SIS 2014

47th Scientific Meeting
of the Italian Statistical Society

Cagliari - June 11/13, 2014

Il comitato organizzatore è lieto di offrire ai partecipanti alla 47^ Riunione Scientifica della SIS l'articolo di Bruno de Finetti

“*Sull'impostazione assiomatica del calcolo delle probabilità*” Annali Triestini, XIX (1949), 29-81

e

“*Aggiunta alla nota sull'assiomatica della probabilità*” Annali Triestini, XX (1950), 5-22

L'articolo e la nota provengono dalle “Opere Scelte”, Vol. I di Bruno de Finetti (ed. Cremonese, 2006) a cura dell'Unione Matematica Italiana e Associazione per la Matematica Applicata alle Scienze Economiche e Sociali.

Il comitato organizzatore ringrazia vivamente l'UMI e AMASES che hanno dato il consenso alla divulgazione gratuita ai soli partecipamenti alla riunione SIS.

Cagliari, 11 Giugno 2014

Il Comitato Organizzatore

Estratto da «*Annali Triestini, sezione 2ª: scienza ed ingegneria*»
XIX (1949), pp. 29-81

SULL'IMPOSTAZIONE ASSIOMATICA DEL CALCOLO DELLE PROBABILITÀ

1. Introduzione

Un' impostazione assiomatica del calcolo delle probabilità può riuscire utile da diversi punti di vista: può prefiggersi semplicemente di elencare in modo esplicito le proprietà formali da cui si parte allo scopo di sviluppare organicamente e rigorosamente la trattazione; può prefiggersi, con ciò, di costruire una teoria formale accessibile a un matematico puro indipendentemente dalle questioni concettuali sulla probabilità tuttora controverse; può anche servire come piattaforma per impostare più chiaramente la discussione su tali questioni, almeno per gli aspetti aventi attinenza con le proprietà formali.

Al primo obbiettivo appare ispirata ad es. la trattazione del Kolmogoroff in Grundbegriffe der Wahrscheinlichkeitsrechnung (1), e al secondo quella del Cantelli sulla «teoria astratta» del calcolo delle probabilità (2); il presente lavoro si prefigge piuttosto il terzo obbiettivo, cercando cioè di esaminare le questioni formali, che vengono messe a fuoco nel modo più limpido mediante l' impostazione assiomatica astratta, collegandole alle impostazioni concettuali ispirate a punti di vista sostanziali sul significato della probabilità.

Naturalmente, anche tali impostazioni possono essere formulate mediante «assiomi», come, per la teoria della frequenza-limite, ha fatto il von Mises (3), e come, per la teoria soggettiva, ha fatto il Keynes (4) e più re-

---

(1) "Ergebnisse der Mathematik", Bd. 2, Heft 3, Springer, Berlin, 1933. Come prototipo di tali assiomatizzazioni va ricordata l'esposizione del Bohlmann, "Enz. Math.Wiss".

(2) F. P. CANTELLI - **Una teoria astratta del Calcolo delle probabilità** - "Giorn. Ist. It. Attuari", A. III, n. 2, 1932 (cfr. anche altri articoli del C. e dell' Ottaviani, ibidem, A. X, nn. 1-2, 1939).

(3) R. von MISES - **Grundlagen der Wahrscheinlichkeitsrechnung** - "Math. Zeitschrift", Bd. 5, 1919 (cfr. anche il suo trattato **Wahrscheinlichkeitsrechnung und ihre Anwendung ecc.**, Deuticke, Wien-Leipzig, 1931, e un' esposizione introduttiva, **Wahrscheinlichkeit, Statistik und Wahrheit,** Springer, Wien, 1936).

(4) J. M. KEYNES - **Treatise on probability** - London, 1921 (trad. ted. di F. M. URBAN, Barth, Leipzig, 1926).

centemente io stesso (5); parlando di «impostazioni assiomatiche» voglio però riferirmi qui soltanto a quelle dell' altro tipo precedentemente nominato, che assiomatizzano le proprietà formali della probabilità concepita astrattamente come funzione reale di enti denominati «eventi», indipendentemente da ogni interpretazione effettiva, e non a quelle che tendono viceversa a tradurre in forma assiomatica una effettiva interpretazione.

2. Piano del lavoro

Nel presente lavoro prenderemo le mosse dagli assiomi di Kolmogoroff, che meglio si prestano a illustrare, punto per punto, dapprima le modificazioni che mi appaiono opportune in sede di impostazione assiomatica astratta, e poi le considerazioni sul merito delle diverse questioni secondo le diverse concezioni della probabilità che verranno all' uopo brevemente richiamate.

Il punto più controverso, nell' ambito delle proprietà formali, è quello concernente l' additività completa, che sarà di conseguenza il più diffusamente trattato. Anzi originariamente il presente lavoro era stato scritto come esposizione su «I diversi punti di vista sull' additività completa delle probabilità» in seguito a un colloquio avuto in merito col prof. Cantelli (6 giugno 1942), e rifatto e ampliato dopo un secondo colloquio tre mesi dopo, ma rimasto inedito causa la sospensione della pubblicazione del «Giorn. Ist. Ital. Attuari» nel 1943; riprendendolo ora (1947-49) mi apparve l' opportunità di sviluppare sistematicamente tutte le questioni relative all' impostazione assiomatica che dovevano comunque entrare sia pur incidentalmente nella trattazione anche a volerne considerare come unico scopo la discussione della additività completa.

3. L' impostazione di Kolmogoroff

Nell' impostazione di Kolmogoroff il calcolo delle probabilità è lo studio dei **campi di probabilità** (E, P) (**Wahrscheinlichkeitsfelder**) così definiti (6):

a) si considera anzitutto una classe C di enti primitivi detti «**casi elementari**»;

(5) In due modi diversi: "schema delle scommesse" e "assiomi sulla disuguaglianza di probabilità"; cfr. **La prévision: ses lois logiques, ses sources subjectives,**"Ann. Inst. Poincaré", T. 7, fasc. I, Paris, 1937, e altri lavori ivi cit.

(6) Notazioni modificate per uniformità col seguito.

b) gli insiemi di «casi elementari», ossia le **sottoclassi** di C, si dicono «**eventi**»;

c) si considera una classe di eventi E e una funzione P sodisfacenti i seguenti

ASSIOMI:

I) E è un **corpo** di insiemi di C (cioè: somma, prodotto, differenza, di insiemi appartenenti ad E appartengono ancor sempre ad E),

II) E contiene C (cioè: fra gli eventi E che si considerano deve figurare l' «evento certo», insieme di **tutti** i «casi elementari»),

III) nel campo E è definita una funzione reale mai negativa P che ad ogni evento E di E fa corrispondere il numero P(E), che si dirà **probabilità** di E (ma si potrebbe anche chiamare ad es. **misura** o **massa** di E o similmente, volendo evitare anche nel nome ogni possibile riferimento a concetti controversi),

IV) $P(C) = 1$ (convenzione sul valore della probabilità dell' evento certo),

V) se E_1 e E_2 sono eventi E disgiunti (senza elementi comuni, ossia incompatibili), è

$$P(E_1 + E_2) = P(E_1) + P(E_2),$$

VI) se $E_1, E_2, \ldots, E_n, \ldots$ è una successione di eventi E, ciascuno contenuto nel precedente, e il cui prodotto è nullo (non esiste cioè alcun caso elementare comune a tutti gli E_n), allora

$$P(E_n) \to 0 \qquad (\text{per } n \to \infty).$$

Questo VI assioma (detto anche «postulato di continuità») che, come è noto e vedremo, è equivalente all' additività completa (o teorema esteso delle probabilità totali), non ha bisogno ovviamente di venire introdotto che quando si considerino campi infiniti (classi E di infiniti eventi).

Finalmente la **probabilità subordinata** $P(E/H)$ (probabilità dell'evento E subordinatamente all'evento H, essendo E ed H eventi della classe E) viene **definita** dal K. in base al «teorema delle probabilità composte» ponendo

$$P(E/H) = P(EH)/P(H),$$

naturalmente sotto la condizione $P(H) \neq 0$.

La **compatibilità** di detti assiomi è provata dal K. considerando E formata dai due soli eventi certo e impossibile con probabilità uno e zero.

4. I punti da discutere

Al riguardo vogliamo discutere i seguenti punti dubbi:

1) è necessario o opportuno considerare sistematicamente gli eventi come insiemi di «casi elementari»?

2) è necessario o opportuno definire congiuntamente il «campo di probabilità» con restrizioni su $\mathcal{E}$ e P?

3) è necessario o opportuno imporre restrizioni sul campo di eventi $\mathcal{E}$, e, in caso affermativo, quella di costituire un **corpo** (Ass. I) va conservata o indebolita o rafforzata?

4) se si abbandonasse tale restrizione, come si potrebbe enunciare una proprietà che tenga il posto della proprietà additiva (Ass. V)?

5) è necessario o opportuno postulare l'additività completa (Ass. VI)?

6) è opportuno assumere il teorema delle probabilità composte come **definizione** della probabilità subordinata?

7) è sodisfacente la sopra accennata dimostrazione della compatibilità?

5. Eventi e «casi elementari»

Per cominciare dal punto 1), vediamo cosa si possa dire in generale a proposito di un campo di eventi $\mathcal{E}$ qualunque, senza definirli come insiemi di «casi possibili». Occorrerà naturalmente ammettere di aver dato un significato all'evento certo e impossibile, alla negazione di un evento ($\overline{E} = \sim E =$ «non E»), al prodotto di due eventi ($E_1 E_2 =$ «E_1 e E_2») (e di conseguenza alla somma: $E_1 + E_2 = \sim (\overline{E}_1 \overline{E}_2) =$ «E_1 o E_2»), o identificando senz'altro «evento» con «proposizione» («affermazione», «Aussage»), o ammettendo comunque analoghe proprietà formali per gli eventi come enti astratti (restando quindi alle teorie sul significato della probabilità il compito di precisare il concetto di evento e delle operazioni logiche sugli eventi).

Se $\mathcal{E}$ è formata di un numero finito di eventi $E_1, E_2, \dots, E_n$, i «casi elementari» o «costituenti» sono gli eventi possibili C$\mathcal{E}$ consistenti nell'affermazione di dati m degli n eventi ($m = 0, 1, 2, \dots, n$) e nella negazione degli altri. E sono in numero di $s \leq 2^n$ (segno = se tutte le combinazioni degli E_h sono compatibili). Le somme di costituenti (2^s) sono tutti e soli gli eventi esprimibili mediante $E_1 \dots E_n$ e le operazioni logiche, e costituiscono il **corpo** K$\mathcal{E}$, minimo corpo contenente $\mathcal{E}$. Poichè $\mathcal{E}$ appartiene a K$\mathcal{E}$, abbiamo $n \leq 2^s$, e quindi per il numero s dei casi possibili C$\mathcal{E}$ e per il numero 2^s degli eventi del corpo K$\mathcal{E}$ si ha rispettivamente

$$\log n / \log 2 \leq s \leq 2^n ,$$
$$n \leq 2^s \leq 2^{2^n} ,$$

ove il valore minimo si ha se e solo se gli $\mathcal{E}$ formano un corpo ($\mathcal{E} =$ K$\mathcal{E}$), e il massimo se e solo se tutte le combinazioni degli $\mathcal{E}$ sono compatibili (7).

Se gli $\mathcal{E}$ sono infiniti, il minimo corpo contenente $\mathcal{E}$ è il corpo K$\mathcal{E}$ degli eventi esprimibili logicamente con un numero finito $E_1, E_2, \dots, E_n$ di eventi di $\mathcal{E}$; la potenza di K$\mathcal{E}$ è la stessa di $\mathcal{E}$: infatti è $\mathfrak{c} \leq \mathfrak{k}$ perchè K$\mathcal{E}$ contiene $\mathcal{E}$, e inversamente $\mathfrak{k} \leq \Sigma_h 2^{2^h} \mathfrak{c}^h = \aleph_0 \mathfrak{c} = \mathfrak{c}$ perchè h eventi tra gli $\mathcal{E}$ si possono scegliere in $\mathfrak{c}^h$ modi e ad ogni scelta corrispondono al più 2^{2^h} eventi, ma $\mathfrak{c}^h = \mathfrak{c}$ (8), ecc.

Ma non è possibile in generale trovare, in K$\mathcal{E}$, dei «casi elementari», nè, coi mezzi introdotti, possiamo uscire da K$\mathcal{E}$.

Per costruire altri eventi dobbiamo supporre introdotta anche l'operazione di **prodotto di infiniti eventi**. Si dicono **corpi boreliani** (9) quelli che insieme ad $E_1, E_2, \dots, E_n, \dots$ contengono anche il prodotto infinito degli E_n (infinità numerabile) (e quindi anche la **somma** di infinità numerabili di eventi del corpo), e con BK$\mathcal{E}$ (10) indicheremo il minimo corpo boreliano contenente $\mathcal{E}$ (e quindi K$\mathcal{E}$), ottenibile prolungando K$\mathcal{E}$ con somme e prodotti di infinità numerabili di eventi.

Diremo **corpi zermeliani** quelli godenti di analoghe proprietà per prodotti infiniti (e somme) anche non numerabili, e con Z$\mathcal{E}$ (= ZBK$\mathcal{E}$ = = ZK$\mathcal{E}$) indicheremo il minimo corpo zermeliano contenente $\mathcal{E}$, ottenibile prolungando BK$\mathcal{E}$ con somme e prodotti di infinità non numerabili.

Soltanto in Z$\mathcal{E}$ siamo certi di incontrare la classe dei «casi elementari» o «costituenti» di $\mathcal{E}$, che indicheremo ancora con C$\mathcal{E}$, e che si costruisce come per $\mathcal{E}$ finita salvo che occorre far uso di prodotti infiniti (benchè nulla vieti che i C$\mathcal{E}$ appartengano, tutti o parte, già a $\mathcal{E}$ o a K$\mathcal{E}$ o a BK$\mathcal{E}$).

Quanto alla **potenza** di C$\mathcal{E}$ e Z$\mathcal{E}$ (che indicheremo con $\mathfrak{s}$ e $\mathfrak{z}$, mentre $\mathfrak{k} = \mathfrak{c}$ è quella di K$\mathcal{E}$ ed $\mathcal{E}$), abbiamo (11)

$$\mathfrak{k} \leq \mathfrak{z} \leq 2^{\mathfrak{s}} \quad , \quad \mathfrak{s} \leq 2^{\mathfrak{k}}$$

(7) Il caso di un numero finito di eventi si trova esposto secondo tali concetti in P. MEDOLAGHI, **La logica matematica e il calcolo delle probabilità**, "Boll. Ass. It. Attuari", n. 18, 1907.

(8) Cfr., per la dimostrazione (che, si badi, richiede naturalmente il principio di Zermelo), a pp. 217 e 233 dell'opera: W. SIERPINSKI - **Leçons sur les nombres transfinis** - Gauthier-Villars, Paris, 1928.

(9) Così anche in Kolmogoroff, op. cit. (1), p. 15.

(10) Anche in Kolmogoroff, op. cit. (1) e (9), B$\mathcal{F}$ indica il prolungamento boreliano del corpo $\mathcal{F}$.

(11) Quanto alla potenza di BK$\mathcal{E}$, è ovviamente compresa tra quelle di K$\mathcal{E}$ e Z$\mathcal{E}$; sembra però che si dovrebbe poter dire al riguardo qualcosa di più preciso.

(perchè una classe di potenza $\mathfrak{n}$ ha $2^{\mathfrak{n}} > \mathfrak{n}$ sottoclassi); poichè non si è mai incontrata l' esistenza di numeri cardinali intermedi fra $\mathfrak{n}$ e $2^{\mathfrak{n}}$ ($\mathfrak{n}$ qualunque purchè infinito), ed anzi secondo l' ipotesi di Cantor (12) non ne esistono, chè $2^{\mathfrak{n}}$ sarebbe il numero successivo ad $\mathfrak{n}$, potremo dire che i soli casi per ora incontrabili (ed anzi, secondo l' ipotesi di Cantor, i soli possibili) sono i tre seguenti (13)

$$\mathfrak{k} = 2^{\mathfrak{s}} \quad , \quad \mathfrak{k} = \mathfrak{s} \quad , \quad \mathfrak{s} = 2^{\mathfrak{k}} \quad ,$$

che sono effettivamente possibili avendosi ad es. il primo se E è già un corpo zermeliano ($E = \mathrm{Z}E$), il secondo se gli E sono incompatibili ($E = \mathrm{C}E$), il terzo se gli E hanno tutte le combinazioni compatibili.

I primi casi, con infinità numerabili ($\mathfrak{n}$), continue ($\mathfrak{c} = 2^{\mathfrak{n}}$) o funzionali ($\mathfrak{f} = 2^{\mathfrak{c}}$) si hanno ad es. per (14):

$$\left.\begin{array}{l} 1)\ E = \text{numeri interi } (\mathfrak{n}) \\ 2)\ E = \text{classi di interi } (\mathfrak{c}) \end{array}\right| \mathrm{C}E = \text{numeri interi } (\mathfrak{n}),\ \mathrm{Z}E = \text{classi di interi } (\mathfrak{c})$$

$$\left.\begin{array}{l} 3)\ E = \text{intervalli con estremi razionali } (\mathfrak{n}) \\ 4)\ E = \text{punti } (\mathfrak{c}) \ \ (\text{o intervalli, id.}) \\ 5)\ E = \text{insiemi lineari } (\mathfrak{f}) \end{array}\right| \begin{array}{l} \mathrm{C}E = \text{punti } (\mathfrak{c}), \\ \quad \mathrm{Z}E = \text{insiemi lineari } (\mathfrak{f}). \end{array}$$

Le considerazioni che possiamo trarre da quanto sopra in relazione alle nostre questioni sono le seguenti:

— comunque vengano dati degli eventi E, è possibile costruire i $\mathrm{C}E$ e considerare gli E come somme di insiemi di $\mathrm{C}E$;

— valgono quindi sempre tutte le proprietà formali come se si partisse dalla definizione basata sui «casi possibili», che non è dunque a tale effetto necessaria;

— e neppure è opportuna per i motivi seguenti:

— i «casi elementari» non sono eventi «privilegiati», «indivisibili» (come i «punti»), chè, ampliando la classe degli eventi E in $E + E_1$, ciascuno dei «casi elementari» $\mathrm{C}E$ si scinde (in generale, cioè salvo incompatibilità o appartenenza di E_1 a $\mathrm{K}E$) nei casi ottenuti come prodotti di esso per i $\mathrm{C}E_1$;

(12) Ogni numero cardinale è un alef, e $\aleph_{\alpha+1} = 2^{\aleph_\alpha}$; in particolare, per $\alpha = 0$, si ha la celebre "ipotesi del continuo" ($\mathfrak{c} = \aleph_1$, ossia: un insieme di potenza inferiore a quella del continuo è finito o numerabile).

(13) Infatti, $\mathfrak{s} \leq 2^{\mathfrak{k}}$ significa (sotto detta ipotesi) $\mathfrak{s} = 2^{\mathfrak{k}}$, o $\mathfrak{s} = \mathfrak{k}$, o $\mathfrak{s} < \mathfrak{k}$; ma questa ultima alternativa è compatibile con la disuguaglianza $\mathfrak{k} \leq 2^{\mathfrak{s}} < 2^{\mathfrak{k}}$ solo se $\mathfrak{k} = 2^{\mathfrak{s}}$.

(14) Manca, per $\mathfrak{s} = \mathfrak{n}$, il caso $\mathfrak{s} = 2^{\mathfrak{k}}$, non esistendo $\mathfrak{k}$ tale che $2^{\mathfrak{k}} = \mathfrak{n}$; ciò avviene naturalmente per ogni $\mathfrak{s} = \aleph_\alpha$ ove α non sia successivo di altro ordinale (ad es. per $\mathfrak{s} = \aleph_\omega$).

— la disomogeneità fra eventi «casi elementari» o «insiemi», e le distinzioni che si potrebbero istituire considerando effettivamente gli eventi come «insiemi» (p. es. finiti, numerabili, ecc.) sono prive di significato;

— il fatto che per giungere ai C bisogna uscire in generale non solo da E, ma da $\mathrm{K}E$ e $\mathrm{BK}E$, pervenendo a $\mathrm{Z}E$ (ossia: non si può far a meno di operazioni transfinite, neppure numerabili), fa apparire desiderabile di non essere obbligati a farvi ricorso sistematicamente;

— tanto più che, in diverse concezioni della probabilità, tale prolungamento del campo degli eventi dà luogo a difficoltà (15).

6. Campi di eventi e probabilità

Sulla base delle nozioni fissate al n. precedente possiamo ora esaminare le questioni 2), 3), e 7) (v. n. 4), che sono strettamente congiunte.

Includere la definizione di evento o di campo di eventi in quella di «campo di probabilità (E, P)», o invece definire le funzioni di probabilità P relative a un campo di eventi E supponendo tale nozione già preventivamente acquisita, è ovviamente solo questione di forma, ma la seconda alternativa appare preferibile per scindere i problemi sui campi di eventi, che sono anteriori e indipendenti dal concetto di probabilità, da quelli riguardanti l' introduzione della probabilità in dati campi di eventi.

Per seguire nel modo più completo tale linea di condotta, è anche opportuno prescindere da ogni restrizione sul campo degli eventi E, non imponendo ad es. che sia un corpo; infatti, se restrizioni dovranno introdursi, ciò non potrà farsi giustificatamente se non in base alle esigenze che potranno apparire in conseguenza dei principi della teoria delle probabilità. Vedremo che, pur di modificare adeguatamente qualche enunciato, si potrà evitare ogni restrizione, e ciò risulterà particolarmente sodisfacente quando rileveremo (nel n. 8) come col pretendere ad es. che gli E formino un corpo ci si precluda di porre il problema conformemente alle sue proprie esigenze.

