria della decisione possano essere utili per la comprensione del problema dell'adozione e del rifiuto di teorie e ipotesi scientifiche. Tale adozione o rifiuto, infatti, può essere considerato come una decisione, del tipo di quelle studiate abitualmente nella teoria della decisione» [Hintikka e Pietarinen 1966, trad. it. p. 143]. Anche la pratica teorica è un tipo di pratica che appare diretta verso uno o piú scopi. Entrambe le parti nella controversia accennata piú sopra sono disposte infatti a riconoscere che i ricercatori scientifici mirano a teorie non solo vere, ma anche informative (con un alto contenuto). (E d'altra parte «il riferimento alla verità o almeno alla probabilità è presente in alcuni dei piú comuni principî della teoria delle decisioni. Si massimizza non l'utilità di ogni risultato particolare, ma l'utilità *attesa*, per un individuo, di una decisione, cioè l'utilità media di tutti i differenti risultati che possono sortire dalla decisione di un individuo, ognuno dei quali pesato con la probabilità che gli compete. È in questo senso che ogni teorico della decisione si sforza di avvicinarsi il piú possibile alla verità» [*ibid.*, p. 145]. Cfr. del resto quanto osservato nel § 5).

In questo nuovo tentativo di sintesi, si tratta allora di cercar d'interpretare gli scopi dei ricercatori scientifici in termini di utilità. Come? Per semplificare, si supponga di avere a che fare con due teorie T e T' incompatibili e tali che $p(T \vee T')=1$. Che conseguenza avrà la scelta di T nel caso in cui T è vera? Che si guadagnerà una quantità d'informazione circa il mondo pari al contenuto di T – d'ora in poi abbreviato in $\mathrm{cont}(T)$. D'altra parte, nel caso in cui T sia falsa si perderà un quantità d'informazione pari al contenuto di T'. La situazione è simmetrica nel caso di scelta di T'. Si ha perciò una matrice dei pagamenti come nella figura 5. S'identificheranno allora le utilità pertinenti alla decisione con $\pm\mathrm{cont}(T)$ e $\pm\mathrm{cont}(T')$. [Per questa proposta, cfr. *ibid.*].

Una condizione che $\mathrm{cont}(T)$ deve soddisfare è naturalmente la seguente:

(1) $\mathrm{cont}(T)$ è tanto maggiore quanto piú grande è la classe di eventi vietati da T.

Tuttavia $p(T)$ risulta tanto minore quanto piú grande è tale classe. *Una* definizione di cont(T) compatibile con (1) è perciò:

(2) $\quad \mathrm{cont}(T) = 1 - p(T).$

Si tralascia di mostrare quali altre condizioni vadano aggiunte a (1) perché (2) risulti *l'unica* definizione compatibile con (1).

In queste circostanze, l'ingrediente che ancora manca per applicare il paradigma bayesiano sono le probabilità di T e T'. Sia E l'evidenza a disposizione. Le probabilità pertinenti saranno allora $p(T|E)$ e $p(T'|E)$. Cosí, l'utilità attesa di T che si abbrevierà $A(T|E)$ sarà pari a

(3) $\quad A(T|E) = p(T|E)\mathrm{cont}(T) - p(T'|E)\mathrm{cont}(T').$

Calcoli elementari stabiliscono che

(4) $\quad A(T|E) = p(T|E) - p(T),$

e, simmetricamente,

(5) $\quad A(T'|E) = p(T'|E) - p(T').$

Il paradigma bayesiano prescrive perciò di scegliere la teoria che massimizza la differenza tra probabilità finali e iniziali. Si tratta di una misura proposta piú volte, nel corso della controversia citata, per le sue molteplici «virtú». In questo contesto, essa gode però di una giustificazione assai piú forte. La sua massimizzazione è infatti la strategia di un soggetto razionale [ma per una generalizzazione cfr. ancora *ibid.*, pp. 158 sgg.].