Lo porremo quindi come segue:

data una classe di eventi E qualunque, vedere se esistono funzioni P(E) definite per ogni E di E e soddisfacenti gli «assiomi» con cui si vorranno caratterizzare le funzioni atte a rappresentare una «distribuzione di probabilità» in E.

Ponendo la questione in tal modo si consegue una maggior chiarezza:

— nel considerare l' insieme di tutte le funzioni P ammissibili in un dato campo E (v. nn. 14, 23, ecc.),

(15) Come si vedrà in seguito; cfr. in partic. il n. 8.

— nel considerare la possibilità di prolungare P, univocamente o più o meno arbitrariamente, dal campo E a un altro campo più ampio qualunque, in particolare al corpo KE, al corpo boreliano B KE, al corpo zermeliano ZE,

— nel concepire la questione della compatibilità dei postulati.

Abbiamo infatti un problema della compatibilità per ogni campo E; la dimostrazione accennata al n. 3 (in fine) dice invece soltanto che gli assiomi del calcolo delle probabilità (secondo K.) non sono contraddittori applicandoli al caso di due soli eventi, uno certo e uno impossibile, caso in cui il calcolo delle probabilità è superfluo, ed è chiaro che ciò non basta affatto per autorizzare a sviluppare la teoria da applicarsi invece in campi E ove interessa, ma ove non sappiamo se gli assiomi sono ancora non contraddittori.

7. Cenno sulle concezioni della probabilità

Occorre ora accennare, sia pur nel modo più sommario possibile, alle diverse concezioni della probabilità, sia per qualche riflesso che ne deriva già per le questioni sugli eventi in parte ormai trattate, sia e più per quelle sulla probabilità.

Distingueremo un po' arbitrariamente (date le infinite diverse sfumature da autore a autore) quattro concezioni:

a) Concezione classica, basata sui casi ugualmente probabili (per ragioni di simmetria) e sulla conseguente definizione della probabilità come rapporto tra il numero dei casi «favorevoli» e «possibili»;

b) Concezione empirica, basata sul concetto di eventi ripetibili la cui frequenza su un gran numero di prove (per la «legge empirica del caso») dà quasi certamente e quasi esattamente la probabilità;

c) Concezione asintotica, che rende più precisa, ma idealizzandola, la precedente, in quanto considera una successione infinita di prove («Kollektiv») definendo la probabilità come limite della frequenza;

d) Concezione soggettiva che considera la probabilità come misura del grado di fiducia di un soggetto determinato nell'avverarsi di un evento (proposizione).

Fra gli autori moderni si possono far rientrare nel tipo b) le posizioni di Castelnuovo, Cantelli, Fréchet, Lévy, e anche di Borel benchè si avvicini al d); nel tipo c) quelle di von Mises, Kamke, Reichenbach, e vi si dichiara favorevole anche Kolmogoroff; nel tipo d) quelle di Keynes, Jeffreys, e dello

scrivente. Per maggiori dettagli occorre rinviare agli scritti ove tali concezioni sono esposte e discusse (16); per dare un' idea delle varianti si avverta tuttavia ad es. (il che ci sarà necessario) che nella c) il von Mises introduce un «Regellosigkeitsaxiom» (assioma di non-regolarità) che il Kamke tralascia ed altri precisano indebolendolo (17), e che nella d) i citati autori inglesi non appaiono tuttavia soggettivisti nel pieno senso del termine.

Ma qui ci basta il poco che occorrerà per stabilire qualche caposaldo dell' impostazione, e maggiori delucidazioni sarebbero più adatte a fuorviare dall' intento che a favorirlo.

8. Riflessi sull' impostazione formale.

Vediamo anzitutto quale diverso significato viene ad assumere l' impostazione formale stabilita, pur restando invariata per sè stessa, a seconda della concezione cui ci si riferisca.

La concezione classica non si applica direttamente che a casi speciali (sulla cui esistenza e delimitazione vi è del resto ampia disparità di vedute (18)); dal punto di vista di un'analisi critica essa non può dare che un incentivo a ricondurre gli assiomi sulla probabilità quantitativa (misura numerica della probabilità) ad assiomi di natura puramente qualitativa (disuguaglianza di probabilità).

---

(16) Oltre le opp. citt. ad (1), (2), (3), (4), accenniamo (tra le molte) alle seguenti: G. CASTELNUOVO, **Calcolo delle probabilità** (diverse ed.), Zanichelli; **La probabilité dans les différentes branches de la science**, "Act. Sci. Ind.", n. 463, Hermann, Paris, 1937; E. BOREL, **Valeur pratique et philosophie des probabilités**, (è il 3° fasc. del 4° tomo del "Traité" diretto da Borel), Gauthier-Villars, Paris 1939; E. KAMKE, **Einführung in die Warhscheinlichkeitstheorie**, Leipzig, 1932; H. REICHENBACH, **Wahrscheinlichkeitslehre**, Sijthoff, Leiden, 1935; H. JEFFREYS, **Scientific Inference**, Cambridge Un. Press, 2.a ed., 1937. Vedasi pure contributi vari negli "Actes du Congrès International de Philosophie Scientifique" (Parigi, 1935) e particolarmente il Fasc. IV ("Act. Sci. Ind.", n. 391, Hermann, Paris, 1936), e negli atti del "Colloque consacré à la théorie des probabilités" (Ginevra, 1937), e particolarmente il Fasc. II ("Act. Sci. Ind." n. 735, Hermann, Paris, 1938); più recentemente discussioni sull' argomento ebbero luogo in America (ma non potei ancora vederne gli atti). Per la mia posizione nei riguardi dei diversi punti di vista, si possono vedere: **Probabilismo**, Perrella, Napoli, 1931; **Statistica e probabilità nella concezione di R. von Mises, Probabilisti di Cambridge, Punti di vista: Emile Borel, Punti di vista: Hans Reichenbach, Punti di vista: D. van Dantzig**, tutti in "Suppl. statistico", poi "Statistica", anni 1936-1941;

(17) Cfr. p. es. A. WALD, **Die Widerspruchsfreiheit des Kollektivbegriffes**, nel cit. n. 735 di "Act. Sci. Ind." (più distesamente, in "Ergebnisse eines math. Kolloquiums", Heft 8, Wien 1937).

(18) Basti accennare alle polemiche del secolo scorso per l' interpretazione "oggettiva" o "soggettiva" **dell' equiprobabilità** (nulla in comune con "soggettivo" nel senso moderno).

Nella concezione empirica un evento ha una probabilità sotto ipotesi mal precisabili di «stabilità della frequenza»; impossibile basarvi analisi rigorose, ma tuttavia, con ragionamento analogo a quello che svilupperemo per la concezione asintotica, si può escludere che gli eventi «dotati di probabilità» debbano necessariamente costituire un corpo.

Nella concezione asintotica la precedente imprecisa condizione di stabilità della frequenza è sostituita da quella di esistenza della frequenza limite (più, eventualmente, una di «non regolarità», cfr. n. 7); anche qui la probabilità non può supporsi definita per ogni evento, e gli eventi per cui la probabilità esiste non formano un corpo. Precisamente, se A e B hanno probabilità determinata, il prodotto AB può non averla, come scende dal seguente esempio.

Consideriamo una successione di prove a testa e croce, che, stando alla concezione asintotica, dovrà ritenersi certamente irregolare e avente frequenza-limite 1/2.

L' individuo A punti sempre su «testa», mentre B alternativamente sempre su testa o sempre su croce negli intervalli per n tra 1! e 2!, 2! e 3!, 3! e 4!, e in generale su testa per $(2h)! < n \leq (2h+1)!$ e su croce per $(2h+1)! < n \leq (2h+2)!$ $(h = 0, 1, 2, \ldots)$. Gli eventi A e B, vincita di A e di B, sono «irregolari» e con frequenza limite 1/2, A per coincidere con «testa», B perchè la frequenza g_n di B per $(h-1)! < n \leq h!$ vale, posto $m = (h-1)!$

$$\text{per } h \text{ pari:} \quad g_n = f_n + \frac{m}{n}\,[g_m - f_m]$$

$$\text{per } h \text{ dispari:} \quad g_n = (1 - f_n) + \frac{m}{n}\,[g_m - (1 - f_m)],$$

da cui, per $n = h!$, $|g_n - f_n| = \frac{1}{h}\,|g_m - f_m| \leq \frac{1}{h}$ (oppure $|g_n - (1 - f_n)| < \frac{1}{h}$), e quindi, fissato a piacere $\varepsilon > 0$, se $n = h!$ è sufficientemente grande perchè sia $|f_k - \frac{1}{2}| < \varepsilon$ per $k \geq n$ e inoltre $1/h < \varepsilon$, abbiamo che in ogni caso, sia h pari o dispari, tanto $|g_n - f_n|$ che $|g_n - (1 - f_n)|$ sono $< 3\varepsilon$. Applicando tale disuguaglianza ai g_m delle formule di partenza, si ricava allora per n qualunque $(> h!)$

$$|g_n - f_n| < \frac{m}{n}\,3\varepsilon < 3\varepsilon \qquad \text{risp.} \qquad |g_n - (1 - f_n)| < 3\varepsilon,$$

da cui

$$\left|g_n - \tfrac{1}{2}\right| < 4\varepsilon \quad , \quad \text{ossia} \quad g_n \to \tfrac{1}{2}, \qquad \text{c.d.d.}$$

Consideriamo ora l' evento AB (vincita di entrambi i giocatori): esso è impossibile nei tratti $(2h+1)! < n \leq (2h+2)!$ dove A e B fanno puntate opposte, e quindi sulle prime $n = (2h+2)!$ prove esso si verifica solo in parte delle

prime $(2h+1)!$ prove e la frequenza è $\varphi_n < \frac{1}{2h+2}$, sicchè la frequenza ha minimo limite zero; nei tratti $(2h)! < n \leq (2h+1)!$ invece AB coincide con A (e con B), e quindi sulle prime $n = (2h+1)!$ prove esso si è verificato tante volte quante A (cioè nf_n) meno al più $(2h)!$, quindi con frequenza $\varphi_n > f_n - \frac{1}{2h+1}$, sicchè la frequenza ha massimo limite 1/2. L' evento AB non ammette probabilità; non vale dunque per gli eventi che ammettono probabilità la proprietà di costituire un corpo.

Nella concezione soggettiva, dove un evento è un caso singolo (singola prova) ben determinato, non sembra invece esista alcun motivo plausibile che precluda la possibilità di esprimere in una valutazione di probabilità il proprio giudizio sull' avverarsi di un evento qualsivoglia, sicchè una distinzione fra eventi che ammettono o no una probabilità appare difficilmente accettabile. Tuttavia la restrizione che la classe d' eventi E ove si suppone noto il valore della probabilità costituisca un corpo non è neppur qui opportuna, perchè il fatto di aver valutato le probabilità di due eventi A e B non determina la probabilità di AB, e, più in generale, supposta nota la probabilità per una classe di eventi E, essa risulta univocamente estensibile non a tutto il corpo KE ma a quello che definiremo (n. 9) come il sistema lineare LE.

9. L' additività semplice.

L' «assioma V» di Kolmogoroff circa l' additività (semplice) della probabilità, ossia il teorema delle probabilità totali, è ammesso in tutte le concezioni (19):

— nella classica, empirica, asintotica, è conseguenza puramente aritmetica della definizione, perchè sono naturalmente additivi sia il numero dei «casi favorevoli» che la «frequenza» e il «limite della frequenza»,

— nella soggettiva come condizione necessaria per la coerenza fra la valutazione di diverse probabilità (secondo diverse considerazioni, p. es. quella basata sullo schema delle scommesse; v. in questo stesso n. 9).

La formulazione di tale assioma si riferisce al caso di eventi incompatibili, e dice che la probabilità della somma di due (e quindi di un numero finito qualunque) di eventi incompatibili è la somma delle loro probabilità; ma se

(19) Un dubbio sull' opportunità di una variante è stato sollevato dal Medolaghi; la variante non avrebbe però significato sostanziale ma farebbe semplicemente sostituire la scala usuale delle probabilità p con altra f (p), f crescente, con convenzione meno opportuna (cfr. B de FINETTI, **Sul significato soggettivo della probabilità**, "Fundamenta Mathematicae", T. XVII, Warszawa 1931, nota (1) a p. 319).

consideriamo degli eventi qualunque $E_1 \dots E_n$, non incompatibili, cosa ci dice la restrizione? Essa è ancora significativa se la si applica ai «costituenti», ma se non si postula che gli E considerati formino un corpo i costituenti sono irraggiungibili, e comunque le restrizioni per gli E_h risulterebbero in forma indiretta.

Se ad es. supponiamo $\mathscr{E}$ formata dei tre soli eventi E_1, E_2, ed $E_3 = E_1 + E_2$ (E_1 e E_2 compatibili) dev' essere ovviamente $p_3 \leq p_1 + p_2$ (precisamente $p_3 = p_1 + p_2 - p_{12}$), ma non possiamo dimostrarlo in base al teorema delle probabilità totali se non ricorrendo alla considerazione di eventi fuori di $\mathscr{E}$ (i prodotti $E_1 E_2$, $E_1 \overline{E}_2$, $\overline{E}_1 E_2$).

E' possibile formulare esplicitamente l' assioma in modo da potersi applicare direttamente a qualunque insieme di eventi $\mathscr{E}$, anche non incompatibili, anche non costituenti un corpo?

Per tale scopo conviene introdurre la nozione di «numeri aleatori $L\mathscr{E}$ linearmente dipendenti da una classe di eventi $\mathscr{E}$» definendoli come combinazioni lineari degli «indicatori» di un numero finito tra gli eventi $\mathscr{E}$, ove indicatore di un evento E si dica il numero aleatorio, che indicheremo con $|E|$, assumente il valore 1 se E si verifica o 0 se non si verifica. Fra gli $\mathscr{E}$ supporremo sempre incluso l'evento certo E_0 (anche se non espressamente menzionato) perchè la sua probabilità $P(E_0) = 1$ è comunque fissata; i numeri aleatori $L\mathscr{E}$ sono quindi quelli esprimibili nella forma

$$X = k_o + k_1 |E_1| + k_2 |E_2| + \dots + k_n |E_n|$$

ove $E_1 \dots E_n$ appartengono ad $\mathscr{E}$ (per l'evento certo E_o è $|E_o| = 1$, sicchè s'è scritto semplicemente k_o anzichè $k_o |E_0|$). Diremo poi che un evento E appartiene ad $L\mathscr{E}$ se vi appartiene il suo indicatore $|E|$; vi appartiene quindi $E_1, E_2, \dots \dots E_n$; una loro somma come $E_1 + E_2$ vi appartiene se E_1 e E_2 sono incompatibili (allora $|E_1 + E_2| = |E_1| + |E_2|$), oppure se il prodotto $E_1 E_2$ è uno degli E_h, p. es. E_3 (allora $|E_1 + E_2| = |E_1| + |E_2| - |E_3|$), ecc.

Come significato intuitivo, possiamo dire che un numero aleatorio appartenente ad $L\mathscr{E}$ è il risultato di un qualunque sistema di scommesse su un numero finito di eventi della classe $\mathscr{E}$, e un evento appartenente ad $L\mathscr{E}$ un evento per il quale è possibile con un tale sistema di scommesse far sì che il risultato sia 1 o 0 a seconda che esso si verifica o no (si pensi agli esempi precedenti); naturalmente non ha ancora alcun senso la distinzione fra «scommesse eque» o no, dipendente solo dall' introduzione della probabilità P(E).

Supponiamo ora definita per gli E di $\mathscr{E}$ una funzione reale non negativa P(E), senza postulare che sia additiva dato che vogliamo appunto analizzare tale restrizione. Diremo che un numero aleatorio X di $L\mathscr{E}$ è equo rispetto

alla funzione P se è possibile scriverlo nella forma $X = k_o + k_1 |E_1| + \dots + k_n |E_n|$ in modo che risulti

$$k_o = -\{k_1 P(E_1) + k_2 P(E_2) + \dots + k_n P(E_n)\};$$

allora infatti, per dire il significato, X è una combinazione di scommesse su $E_1 \dots E_n$ colle quote $P(E_1) \dots P(E_n)$ (cioè: una lira se E si verifica, vale P (E)).

Dimostriamo ora che, se $\mathscr{E}$ è un corpo, la condizione di additività (semplice)

(A) $P(E_1 + E_2) = P(E_1) + P(E_2)$ (per E_1 e E_2 incompatibili)

è equivalente a ciascuna delle forme seguenti

(A') se un numero aleatorio di $L\mathscr{E}$ è certamente positivo, esso non è equo rispetto alla funzione P, o addirittura

(A'') non ogni numero aleatorio $L\mathscr{E}$ è equo rispetto alla funzione P.

Dalla (A) scende infatti notoriamente la possibilità di definire per X il valor medio o speranza matematica $\mathscr{M}(X) = k_o + k_1 P(E_1) + \dots + k_n P(E_n)$ e la definizione data di equità significa $\mathscr{M}(X) = 0$; se X è certamente positivo (se riescono positivi, cioè, tutti i valori possibili) è $\mathscr{M}(X) \neq 0$, sicchè vale la (A'), e, poichè di tali numeri ne esistono certamente (p. es. se $k_o > \Sigma |k_i|$), vale anche la (A'').

Se invece la (A) è falsa, ossia esistono E_1 e E_2 incompatibili per cui $P(E_1 + E_2) \neq P(E_1) + P(E_2)$, risulta «equo rispetto a P» il numero aleatorio $k\{|E_1 + E_2| - |E_1| - |E_2| - P(E_1 + E_2) + P(E_1) + P(E_2)\}$ che è una qualunque costante (perchè $|E_1 + E_2| - |E_1| - |E_2| \equiv 0$, trattandosi di eventi incompatibili, e resta la costante $k\{P(E_1) + P(E_2) - P(E_1 + E_2)\}$); ma allora se è equo X lo è anche X + k e quindi qualunque $L\mathscr{E}$.

L' enunciato (A') implica di per sè anche la non-negatività di P (E), sicchè esso può assumersi come unico assioma (in luogo del III e V di K.), ed è esprimibile per una classe $\mathscr{E}$ qualunque (anche non costituente un corpo), in quanto vale il seguente

TEOREMA:

una funzione reale P(E) definita in una classe di eventi $\mathscr{E}$ qualsiasi può essere prolungata a tutto il corpo $K\mathscr{E}$ (e allora anche a qualsiasi corpo $\mathscr{K} \supset K\mathscr{E}$) in modo che vi risulti additiva e mai negativa, se e soltanto se in $\mathscr{E}$ soddisfa la (A').

Che la condizione sia necessaria è ovvio: la (A) in $K\mathscr{E}$ equivale alla (A') che non può valervi se già non vale in $\mathscr{E}$. Che sia sufficiente dimostriamo per induzione (transfinita, se K è infinito non numerabile) supponendo K bene-ordinato (cominciando con E, che sarà un segmento K_ε di K): basterà dimostrare che se P si è potuto prolungare, rispettando la (A'), ad un segmento K_α, lo stesso si può fare anche in $K_{\alpha+1}$ (α numero ordinale qualunque), ossia basta dimostrare che se P soddisfa la (A') in una classe E qualsiasi, è possibile anche per un nuovo evento E qualsiasi attribuire a P (E) un valore in modo che la (A') continui ad essere soddisfatta in $E + (E)$.

Consideriamo dunque un evento E e vediamo per quali p ($0 \leq p \leq 1$) la posizione P (E) = p risulti compatibile con la (A') e coi valori di P supposti già definiti in E: si ha incompatibilità se esistono degli eventi $E_1, E_2, \ldots, E_n$ di E (in numero finito) le cui probabilità indichiamo $p_1, p_2, \ldots, p_n$, e delle costanti $k_1, k_2, \ldots, k_n$ tali che

$$X = |E| - p + \sum_{1}^{n}{}_h k_h \{|E_h| - p_h\}$$

risulti sempre positivo (e allora p è un valore troppo piccolo per P(E)) oppure sempre negativo (e allora p è troppo grande); indichiamo con p' l' estremo superiore dei valori «troppo piccoli» e con p'' l' estremo inferiore dei valori «troppo grandi» (ovviamente: risultano risp. «troppo grandi» e «troppo piccoli» tutti i $p > p''$ e tutti i $p < p'$); se è $p' < p''$, risultano compatibili tutti i valori $p' \leq p \leq p''$ (che gli estremi vadano inclusi risulta per continuità (20)); se è $p' = p''$, solo il valore $p = p' = p''$; si ha invece incompatibilità se risulta $p' > p''$, ossia se esiste un valore $p'' < p < p'$ tale che dà luogo sia a un numero aleatorio X sempre positivo che ad un altro X' (con altri eventi $E'_1 \ldots E'_{n'}$ e costanti $k'_1 \ldots k'_{n'}$) sempre negativo. Ma in tal caso $X - X'$ è sempre positivo, e facendo la differenza si elide l'addendo $|E| - p$ sicchè $X - X'$ risulta linearmente dipendente da $E_1 \ldots E_n E'_1 \ldots E'_{n'}$, ossia da E: la (A') non vale già in E.

Possiamo notare incidentalmente che le infinite scelte arbitrarie (di un p tra p' e p'') si possono anche eseguire in modo determinato (p. es. $p = p'$, $p = p''$, $p = \frac{1}{2}(p' + p'')$, in generale $p = \lambda p' + (1 - \lambda) p''$ con λ numero ($0 \leq \lambda \leq 1$) costante o funzione dell'evento).

10. Campo di univocità del prolungamento.

Il ragionamento del n. precedente consente anche di precisare quali sono gli eventi E la cui probabilità risulta univocamente determinata quando sia nota P in E.

Avendo ormai accertato che la (A') implica l' additività per P e quindi la possibilità di definire in LE il «valor medio» $M(X)$, converrà indicare addirittura con P(X) tale valor medio intendendo così di estendere la funzione $P(|E|) = P(E)$ per linearità ad ogni X combinazione lineare degli $|E|$ (21).

(20) L' ovvia dimostrazione si trova del resto in op. cit. (5).

La condizione può allora esprimersi

(A''') P è funzione additiva per gli X di LE, e non negativa nel senso che

se $X \geq 0$ (certamente), $P(X) \geq 0$.

Per un X non appartenente ad LE, è possibile attribuire a P(X) un qualunque valore purchè $P(X) \leq P(X')$ per ogni X' di LE tale che $X \geq X'$ (certamente), e $P(X) \leq P(X')$ nel caso opposto. In particolare le limitazioni per P(E) risultano dal confronto colle combinazioni X' di LE sempre $\geq |E|$ (cioè ≥ 1 se E si verifica e ≥ 0 comunque), risp. $\leq |E|$ (cioè ≤ 1 comunque, e ≤ 0 se E non si verifica).

In tal modo potremo, con vantaggio di omogeneità, impostare ogni questione rispetto ai numeri aleatori di LE senza preoccuparsi di distinguere e riconoscere a priori quelli che rappresentano (come «indicatori») gli eventi appartenenti ad LE, che sono quelli che ammettono i due soli valori possibili 0 e 1.

Indichiamo con L_pE l' insieme (ovviamente sistema lineare) dei numeri aleatori X per i quali P(X) risulta univocamente determinato quando si assegni in E (e quindi ovviamente in LE) la funzione P; in particolare, gli eventi E appartenenti a L_pE saranno quelli la cui probabilità P(E) risulta univocamente determinata; in altre parole L_pE è il campo in cui il prolungamento della P risulta univoco.

Ovviamente LE appartiene ad L_pE qualunque sia la funzione P, ma interessa vedere se e sotto quali condizioni L_pE coincide con LE, e altrimenti quali nuovi numeri aleatori ed eventi appartengono ad L_pE, e se ne esistono che appartengono ad L_pE qualunque sia la funzione P, pur senza appartenere ad LE, ecc.

Occorre anzitutto distinguere i due casi che, secondo (A'''), possono darsi per una $X \geq 0$: può essere $P(X) > 0$, oppure $P(X) = 0$; il valore 0 potrà aversi o solo per $X \equiv 0$, o anche per altri X di LE, che diremo insieme alle loro combinazioni lineari «numeri aleatori di LE trascurabili per rispetto a P»; essi formano un sistema lineare T_pE. Il significato riesce particolarmente intuitivo osservando che gli X trascurabili sono quelli che differiscono da 0

(21) Con ciò non voglio senz'altro pronunciarmi circa l'opportunità di adottare sistematicamente la notazione P(X) per $M(X)$; per le presenti considerazioni astratte è certo che l' uso di due simboli distinti è scomodo e non giustificato da ragioni di praticità altrove sensibili.

in casi trascurabili (casi E per cui $|E|$ è trascurabile), i quali hanno probabilità nulla; ciò è vero infatti per un $X \geq 0$ con $P(X) = 0$ (perchè anche $P(kX) = kP(X) = 0$, e se k è abbastanza grande tutti i valori possibili di kX, che sono in numero finito, divengono > 1, sicchè $kX > |X \neq 0| =$ indicatore dell'evento consistente nell'essere $X \neq 0$, ossia numero che vale 1 se $X \neq 0$ e 0 se $X = 0$, e dunque $P(X \neq 0) = 0$), e quindi lo è per una combinazione lineare di tali X (che è non nulla, al più, nella somma dei casi ove è non nullo uno degli addendi), e quindi ogni X compreso fra due tali combinazioni X' e X'' (dato che, salvo casi di probabilità nulla, è $X' = X'' = 0$, e quindi $X' \leq X \leq X''$ significa $X = 0$).

Se $\mathcal{E}$ è classe finita vale senz' altro anche la reciproca: un evento a probabilità nulla è trascurabile (infatti: se $P(E) = 0$ vuol dire che, qualunque sia $\varepsilon > 0$, esiste un $X \geq |E|$ con $P(X) < \varepsilon$, ma se $\mathcal{E}$ è finita X e quindi P(X) sono funzioni (continue) di $k_0, k_1, \ldots, k_n$ in numero finito, e nel campo chiuso delle k_i soddisfacenti la $X \geq |E|$ esiste, per il teorema di Weierstrass, una combinazione X che realizza il minimo, cioè $P(X) = 0$).

Il medesimo ragionamento dimostra, più in generale, che se P(E) (od anche P(X)) è univocamente determinato, cioè E appartiene a $L_P\mathcal{E}$ ($\mathcal{E}$ finita) senza appartenere ad $L\mathcal{E}$ esistono, in $L\mathcal{E}$, X' e X'' tali che $X' \leq |E| \leq X''$ (oppure $X' \leq X \leq X''$) con $P(X'' - X') = 0$, ossia esiste $X'' - X' \geq 0$ (non $\equiv 0$) appartenente a $T_P\mathcal{E}$ (ed è quindi $X' = X'' = |E|$ (oppure $= X$) salvo casi trascurabili).

Per $\mathcal{E}$ finita si ha dunque la conclusione completa: il prolungamento di una P definita in $\mathcal{E}$ è univoco soltanto su $L\mathcal{E}$ se non esistono in $L\mathcal{E}$ degli $X \geq 0$ (non $\equiv 0$) per cui $P(X) = 0$; se invece ne esistono, il prolungamento è univoco sul sistema $L_P\mathcal{E}$ più ampio, che, oltre gli $L\mathcal{E}$, comprende gli X trascurabili rispetto a P (ossia, oltre gli X di $L\mathcal{E}$ gli $X + X'$ con $P(X' \neq 0) = 0$). Come corollario (osservando che in $\mathcal{E}$ finita si può sempre scegliere P in modo che $P(X) \geq 0$ se $X \geq 0$ (non $\equiv 0$), ad es. attribuendo probabilità $1/s$ a tutti gli s costituenti): se $\mathcal{E}$ è finito, nessun E (od X) fuori di $L\mathcal{E}$ appartiene ad $L_P\mathcal{E}$ qualunque sia P (ossia l'intersezione di tutti gli $L_P\mathcal{E}$ si riduce al solo $L\mathcal{E}$).

Se $\mathcal{E}$ è infinita i problemi si complicano: viene a mancare la «compattezza in sè» (22) di $L\mathcal{E}$ che legittimava l' applicazione del teorema di Weierstrass, e può risultare pertanto P(E) (o P(X)) univocamente definita in quanto, per ogni $\varepsilon > 0$, esistano, in $L\mathcal{E}$, X' e X'' per cui sia $X' \leq |E| \leq X''$ e $P(X'') - P(X') \leq \varepsilon$ senza che ciò valga più per $\varepsilon = 0$. Si può dunque enunciare

(22) M. FRECHET, **Les espaces abstraits**, Gauthier-Villars, Paris, 1928.

analogamente che: P(X) è univocamente determinato se e solo se, per ogni $\varepsilon > 0$, può darsi una scomposizione $X = X' + \Delta$ con X' appartenente a $L\mathcal{E}$ e Δ tale che $P(\Delta) \leq \varepsilon$ ($\Delta = X - X' \leq X'' - X'$, $P(\Delta) \leq P(X'') - P(X') \leq \varepsilon$).

Ma non avviene più sempre che esistano P tali che $L_P\mathcal{E} = L\mathcal{E}$: ciò dipende dalla classe $L\mathcal{E}$. Se ad es. gli eventi $\mathcal{E}$ sono illimitatamente compatibili ciò è vero sempre anche se il loro numero è infinito (potenza qualunque): basta ad es. porre $P(E) = 1/2$ per ogni E di $\mathcal{E}$, e per ogni altro E' la probabilità sarà indeterminata potendo assumere ogni valore tra 0 e $1/2$, o tra $1/2$ e 1, o tra 0 e 1 (estremi inclusi) a seconda che esiste un E per cui $E \supset E'$, o $E' \supset E$, o nè l'uno nè l'altro (in ciascun caso possono anche esistere più E, ma i due casi si escludono perchè da $E_1 \supset E'$ e $E' \supset E_2$ scenderebbe $E_1 \supset E_2$ contro la ipotesi della compatibilità).

Che ciò non sia però vero in generale si vede considerando l' esempio di una classe $\mathcal{E}$ numerabile di eventi E_r ove r percorre i razionali di (0,1) ed $E_r \supset E_s$ per $r > s$; sarà $P(E_r) = F(r)$ con F(x) funzione mai decrescente e quindi continua in $0 \leq x \leq 1$ salvo al più un numero finito o un' infinità numerabile di salti. Se x è un punto di continuità di F(x) l' evento C_x prodotto degli E_r $(r > x)$ e delle negazioni degli E_r $(r < x)$ ha $P(C_x) = 0$, l' evento B_x prodotto degli E_r $(r > x)$ ha $P(B_x) = F(x)$, ecc.; i C_x, i $B_x, \ldots$ appartengono quindi ad $L_P\mathcal{E}$ pur senza appartenere a $L\mathcal{E}$ (che corrisponde alle somme di intervalli a estremi razionali in numero finito); vediamo anzi che mentre $L\mathcal{E}$ è numerabile, $L_P\mathcal{E}$, qualunque sia P, ha almeno (23) la potenza del continuo.

Sembra tuttavia evidente che, anche per $\mathcal{E}$ infinito, soltanto gli $L\mathcal{E}$ appartengano ad $L_P\mathcal{E}$ per P qualunque (nell' es., preso un C_x o B_x qualunque, basta scegliere P che dia una discontinuità in x, e $P(C_x)$ e $P(B_x)$ non risultano più univocamente determinate). Sarebbe interessante dimostrarlo.

11. Interpretazioni geometriche.

Dal punto di vista formale, le considerazioni svolte costituiscono, mi sembra, l' impostazione più generale ed astratta che si possa dare alla teoria delle funzioni d' insieme semplicemente additive: dico «funzioni d' insieme» perchè, fissato il corpo totale zermeliano $\mathcal{Z}$ degli eventi che si vogliono considerare, ogni evento corrisponde a un insieme di «casi possibili» $C\mathcal{Z}$, e ciò corrisponde all'immagine e alla nomenclatura più usuali che può esser opportuno richiamare.

(23) Se esiste un intervallo sia pur piccolo ove F(X) rimane costante, la potenza è non $\mathfrak{c}$ (continuo) ma $2^{\mathfrak{c}} = \mathfrak{f}$, perchè ogni E corrispondente a un qualunque insieme di punti di quell' intervallo ha $P(E) = 0$.

Agli eventi corrispondono gli insiemi, alla probabilità una funzione additiva che diremo misura, alla classe E gli insiemi per cui la misura è definita direttamente (come nel caso usuale gli intervalli, di misura = lunghezza), ecc.

Quanto si è stabilito significa, riassumendo: che il problema della misura (semplice additività!) è possibile in ogni campo, ed è possibile dopo aver assegnato ad arbitro la misura in una qualunque classe d' insiemi E (rispettando le (A')): gli insiemi la cui misura risulta allora determinata sono gli LE (nel caso usuale: le somme d' intervalli in numero finito) e, per P assegnata, gli $L_P E$ (nel caso usuale: gli insiemi misurabili secondo Jordan-Peano): la maggior generalità della trattazione risiede però non solo nel considerare un qualunque campo CZ (non lineare o superficiale, nè dotato di proprietà topologiche o altro) e qualunque misura (se si preferisce: qualunque distribuzione di massa, cioè non necessariamente uguale per intervalli uguali) ma soprattutto nel non postulare che E sia un corpo (vengono a cadere molte semplificazioni comode per la trattazione ma dannose per la visuale del problema che viene in parte nascosta).

I numeri aleatori X di LZ corrispondono alle funzioni definite in CZ (come nella «teoria astratta» di Cantelli); se anzichè allo spazio (non topologico) CZ ci si riferisce allo spazio LZ (sistema lineare delle X) si ha un' interpretazione geometrica ancor più espressiva: anzichè una funzione additiva degli insiemi di CZ la P è semplicemente una funzione lineare nel sistema lineare LZ, individuata (posto che $P(1) = 1$) dal piano $P(X) = 0$, e la condizione $P(X) \geq 0$ per $X \geq 0$ significa che il piano $P = 0$ non deve tagliare Z (ossia l' insieme convesso determinato dai Z). Dare P in E significa dare il piano di LE che è sostegno del fascio di piani di LZ corrispondenti ai possibili prolungamenti di P a LZ; se il sostegno non taglia Z, può analogamente scegliersi il piano; se, causa tale restrizione, il sostegno risulta allargato, si ha $L_P E$ più esteso di LE (analogia geometrica: i piani per un punto A che non taglino un poliedro convesso C hanno in comune solo A se è esterno a C o se è un vertice, ma se A è su uno spigolo del poliedro detti piani hanno per sostegno la retta dello spigolo, e se A è su una faccia il piano non può essere che quello della faccia. Ecc. ecc.

Non so se una trattazione così astratta sia stata compiutamente svolta; se così non fosse potrebbe avere interesse intraprendere un tale lavoro.

12. La definizione di probabilità subordinata

Prima di addentrarsi nella questione dell' additività completa, conviene soffermarci su quella concernente la definizione di probabilità subordinata. Ci eravamo chiesti, precisamente (punto 6 del n. 4), se era opportuno definirla, come il Kolmogoroff, in base al teorema delle probabilità composte, dicendo cioè per definizione «probabilità dell'evento E subordinatamente all'evento H» il numero P(E/H) dato da

$$P(E/H) = P(EH)/P(H) \qquad (\text{supposto } P(H) \neq 0).$$

Sussistonó tre obiezioni riguardanti

— l' assenza di significato concettuale,

— la natura del «teorema delle probabilità composte»,

— l' esclusione del caso $P(H) = 0$.

La prima è pregiudiziale, e senza averla discussa le altre non hanno senso. Anche se si ammettesse - prescindendo dalle obiezioni successive - che la relazione scritta dia sempre il valore dovuto, rimane il problema logico di distinguere se essa vada assunta come definizione o come teorema. Beninteso, ogni definizione è arbitraria, e chi volesse attribuire al rapporto P(EH)/P(H) una denominazione qualunque finora priva di significato è padrone di farlo; ma se la denominazione di cui si parla - nel nostro caso la «probabilità subordinata» - ha già un significato intuitivo bisogna poi o ignorarlo o dimostrare che la definizione vi corrisponde. Non possiamo cioè far sì che il teorema delle probabilità totali sia vero per definizione, ma possiamo o definire le probabilità subordinate precisando il significato intrinseco, concettuale, intuitivo, del termine, e allora resta a dimostrare che esse soddisfano a quel teorema, o definire «probabilità subordinata» il numero che rende soddisfatto quel teorema, e allora resta a dimostrare che è lecito interpretarlo e applicarlo come «probabilità subordinata nel senso intuitivo».

Poste così le due alternative, appare senz' altro più limpida e significativa la prima; vediamo come si possa seguirla nelle diverse concezioni.

Nella concezione asintotica, la definizione di probabilità subordinata è spontanea: basta ripetere la definizione della probabilità (come frequenza-limite) limitandosi alle prove in cui è sodisfatta l' ipotesi, ossia P(E/H) = limite del rapporto tra frequenza di (EH) e frequenza di H. Se la frequenza-limite di H esiste ed è non nulla, la definizione è matematicamente equivalente al teorema delle probabilità composte $P(E/H) = P(EH)/P(H)$, ma anche se la frequenza di H non tende a un limite o ha limite nullo può tuttavia esistere P(E/H) (esempio banale: è sempre $P(H/H) = 1$).

Nella concezione empirica è anche qui difficile concretare qualcosa; cosi pure nella concezione classica, dove però, accogliendo le esemplificazioni in cui si parla di casi di probabilità nulla «ugualmente probabili» (p. es. dell' «intero scelto a caso», o «punto (di un segmento) scelto a caso», ecc.) appare spontaneo che vi s' intenda lecito parlare di probabilità subordinata anche per eventi di probabilità nulla (p. es., dire che, scegliendo a caso un punto di un segmento, la probabilità che esso sia A subordinatamente all' ipotesi che sia o A o B, vale 1/2, essendo A e B due punti qualunque). La legittimità di parlare della probabilità di E subordinatamente ad H anche per $P(H) = 0$ appare poi difficile a negarsi almeno immaginando il caso in cui si definisca E facendolo consistere, nel caso che H si verifichi, nell' avverarsi di una prova di probabilità nota.

Nella concezione soggettiva, la probabilità subordinata si definisce direttamente come la probabilità; se ad es. si presceglie la definizione basata sulle quote di scommessa, $P(E/H)$ è la quota per una scommessa sull' evento E che è *nulla* se non si verifica l' ipotesi H. Il teorema delle probabilità composte vi appare come una restrizione necessaria per la coerenza; definizione e teorema sussistono indipendentemente dalla condizione $P(H) = 0$; se essa vale, c' è di più solo la possibilità di ricavare $P(E/H)$ in funzione di $P(EH)$ e $P(H)$.

Concludendo: secondo tutte le concezioni, in quanto consentano di dire qualcosa, risultano non giustificate l' adozione del teorema delle probabilità composte come definizione della probabilità subordinata, e la restrizione $P(H) \neq 0$; secondo la concezione soggettiva si aggiunge anche l'ultima delle tre obiezioni, perchè il teorema delle probabilità composte non viene ivi a significare che $P(E/H)$ coincide necessariamente con $P(EH)/P(H)$ ma solo che chi non seguisse tale regola di valutazione sarebbe incoerente.

13. Interpretazione analitica

Dal punto di vista formale, la sola obiezione che porta modificazioni è quella concernente il caso $P(H) = 0$; formalmente, le sue conseguenze si possono interpretare concependo la probabilità come una grandezza non archimedea, espressa da un numero reale quando è non nulla, ma da doversi concepire come un «infinitesimo attuale» per gli eventi «a probabilità nulla» ma non impossibili (24). A prescindere da tale terminologia, che può soddisfare più o meno, si tratta di stabilire che è sempre possibile considerare il rapporto fra le probabilità di due eventi, E_1 e E_2, anche se una o entrambe sono nulle (rapporto che è un numero reale positivo, o 0, o ∞); basta infatti definire $P(E_1)/P(E_2) = P(E_1/H)/P(E_2/H)$ ove $H = E_1 + E_2$ (allora infatti $P(E_1/H) + P(E_2/H) \geq 1$, sicchè il secondo membro non può avere la forma 0/0, ma a/b = positivo o al più $0/b = 0$ o $a/0 = \infty$).

Dal punto di vista dell' impostazione assiomatica, volendo introdurre anche le probabilità subordinate occorrerebbero naturalmente nuovi assiomi o formulazioni più ampie delle precedenti.

(24) Ne ho trattato ampiamente in **Les probabilités nulles**, "Bull. Sci. Math.", 1926.

Per esempio, attenendosi all' esposizione del Kolmogoroff:

— anzichè solo $P(E)$ (Assioma III) postulare che sia definito $P(E/H)$ per le coppie di eventi E, H (in particolare $P(E) = P(E/C)$ per C = evento certo) (25);

— anzichè $P(C) = 1$ (Assioma IV) postulare che $P(H/H) = 1$ (in particolare $P(C/C) = P(C) = 1$);

— aggiungere (nuovo assioma) il teorema delle probabilità composte nella forma

$$P(H/K)P(E/HK) = P(EH/K);$$

per K = «evento certo» si ha la forma usuale $P(H)P(E/H) = P(EH)$, non sufficiente allo scopo (26).

14. La questione fondamentale

Avviandoci all' esame della questione dell' additività completa, conviene ora far il punto sulla visione del problema quale ci appare dopo le considerazioni fin qui svolte.

La questione consiste nell' alternativa di considerare sufficienti gli assiomi finora posti, o di aggiungere ancora quello dell' additività completa, e occorre anzitutto essere in chiaro sul significato che ha l' affermare che questo o quel gruppo di assiomi è sufficiente.

Affermare che un gruppo di assiomi è sufficiente dovrebbe significare che non solamente la *probabilità* $P(E)$ *deve soddisfare quegli assiomi*, ma inversamente *ogni funzione* $P(E)$ *che li soddisfa può essere assunta come probabilità nel campo* E *considerato*.

Tale significato vale senz' altro nella teoria soggettiva (almeno nella accezione dello scrivente), in quanto l' ufficio della logica e quindi della matematica si limita al compito di distinguere le funzioni che possono o non possono essere assunte a rappresentare le probabilità senza dar luogo a incoerenza, mentre la scelta di quella qualsiasi, tra le infinite opinioni coerenti, che egli preferisce, o meglio che *sente*, è affare personale di ogni soggetto (27).

(25) Per la definizione di E/H è infatti $E/C = E$ (se C = certo).