Tutto questo dà naturalmente per scontato il fatto che T e T' abbiano dell'evidenza *in comune*. Ma questa assunzione è realistica? Come Feyerabend ha enfatizzato, teorie tra loro incommensurabili possono essere confutate solo con le esperienze che sono loro proprie [cfr. per esempio Feyerabend 1970, trad. it. p. 309]. Ma allora, se T e T' non condividono alcuna evidenza e quindi E si riduce a semplice tautologia, la proposta precedente non dà alcuna risposta al problema poiché in tal caso dalle (4) e (5) si ottiene immediatamente

(6) $\quad A(T|E) = A(T'|E) = 0.$

Questo non vieta però di ricorrere in un caso del genere a un confronto indiretto (tale confronto, ovviamente, avrà un senso solo nel caso in cui delle due

	T è vera	T' è vera
Si sceglie T	$+\mathrm{cont}(T)$	$-\mathrm{cont}(T')$
Si sceglie T'	$-\mathrm{cont}(T)$	$+\mathrm{cont}(T')$

Figura 5.
Matrice di pagamento per una decisione teorica.

teorie «l'una è formulata in un linguaggio tale che alcune delle sue regole d'uso implicite sono inconsistenti con quelle della seconda» [Feyerabend 1962, p. 74]). Per ciascuno dei due programmi si considera l'evidenza «interna» disponibile, E ed E' rispettivamente, si calcolano le previsioni rispetto a tale evidenza e si sceglie la teoria che la massimizza rispetto alla propria evidenza interna. L'ostacolo rappresentato dalla «varianza di significato» risulta cosí aggirato.

In conclusione, con o senza una soluzione di quest'ultimo problema, il paradigma bayesiano è in grado di dare una risposta interessante al problema della scelta tra programmi di ricerca rivali. La stessa manovra suggerita per affrontare il caso dell'evidenza vuota può infatti essere replicata anche relativamente alla misura di probabilità p. Se infatti – com'è plausibile supporre – nei due programmi di ricerca rivali si scegliessero metriche diverse per p, la prescrizione bayesiana potrebbe essere reinterpretata nei termini della scelta del programma che massimizza A rispetto alla propria misura di probabilità e rispetto alla propria evidenza.

6.3. Teoria «bayesiana» della pratica contro «filosofia della prassi».

«Se il problema di identificare teoria e pratica si pone, si pone in questo senso: di costruire, su una determinata pratica, una teoria che coincidendo e identificandosi con gli elementi decisivi della pratica stessa, acceleri il processo storico in atto, rendendo la pratica piú omogenea, coerente, efficiente in tutti i suoi elementi, cioè potenziandola al massimo; oppure, data una certa posizione teorica, di organizzare l'elemento pratico indispensabile per la sua messa in opera» [Gramsci 1933, p. 1780]. Cosí Antonio Gramsci struttura la sua versione della filosofia dell'identità: «L'identificazione di teoria e pratica è un atto critico, per cui la pratica viene dimostrata razionale e necessaria o la teoria realistica e razionale» [*ibid.*]. Questa mossa acquista il suo pieno senso nel quadro concettuale della cosiddetta «filosofia della prassi», ove il termine 'prassi' è slittato dall'antico significato aristotelico (cfr. sopra, p. 170) a denotare l'attività umana che *realizza* nella storia ciò che è implicito nel «processo»: la figura di agente «razionale» che Gramsci qui disegna è infatti impensabile senza lo sfondo della transizione [cfr. ancora *ibid.*]. Per questa via è agevole risalire all'idea marxiana del comunismo come «movimento reale che abolisce lo stato di cose esistente» e non come risultato di una molteplicità di scelte individuali.

Ma sotto questo profilo il marxismo risulta doppiamente insufficiente. Non solo suppone che tutte le interazioni entro le moderne società capitalistiche possano essere ridotte, in ultima analisi, a un unico fattore (è questa una forma del dogma dei sistemi centrati: cfr. l'articolo «Centrato/acentrato» in questa stessa *Enciclopedia*) ma assume anche che il loro «superamento» sia un semplice risultato del suddetto «movimento reale che abolisce lo stato di cose esistente».

L'approccio (neo)bayesiano invece implica uno spostamento dell'enfasi sulla molteplicità degli individui (e delle loro coalizioni) – ciascuno dei quali è caratterizzato da una particolare funzione di utilità – e sul fatto che la stessa scelta collettiva («Il movimento reale» di marxiana memoria!) deve essere determinata dalla massimizzazione della media aritmetica di tali funzioni (eventualmente soggetta a opportuni vincoli).