(26) Mostriamo la ragione della non-sufficienza su un esempio. Siano E_1, E_2, E_3 incompatibili ugualmente probabili e di probabilità nulla; posto $H = E_1 + E_2$, $K = E_1 + E_2 + E_3$, sarà $P(E_1/H) = P(E_2/H) = 1/2$, $P(E_1/K) = P(E_2/K) = 1/3$, $P(H/K) = 2/3$; in base alla sola formulazione usuale potrei ad es. valutare ad arbitrio $P(E_1/K)$ e $P(E_2/K)$ (anche diversi tra loro) pur tenendo ferme le valutazioni restanti.

(27) Non s' intende con ciò misconoscere o negare i motivi ed elementi di giudizio che in molti campi e casi conducono a una generale più o meno stretta concordanza di opinioni; tali motivi e il meccanismo psicologico del loro influsso vengono anzi studiati e analizzati, appunto in quanto si considerano motivi (di cui ciascuno può tenere il conto che crede) e non dogmi o definizioni o teoremi o leggi.

Secondo lo «schema delle scommesse» (di cui al n, 9), i numeri aleatori di $L\mathscr{E}$ equi rispetto a P sono le scommesse possibili in base alla valutazione di probabilità P; l'assioma (A''') dice allora che sarebbe un' incoerenza ritenere equa una scommessa se *in ogni caso* dà perdita. Dire che altri assiomi non occorrono significa ritenere che tutte le opinioni che non danno luogo a tale incoerenza sono ammissibili.

Diversa è la situazione per le teorie che considerano il valore della probabilità come univocamente determinato da un significato logico o statistico o fisico o comunque obbiettivo; da tale punto di vista una costruzione assiomatica non può che apparire insufficiente in quanto non può definire *una* funzione P, l' unica vera autentica probabilità, ma la classe delle funzioni che condividono con essa certe proprietà formali.

In qualche modo è tuttavia possibile anche nelle concezioni empirica e asintonica dare un senso all' insieme di tutte le funzioni P formalmente ammissibili. Anzitutto si può immaginare che le frequenze o frequenze-limiti siano diverse da quelle osservate, e dire P ammissibile o no a seconda che si può o no fare in modo di ottenerla; al solito, ciò ha un senso più preciso nella concezione asintotica ove P si può dire ammissibile se si riesce a costruire un «Kollektiv» in modo che la frequenza di ogni E di $\mathscr{E}$ abbia per limite P (E).

Altro senso può essere il seguente (per qualunque concezione). L' opinione sugli $\mathscr{E}$ espressa dalla funzione P (E) si modifica nella P (E/H) considerando il giudizio subordinato all' ipotesi che si verifichi un evento H. Si potrà dire che le funzioni P ammissibili sono quelle che, data la P «effettiva», ne discendono come probabilità subordinate alle diverse ipotesi H possibili ?

Se $\mathscr{E}$ è un corpo finito, e P assegna probabilità non nulla ad ogni costituente $C\mathscr{E}$, ciò è possibile senz' altro: dato che $P(E/H) = P(E)\,P(H/E)/ /P(H)$, basta costruire l' evento H in modo che per ogni costituente C risulti P (H/C) proporzionale a $P_1(C)/P(C)$ (28), e P (E/H) risulterà uguale a quella qualunque funzione additiva $P_1(E)$ che si desideri.

15. L' additività completa.

Mentre sugli altri punti si avevano solo divergenze metodologiche (opportunità di precisare in un modo o nell' altro qualche dettaglio) o di concezione (che non toccano gli assiomi ma il significato della teoria), s'incontra

(28) Basta p. es. far consistere H nel fatto che si verifichi uno dei costituenti per i quali risulta $P_1(C)/P(C) > \xi$, essendo ξ un numero il cui valore ci è tenuto segreto, ma che si sa essere stato "scelto a caso" (cioè: con densità di probabilità costante) nell' intervallo $(0, k)$ $(k = \max P_1(C)/P(C))$.

ora il solo punto cruciale su cui il disaccordo concerne l' inclusione o meno di un nuovo assioma: precisamente dell' «assioma VI» di Kolmogoroff, equivalente, come si vedrà tosto, all' additività completa che anzitutto definiremo. L' additività semplice dice che se $E_1\,E_2 \ldots E_n$ sono eventi incompatibili in numero finito, ed E è la loro somma, la probabilità p di E è la somma delle probabilità p_h degli E_h:

(A) $\quad p = p_1 + p_2 + \ldots + p_n$;

ne segue senz' altro che se gli E_h costituiscono un' infinità numerabile è

(B) $\quad p \geq p_1 + p_2 + \ldots + p_n + \ldots$

perchè, qualunque sia m, è sempre $p = p_1 + p_2 + \ldots + p_m + p_m^*$ ove con p_m^* si indichi la probabilità (necessariamente non negativa) dell' evento E_m^*, somma logica di $E_{m+1}, E_{m+2}, \ldots, E_{m+n}, \ldots$

Si dice che sussiste l'*additività completa* quando nella (B) vale il segno di uguaglianza, ed è cioè

(C) $\quad p = p_1 + p_2 + \ldots + p_n \ldots ,$

e la questione controversa è la seguente:

deve assumersi come necessariamente e sempre valida la (C) (teorema esteso delle probabilità totali, o proprietà della additività completa), oppure in generale può affermarsi soltanto che sussiste la (B) cosicchè la (C) costituisce una proprietà o restrizione particolare che può valere o no?

Rileviamo subito che la proprietà equivale al «postulato di continuità» (Assioma VI di K.), perchè vale se e solo se per la successione E_m^* (soddisfacente le condizioni del detto assioma) le probabilità $P(E_m^*)$ tendono a zero (la ragione della denominazione di «assioma di continuità» apparirà meglio quando, nel n. 24, riprenderemo più di proposito l' argomento).

16. Osservazioni

Nella discussione di tale questione s' intrecciano e sovrappongono considerazioni di natura tanto svariata, che riesce difficile scegliere una via di esposizione atta a rendere chiara la dipendenza fra i diversi argomenti e il nesso che li coordina — e in senso diverso a seconda dei diversi punti di vista — alla presa di posizione rispetto all' additività completa.

Per agevolare l' orientamento indichiamo fin d' ora succintamente l' itinerario prescelto. Ci occuperemo dapprima del caso più semplice, quello ove si abbia solo un' infinità numerabile di «casi elementari»: tutte le questioni acquistano allora la massima evidenza. Il vantaggio di acquisire in tal modo una prima visione intuitiva e di rispondere subito col minimo di mezzi a qualche dubbio pregiudiziale compensa lo svantaggio di dover poi riprendere vari argomenti, completarne l' analisi, correggere certe impressioni, integrarle con il risultato di nuove constatazioni.

Solo in questa seconda fase, tornando a impostare la trattazione secondo la linea della generalità finora seguita (classe di eventi qualunque), molti aspetti appariranno in luce o riappariranno in piena luce, ma si richiedono troppe considerazioni preliminari indipendentemente dalle quali possiamo subito affrontare il problema nel caso più semplice.

17. Esempi

Per la prima analisi, intesa a fissare anzitutto la visione del problema nelle sue grandi linee, sarà utile indicare e tener presenti due esempi per veder subito cosa implichi e significhi il decidersi per l' una o l' altra alternativa.

I «casi elementari» C corrispondono, in entrambi gli esempi, ai numeri interi positivi (più, nel secondo, il valore «$+\infty$»).

Il primo esempio è quello notissimo del cosiddetto «intero scelto a caso», cioè di un numero aleatorio X avente probabilità nulla di assumere un qualunque determinato dei valori possibili $X = 1, 2, 3, \ldots, n, \ldots$, ma ad es. probabilità $1/2$ di essere dispari, $1/10$ di terminare per 7, probabilità nulla di essere primo, o quadrato, o privo di cifre «7», ... e in generale avente probabilità $\lim \frac{1}{n} a_n$ di appartenere a una classe di numeri A cui appartengono a_n numeri sui primi n (e sia quindi $\frac{1}{n} a_n$ la probabilità di avere un A scegliendo «a caso» (cioè con uguale probabilità) un numero fra i primi n). Naturalmente bisogna limitarsi alle classi A per cui il limite esiste.

In questo esempio l'a d d i t i v i t à c o m p l e t a n o n v a l e, perchè i singoli casi elementari hanno probabilità nulla, $p_1 = p_2 = \ldots = p_n = \ldots = 0$, e la loro somma è ancora zero, mentre la probabilità della somma di tutti i casi possibili, ossia dell' evento certo, è uno. Altrettanto si può dire, per fare una constatazione meno banale, per le probabilità che X sia quadrato o dia residuo $1, 2, \ldots, n, \ldots$ detraendo il massimo quadrato contenutovi.

Ma in altri casi ci si basa invece su questa osservazione, che $p_1 = p_2 = \ldots = p_n = \ldots = 0$, per concludere, b a s a n d o s i s u l l' a d d i t i v i t à c o m p l e t a, in modo del tutto opposto. Facciamo un esempio semplice, pur notando che sostanzialmente sullo stesso presupposto e secondo lo stesso schema si sviluppano dimostrazioni fondamentali per certe trattazioni, come probabilità di convergenza (29), probabilità per i numeri aleatori nei punti di discontinuità della funzione di ripartizione (30), ecc.).

Consideriamo una successione di prove a testa e croce, e diciamo X il numero così definito: $X = \infty$ se si verificano infinite volte entrambi i risultati, $X = n$ se a cominciare dall' n-esima prova si verifica sempre il medesimo risultato. Per ogni n è ovviamente $p_n \leq 1/2^h$ qualunque sia h (perchè $1/2^h$ è la probabilità che le prime h prove dopo l' n-esima sodisfino la condizione di dare il suo stesso risultato) e quindi $p_n = 0$. Se ne conclude, postulando l' additività completa, che è nulla la probabilità che X sia finito, ossia che la successione finisca per dare definitivamente sempre testa o sempre croce.

E' innegabile che tale conclusione sembra ovvia, ma è facile vedere che la sua deduzione non solo non risponde a necessità ma anzi è senz' altro falsa se non si esclude la legittimità di considerare il precedente esempio di intero X scelto a caso. Se infatti supponiamo, per materializzare l' esempio, di sapere che ad un certo momento la moneta verrà sostituita con una avente «testa» su entrambe le faccie, ma ciò avverrà dopo un numero X di colpi che sarà determinato «a caso», tutte le probabilità relative a un numero finito n qualunque di prove rimarranno inalterate (chè l'eventualità $X \leq n$, che la sostituzione della moneta vi abbia già avuto luogo, ha probabilità nulla, e non influisce quindi sul risultato) e quindi è ancora $p_1 = p_2 = \ldots = p_n = \ldots = 0$ e tuttavia è $p = 1$ la probabilità che X sia finito.

18. Considerazioni preliminari

Per molto tempo si sono considerate promiscuamente applicazioni dei due opposti concetti, come in tali esempi, senza tuttavia giungere a conflitti o avvertirli, finchè si riferivano a trattazioni staccate di argomenti distinti. Ma al momento di costruire il calcolo delle probabilità come una teoria dai principi ben determinati si presentò ineluttabile il dilemma: o r i f i u t a r e p r o b l e m i s u l t i p o d i q u e l l o d e l l'«i n t e r o s c e l t o a c a s o» (e a l l o r a t r o v a r e m o t i v i p e r n e g a r n e l a l e g i t t i m i t à), o r i n u n c i a r e a c o n c l u s i o n i b a s a t e s u l l' a d d i t i v i t à c o m p l e t a (e a l l o r a a c c e t t a r e qual-

(29) La probabilità cioè che converga una serie $X_1 + X_2 + \ldots + X_n + \ldots$ con X_h numeri aleatori (Borel, Cantelli).

(30) Se per la funzione di ripartizione $F(x)$ di un numero aleatorio X si ha per $x = x_0$ una discontinuità, $F(x_0 + 0) - F(x_0 - 0) = p > 0$, possiamo dire che si ha probabilità p che risulti $X = x_0$? La conclusione sussiste si si ammette l' additività completa, ma altrimenti si può solo dire che tale probabilità è $\leq p$ (eventualmente anche nulla: basta pensare al caso in cui si considerino possibili ed equiprobabili i soli valori $x_0 + 1/n$).

che complicazione maggiore). E' questa, vista nelle sue conseguenze, l'alternativa in oggetto.

Avvertito il dilemma, si manifestò da parte della maggioranza degli Autori una tendenza ad accogliere l' additività completa, senza vagliare eccessivamente le ragioni che potevano esservi pro o contro. Ad es. il Cramèr, nella sua trattazione pur così acuta e rigorosa nello svolgimento, a tale premessa non dedica altra giustificazione che la frase seguente: "It is natural to require that this relation (il teorema delle probabilità totali) should hold even as $n \to \infty$" (31). La ragione più intima di tale tendenza sta certamente nel desiderio di modellare il calcolo delle probabilità sulla falsariga della teoria della misura per potervi trasportare i risultati ivi stabiliti e anzitutto l' integrale di Lebesgue. Altra ragione: la riluttanza ad ammettere che un' infinità numerabile di casi di probabilità nulla possa dare come somma un caso di probabilità non nulla, o in particolare l' evento certo di probabilità uno

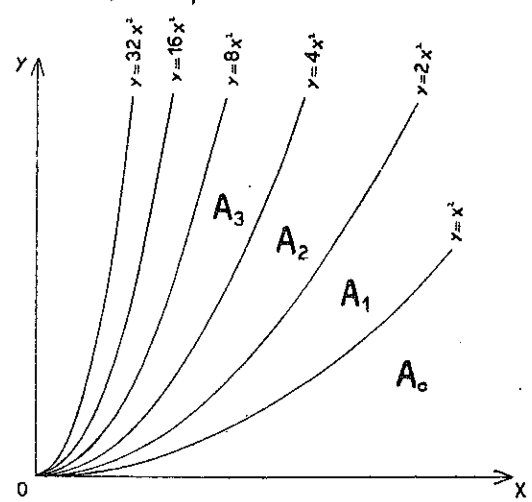

Figura 1.

(come nell'esempio dell' «intero scelto a caso» (n. 17); sembra addirittura che taluno cada nell' abbaglio di considerare equivalenti le due frasi «qualunque sia n, la probabilità che sia $X > n$ vale 1» e «la probabilità che sia $X > n$ per qualunque n (ossia: che $X = \infty$) vale 1», o cerchi comunque di avvalorare a parole l' additività completa come se il pensiero fosse influenzato da un' impressione del genere. Eppure esempi come come quello della fig. 1 (regioni $A_0, A_1, \ldots, A_n, \ldots$ formati in O un angolo nullo, e la cui somma è il quadrante formante ivi un angolo retto mostrano che anche per

(31) H. CRAMER, Random variables and probability distributions, Cambridge Univ. Press, 1937.

grandezze geometriche può verificarsi senza inconvenienti quello che per la probabilità si vorrebbe escludere come paradosso (per certuni tuttavia l'obiezione sembra solo di parole: superabile cioè sol che si dica «probabilità infinitesime» anzichè «nulle»: cfr. p. es. op. cit. (40)).

Vi sono poi molti tentativi di giustificare effettivamente l' accoglimento dell' additività completa, consistenti nel respingere, mediante motivazioni di vario genere, questa o quella delle conseguenze della sua negazione, così come ve ne sono in senso opposto. Sarebbe impossibile e fuori luogo enumerare fin d' ora tali argomenti, che incontreremo e discuteremo proseguendo.

Vogliamo però osservare pregiudizialmente che ai motivi consistenti nel rilevare questo o quel vantaggio o inconveniente dell' adozione di questa o quella alternativa non dovrebbesi attribuire un valore determinante, a meno che non si voglia considerare la teoria delle probabilità come una costruzione analitica arbitraria da istituire liberamente nel modo che riesce più comodo o elegante. Motivi del genere andranno piuttosto giudicati in quanto possano avere un nesso con questioni concettuali attinenti alla probabilità.

La via maestra però dovrebbe consistere nell' esaminare, fin dove le premesse sono abbastanza precise per consentirlo, quale delle due alternative discende dalla definizione di probabilità di ogni data concezione, o risulti in armonia col suo spirito informatore.

19. L' additività completa e la concezione asintotica

Cominciamo dalla concezione asintotica, che consente la risposta più immediata e categorica: l' additività completa **non** sussiste.

Se i casi possibili ad ogni prova sono $A_1, A_2, \ldots, A_m, \ldots$ e le loro frequenze dopo n prove sono

$$\frac{1}{n} a_n^{(1)} + \frac{1}{n} a_n^{(2)} + \frac{1}{n} a_n^{(3)} + \ldots + \frac{1}{n} a_n^{(m)} + \ldots = 1$$

(somma di un numero finito, $\leq n$, di termini non nulli), non segue, analiticamente, che per le frequenze-limiti sia

$$a^{(1)} + a^{(2)} + a^{(3)} + \ldots + a^{(m)} + \ldots = 1 ;$$

potrebbero benissimo ad es. esser tutte nulle:

$$\frac{1}{n} a_n^{(m)} \to a^{(m)} = 0,$$

come ad es. se ogni A_m si verificasse una sola volta nel «Kollektiv» (o un numero finito di volte). Poichè tale esempio di «Kollektiv» sembrerà forse un caso limite inaccettabile (difficile prevedere ogni punto di vista !), ne costruiamo uno diverso: partiamo da un «Kollektiv» di successive estrazioni di cifre (0, 1, 2, . . . , 9) e dividiamolo in gruppi successivi di tante cifre quante ne conta $1!, 2!, 3!, \ldots, n!, \ldots$; il gruppo n-esimo formerà un numero m, e in base ad esso scegliamo la m-esina (in ordine crescente) fra le n! permutazioni dei numeri $1, 2, 3, \ldots, n$ (se $m > n!$, prendendo $m-n!$), e formiamo un nuovo Kollektiv di numeri disponendo una di seguito all'altra tutte queste permutazioni. Ogni numero vi figura infinite volte, ma sui primi n posti non più di $\sqrt{2n}$ volte e quindi con frequenza-limite nulla.

E' curioso osservare che Reichenbach (32), che si basa sulla concezione di von Mises, e Kolmogoroff, che dichiara di aderirvi, sostengono tuttavia l'additività completa (quanto al von Mises stesso, non so se se ne sia occupato (33)). Ciò mi dà la stessa impressione che mi ha fatto lo scarso peso dato da von Mises ad osservazioni del Marbe (34): che cioè la definizione mediante il «Kollektiv» sia in fondo considerata dai suoi stessi fautori come una facciata per illudersi d'aver dato alla probabilità una definizione rigorosa e fondata su base sperimentale, ma non come un effettivo fondamento cui confidano potersi riferire nel discutere problemi precisi. Così come il von Mises prescinde dall'asserito fondamento sperimentale della propria costruzione nel rispondere a Marbe, così i suoi seguaci prescindono dalla definizione stessa nel prendere posizione sull'additività completa.

20. L'additività completa e la concezione soggettiva

Nella concezione empirica è difficile impostare questioni del genere dato il significato sfuggente della definizione; i suoi fautori sono in genere favorevoli all'additività completa per motivi pure alquanto sfuggenti.

Nella concezione classica l'esempio dell'«intero scelto a caso» appare come spontanea estensione dei «casi ugualmente probabili», ma sarebbe eccessivo farne un'affermazione precisa.

Nella concezione soggettiva si tratta di vedere se, nell'attribuire delle probabilità $p_1, p_2, \ldots, p_n, \ldots$ a una classe completa numerabile di eventi incompatibili, sia necessario per la c o e r e n z a che la somma dei p_n dia 1 o possa invece anche essere < 1. Anche qui la questione non è posta con ciò in forma del tutto univoca, poichè la definizione stessa di coerenza potrebbe venire ritoccata, ma è tuttavia ben chiarito il senso in cui la questione deve essere concepita.

(32) Cfr. op. cit. (16).

(33) Se interpreto bene le considerazioni in merito al principio del suo trattato, v. cit. (3), pare egli riconosca che l'additività completa non è una proprietà necessaria delle distribuzioni di probabilità ma solo una particolarità che può valere o meno.

(34) K. MARBE, **Das Ausgleichsprinzip in der Statistik**, Beck, München, 1938; per la sua discussione col von Mises, cfr. **Punti di vista: Karl Marbe** (cit. (16)).

Ecco alcune considerazioni al riguardo, il cui significato e peso non sono del resto strettamente legati e condizionati a quella particolare concezione.

Un motivo che tenderebbe ad avvalorare l'additività completa: se le probabilità p_n hanno somma $p < 1$, stipulando tutte le infinite scommesse posso ricevere in ogni caso la somma 1 pagando complessivamente la somma p, e quindi avrei un'incongruenza. Ma si tratta in realtà di un circolo vizioso, perchè solo se sapessi valida l'additività completa potrei pensare di estendere la nozione di «combinazione di scommesse equa» a combinazioni di i n f i n i t e scommesse, e di basarle sulla s e r i e delle quote di scommessa.

Altro argomento: anche limitandosi a combinazioni finite, se i casi considerati hanno probabilità nulla, risulta equa una scommessa in cui la posta viene pagata contro una puntata nulla, e in cui quindi il guadagno è sempre ≥ 0. Ma con tale argomentazione si verrebbe a rigettare in genere la possibilità che un evento non impossibile abbia probabilità nulla, e ciò invece è certamente inevitabile (basti pensare che se si considera un'infinità non numerabile di casi elementari essi hanno necessariamente probabilità nulla salvo al più un'infinità numerabile tra essi (35)). Non si può quindi pensare di rafforzare la definizione di coerenza pretendendo che il guadagno debba non essere ≥ 0 (anzichè > 0), perchè tale pretesa sarebbe inadempibile.

Un motivo contro l'additività completa: se $p_1 + p_2 + \ldots + p_n + \ldots = 1$, vuol dire che, qualunque sia $\varepsilon > 0$, esiste un numero f i n i t o di casi elementari la cui probabilità complessiva $p_{h_1} + p_{h_2} + \ldots + p_{h_n}$ supera $1 - \varepsilon$. Una valutazione delle probabilità completamente additiva nel campo C rappresenta quindi un'opinione e s t r e m a m e n t e a s i m m e t r i c a della probabilità dei diversi casi; se il soggetto non trova ragione per una valutazione così asimmetrica, o se magari non vede motivo di alcuna preferenza tra caso e caso, come potremmo includere nella definizione della coerenza (di una mera formale coerenza) una condizione che non gli permetterà di valutare tutte uguali (e quindi nulle) le probabilità p_n, ma violenti le sue facoltà di giudizio obbligandolo addirittura ad attribuire praticamente tutta la probabilità a un gruppo finito di eventi quale che sia, magari scegliendoli a vanvera? Simile imposizione riguardante il merito della valutazione riesce del tutto estranea al significato della condizione relativa alla c o e r e n z a.

Altri motivi. Contro l'additività completa, il fatto che, limitandosi alle distribuzioni di probabilità che la sodisfano, il limite di una distribuzione di

(35) Infatti n al più hanno probabilità $\geq 1/n$, e quindi un'infinità numerabile al più ha probabilità > 0.

probabilità variabile non sarebbe più una distribuzione di probabilità (v.n. 22). A favore, il dubbio, che potrebbe sorgere, circa l'impossibilità di estendere la distribuzione di probabilità dell'«intero scelto a caso» (o simili), senza intrinseca contraddizione, ad ogni insieme di numeri (per cui non esista il limite di $\frac{1}{n}a_n$): ma tale dubbio è per noi già eliminato dalla dimostrazione data nel n. 9, secondo la quale il prolungamento rispettando l'additività semplice è sempre possibile. L'argomento si ritorcerà anzi contro l'additività completa quando vedremo (nel prossimo n. 21) che essa invece rende generalmente impossibile il prolungamento.

21. L'additività completa e l'additività perfetta

Veniamo ora all'esame particolareggiato delle conseguenze dell'additività completa.

Anzitutto rileviamo che essa in generale non è neppure possibile; la possibilità di definire una P completamente additiva sussiste soltanto in particolari classi d'eventi.

E' noto infatti che il Vitali ha dimostrato (36) l'impossibilità di estendere la definizione di «misura» a tutti gli insiemi; tale risultato è stato esteso a casi più generali e sotto ipotesi meno restrittive da Banach, Kuratowski e Ulam (37), e possiamo così enunciarlo con riferimento al nostro problema (rinunciando a precisare una restrizione che va posta relativamente alla potenza dell'insieme dei casi possibili):

una P(E) completamente additiva non esiste quando E sia un corpo zermeliano infinito (ossia comprenda tutti gli eventi formati partendo da un'infinità di «casi possibili»), a meno che la probabilità non sia concentrata su un numero finito o un'infinità numerabile di tali casi cui competano probabilità positive di somma uguale ad 1.

In quest'ultimo caso diremo che la P(E) è perfettamente additiva.

Le tre ipotesi dell'additività semplice, completa, perfetta, si possono caratterizzare succintamente riferendosi alla possibilità di ottenere una probabilità finita per raggruppamento di casi a probabilità nulla (o infinitesima, se si preferisce tale locuzione): un tale fatto è possibile già per infinità numerabili nell'ipotesi dell'additività semplice, lo è solo per infinità non numerabili nell'ipotesi dell'additività completa, non lo è mai nell'ipotesi di additività perfetta.

(36) G. VITALI, **Sul problema della misura dei gruppi di punti di una retta**, Bologna, 1905.

(37) "Fundamenta mathematicae", T. XIV, T. XV, T. XVI, Warszawa, 1928-30.

Prospettando in tal modo le tre varianti, mi sembra appaiano senz'altro più persuasive le due estreme, perchè una diversità di comportamento è più naturale abbia a presentarsi se mai nel passaggio dal finito all'infinito che dall'infinito numerabile al non numerabile.

22. La chiusura e le diverse concezioni

Un'altra difficoltà derivante dall'accoglimento dell'additività completa consiste nel fatto che, a differenza delle distribuzioni semplicemente additive, quelle completamente additive (su una certa classe di eventi E) non costituiscono un insieme chiuso (e altrettanto avviene per quelle perfettamente additive). Ossia: se $P_n(E)$ è una distribuzione variabile con n, definita su E, ed esiste, almeno per gli E di una certa sottoclasse E', il limite $P(E) = \lim P_n(E)$, P (E) è semplicemente additiva se lo è ogni $P_n(E)$, ma può non essere completamente additiva anche se tutte le $P_n(E)$ sono completamente o magari perfettamente additive (in E' o anche in E). In altre parole: il limite di una distribuzione semplicemente additiva è una distribuzione semplicemente additiva, ma il limite di una distribuzione completamente o perfettamente additiva può essere una distribuzione solo semplicemente additiva, perchè delle probabilità concentrate per tutte le $P_n(E)$ possono trasformarsi in probabilità diffuse nel passaggio al limite.

Anzi, il metodo più facile per creare esempi di distribuzioni non completamente additive consiste appunto nel partire da distribuzioni completamente additive ed eseguire un passaggio al limite. Il caso più noto è quello della già accennata (n. 17) distribuzione di probabilità definita nel campo dei numeri interi come limite delle distribuzioni $P_n(E)$, ove $P_n(E)$ sia la distribuzione che attribuisce probabilità $1/n$ a ciascuno dei primi n numeri.

Esempi analoghi si possono costruire nel caso di un punto aleatorio di una retta o piano dove possiamo attribuire con analogo passaggio al limite la probabilità p ad un insieme A qualunque per cui il limite esista ponendo $p = \lim p_n$, $p_n = \frac{1}{n} \times$ misura di A_n, A_n = parte di A compresa nell'intervallo di lunghezza n (per il caso della retta, risp. nel cerchio di area n per il caso del piano) e centro nell'origine. Si può però anche evitare (come nel caso dell'angolo, v. fig. 1), ogni passaggio al limite; come esempio definiamo una distribuzione di probabilità nel campo aperto interno ad una circonferenza o contorno qualunque (la cui lunghezza assumeremo unitaria) assegnando come probabilità di un qualunque insieme A appartenente a detto campo la lunghezza della parte del contorno toccata da A (esattamente: appartenente all'insieme derivato A'). La probabilità risulta determinata per A in modo accettabile purchè nessun tratto del contorno appartenga contemporaneamente al derivato di A e del suo complementare: se ciò avvenisse la somma delle due probabilità opposte sarebbe > 1: ciò escludendosi, eventi incompatibili non hanno mai sovrapposizioni delle rispettive parti di contorno e sussiste l' additività semplice. Non però l' additività completa, perchè l' intero campo, ossia l' evento certo di probabilità 1, è la somma di infinite corone circolari (basta suddividerlo con circonferenze di raggio sempre più grande e tendente a quello dell' intero campo), e ciascuna di esse, non toccando il contorno, ha probabilità nulla. Per rendersi ragione intuitivamente della distribuzione potremo dire sinteticamente che essa è tutta addensata in aderenza al contorno (che però non appartiene al campo: quindi non su di esso); volendo, si può anche interpretare tale esempio come una immagine al finito della precedente distribuzione sul piano, ottenuta rappresentando il piano

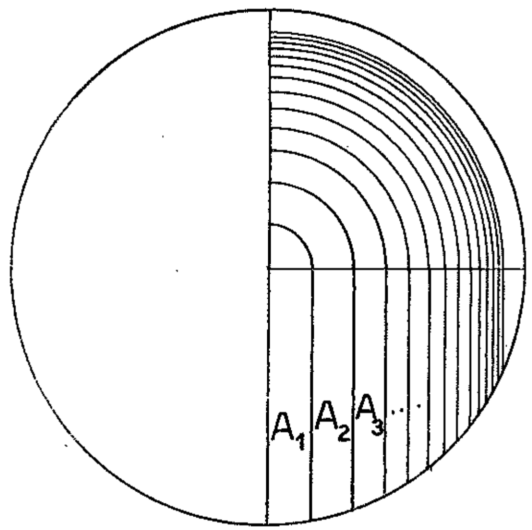

Figura 2.

sull' interno di una circonferenza ad es. mediante la trasformazione di coordinate polari $\rho' = 1 - e^{-\rho}$, $\theta' = \theta$: ogni area interna al contorno (che nell' immagine corrisponde a un' area finita del piano) ha probabilità nulla, mentre ogni strisciolina per quanto sottile aderente al contorno (o una parte di esso di lunghezza λ) ha probabilità $p = 1$ (risp. $p = \lambda$).

Sugli esempi considerati, e particolarmente sull' ultimo, è istruttivo persuadersi come nella somma di infiniti eventi incompatibili possa prodursi una probabilità totale superiore alla somma della serie; considerando una qualunque area A suddivisa in parti $A_1, A_2, \ldots, A_n, \ldots$ (come p. es. il semicerchio A a destra nella fig. 2, di cui si deve immaginare proseguita indefinitamente la suddivisione in figure A_n costituite da un quarto di corona circolare nel quadrante superiore e dalla striscia verticale che lo prolunga nel quadrante inferiore, in modo da esaurire al limite l' intero semicerchio A) avverrà in generale che una parte del contorno è toccata dalle singole parti $A_1, A_2, \ldots, A_n, \ldots$ e la sua lunghezza è data dalla serie delle lunghezze relative alle singole A_n, ossia dalla serie delle singole probabilità (concentrate), ed un' altra parte del contorno è progressivamente avvicinata ma mai raggiunta, e la sua lunghezza rappresenta il totale delle probabilità «diffuse» fra gli infiniti insiemi A_n, ossia quel limite che il «postulato di continuità» (forma ristretta) vorrebbe sempre nullo: l' area residua di A dopo levate $A_1 \ldots A_n$ svanisce (è vuota al limite per $n \to \infty$), ma, finchè il processo non è spinto al limite, essa occupa sempre quel certo tratto di contorno su cui si schiaccia.

Se, in particolare, tale tratto manca, la distribuzione è completamente additiva (come nell' es. che si ha considerando nella figura il solo quadrante inferiore); se invece manca il contatto delle A con tratti di contorno si ha un evento a probabilità non nulla composto di un' infinità numerabile di eventi a probabilità nulla (id., quadrante superiore). Considerazioni analoghe si possono fare sul caso degli angoli (fig. 1) e sugli altri esempi.

23. Argomenti in favore della chiusura.

Rimane da chiedersi se esistano dei motivi per avvalorare una propensione a ritenere che il limite di una distribuzione di probabilità ammissibile debba ancora costituire una distribuzione di probabilità ammissibile.

Nella concezione asintotica è un fatto di analisi che la chiusura sussiste almeno sotto qualche restrizione. Abbozziamo la dimostrazione per un campo d' eventi $\mathcal{E}$ numerabile $E_1, E_2, \ldots, E_n, \ldots$ Avremo un «Kollektiv» $a^h_{(1)}$ che dà come frequenze limiti $P_1(E_n)$, un secondo $a^{(2)}_h$ che dà $P_2(E_n)$, ecc.; potremo costruire un «Kollektiv», a_h, prendendo opportunamente n_1 termini del primo Kollektiv, poi n_2 del secondo, ..., n_k del k-esimo, ecc. in modo che, se per $k \to \infty$ è $P_k(E_n) \to P(E_n)$ per ogni n, risultino $P(E_n)$ le frequenze-limiti sul nuovo «Kollektiv». Basterà prendere dal k-esimo «Kollektiv» un numero n_k di termini sufficientemente grande e a partire da un indice sufficientemente elevato per assicurare che nel corso degli n_k termini le frequenze degli eventi $E_1, E_2, \ldots, E_k$ si avvicinino sufficientemente a $P_k(E_1)$, $\ldots, P_k(E_k)$, per poter garantire che al limite qualunque E_n sarà incluso nel procedimento e la sua frequenza differirà da $P(E_n) = \lim P_k(E_n)$ per meno di un $\varepsilon_k \to 0$.

Nella concezione soggettiva considerazioni tipo frequenza-limite non hanno valore (nè, vedremo in seguito, n. 27, ha valore un' interpretazione possibile basata sul «passaggio al limite rispetto all' ipotesi»), ma v'è un concetto tipicamente soggettivo e del tutto diverso che lascia interpretare il procedimento. Supponiamo che le funzioni di probabilità $P_1(E), P_2(E), \ldots, P_n(E), \ldots$ rappresentino le opinioni di un' infinità di individui $O_1, O_2, \ldots, O_n, \ldots$ sugli eventi di una classe $\mathcal{E}$; supposto che (almeno in una sottosclasse $\mathcal{E}'$ di $\mathcal{E}$) $P_n(E)$ tenda a un limite $P(E)$ per $n \to \infty$, perchè non potrebbe un individuo O assumere tale funzione-limite come propria valutazione di probabilità? Più in generale, è certamente ancora ammissibile ogni combinazione lineare delle P_n

$$P(E) = \lambda_{h_1} P_{h_1}(E) + \lambda_{h_2} P_{h_2}(E) + \ldots + \lambda_{h_m} P_{h_m}(E)$$

(coefficienti λ positivi e di somma $= 1$); ammettendo la chiusura dell' insieme delle funzioni P, risultano ammissibili anche le funzioni limiti di tali combinazioni lineari, p. es. le serie $P(E) = \sum_1^{\infty}{}_h \lambda_h P_h(E)$ (con $\lambda_h \geq 0$ e $\sum_1^{\infty}{}_h \lambda_h = 1$), il limite di $P_n(E)$ nel senso di Cesaro e simili, ecc.

Queste considerazioni non possono in alcun modo costituire un argomento logico a favore della chiusura — e quindi contro l' additività completa — ma tuttavia illuminano un aspetto che mi sembra contribuisca a far apparire psicologicamente più plausibile detta tesi.

24. Il «postulato di continuità».

Abbiamo rilevato, nel n. 15, come l' additività completa sia equivalente al «postulato di continuità» (assioma VI di K.); occorre riprendere l'argomento e completarlo, giungendo a una formulazione ancor più significativa, indispensabile per il seguito (n. 33) e atta a metter meglio in evidenza l' appropriatezza del termine di «continuità».

A tal fine occorre introdurre la definizione (38) di massimo e minimo limite (ed eventualmente limite) di una qualunque successione di eventi $E_1, E_2, \ldots, E_n, \ldots$: l' evento $E' = \max \lim E_n$ significa che, dato N comunque grande, si verifica almeno un E_n con $n \geq N$, e l'evento $E'' = \min \lim E_n$ che esiste un N tale che tutti gli E_n si verificano da $n = N$ in poi. Si può anche dire che E' significa che si verificano infiniti tra gli E_n, ed E'' che si verificano tutti tranne al più un numero finito (proprietà queste però legate al fatto che gli E costituiscono una successione numerabile). Se massimo e minimo limite coincidono si dirà, al solito, limite di E_n lo evento $E = E' = E''$ (39).

La proprietà della «continuità» si esprime allora più in generale: se gli eventi $E_1, E_2, \ldots, E_n, \ldots$ tendono ad un limite E, è $P(E) = \lim P(E_n)$; brevemente: P e lim sono permutabili: $P(\lim E_n) = \lim P(E_n)$.

Il termine di «continuità» usato per indicarla risulta con ciò più palesemente giustificato.

La dimostrazione (naturalmente, basata sull' additività completa e ad essa subordinata) è ovvia: ponendo $E_n' = E_n + E_{n+1} + E_{n+2} + \ldots$, si ha manifestamente che sia E_n che E' sono contenuti in E_n', che potrà essere scisso in eventi incompatibili in entrambi i modi seguenti

$$E_n' = E_n + A_n = E' + B_n$$

da cui

$$P(E_n) \leq P(E') + P(B_n) \quad , \quad \max \lim P(E_n) \leq P(E')$$

poichè $B_n \to 0$. Analogamente

$$\min \lim P(E_n) \geq P(E''),$$

ossia

$$P(\min \lim E_n) \leq \min \lim P(E_n) \leq \max \lim P(E_n) \leq P(\max \lim E_n)$$

sicchè in particolare, se è $E' = E'' = E$, risulta necessariamente

$$\lim P(E_n) = P(E).$$

25. Osservazioni

La dimostrata equivalenza dell' additività completa con la continuità richiede due osservazioni.

La prima di carattere storico e psicologico, per notare che da diverse parti e per molto tempo si è creduto di poter dimostrare l'additività completa in base all' assioma di continuità, considerando questo (o almeno la sua formulazione ristretta) come una proprietà necessaria. Anche ora che la sua natura di ulteriore assioma, indipendente dai precedenti, non sembra venir più da alcuno misconosciuta, la preferenza data spesso, ad es. nella tratta-

(38) La definizione è in stretta relazione con quella usuale dell' analisi, come risulta ovvio osservando che per gli indicatori è:

$$|\max \lim E_n| = \max \lim |E_n|, \; |\min \lim E_n| = \min \lim |E_n|.$$

(39) V. (38); quindi (se uno e quindi l'altro esistono) $|\lim E_n| = \lim |E_n|$. Si osserverà ancora che il massimo limite della negazione è la negazione del minimo limite (e viceversa), che il massimo limite di una somma logica (o minimo di un prodotto) è la somma logica dei massimi limiti (o prodotto dei minimi), e quindi il limite (quando esista) è invertibile con negazione, somma, prodotto.

zione cui ci riferiamo del Kolmogoroff, all' assunzione come assioma della continuità piuttosto che direttamente dell' additività completa denota che si ravvisa un più marcato favore psicologico per il suo accoglimento.

Effettivamente è comprensibile che gli stessi motivi per cui a prima vista sembra spontaneo accogliere l' additività completa appaiano sotto un aspetto ancor più seducente nella diversa formulazione. Ma per persuadersi quanto tale impressione sia illusoria, basta osservare che essa avrebbe un carattere ancor più marcato di apparente evidenza parlando di classi e loro numero di elementi, o di figure e loro area, anzichè di eventi e loro probabilità. Eppure in quei casi, ove manca ogni nebulosità che possa esser fonte di perplessità, la proprietà non sussiste affatto. In una classe infinita, considerata una successione di elementi $a_1, a_2, \dots, a_n, \dots$ e detta A_h la classe formata da a_{h+n} $(n = 0, 1, 2, \dots, n, \dots)$, la successione di classi A_h è decrescente e tendente a zero, eppure il suo numero di elementi è sempre l' infinità numerabile. In un piano, considerata la regione esterna al cerchio di centro O e raggio r, al crescere di r si ha una successione di aree decrescenti e tendente a zero (nel senso che la parte comune è vuota), eppure la loro area è sempre infinita anzichè tendere a zero come vorrebbe l' «assioma». Anche le considerazioni relative agli esempi delle figg. 1 e 2 chiariscono la cosa.

La seconda osservazione è costruttivamente più importante. Le due proprietà, dell' additività completa e della continuità, sono bensì equivalenti quando si pensi la probabilità P definita per tutti gli eventi di un corpo boreliano (n. 5), ma la continuità ha il vantaggio di esser esprimibile in un campo di eventi E qualunque, anche senza la restrizione, incompatibile come s' è detto con la generalità dell' impostazione (n. 8), che essi costituiscano un corpo (o, peggio, un corpo boreliano). Grazie a ciò potremo cercar di impostare la trattazione relativa all' additività completa nel medesimo spirito di generalità con cui lo si è fatto (n. 9) per l' additività semplice; su ciò torneremo col n. 33 dopo aver esaminato altri aspetti della questione, forse logicamente posponibili ma necessari per completare intanto la trattazione nelle parti più elementari e più essenziali a una presa di posizione psicologica sull' alternativa pro o contro l' additività completa.

26. Paradosso della non-conglomerabilità

Occorre a tal fine ancora esaminare se l' introduzione delle probabilità subordinate e del teorema delle probabilità composte, ad esse relativo, possa portare nuovi elementi ed argomenti nella discussione del nostro problema. Si è infatti ritenuto che ciò porti a incontrare un paradosso; vedremo che

ciò deriva dall' ammettere tacitamente per le probabilità subordinate una proprietà (proprietà conglomerativa) equivalente alla continuità.

Ecco il presunto paradosso (che il Cantelli (40) attribuisce al Lévy): siano X e Y due «interi scelti a caso» (nel senso del n. 17) e «indipendentemente l' uno dall' altro» (ciò significa: «ogni probabilità riguardante X è la stessa sia non subordinatamente che subordinatamente a qualunque ipotesi riguardante Y, e viceversa»). Allora - argomenta il Lévy - qualunque sia il valore x assunto da X, la probabilità che sia $Y \leq x$ è nulla; analogamente, qualunque sia il valore y assunto da Y, è nulla la probabilità che sia $X \leq y$; quindi è nulla sia la probabilità che $Y \leq X$ che quella che sia $X < Y$, assurdo, trattandosi di due eventi contrari.

Ma è chiaro (si noti nel nostro enunciato la sostituzione di x e y con X e Y nel passaggio conclusivo!) che si è fatta un' ammissione implicita: si è ammesso cioè che se l'evento $Y \leq X$ ha probabilità nulla subordinatamente a ciascuna delle ipotesi possibili e incompatibili $X = 1, X = 2, X = 3, \dots$ $\dots, X = x, \dots,$ esso stesso deve avere probabilità nulla. Tale proprietà sussiste senz' altro quando le diverse ipotesi possibili sono in numero finito, ma per poterla estendere al caso di un' infinità di ipotesi è necessario e sufficiente ammettere appunto l' additività completa.