Da questo punto di vista cade ogni identità di «teoria» e «prassi». Una teoria della pratica (cfr. in particolare il § 6.1) richiede precisamente la loro netta distinzione.

6.4. Una teoria «quasi-empirica»?

In conclusione: la classica dicotomia teoria/pratica si è risolta, nel quadro concettuale qui abbozzato, in una nuova teoria. Una teoria del comportamento razionale, individuale o per *n* persone, in condizioni di certezza o in condizioni di incertezza, fino a «catturare» anche la stessa «pratica teorica». (E anche questo stesso esito è per un certo verso antitetico a quello della «filosofia della prassi» proprio in quanto, come si vedrà tra poche righe, consente di rivedere la pratica alla luce della teoria senza sottomettere – come pretende invece l'altra tradizione – le teorie a un qualche tipo di «criterio della prassi»).

Come si è sottolineato in altri articoli di questa stessa *Enciclopedia* (in particolare «Modello» e «Teoria/modello»), una teoria o paradigma del genere (non si è assunta qui una differenza rilevante tra i due termini, ma li si è considerati interscambiabili, preferendo ora l'uno ora l'altro a seconda dei particolari aspetti epistemologici che piú si volevano enfatizzare) non è altro che un opportuno «generatore di modelli»: per quelle situazioni problematiche, nella fattispecie, per cui Aristotele utilizzava il termine προαίρησις. Sotto questo profilo, la teoria «spiega» vari tipi di comportamento nel senso della usuale spiegazione scientifica: i modelli che essa genera non mirano a fornire delle certezze, ma a organizzare in una struttura semplice e intelligibile una massa di dati che precedentemente lasciava perplessi, a ridurre, per cosí dire, l'arbitrario delle descrizioni. Poiché obiettivo di tale modellizzazione è comunque il mondo dei desideri, delle preferenze e dei conflitti degli uomini, questi modelli – nella misura in cui «simulano» mediante opportuni strumenti matematici situazioni psicologiche, sociali, storiche, ecc. (valgano per tutti gli esempi, dati nel § 5, di teoria delle decisioni e teoria dei giochi, come il gioco perturbato o la *tracing procedure*, ecc.), potranno almeno contribuire ad allentare la rigida contrapposizione tra «spiegazione» e «comprensione» (per cui si veda, in particolare, l'articolo «Spiegazione» in questa stessa *Enciclopedia*). Dopotutto, i vari «assiomi di razionalità» che via via sono stati introdotti per sofisticare sempre piú la «teoria dell'utilità» qui delineata, possono venir interpretati anche come «leggi generali» per certe forme di comportamento umano.

L'eccezione e la regola: non diversamente dai modelli impiegati nelle scienze della natura, anche questo tipo di modellizzazione ha le sue tecniche di «validazione», cioè di controllo empirico. Ma il banco di prova di modelli che pretendono di spiegare la pratica non può essere che la pratica stessa. Il confronto con la tradizione di cui al § 6.3 sarà anche qui interessante: niente «unità dialettica» di teoria e prassi, ma interazione tra spiegazione e prescrizione: una teoria del comportamento razionale che spieghi una vasta gamma di azioni finirà per diventare prescrittiva fino al punto di correggere il comportamento degli agenti in conformità con le proprie leggi generali. Come guarderà, dunque, questa teoria della pratica alle sue «anomalie»? Se queste sono «recalcitranti» – cioè se sul lungo periodo gli agenti insistono nel *non* correggere il loro comportamento, a tali anomalie si guarderà nello stesso spirito con cui si è qui considerato qualche particolare «paradosso» o «patologia»: verranno viste, cioè, come indicazioni che un certo modello va sostituito da uno piú sofisticato senza necessariamente concludere che è «falsificata» la teoria (cfr. del resto quanto è stato osservato nel piú volte citato articolo «Induzione statistica», VII, pp. 385-386, 391).