L' argomento del Lévy si basa su una proprietà e un esempio che avevo considerato già nel 1930 nella nota «Sulla proprietà conglomerativa delle probabilità subordinate» (41) dimostrando che tale proprietà conglomerativa non vale nel caso di infinite ipotesi.

La formulazione generale della proprietà predetta è la seguente: se si ha un'infinità che per l'attuale proposito dobbiamo supporre numerabile (ma per il concetto di conglomerabilità per sè stesso potrebbe avere potenza qualunque) di ipotesi possibili (incompatibili) $H_1, H_2, \dots, H_n, \dots,$ e subordinatamente a ciascuna di esse la probabilità di un evento E rimane compresa tra due limiti p' e p'': $p' \leq P(E/H) \leq p''$, si ha anche $p' \leq P(E) \leq p''$. La proprietà è equivalente all' additività completa; infatti se vale l' additività completa vale anche la conglomerabilità, perchè $P(EH_1 + EH_2 + \dots + EH_n) = P(H_1)P(E/H_1) + \dots + P(H_n)P(E/H_n) = \bar{p}\{P(H_1) + \dots + P(H_n)\}$ con $\bar{p}$ compreso tra p' e p''; scelto n abbastanza grande perchè $P(H_1) + \dots + P(H_n) \geq 1 - \varepsilon$ (e a maggior ragione quindi $P(EH_1) + \dots + P(EH_n) \geq P(E) - \varepsilon$) risulta $P(E) = p(1 - \theta_1 \varepsilon) + \theta_2 \varepsilon$ (con $0 \leq \theta_{1,2} \leq 1$); viceversa, se non vale l'additività

(40) F. P. CANTELLI, Sulla estensione del principio delle probabilità totali ad una successione illimitata di eventi incompatibili, "Giorn. Ist. It. Attuari", A. VI, n. 4, 1935. Si veda tale lavoro anche per conoscere le opinioni del Cantelli (sostenute pure nel colloquio da cui ebbe origine la presente esposizione).

(41) "Rend. R. Ist. Lombardo", Vol. LXIII, 1930.

completa, data una classe completa di eventi incompatibili $A_1\ B_1\ A_2\ B_2\ A_3\ B_3 \ldots$, potremo attribuire ad es. una probabilità p_1 ad A_1 e B_1, p_2 ad A_2 e B_2, e così via, con p_n tutte non nulle e tali che $\Sigma_n\, p_n = 1/3$, attribuendo però ad $A = \Sigma_n\, A_n$ la probabilità $\Sigma_n\, p_n = 1/3$ (supponendo cioè la probabilità completamente additiva sulle A_n) e invece a $B = \Sigma_n\, B_n$ la probabilità 2/3. Subordinatamente a ciascuna delle seguenti ipotesi possibili: $H_1 = A_1 + B_1$, $H_2 = A_2 + B_2$, ecc. la probabilità di A è

$$P\,(A/H_n) = P\,(AH_n)\,/\,P\,(H_n) = P\,(A_n)\,/\,\{P\,(A_n) + P\,(B_n)\} = p_n/2p_n = 1/2\,,$$

e tuttavia $P\,(A) = 1/3$.

Tale esempio mette in luce come la non-conglomerabilità s' incontri anche senza considerare probabilità subordinate ad ipotesi di probabilità nulla: il paradosso, se così lo si vuol chiamare, subentra sempre che non ci si limiti all' additività completa. Ma si tratta di un paradosso nel senso proprio della parola: non assurdità, ma verità che appare strana; basta osservare che tale «paradosso» probabilistico non è se non la traduzione in termini di probabilità di ovvi fatti analitici e geometrici (come illustrato nel n. 30).

27. Continuità rispetto alla subordinazione.

E' utile a questo punto osservare, per meglio chiarire il parallelismo tra l'*additività completa* e la *conglomerabilità*, che, come quella si riduce alla «continuità» (nel senso del n. 24), così questa si riduce facilmente alla formulazione perfettamente simmetrica e precisamente ancora alla «continuità» di $P\,(E/H)$, però rispetto ad H anzichè alla E.

Volendo escludere per ora il caso di probabilità subordinata ad eventi di probabilità nulla, sarà sempre $P\,(E/H) = P\,(EH)\,/P\,(H)$; supposto che $H_1\ H_2 \ldots H_n \ldots$ sia una successione di eventi con limite H, la successione $EH_1, EH_2, \ldots, EH_n, \ldots$ ha limite EH (cfr. nota (39)) e nell' ipotesi dell' additività completa sarà $P\,(EH_n) \longrightarrow P\,(EH)$, $P\,(H_n) \longrightarrow P\,(H)$, $P\,(E/H_n) = P\,(EH_n)\,/\,P\,(H_n) \to P\,(EH)\,/\,P\,(H)$ purchè $P\,(H) \neq 0$. In forma più generale, si ha quindi la continuità rispetto ad E e H; se $E_n \longrightarrow E$ ed $H_n \longrightarrow H$, purchè $P\,(H) \neq 0$, è $P\,(E_n/H_n) \to P\,(E/H)$.

E inversamente (come al n. 26).

28. Caso di ipotesi a probabilità nulla.

Se, supponendo sempre non nulle le probabilità delle ipotesi H_n, è $P\,(H) = 0$, non si può concludere che $P\,(E/H_n) \longrightarrow P\,(E/H)$ pur continuando a supporre valida l' additività completa. Per convincersene basta pensare che

sostituendo E con E′ (risp. E″), coincidente con E salvo nel caso H in cui è sempre *vero* (risp. *falso*) indipendentemente da E (42), si ha, per ogni n, $P(E'/H_n) = P(E''/H_n) = P(E/H_n)$ (perchè $P(E^*H_n) = P(E^*H_n H) + P(E^*H_n \overline{H}) = 0 + P\,(EH_n \overline{H})$, ponendo sia $E^* = E'$, che $E^* = E''$ o $E^* = E$; quindi $P\,(E'\,H_n) = P\,(E''\,H_n) = P\,(E\,H_n)$, c. d. d.); è quindi invariabile anche $\lim P(E^*/H_n)$ (oppure i limiti massimo e minimo), ma $P(E'/H) = 1$, $P(E''/H) = 0$, e non può quindi il limite essere uguale sia a zero che ad 1, come vorrebbe la continuità.

Ammettendo le probabilità subordinate a ipotesi di probabilità nulla, la proprietà predetta, valida nell' ipotesi di additività completa, va espressa in generale come segue: se $E_n \to E$ e $H_n \to H$, è $P\,(E_n/H_n) \to P\,(E/H)$ purchè la probabilità di H non sia incommensurabilmente minore di quella degli H_n, ossia, precisamente, purchè $P\,(H_n)/P\,(H)$ risulti limitato (almeno da un certo n in poi). Il caso opposto, in cui la probabilità di H fosse incommensurabilmente maggiore di quelle degli H_n, è di per sè escluso, nell' ipotesi di additività completa, perchè in quel caso $P\,(HH_n)/P\,(H) = P\,(H_n/H)$ avrebbe minimo limite nullo, anzichè dare $\lim P\,(H_n/H) = P\,(H/H) = 1$.

29. Chiarimenti sulle probabilità definite da passaggio al limite.

Possiamo ora riprendere e chiarire le questioni accennate al n. 9 sul valore di un passaggio al limite come quello che definisce il cosiddetto esempio di «un intero scelto a caso». Avevamo ivi chiarito che nessun riferimento veniva da noi fatto all' interpretazione come frequenza limite; aggiungiamo ora che nessun riferimento viene parimenti fatto all' interpretazione che potrebbe dirsi della probabilità-limite. Essa consisterebbe nel ragionare in base alla «continuità rispetto all' ipotesi»: dato che la probabilità subordinata all' ipotesi H_n che si verifichi uno dei primi n casi è $P_n(E) = P(E/H_n)$ *ne segue* che la probabilità non subordinata, ossia la probabilità subordinata all' evento certo $H = \lim H_n$, sarà $P(E) = \lim P_n(E)$.

Ora, è proprio vero il *contrario* e cioè che abbandonando l'additività completa (senza di che andrebbe esclusa la considerazione degli stessi esempi in oggetto), la *continuità in tal senso non solo non viene postulata ma anzi viene negata*. Può avvenire solo casualmente che $P(\lim E_n) = \lim P(E_n)$, oppure $P(E/\lim H_n) = \lim P(E/H_n)$, perchè *tali relazioni costituirebbero dei teoremi soltanto se si ammettesse l'additività completa*.

Tutto ciò che si può dire è che, supponendo assegnate le probabilità $P(E/H_n)$ per E appartenente a una certa classe di eventi $\mathcal{E}$ ed H_n appartenente

(42) Ponendo cioè: $E' = E + H$, $E'' = E\,.\,\overline{H}$, da cui $E'\,.\,\overline{H} = E''\,.\,\overline{H} = E\,.\,\overline{H}$.

a una certa successione d'eventi (tendente oppure no all'evento certo), la funzione $P(E) = \lim P(E/H_n)$, definita per gli E' per i quali il limite esiste, rappresenta una distribuzione additiva, ossia una delle infinite possibili distribuzioni di probabilità; quindi nulla obbliga ma nulla impedisce di assumere $P(E)$ come valutazione per le probabilità.

In sostanza si tratta sempre di un passaggio al limite che ha il solo scopo di esprimere analiticamente una certa distribuzione.

Appunto per mostrare che il passaggio al limite non è che strumento di rappresentazione analitica, ho preferito costruire nel n. 22 l' esempio della figura dove il risultato di definire una distribuzione non completamente additiva è ottenuto senza far figurare una distribuzione variabile, e appunto per tale vantaggio ci riferiremo preferibilmente a tale esempio. Ma non con altro spirito o significato si deve pensare ci riferiremo, quando risulti opportuno, a una distribuzione nella cui definizione si utilizzi un passaggio al limite, come facciamo nell' esempio che segue, in cui sfruttiamo appunto, per precisare le probabilità nel caso cui si riferisce il paradosso di Lévy, il procedimento consistente nel definire direttamente $P(E/H_n)$ per certi eventi $H_1\, H_2 \ldots H_n \ldots$ e nel porre per definizione $P(E/H) = \lim P(E/HH_n)$ ogni qual volta tale limite esiste.

30. Esempi per la non-conglomerabilità.

Nel caso del Lévy, (n. 26) sul quale vogliamo ora illustrare in alcuni modi la non conglomerabilità, detta $H_{m,n}$ l' ipotesi che risulti $X = 1, 2, \ldots n$, $Y = 1, 2, \ldots, m$ (ossia $H_{m,n} = [(X < n) . (Y \leq m)]$), si può precisare la distribuzione di probabilità da considerarsi, ponendo, per ogni E, $P(E/H_{n,m}) = \frac{1}{nm} \times$ «numero delle coppie (x, y), fra le nm di $H_{n,m}$, soddisfacenti E», e $P(E/H) = \lim P(E/HH_{n,m})$ quando n e m tendono ad ∞ restando $n = m$. L'evento $E = [Y < X]$ considerato dal Lévy corrisponde alle coppie (x, y) con $y < x$, e nel rettangolo $H_{n,m}$ ne esistono $\frac{1}{2} n(n-1)$ se $n \leq m$, risp. $\frac{1}{2} n(n-1) + m(n-m)$ se $n > m$; dividendo per nm si ha la probabilità

$$P(E/H_{n,m}) = \frac{1}{2}\frac{n-1}{m}, \text{ risp. } = 1 - \frac{1}{2}\frac{m+1}{n}$$

facendo tendere ad ∞ sia n che m, in modo che $n/m \to 1$, si ha $P(E) = 1/2$; considerare invece probabilità di E subordinate ad ipotesi corrispondenti a valori limitati di X (o di Y), significa far tendere ad ∞ prima la sola m (la sola n), mentre i due passaggi al limite non sono invertibili. Il «paradosso probabilistico» della non conglomerabilità corrisponde quindi al «paradosso analitico» di due passaggi al limite non invertibili.

Geometricamente, se rappresentiamo cartesianamente X e Y (che ora, per variare un po' l' esempio, supporremo suscettibili di assumere tutti i valori reali), lo stesso evento E corrisponde all' angolo tra il semiasse x e la bisettrice del quadrante positivo, ed ogni evento consistente in una limitazione per la X (o la Y) in una striscia verticale (orizzontale). La probabilità corrisponde, mediante il passaggio al limite, all' area, sicchè il paradosso probabilistico equivale al paradosso geometrico per cui dividendo il quadrante positivo in tante striscie verticali, l' area appartenente all' angolo E è in ciascuna di esse infinitamente minore dell' altra, mentre dividendo in striscie orizzontali avviene il contrario. Sicchè, sembrerebbe di concludere, ciascuna delle due aree è infinitamente maggiore dell' altra (v. fig. 3).

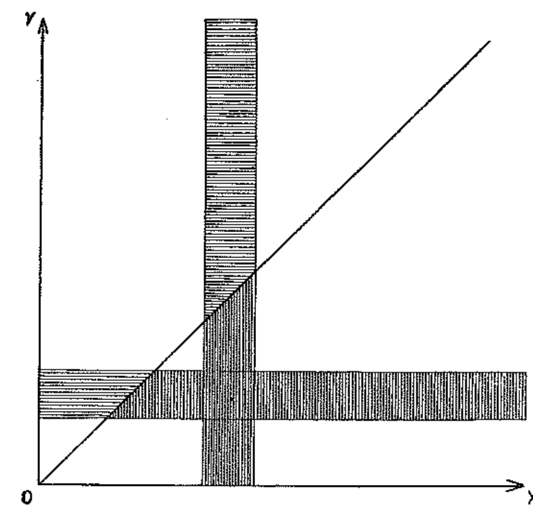

Figura 3.

Per evitare anche qui i passaggi al limite, e insieme per non far intervenire eventi a probabilità nulla mostrando così che non dalla loro introduzione nasce il paradosso, ricorriamo ancora al medesimo esempio del n. 22; per maggior comodità considereremo però un quadrato anzichè un cerchio, e supporremo che la probabilità sia data dalla lunghezza dell'orlo sui due soli lati segnati con linea ingrossata nella fig. 4. Le suddivisioni verticali e orizzontali dividono il quadrato in rettangoli $E_{n,m} = A_n B_m$, ove si indichino con $A_1 \ldots A_n \ldots$ le successive striscie verticali, e con $B_1 \ldots B_m \ldots$ quelle orizzontali. L'evento E consiste, analogamente agli es. precedenti, dei casi ove $m < n$, ossia dei rettangoli al di sotto della diagonale (ascendente); la probabilità di E è 1/2, perchè si appoggia a uno dei due lati costituenti il contorno utile.

La probabilità di E subordinatamente a un qualunque A_n è però nulla (perchè il contorno utile di A_n è sul lato superiore appartenente a non-E), e subordinatamente a un qualunque B_n è sempre 1 (perchè il contorno utile di B_n è sul lato verticale, e appartiene totalmente ad E).

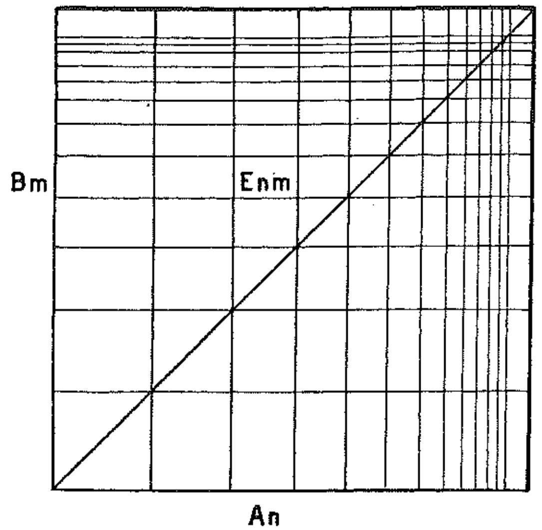

Figura 4.

Su quest' ultimo esempio è facile rendersi ragione, come al n. 22, del modo in cui compaiono i presunti paradossi: le probabilità subordinate sono rapporti tra lunghezze di contorni, e, se si considera una suddivisione in infinite ipotesi incompatibili $H_1 \ldots H_n \ldots$, oltre i tratti di contorno cui si appoggia una delle aree H_n (per la cui suddivisione in parti appartenenti ad E e non-E vale la proprietà di media espressa dalla conglomerabilità) esiste in generale un tratto che viene approssimato dagli H_n per $n \to \infty$ (nel nostro esempio, il lato verticale e quello orizzontale a seconda che si ci riferisce alle ipotesi A_n oppure alle B_m), e sul modo in cui esso influisce sulla media nulla può dirsi in base alla proporzione determinata sull' altra parte del contorno (ed è ciò che vorrebbe la proprietà conglomerativa).

31. Osservazioni critiche sul concetto di « probabilità subordinata ».

Al fine di renderci ragione dei motivi psicologici per cui il paradosso di Lévy appare un paradosso, giovano forse alcune considerazioni sul significato delle probabilità subordinate. Dicendo che $P(E/H) = p$, si afferma di valutare p

la probabilità « che E si verifichi subordinatamente al verificarsi di H», e l' ente E/H o la sua spiegazione tra virgolette dev' essere considerata un tutto unico. In termini di scommesse, p è la probabilità che servirà di base per una scommessa da stipularsi attualmente, e che rimane nulla se H non si verifica, essendo vinta o persa, se H si verifica, a seconda che E si verifichi o no. Nel linguaggio comune (e forse anche in qualche concezione oggettiva della probabilità) il nesso della frase si spezza, assumendo il seguente significato ben diverso «supposto che H si verifichi, allora la probabilità di E è p», e, assunto tale aspetto di deduzione logica, sarebbe indubbiamente legittimo dedurre dall' enunciato con perfetto sillogismo che se da un lato è certo che deve verificarsi uno degli eventi (sia pure infiniti) $H_1\ H_2 \ldots H_n \ldots$, e se in ciascuno di tali casi possibili è accertato che la probabilità di E è p, allora è senz' altro $P(E) = p$.

Che tale modo di intendere sia logicamente inesatto è facile persuadersi; basta del resto un facile esempio per ridurlo visibilmente all' assurdo. Seguendo tale interpretazione la condizione dell' incompatibilità per gli H_n non sarebbe necessaria, e si potrebbe ad es. concludere che la probabilità che un estratto al lotto sia un numero pari è 1/3 perchè tale è subordinatamente a ciascuna delle ipotesi di ottenere (1, 2, 3) o (3, 4, 5) o (5, 6, 7)..., e una di tali ipotesi è necessariamente vera. Seguendo tale interpretazione ancora più macchinalmente si potrebbe concludere che la probabilità di testa e croce non è 1/2, perchè, considerando tre prove, subordinatamente alle quattro ipotesi possibili che testa si presenti 0, 1, 2, 3 volte la probabilità è 0, 1/3, 2/3, 1, ossia in nessun caso 1/2.

Considerando invece la «probabilità-subordinata» come un concetto nuovo e unico, nulla vi è di contraddittorio nell' ammettere che le scommesse su un evento E subordinatamente a ciascuna di singole ipotesi, di cui ciascuna per sè stessa pressochè impossibile, vengano fatte a condizioni diverse da una scommessa non subordinata; può tutt' al più persistere forse quella solita impressione di disagio che danno le proprietà nuove caratteristiche dell' infinito in tutti i campi della matematica finchè l' assuefazione e la meditazione non l' acquietino.

32. Considerazioni per un' impostazione generale.

Le considerazioni e le conclusioni finora svolte o stabilite in merito all' additività completa dimostrano sufficientemente che l' accettazione di tale assioma non va scevra da dubbi e da difficoltà.

Volendo abbozzare uno schema d' impostazione e trattazione generale, converrà pertanto non presupporre senz' altro valida l' additività completa precludendosi l' esame non solo dei casi ove non valesse ma anche delle circostanze e modalità in cui può valere: converrà al contrario prescindere dal considerare quelle proprietà come un presupposto per circoscrivere a priori l'ambito dello studio, proponendosi come oggetto di ricerca il caso generale delle funzioni semplicemente additive per occuparsi di vedere quali e in qual campo e senso risultino completamente additive.

Sull'opportunità di un simile modo di procedere mi sembra dovrebbe concordare non solo chi si trovasse nel primo o secondo dei possibili atteggiamenti di seguito elencati, ma anche chi aderisse al terzo:

1) chi fosse convinto che l'additività completa è insostenibile (come è mia opinione), non potrebbe che procedere così, pur trovando interessante studiare quelle particolari distribuzioni di probabilità che in certi campi risultino completamente additive;

2) chi ritiene impregiudicata la questione, vedrebbe nella via indicata il mezzo migliore per esaminare in tutte le loro conseguenze entrambe le eventualità, e poter poi scegliere o prospettare obiettivamente ad altri la alternativa;

3) chi infine rimanesse tuttavia convinto della necessità dell'additività completa per la teoria delle probabilità, potrebbe ugualmente riconoscere come più significativo e elegante inquadrare lo studio delle funzioni completamente additive in un dato campo nello studio del comportamento rispetto all'additività completa delle funzioni semplicemente additive.

Allo stesso modo come sarebbe pur sempre più istruttivo definire e studiare le funzioni in generale, e la continuità, per soffermarsi poi sulle funzioni continue come caso particolare, anche per chi ritenesse di non aver mai a incontrare in pratica altro che quest'ultimo. E vedremo che l'analogia è molto stretta e condurrà a molte osservazioni chiarificatrici che altrimenti potrebbero non venire alla mente, come del resto è prevedibile dato il significato dell'additività completa come «continuità».

33. Numeri aleatori e additività completa.

Per impostare la trattazione in generale giova anche qui come nel n. 9 considerare definita la funzione P non solo per gli eventi E di una certa classe di eventi $\mathcal{E}$ ma anche per i numeri aleatori X del sistema lineare $L\mathcal{E}$ o di quello più ampio $\mathcal{L}$ cui risulta possibile estenderla.

Le condizioni cui una funzione P andava assoggettata, come visto nel n. 10 erano le seguenti:

1) P dev'essere additiva ($P(X+Y) = P(X) + P(Y)$) e

2) non negativa ($P(X) \geq 0$ se è certamente $X \geq 0$).

Supposto assegnato il valore di $P(X)$ per gli X di un certo sistema lineare di numeri aleatori $\mathcal{L}$, si potrà considerare per ogni altro numero aleatorio X_0 gli estremi inferiore $P_<(X_0)$ e superiore $P_>(X_0)$ dei valori attribuibili a $P(X_0)$ senza che P cessi d'essere additiva e non-negativa nel campo $\mathcal{L} + (X_0)$; essi sono rispettivamente l'estremo superiore di P(X) per gli X di $\mathcal{L}$ certamente $\leq X_0$, e l'estremo inferiore di P(X) per gli X di $\mathcal{L}$ certamente $\geq X_0$. In particolare $P(X_0)$ rimane univocamente definito se $P_<(X_0) = P_>(X_0)$, ossia se, fissato a piacere $\varepsilon > 0$, esistono in $\mathcal{L}$ due numeri aleatori X' e X'' tali che certamente $X' \geq X \geq X''$ e $P(X'' - X') < \varepsilon$. E' quanto, sostanzialmente, s'era detto in merito nel n. 9 e dimostrato (riferendosi, senza sostanziali restrizioni, ad eventi) nel n. 10; non facciamo ora in più se non considerare e definire esplicitamente gli estremi inferiore e superiore.

Per essi vale una proprietà di concavità (o convessità); infatti ovviamente

$$P_<(X+Y) \geq P_<(X) + P_<(Y), \quad P_>(X+Y) \leq P_>(X) + P_>(Y),$$

e

$$P_<(\lambda X) = \begin{cases} \lambda P_<(X) & \text{se} \quad \lambda > 0 \\ \lambda P_>(X) & \text{se} \quad \lambda < 0 \end{cases}$$

In particolare si può concludere che i numeri aleatori per cui lo scarto tra estremo superiore e inferiore di P non supera un valore ρ assegnato costituiscono un insieme convesso, nel senso che se vi appartengono $X_1, X_2, \ldots$ $\ldots, X_n$ vi appartiene anche ogni $X = \Sigma_h \lambda_h X_h$ con $\lambda_h \geq 0$, $\Sigma_h \lambda_h = 1$: se infatti è, per ogni h, $P_>(X_h) - P_<(X_h) < \rho$, è

$$P_>(X) - P_<(X) \leq \Sigma_h \lambda_h P_>(X_h) - \Sigma_h \lambda_h P_<(X_h) =$$
$$= \Sigma_h \lambda_h [P_>(X_h) - P_<(X_h)] < \rho \Sigma_h \lambda_h = \rho.$$

L'insieme per cui $\rho = 0$ è il sistema lineare $L_P\mathcal{L}$ su cui il prolungamento di P è univoco (n. 10); un prolungamento a qualunque sistema più ampio è sempre possibile e in infiniti modi (col procedimento d'induzione transfinita del n. 9).

Giova tener presenti queste circostanze per vedere se e fin dove sia possibile seguire la stessa traccia per l'additività completa.

Occorre anzitutto formulare le condizioni di additività completa riferendoci ai numeri aleatori anzichè agli eventi; la forma suggerita da una spontanea estensione delle formule relative agli eventi sarebbe

$$P(X_1 + X_2 + \ldots + X_n + \ldots) = P(X_1) + P(X_2) + \ldots + P(X_n) + \ldots$$

(traduzione dell'additività completa) oppure

$P(X_n) \to 0$ quando $X_n \to 0$ (traduzione dell'assioma di continuità);

ma evidentemente tali enunciati non hanno un significato univoco fintanto che non vengano precisati.

Le restrizioni all' enunciato dell' additività completa quale è noto dalle trattazioni che la postulano (Kolmogoroff, IV, § 2, p. 35) e quale risulta dall' analoga proprietà per gli integrali (cfr. ibid. p. 34) è la convergenza della serie $\Sigma P(|X_n|)$ relativa ai valori assoluti: sotto tale condizione vi si dimostra infatti che la serie $\Sigma_n |X_n|$, e quindi a maggior ragione la $\Sigma_n X_n$, converge salvo in casi di probabilità nulla.

Ripetiamo il ragionamento colle varianti dovute alla presente impostazione. Se $\Sigma_n P(|X_n|)$ converge, scelti comunque ε e λ positivi, esiste un N tale che, qualunque sia q, risulta $\sum_{N}^{N+q} {}_n P(|X_n|) < \lambda\,\varepsilon$ ossia $P(\sum_{N}^{N+q} {}_n |X_o|) < \lambda\,\varepsilon$, e a maggior ragione $P(\sum_{N}^{N+q} {}_n |X_n| > \lambda) < \varepsilon$, o, scrivendo brevemente, $P(\Sigma_q > \lambda) < \varepsilon$ (perchè il numero aleatorio X definito col porre $X = 0$ o $X = \lambda$ a seconda che $\Sigma_q \leq \lambda$ o $\Sigma_q \geq \lambda$, ossia l'indicatore $|\Sigma_q > \lambda|$, è sempre $\leq \Sigma_q$, e il suo valor medio $\lambda P(\Sigma_q > \lambda)$ è quindi $\leq P(\Sigma_q) < \lambda\,\varepsilon$).

Se vale per gli eventi $\Sigma_q > \lambda$ l'assioma di continuità, potremo dire che l' evento limite $\Sigma > \lambda$, ossia $\Sigma_n |X_n| > \lambda$, ha ancora probabilità $< \varepsilon$ ossia (a fortiori) che la probabilità che $\Sigma_n |X_n|$ diverga è $< \varepsilon$, od ancora, attesa l'arbitrarietà di ε, che essa è nulla: $P(\Sigma_n |X_n| = \infty) = 0$. A maggior ragione è nulla la probabilità che sia non convergente $\Sigma_n X_n$, e, poichè il resto di tale serie è inferiore a quello della $\Sigma_n |X_n|$, possiamo concludere nelle condizioni precedenti che ponendo $X = \Sigma_n X_n$ coll' intesa di considerare $X = 0$ qualora $\Sigma_n X_n$ risulti non convergente, risulta $|P(X) - \sum_{1}^{N-1} {}_n P(X_n)| < \lambda\,\varepsilon$ ossia $P(X) = P(\Sigma_n X_n) = \Sigma_n P(X_n)$.

Quanto alla seconda formulazione, $P(X_n) \to 0$ se $X_n \to 0$, è evidente che essa non è ammissibile se s' intende riferirsi alla convergenza semplice dei numeri aleatori X_n verso lo 0 (impossibilità che la successione $X_1 X_2 \ldots X_n \ldots$ non tenda a zero, o, considerando gli X_n come funzioni dei «casi elementari possibili» C, $X_n(C) \to 0$ per ogni C). Basta pensare al caso di una classe numerabile completa di eventi incompatibili $E_1 E_2 \ldots E_n \ldots$: pur valutandone le probabilità $p_1 p_2 \ldots p_n \ldots$ in modo da rispettare l' additività completa ($\Sigma_n p_n = 1$), la successione di numeri aleatori $X_n = |E_n| / p_n$, convergente verso 0 (in ogni caso E_h è infatti $X_n = 0$ per $n > h$), dà $P(X_n) = p_n/p_n = 1 \to 1$ anzichè $\to 0$ (ed $|E_n| / p_n^2$ darebbe $\to \infty$). La condizione per i numeri aleatori equivalente alla $P(E_n) \to 0$ per $E_n \to 0$ va espressa invece:

$P(X_n) \to 0$ se $X_n \to 0$, essendo gli X_n uniformemente limitati. Infatti, gli $|E_n|$ sono uniformemente limitati ($|E_n| \leq 1$) e se ne deduce quindi la condizione usuale; viceversa, se essa vale, e se è sempre $|X_n| < K$, è

$$|P(X_n)| < \varepsilon + K\,P(|X_n| > \varepsilon) \quad (\text{qualunque sia } \varepsilon > 0),$$

ma se $X_n \to 0$ anche $[|X_n| > \varepsilon] \to 0$ (ciò significa - si rammenti - che è impossibile che vi siano infiniti $|X_n| > \varepsilon$), e quindi

$$\lim P(|X_n| > \varepsilon) = 0 \quad , \quad \lim |P(X_o)| < \varepsilon,$$

e, per l' arbitrarietà di ε, $\lim P(X_n) = 0$, c.d.d.

34. Continuità sui sistemi lineari.

Una somma di infiniti numeri aleatori si può interpretare come una scommessa composta di infinite scommesse; volendosi ricollegare alla definizione di coerenza fondata su tale ordine di idee, la condizione dell' additività completa e della continuità significherebbe che l' impossibilità di assicurarsi un guadagno certo mediante combinazioni di scommesse eque va estesa a combinazioni numerabili (opportunamente definite e delimitate).

La condizione del n. 9 per la c o e r e n z a si trasforma allora nella seguente più restrittiva condizione per la c o e r e n z a e c o n t i n u i t à di una funzione F sul sistema lineare di numeri aleatori $\mathcal{L}$: considerati i numeri aleatori X della forma

$$X = k_1 \{X_1 - P(X_1)\} + k_2 \{X_2 - P(X_2)\} + \ldots + k_n \{X_n - P(X_n)\} + \ldots$$

ove i k_h siano numeri reali qualunque, gli X_h appartengano ad $\mathcal{L}$, la serie sia convergente e i resti siano dei numeri aleatori ugualmente limitati ($|R_n| < K$), nessuno di tali X risulta certamente positivo.

Tale formulazione ha il pregio di mettere in luce il vero senso del problema, e cioè il fatto che la proprietà dell' additività completa o non completa, ossia della continuità o discontinuità, riguarda il comportamento della funzione P su un sistema lineare $\mathcal{L}$; studiare completamente il comportamento della P rispetto alla continuità significa pertanto distinguere quali sistemi lineari $\mathcal{L}$ appartengano al complesso Λ_P dei sistemi lineari su cui P è continua, e quali no.

Alcune proprietà del complesso Λ_P vanno subito stabilite.

Anzitutto, se un sistema lineare $\mathcal{L}$ appartiene a Λ_P, vi appartiene pure (ovviamente) ogni sistema lineare contenuto in $\mathcal{L}$, ed anche il sistema lineare $\mathcal{L}'$ formato da tutti i numeri aleatori ottenibili dagli $\mathcal{L}$ mediante passaggio al limite nel senso del n. precedente ($X_n \to X$, $|X_n - X| < K$).

Se i sistemi lineari $\mathcal{L}_1$ e $\mathcal{L}_2$ appartengono a Λ_P, lo stesso è del sistema lineare $\mathcal{L}_1 + \mathcal{L}_2$ formato dalle somme $X_1 + X_2$ di numeri aleatori X_1 di $\mathcal{L}_1$ e X_2 di $\mathcal{L}_2$ (infatti ogni serie formata con gli $\mathcal{L}_1$ e $\mathcal{L}_2$ è somma di una serie formata con $\mathcal{L}_1$ e una formata con $\mathcal{L}_2$, ecc.). Però la proprietà non vale se anzichè due o un numero finito di sistemi $\mathcal{L}_1 \mathcal{L}_2 \ldots \mathcal{L}_n$ ne consideriamo una infinità.

Consideriamo ad es. una classe numerabile completa di eventi incompatibili E_{hk} $(h, k = 1, 2, \ldots, n, \ldots)$ di probabilità p_{hk}, e indichiamo con $E_h = \Sigma_k E_{hk}$ la somma degli eventi con primo indice h, e con p_h la sua probabilità. Supponiamo ora che sia $p_h = \Sigma_h p_{hk}$ (per ogni h), ma $\Sigma_h p_h < 1$ (ossia $\Sigma_{hk} p_{hk} < 1$). Sui sistemi lineari L_h definiti dagli E_{hk} e da E_h, la P è continua e quindi lo è su ogni sistema lineare L determinato da un numero finito di L_h; ciò non sussiste più però quando si consideri il sistema L determinato da tutti gli infiniti L_h.

L' esempio ora dato dimostra poi che non è vera in generale quella che sarebbe stata l' ipotesi più semplice (e che le precedenti proprietà potevano far sperare): che cioè il complesso Λ_P fosse costituito da tutti e soli i sistemi lineari appartenenti a un certo sistema lineare L^*, che avrebbe in tal caso assunto il significato di «campo totale di continuità».

35. Il prolungamento continuo

Supposto di aver definito P per un sistema lineare L_0 in modo da risultarvi continua, si pone il problema della possibilità di estendere P a un campo ampliato conservando la continuità, e di riconoscere se ciò si possa fare in un unico modo o con quale grado di arbitrarietà.

Consideriamo anzitutto il campo di prolungamento continuo univoco di una P definita su L_0 e ivi continua: intendiamo definirlo dicendo che un numero aleatorio X vi appartiene se esiste un unico valore attribuibile a P (X) in modo che P risulti additiva e continua anche sul sistema $L_0 + (X)$.

Non possiamo concludere senz'altro che prolungando in tal modo la P al campo di prolungamento continuo univoco si ottenga ivi una funzione additiva e continua: i prolungamenti per continuità ai diversi elementi X potrebbero essere incompatibili. Ciò dovrebbe tuttavia risultare escluso per le stesse ragioni sussistenti nel caso della misura, e per le quali, deducendo mediante prolungamento basato sull' additività completa la misura di Borel dalla lunghezza dei segmenti, la stessa misura di Borel risulta completamente additiva nel nuovo campo (degli insiemi misurabili - B).

Considerando allora la funzione P nel campo L_1 di prolungamento continuo univoco, potremo dire che essa vi è continua ma non ulteriormente prolungabile per continuità.

All' infuori del campo L_1 del prolungamento continuo univoco, il prolungamento continuo sarà possibile o no a seconda del sistema lineare nel quale lo si voglia effettuare. Sappiamo già che il prolungamento non è sempre possibile, come si è rilevato nel n. 21, a differenza che per le funzioni P semplicemente additive, prolungabili sempre a qualunque campo come detto nel

n. 9. Tuttavia giova notare che il prolungamento ad un nuovo elemento (evento o numero aleatorio) risulta ancora possibile: il medesimo ragionamento del n. 9 mostra infatti che se un' incongruenza scaturisse dall' attribuire un qualunque valore a P(X) per una X fuori dell'attuale campo di definizione L_1, l' incongruenza preesisterebbe tra le valutazioni entro L_1. E' questo del resto il caso in cui quel ragionamento era stato fatto dal Lévy (43), ed esso mostra effettivamente la possibilità di prolungare la P a nuovi elementi in numero finito comunque grande, ossia di aumentare di un numero finito le dimensioni del sistema lineare ove P è definito. L' impossibilità di prolungamento incondizionato non deriva dal raggiungimento di una barriera al di là della quale non si possa più fare un passo, ma dall' incompatibilità che può sorgere dal considerare insieme infiniti passi di cui ciascuno lecito.

Tale situazione è tutt' altro che stravagante: anche per i sottoinsiemi finiti (di un dato insieme infinito) si ha il medesimo fatto, che aggiungendovi altri elementi in numero finito non cessano di essere sottoinsiemi finiti; essa però complica tutte le questioni in quanto s' incontrano problemi di compatibilità che non hanno riscontro nel caso dell' additività semplice.

A cominciare dal problema (analogo ed in un certo senso inverso a quello del precedente n. 34) consistente nello studiare il nuovo complesso dei sistemi lineari Λ_{P, L_1} per i quali sussiste la possibilità di prolungamento continuo della funzione P definita in L_1 ed ivi completamente additiva, si presentano, come si vede, numerosi problemi, probabilmente non facili, di cui mi sono limitato ad indicarne qualcuno a titolo esemplificativo. Si tratta sostanzialmente di problemi di teoria degli insiemi che ignoro se e fino a qual punto siano stati già studiati (44).

Per indicare il significato analitico dell' impostazione considerata, diamo ancora, nel prossimo n. 36, un cenno sulla sua interpretazione come teoria dell' integrazione astratta (e in particolare teoria della misura, se le funzioni integrande sono le «funzioni caratteristiche» di insiemi: $f(x) = 1$ se $x \in I$, e $f(x) = 0$ se $x \notin I$).

36. Saggio d'impostazione per l' integrazione astratta.

Consideriamo degli enti X costituenti un sistema lineare L, comprendente le costanti (numeri reali) c, e pei quali risultino definite:

(43) P. LÉVY - Calcul des probabilités, Gauthier-Villars, Paris, 1928, Appendice.

(44) Ho ritardato la pubblicazione della presente memoria anche per poter vedere dei lavori che sembra si occupino di argomenti del genere; finora però non mi è riuscito di procurarmeli.

a) una struttura d'ordine $X_1 \leq X_2$,
b) una struttura di limite $X_n \to X$,

da assoggettarsi alle restrizioni che occorreranno.

Per riferirsi al caso usuale dell'integrazione, ivi si avrebbe:

$X = f(x)$: funzioni reali finite degli enti x (numeri, punti, funzioni, ...) del campo d'integrazione C,

$X_1 \leq X_2$: è $f_1(x) \leq f_2(x)$ per ogni x di C,

$X_2 \to X$: è $f_n(x) \to f(x)$ per ogni x di C.

Una funzione F(X) di X si dice (in un certo campo di L):

positiva, se è $F(X) \geq 0$ per $X \geq 0$,
additiva, se è $F(X+Y) = F(X) + F(Y)$, e $F(\lambda X) = \lambda F(X)$,
continua, se è $F(X_n) \to F(X)$ quando $X_n \to X$,
completamente additiva se è additiva e continua.

Per una teoria dell'integrazione nel senso più ampio basta richiedere che l'«integrale» sia una funzione F(X) additiva e positiva (impostazione tipo misura di Jordan-Peano e integrale di Riemann), e come problema subordinato si presenta, come se si postulassero tali restrizioni secondo la moda invalsa dopo Lebesgue, ed anzi in modo più ampio, lo studio dei campi (in particolar modo, dei sistemi lineari) sui quali la funzione F(X) sia completamente additiva (o sia ivi prolungabile in modo da rimanervi tale).

Sotto l'ulteriore condizione dell'additività completa, si hanno possibilità di particolarizzazioni secondo i concetti di Baire-Borel e di Lebesgue, precisabili come segue.

Dato in L un insieme qualunque I, diremo I' l'insieme derivato di I, formato dagli X ottenibili con passaggio al limite da elementi di I: se gli X_n appartengono ad I, ed è $X_n \to X$, X appartiene ad I'. Il derivato di I' si dice derivato secondo, I'', e così via per induzione transfinita.

Nel caso delle funzioni: le I' sono le funzioni ottenibili come limiti delle funzioni di un dato insieme I, e così di seguito; se in particolare I è l'insieme delle funzioni continue, $I', I'', \ldots, I^{\alpha}, \ldots$ sono gli insiemi delle funzioni di Baire di ordine fino a $1, 2, \ldots, \alpha, \ldots$, e $I^{(\Omega)}$ è l'insieme di tutte le funzioni di Baire.

Ciò posto, si può estendere univocamente la F(X) al campo $L_o^{(\Omega)}$ in modo che vi risulti continua applicando ripetutamente la $F(X_n) \to F(X)$, seguendo il concetto di Baire e Borel.

Seguendo invece il concetto di Lebesgue si può procedere in modo più elementare ed estenderla maggiormente: estenderla cioè al campo totale L_o^* dell'univocità del prolungamento.

Con riferimento alla F, diremo «intorno ε» di un insieme I di L, e indicheremo con $I_{(\varepsilon)}$, l'insieme degli elementi X di L per cui esistono X′ e X″ appartenenti ad I e tali che $X' \leq X \leq X''$, $F(X'' - X') < \varepsilon$. Diremo «intorno 0» di I, e indicheremo I^*, l'insieme comune a tutti gli $I_{(\varepsilon)}$. Se X appartiene a I^*, F(X) è univocamente determinato potendosi per ogni $\varepsilon > 0$ scegliere X′ e X″ di modo che $F(X') \leq F(X) \leq F(X'') \leq F(X') + \varepsilon$.

Nel caso dell'integrazione (per limitare l'esemplificazione al caso di una sola variabile) le funzioni del campo primitivo I sono le funzioni caratteristiche degli intervalli, cioè le funzioni $f_{p,q}(x)$ definite dall'assumere il valore 1 o 0 a seconda che

$(x-p)(x-q) < 0$ o > 0 (e quindi quelle dei boreliani, che ne sono combinazioni lineari), e la F è definita da

$F(f_{p,q}) = q - p$ (ossia $\int_p^q dx = q - p$). Salvo lievi varianti la trattazione risulta allora la stessa del Baire seguendo il primo schema e la stessa del Lebesgue seguendo il secondo.