Se infine si conviene di definire falsificatori potenziali di una teoria di questo tipo quei casi in cui essendo preteoricamente ovvio quali linee d'azione non sono ottimali, la teoria genera modelli in cui l'agente sceglie una di queste linee, la teoria viene a essere dotata anche di una sorta di «contenuto empirico». Naturalmente, dato che in svariati casi si è incerti di fatto su quale decisione prendere e non è affatto ovvio preteoricamente che una particolare linea d'azione non è ottimale, si tratta di «contenuto empirico» in un senso abbastanza pickwickiano. A parere di chi scrive, è forse piú adeguato parlare di teoria «quasi-empirica», nello stesso senso con cui Lakatos si è riferito al carattere «quasi-empirico» delle metodologie scientifiche (queste ultime sono, sostanzialmente, anch'esse delle «teorie della razionalità» che possono eventualmente rivelarsi in disaccordo con i giudizi di valore dei ricercatori militanti). Se la teoria descrive sufficientemente bene la classe di casi in cui è completamente determinata, organizzandoli in una struttura semplice e coerente, essa potrà indurre gli agenti in disaccordo a rivedere le proprie posizioni. In questo senso la teoria acquista un valore normativo: e il fatto che degli agenti conformino i loro comportamenti ad essa è a sua volta interpretabile come un segno della sua adeguatezza (continuando il parallelo con la metodologia: il caso di uno scienziato che riorienta la sua pratica di ricerca alla luce di certi principî di razionalità non è poi cosí raro come certi approcci a-teorici o anti-teorici alla metodologia vorrebero far credere). Si tratta, ovviamente, di un circolo: ma, per dirla con Goodman, di un circolo *virtuoso* (o, per finire come si è cominciato, di un equilibrio riflessivo). [S. M.].

Albert, H.
1978 *Traktat über rationale Praxis*, Mohr, Tübingen.

Allais, M.
1953 *Le comportement de l'homme rationnel devant le risque: critique des postulats et des axiomes de l'école américaine*, in «Econometrica», XXI, pp. 503-46.

Anscombe, F. J., e Aumann, R. J.
1963 *A definition of subjective probability*, in «Annals of Mathematical Statistics», XXXIV, pp. 199-205.

Arendt, H.
1958 *The Human Condition*, University of Chicago Press, Chicago (trad. it. Bompiani, Milano 1964).

Boltzmann, L.
[1899] *Über die Entwicklung der Methoden der theoretischen Physik in neuerer Zeit*, in *Populäre Schriften*, Barth, Leipzig 1905; ora in *Theoretical Physics and Philosophical Problems*, Reidel, Dordrecht-Boston 1974, pp. 77-100.

Carnap, R.
1950 *Logical Foundations of Probability*, University of Chicago Press, Chicago.

Debreu, G.
1959 *Theory of Value. An Axiomatic Analysis of Economic Equilibrium*, Wiley, New York.

Duhem, P.
1906 *La théorie physique, son objet et sa structure*, Chevalier et Rivière, Paris (trad. it. Il Mulino, Bologna 1978).

Feyerabend, P. K.
1962 *Explanation, Reduction and Empiricism*, in H. Feigl e G. E. Maxwell (a cura di), *Minnesota Studies in the Philosophy of Science*, III. *Scientific Explanation, Space, and Time*, University of Minnesota Press, Minneapolis, pp. 28-97.
1970 *Consolations for the Specialist*, in I. Lakatos e A. M. Musgrave (a cura di), *Criticism and the Growth of Knowledge*, Cambridge University Press, London, pp. 197-230 (trad. it. Feltrinelli, Milano 1976, pp. 277-312).
1978 *Science in a Free Society*, New Left Books, London.
1980 *Erkenntnis für freie Menschen*, Suhrkamp, Frankfurt am Main (trad. it. Feltrinelli, Milano 1981).

Gadamer, H. G.
1960 *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik*, Mohr, Tübingen 1965² (trad. it. Fabbri, Milano 1972).

Giorello, G., e Mondadori, M.
1978 *Dinamica della conoscenza scientifica e dialettica*, in U. Curi (a cura di), *La razionalità scientifica*, Francisci, Padova, pp. 107-49.

Gramsci, A.
[1933] *Introduzione allo studio della filosofia*, in *Quaderni del carcere*, Einaudi, Torino 1975, p. 1780.

Habermas, J.
1971 *Theorie und Praxis*, Suhrkamp, Frankfurt am Main 1971² (trad. it. Il Mulino, Bologna 1973).

Harsanyi, J. C.
1975 *The tracing procedure: a Bayesian approach to defining a solution for n-person non-cooperative games*, in «International Journal of Game Theory», IV, pp. 61-94.
1976 *Essays on Ethics, Social Behavior, and Scientific Explanation*, Reidel, Dordrecht-Boston.
1977a *On the Rationale of the Bayesian Approach: Comments on Professor Watkins' Paper*, in R. E. Butts e J. Hintikka (a cura di), *Foundational Problems in the Special Sciences*, Reidel, Dordrecht-Boston, pp. 381-92.
1977b *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press, New York.
1977c *Rule utilitarianism and decision theory*, in «Erkenntnis», XI, pp. 25-53.

Hintikka, J.
1971 *Unknown Probabilities, Bayesianism, and de Finetti's Representation Theorem*, in R. C. Buck e R. S. Cohen (a cura di), *Boston Studies in the Philosophy of Science*, VIII. *In Memory of Rudolf Carnap*, Reidel, Dordrecht-Boston, pp. 325-41.

Hintikka, J., e Pietarinen, J.
1966 *Semantic Information and Inductive Logic*, in J. Hintikka e P. Suppes (a cura di), *Aspects of Inductive Logic*, North-Holland, Amsterdam, pp. 96-112 (trad. it. in J. Hintikka, *Induzione, accettazione, informazione*, Il Mulino, Bologna 1974, pp. 143-60).

Hume, D.
1739 *A Treatise of Human Nature*, Noon, London 1739-40 (trad. it. in *Opere*, vol. I, Laterza, Bari 1971, pp. 5-665).

Joachim, H. H.
[1902-17] *Aristotle. The Nicomachean Ethics*, Clarendon Press, Oxford 1951.

Kant, I.
1787 *Kritik der reinen Vernunft*, Hartknock, Riga 1787² (trad. it. Utet, Torino 1967).
1788 *Kritik der praktischen Vernunft*, Hartknock, Riga (trad. it. in *Scritti morali*, Utet, Torino 1970, pp. 127-315).
1793 *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis*, in «Berlinische Monatsschrift», XXII, pp. 201-84 (trad. it. in *Scritti politici e di filosofia della storia e del diritto*, Utet, Torino 1965², pp. 237-81).

Kuhn, Th. S.
1962 *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago 1970² (trad. it. Einaudi, Torino 1978⁴).

Lakatos, I.
[1965] *Changes in the Problem of Inductive Logic*, in I. Lakatos (a cura di), *International Colloquium in the Philosophy of Science. Bedford College, 1965. The Problem of Inductive Logic*, North-Holland, Amsterdam 1968, pp. 315-417.
1970 *Falsification and the Methodology of Scientific Research Programmes*, in I. Lakatos e A. M. Musgrave (a cura di), *Criticism and the Growth of Knowledge*, Cambridge University Press, London, pp. 91-196 (trad. it. Feltrinelli, Milano 1976, pp. 164-276).

Lewis, D. K.
1969 *Convention: a Philosophical Study*, Harvard University Press, Cambridge Mass.

Luce, R. D., e Raiffa, H.
1957 *Games and Decisions. Introduction and Critical Survey*, Wiley, New York 1967⁷.

Mackie, J. L.
1980 *Hume's Moral Theory*, Routledge and Kegan Paul, London.

Mondadori, M.
1979 *Sulla pratica induttiva*, in «Materiali filosofici», nuova serie, n. 2-3, pp. 27-46.

Morgenstern, O.
[1966] *L'attitude de la nature et le comportement rationnel*, in E. M. Claassen (a cura di), *Les fondements philosophiques des systèmes économiques. Textes de Jacques Rueff et essais rédigés en son honneur, 23 août 1966*, Payot, Paris 1967, pp. 131-41.

Nash, J. F.
1950a *Equilibrium points in n-persons games*, in «Proceedings of the National Academy of Sciences of the United States of America», XXXVI, pp. 48-49.
1950b *The bargaining problem*, in «Econometrica», XVIII, pp. 155-62.
1951 *Non-cooperative games*, in «Annals of Mathematics», LIV, pp. 286-95.

Neumann, J. von, e Morgenstern, O.
1947 *Theory of Games and Economic Behavior*, Princeton University Press, Princeton N.J. 1947².

Niiniluoto, I., e Tuomela, R.
1973 *Theoretical Concepts and Hypothetico-Inductive Inference*, Reidel, Dordrecht-Boston.

Ostrom, V.
1964 *Culture, Science and Politics*, in W. J. Gore e J. W. Dyson (a cura di), *The Making of Decisions; a Reader in Administrative Behavior*, Free Press, Glencoe Ill., pp. 85-92.

Popper, K. R.
1974a *Autobiography*, in P. A. Schilpp (a cura di), *The Philosophy of Karl Popper*, vol. I, Open Court, La Salle Ill., pp. 3-181 (trad. it., della sola *Autobiography*, Armando, Roma 1976).
1974b *Replies to my Critics*, *ibid.*, vol. II, pp. 961-1197.

Rawls, J.
1971 *A Theory of Justice*, The Belknap Press of Harvard University Press, Cambridge Mass.

Schelling, T. C.
1960 *The Strategy of Conflict*, Harvard University Press, Cambridge Mass.

Selten, R.
1965 *Spieltheoretische Behandlung eines Oligopolsmodells mit Nachfrageträgheit*, in «Zeitschrift für die Gesamte Staatswissenschaft», CXXI, pp. 301-24, 667-89.

Simon, H. A.
1960 *The New Science of Management Decision*, Harper, New York.

Stegmüller, W.
1973 *Theorienstrukturen und Theoriendynamik*, in *Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie*, vol. II, tomo II, Springer, Berlin - New York.

Thom, R.
1980 *Parabole e catastrofi: intervista su matematica, scienza e filosofia*, Il Saggiatore, Milano.

Vico, G.
[1708] *De nostri temporis studiorum ratione*, Mosca, Napoli 1709 (trad. it. in *Opere*, Ricciardi, Milano-Napoli 1953, pp. 169-242).

Watkins, J. W. N.
[1965] *Non-Inductive Corroboration*, in I. Lakatos (a cura di), *International Colloquium in the Philosophy of Science. Bedford College, 1965. The Problem of Inductive Logic*, North-Holland, Amsterdam 1968, pp. 61-66.
1977 *Towards a Unified Decision Theory: a Non-Bayesian Approach*, in R. E. Butts e J. Hintikka (a cura di), *Foundational Problems in the Special Sciences*, Reidel, Dordrecht-Boston, pp. 346-79.

La classica **opposizione/contraddizione** (cfr. anche **coppie filosofiche, dialettica**) di «teoria» e «pratica» può dare luogo, nel quadro concettuale di una **scienza** congetturale e perpetuamente rivedibile (cfr. **conoscenza, ricerca**) a una **teoria/modello** del comportamento (cfr. **comportamento e condizionamento**) razionale (cfr. **ragione, razionale/irrazionale**) dell'**uomo**: si tratta infatti d'investigare le modalità di **decisione** individuale e collettiva, in condizioni sia di certezza sia d'incertezza (cfr. **certezza/dubbio**). Rilevante è allora il **paradigma** bayesiano (cfr. **probabilità**) che dal caso individuale (cfr. **induzione statistica**) può venir esteso a situazioni di **conflitto** tra vari individui, secondo le modalità della teoria dei **giochi** (cfr. anche **tattica/strategia**), e può fornire il supporto di un approccio all'**etica** stessa, nonché a questioni di teoria della **giustizia**, e piú in generale, di **politica** (cfr. anche **società, stato**). Nato nel contesto dell'**economia** (in relazione alla determinazione di punti di equilibrio: cfr. **equilibrio/squilibrio**) il principio che sottende questo paradigma, cioè il principio di massimizzazione dell'**utilità** attesa, può essere proficuamente applicato alla stessa «pratica teorica», cioè alla scelta fra varie **ipotesi** (cfr. anche **metodo**) in una valutazione comparata dell'**informazione** che esse arrecano. In questa prospettiva l'**unità** fra teoria e pratica non viene comunque postulata aprioristicamente nel cielo dell'**ideologia**, ma viene progressivamente realizzata per tentativi ed errori (cfr. **errore**) secondo le tipiche modalità di controllo (cfr. **empiria/esperienza, esperimento, verificabilità/falsificabilità**) che sono abituali per ogni **modello** scientifico.