L'equivalenza dei due procedimenti andrebbe analizzata in generale, determinando quali proprietà sia necessario e sufficiente postulare per le strutture (a) d'ordine e (b) di limite affinchè essa sussista (ripercorrendo l'usuale dimostrazione dell'equivalenza di misura-B e misura-L ove entrambe definite).

Non mi proponevo qui di affrontare simili questioni; il cenno può solo servire a chiarire il punto di vista dal quale mi sembrerebbe opportuno porsi nell'esaminarle. E' mia convinzione che vi sia qualcosa d'ingiustificato nell'esclusivo favore ormai rivolto alle trattazioni imperniate sull'additività completa (misura e integrale di Lebesgue) a scapito di quelle basate sul concetto finitista (misura di Jordan-Peano e integrale di Riemann). Lo stesso studio della additività completa risulta mutilato quando essa divenga un presupposto, tanto più che per renderlo accettabile occorre limitare sistematicamente la trattazione a un campo pregiudizialmente ristretto (insiemi e funzioni misurabili).

37. Cenni su altri problemi.

Un altro vasto campo di problemi cui si giungerebbe nell'indirizzo indicato è quello consistente nello studio e classificazione dell'insieme delle funzioni additive possibili in un dato campo.

Tale insieme P costituisce un insieme convesso chiuso; fra le funzioni P di P figurano in particolare quelle che concentrano tutta la probabilità in un unico «caso elementare» C, e le loro combinazioni lineari, corrispondenti a distribuzioni in cui tutta la probabilità risulta concentrata in un numero finito di «casi elementari».

Interessante poi sarebbe lo studio sistematico di quelle distribuzioni P che si possono ottenere mediante passaggio al limite da quelle predette, naturalmente definibili solo per quegli eventi o numeri aleatori sui quali si ha convergenza. Ciò interesserebbe anche perchè è a questo particolare artificio di definizione che si ricorre in molti esempi (come in quello più volte citato dell'«intero scelto a caso»: distribuzione limite di quella con probabilità 1/n per ciascuno dei primi n casi). Ma questo è certo un tipo di distribuzione molto speciale, e sarebbe bene sapere in cosa precisamente consistano le particolarità di questo caso, in modo anche di poter dire a priori se una certa distribuzione altrimenti definita o supposta tale possa o meno pensarsi ottenuta con siffatto passaggio al limite.

Un caso limite del tutto opposto, e che non mi consta sia stato mai preso in considerazione, è quello cui darebbe luogo la considerazione di un «ultrafiltro» (45): cioè di un insieme di insiemi tale che (a) di due insiemi complementari, vi appartiene sempre uno si e l'altro no; (b) vi appartiene ogni intersezione d'insiemi che vi appartengono e (c) ogni insieme contenente un altro che vi appartiene.

Dato un ultrafiltro, sull'insieme dei casi elementari C, si può definire una distribuzione di probabilità (estesa a tutti gli eventi) assumendo P(E) = 1 se E (come insieme di casi elementari) appartiene all'ultrafiltro, e altrimenti P(E) = 0. Infatti la P è additiva, perchè in ogni classe completa di eventi incompatibili ve n'è sempre uno e uno solo appartenente all'ultrafiltro.

Nel caso più banale di ultrafiltro, quello costituito da tutti e soli gli insiemi contenenti un dato elemento, si ha così la distribuzione di probabilità concentrata in un unico elemento (caso già considerato poco sopra).

Altri casi non si possono costruire effettivamente (46), ma se ne può provare l'esistenza mediante il teorema di Zorn ossia basandosi sull'assioma della scelta. E in tali casi si avrebbe questa specie di logica in un certo senso intermedia fra la logica del certo e quella del probabile: come nella logica del certo, gli eventi si distinguerebbero in due sole categorie, quelli a probabilità uno o zero, se così si vuol dire «ultraprobabili» e «ultraimprobabili», in luogo di «veri» e «falsi»; a differenza che ivi, la somma logica di più eventi «ultraimprobabili» può però essere «ultraprobabile» (se si tratta di infiniti addendi)!

(45) N. BOURBAKI, **Topologie générale,** Ch. I, "Act. Sci. Ind." n. 858, Hermann, Paris, 1940, pag. 25.
(46) loc. cit. (45).

Proprietà analoghe si avrebbero per funzioni P date da combinazioni lineari di funzioni di tale «tipo-ultrafiltro»: notevole la presenza di quelle che si potrebbero chiamare «probabilità aggiutinate», in quanto è impossibile suddividerle in parti inferiori a un certo limite (nel caso ultrafiltro, addirittura, ad uno), benchè non si tratti di probabilità concentrate (non sono concentrate in nessun «caso elementare») (47).

Bruno de Finetti

(47) Al momento di licenziare le ultime bozze ho potuto prendere visione del volume: GARRETT BIRKHOFF - **Lattice Theory** - "Amer. Math. Soc. Colloquium Publ.", Vol. XXV, N. Y. 1940, in cui un capitolo è dedicato alle applicazioni al calcolo delle probabilità. Il concetto della impostazione assiomatica corrisponde al punto di vista qui sostenuto (additività semplice; additività completa come caso particolare); per il resto, i problemi ivi considerati non sembrano però, almeno a prima vista, aver attinenza con le questioni qui trattate o poste.

Estratto da «*Annali Triestini, sezione 2ª: scienza ed ingegneria*»
XX (1950), pp. 5-22

AGGIUNTA ALLA NOTA SULL'ASSIOMATICA DELLA PROBABILITA'

1. Introduzione.

Nella mia nota "Sull' impostazione assiomatica del calcolo delle probabilità" (1), nonostante il desiderio di toccare tutti gli aspetti della questione, non sarebbe certo stato possibile discutere tutte le diverse sfumature di opinione espresse da numerosi Autori. Tra le molte omissioni inevitabili ve n'è certamente anche alcune che si sarebbero dovute evitare, e tra esse mi riferisco particolarmente ad una che rende incompleta per il lettore la visione del pensiero del Cantelli sull' argomento.

Oltre la nota già citata (2) il Cantelli ne dedicò infatti alla medesima questione un' altra (3) in cui si proponeva di lumeggiare un' altro aspetto: quello della legittimità degli esempi contradicenti l' additività completa considerandoli non più da un punto di vista meramente astratto ma dal punto di vista di una realizzabilità nel senso empirico.

Mi propongo qui di rimediare all' involontaria omissione, sia per chiarire il riferimento al punto di vista del Cantelli, sia per precisare correlativamente il mio punto di vista secondo i nuovi aspetti introdotti.

Voglio infine accennare ad un' altra questione, in senso lato analoga, sollevata dal Borel (4): si tratta di considerazioni, in cui l' interpretazione probabilistica ha in effetti una funzione di "coloritura" più che di sostanza, tendenti a rigettare il tanto discusso principio di Zermelo sulla base del noto paradosso di Hausdorff relativo a sovrapponibilità per rotazione di insiemi

1) Bruno de Finetti, *Sull' impostazione assiomatica del calcolo delle probabilità*. "Annali Triestini", Vol. XIX, 1949, Sez. 2.a.

2) F. P. Cantelli, *Una teoria astratta del Calcolo delle probabilità*, "Giorn. Ist. Ital. Attuari", A. III, n. 2, 1932 (cfr. anche altri articoli del C. e dell' Ottaviani, ibidem, A. X, nn. 1 - 2, 1939) (questa è la cit. (2) nella nota predetta).

3) F. P. Cantelli, *Sulla estensione del principio delle probabilità totali ad una successione illimitata di eventi incompatibili*, "Scritti matematici offerti a Luigi Berzolari", Pavia 1936.

4) E. Borel, *Les paradoxes de l'infini*, Gallimard, Paris, 1946; v. il cap. X e partic. i nn. 83 - 86.

su di una sfera (5); la mia opinione al riguardo collima, come si vedrà, con quella espressa dal Lévy (6) nel discutere le argomentazioni di Borel.

2. - Punto di vista "astratto" e "fisico"

E' opportuno cominciare, preliminarmente, con qualche precisazione in merito alla distinzione fra "punto di vista astratto" e "punto di vista fisico" nel trattare di probabilità: il Cantelli ha sempre tenuto (nelle Note citate e altrove, e particolarmente lo sottolinea in una recente lettera) "a mettere in evidenza tale distinzione, pure occupandosi del Calcolo delle probabilità da entrambi i punti di vista, perchè si tratta di considerazioni di carattere diverso".

E' ben noto infatti come il Cantelli abbia proposto una formulazione astratta del Calcolo delle probabilità atta a rendere accessibili ad ogni matematico i problemi matematici che esso pone, indipendentemente da ogni questione controversa inerente al significato della nozione di probabilità, e non penso nè potrei contestare la legittimità di tale modo di procedere o l' utilità di tale intendimento ai fini suddetti. Di ciò ho fatto cenno, con riferimento alla nota (2), nelle prime righe della (1).

E nulla può vietare, a chi volesse lanciarsi su tale via con la tranquillità di chi fa della matematica pura senza curarsi del significato originario dei termini che vi introduce, una scelta arbitraria di assiomi e convenzioni, e in definitiva la creazione di una teoria che col Calcolo delle probabilità non conservi che tenui somiglianze, come le varie specie di "spazi astratti" (p. es. lo spazio hilbertiano, gli spazi topologici, ecc.) con lo "spazio fisico".

Devo dire però che non mi sembra si sia fatto finora alcunchè di simile, nè da parte del Cantelli nè di altri, e che non credo ciò sarebbe opportuno. La validità dell' opera del Cantelli in tale direzione (e la sua rispondenza allo scopo dichiarato e surriportato) sta proprio nel non scostarsi, nella struttura, da quello che egli chiama "Calcolo delle probabilità nel senso fisico" (o, preferirei dire, "nel senso interpretativo", per non pregiudicare la natura dell' interpretazione che si voglia considerare).

Comunque, per quanto mi riguarda, tengo a precisare che nelle mie considerazioni anche di natura astratta non ho mai inteso di distinguere una "teoria astratta" come tale, come qualcosa di avente un "carattere diverso"

5) F. Hausdorff, *Grundzüge der Mengenlehre;* cfr. anche G. Vitali, *Un risultato di F. Hausdorff e la compressibilità della materia,* Rend. Lincei, 21.6.1931.

6) P. Lévy, *Les paradoxes de l'infini et le calcul des probabilités,* "Bull. Sci. Math.", 2e S., 73, I, 1949; v. anche (in altro senso) G. Bouligand, *Les speculations sur l'infini,* "Revue gén d. Sci.", T. LIV, n. 1, 1947.

dal Calcolo delle probabilità nel senso interpretativo. Tant' è vero che tutte le discussioni sulle questioni assiomatiche sviluppate in (1) si riferiscono al significato effettivo di "probabilità" quale interpretato nelle diverse concezioni (elencate ivi nel n. 7), o in particolare a quella concezione cui aderisco personalmente (probabilità soggettiva).

Non vorrei che fra il dire che c' è o che non c' è una "distinzione" fra teoria astratta e interpretativa si vedesse solo una maggiore o minore inclinazione a sottolineare ciò che vi ha di diverso o di comune fra considerazioni astratte e interpretative. Si tratta di cosa ben precisa. Ammetterei la distinzione di una "teoria astratta" se considerassi una teoria matematica suscettibile di risultare inapplicabile in qualche caso nella interpretazione concreta (come sarebbe il caso se ad es. il "teorema delle probabilità totali", o quello "esteso", fosse ammesso come "assioma" nella teoria astratta mentre si ammettesse che in una certa concezione interpretativa tale proprietà potrebbe sussistere solo approssimativamente, o magari non sussistere affatto). Dicendo che la distinzione di una "teoria astratta" è estranea al mio punto di vista, intendo dire viceversa che tutto quanto dico "in senso astratto" intendo dirlo anche "in senso interpretativo", prescindendo soltanto da questioni contingenti circa i valori numerici delle probabilità che considero. Allo stesso senso che il semplice conteggio "se oggi ho comperato 2 litri di olio a 40 L. al litro ho speso 80 L." ha senso astratto o concreto a seconda che la quantità e il prezzo rispecchiano fatti reali o ipotetici, senza che ciò implichi l' esistenza di due diverse teorie, una astratta a una commerciale, per fondamento di tale conteggio.

Così nel calcolo delle probabilità, per me è in tale senso astratta ogni affermazione sul gioco di testa e croce con una moneta "perfetta", in quanto l' aggettivo "perfetta" esprime sinteticamente l' assunzione che le due facce siano ugualmente probabili e le prove indipendenti; la stessa diviene concreta quando qualcuno gioca con una moneta per la quale (nel senso della concezione interpretativa che egli adotta, e in base a circostanze che essa può suggerire e che egli avrà osservato) ritiene verificate tali condizioni (di equiprobabilità e indipendenza) e pertanto appropriato l'appellativo di "perfetta" che le sintetizza.

In generale: un' affermazione del calcolo delle probabilità avrà un significato astratto solo in quanto posso esimermi dallo specificare se i valori delle probabilità sono quelli che corrispondono a questa o quella concezione della probabilità in senso interpretativo (dato che di concezioni sostenute dai diversi Autori ce n' è parecchie), e se vi corrispondono effettivamente in una concreta applicazione o solo ipoteticamente in un esempio immaginario. Un' affermazione astratta potrà anche non trovare rispondenza reale perchè vi intervengono eventi praticamente indecidibili (p. es. se una distanza è multiplo razionale o irrazionale del metro) o la cui probabilità non abbia senso in una certa concezione, o, secondo altre avendolo, non sia stata valutata; ma si tratterà sempre solo di circostanze contingenti di distinzione (come, nel precedente caso dei prezzi, per un bene privo di una valutazione corrente, come ad es. un oggetto d' arte unico in un periodo nel quale non si discute di una possibile vendita).

Ciò occorreva dire perchè da tale atteggiamento deriva una conseguenza essenziale: che respingo ogni argomentazione del tipo molto frequente "questo andrà bene in teoria ma non in pratica", "in senso astratto ma non in senso fisico", e così via. Tutto ciò che ho scritto mi sarei risparmiato di scriverlo se pensassi che vale solo "in senso astratto"; per me vale sempre, in principio, nel senso "interpretativo", pur senza pretendere con ciò che si debba o si possa misconoscere e sormontare i motivi che rendono in pratica inutile o delicata la considerazione effettiva di casi che possano apparire eccessivamente complicati o artificiosi.

Da ciò deriva una fondamentale diversità di atteggiamento di cui è bene aver così individuato la profonda scaturigine (e si spiega quindi come gli argomenti matematici possano solo in parte far presa nelle discussioni fra sostenitori di opposti punti di vista). Chi sostiene la sostanziale identità di "teoria astratta" e "teoria interpretativa" è assai più circospetto nel fare delle assunzioni gratuite sia pure in sede di quei problemi che altri attribuirebbe all' ambito della teoria astratta (p. es. ad ammettere l' additività completa), ma è disinvolto e tranquillo nell' accettare le più estreme conseguenze (anche nel senso interpretativo e nelle applicazioni) degli assiomi che ha fissato in quanto accertatane la "necessità" (logica !) in nesso alla precisa nozione di probabilità presa in considerazione (sia essa la propria per difenderla, o quella di altri studiata come possibile variante o per controbatterla).

Chi invece sostiene la sostanziale diversità di "teoria astratta" e "teoria interpretativa" è tratto a considerare come questione di mera convenzione la scelta di questo o quell' atteggiamento su molti quesiti considerati "in sede teorica" (7), mentre si riserva di accettarne o meno la validità in

7) Devo tuttavia fare un' eccezione, precisando in quale senso potrebbe sfuggire a tale critica qualche espressione del genere purchè debitamente intesa. La spiegazione risulterà più chiara riferendosi a un esempio concreto.

Consideriamo l' affermazione, che, sotto condizioni che è inutile qui specificare, si ha probabilità "uno" che la frequenza f_n tenda a un valore p per $n \to \infty$. Supponiamo, naturalmente, che si tratti di un problema in cui valga la proprietà seguente: la probabilità che $|f_n - p|$ sia minore di un prefissato $\varepsilon > 0$ per tutti gli n tra m ed $m + q$, diviene piccola quanto si vuole pur di prendere m sufficientemente grande.

L' enunciato sulla frequenza limite può considerarsi:

sede applicativa per lo più in base addirittura a mere personali sfumature di gusto (difatti variabilissime fra Autori pure sostanzialmente aderenti a tale modo di vedere), tradotte in argomentazioni asistematiche presentate ed escogitate "ad hoc" caso per caso come giustificazione a posteriori del fatto di apprezzare qualcosa come conforme o contrario al loro "buon senso", di giudicare che qualcosa "persuade" o "non persuade", che piace o ripugna.

Dal primo punto di vista, tutte le obiezioni del genere non costituiscono che, tutt' al più, difficoltà di mero carattere tecnico o paradossi nel senso etimologico (verità contrarie a ciò che ci si aspetterebbe a prima vista), prive di ogni possibilità di ripercussione sull' impostazione teorica (se non nel senso di chiarirne taluni aspetti).

3. La "concepibilità" e la "logica del press' a poco"

Nel riprendere gli argomenti sviluppati dal Cantelli nella nota citata ad (3), tralascio naturalmente quelli già discussi in (1) in quanto figuravano anche nella nota (2). Rimane pertanto da esaminare:

1) come una conseguenza dell' enunciato in termini finiti e dell' additività completa: in tal senso sarebbe accettabile per chi ammettesse (come io ammetto) la legittimità di parlare della probabilità relativa alla frequenza limite (probabilità di un evento come qualunque altro, che non richiede alcuna nuova definizione e specificazione) ed anche l' additività completa (che invece non trovo argomenti per accettarla);

2) come una definizione della probabilità di convergenza verso p della frequenza-limite: in tal senso sarebbe accettabile per chi ammettesse che per un tale evento la probabilità si possa introdurre ma a patto di nuove appropriate convenzioni, non rientrando il caso in quegli eventi la cui probabilità risulti "definita" da convenzioni precedenti (opinione su cui non posso consentire, dato che la probabilità come nozione non può avere che un significato unico per qualsivoglia tipo di eventi, e posso variare da evento ad evento solo i procedimenti ausiliari per la valutazione pratica, numerica, di tale probabilità);

3) come una frase sintetica per esprimere in linguaggio suggestivo benchè improprio ciò che in modo esatto è precisato nell' enunciato in termini finiti (ossia una proprietà asintotica di probabilità relative a frequenze finite).

Quest' ultima interpretazione corrisponde forse a certe spiegazioni del Lévy e magari anche del Fréchet (mi riferisco anche a conversazioni private), benchè non sia certo di sfumature che possono far pendere tra questa e la precedente (o le precedenti). L' ultima interpretazione sarebbe lecita per chi ammettesse che della probabilità dell' evento "frequenza-limite = p" non abbia mai senso parlare, e quindi la locuzione "probabilità di detto evento", mancando di significato letterale, sia utilizzabile in altro senso convenzionale. Occorrerebbe, naturalmente, tener sempre presente che il senso è puramente convenzionale (così come dire che due curve hanno "contatto tripunto" non significa alla lettera che abbiano in comune tre punti ma indica sinteticamente una proprietà che, come qui, andrebbe espressa a rigore mediante somma e prodotto logico di infinite disuguaglianze).

1) il concetto di "concepibilità" quale norma discriminativa per accettare o respingere esempi, e la fondatezza di un criterio di dimostrazione appoggiato a tale base;

2) le conseguenze di una "indistinguibilità" nei risultati derivante dall' irraggiungibilità di misure esatte;

3) la "verificabilità" di ogni valutazione di probabilità nel senso delle "relazioni fra probabilità e frequenza".

Cominciando dalla "concepibilità", è ben noto da tutti i campi della matematica che tale termine non ha un significato o quanto meno un' estensione univocamente determinati per tutti gli Autori che vi fanno appello. E' questo senz' altro anche il caso del calcolo delle probabilità; ma, quand' anche così non fosse, sarebbe pur sempre necessario, prima di fare uso di tale concetto e soprattutto prima di basarvi delle conclusioni, delimitarlo in modo preciso. Così come un geometra può decidere di limitarsi a considerare (e dire, se vuole, che li considera i soli "concepibili") i metodi di costruzione implicanti la sola riga, o solo riga e compasso, o solo procedimenti algebrici, ecc.

Nel nostro caso, cioè per il calcolo delle probabilità, non mi consta che una precisazione del genere sia mai stata tentata, e, per discuterne, non c' è quindi che tentare di intravvederne la possibile natura in base a un' interpretazione a ritroso di argomentazioni usate frammentariamente. Sembra legittimo desumere, da un tentativo del genere, che la limitazioni cui pensano il Cantelli e gli Autori che ragionano in modo analogo sono di tre tipi: 1) esclusione di esempi in cui gli eventi che si considerano (indipendentemente da ogni questione sulla loro probabilità) hanno definizione "troppo complicata", o in particolare tale da non potersi costruire senza infinite scelte; 2) l' ulteriore difficoltà derivante dall' inesattezza delle misure, e quella 3) inerente alla valutazione delle probabilità. Poichè queste due ultime questioni vanno trattate a sè sotto altri aspetti (le avevamo già elencate come questioni distinte), ci limiteremo qui ad accennare al primo concetto, e a sviluppare le considerazioni generali del caso che, per tale loro carattere, sussistono senza modificazione anche con riferimento a nozioni di "concepibilità" basate sui due ulteriori concetti.

Se ci si allontana dal caso elementare, di un numero finito di casi possibili tutti con probabilità positiva, si possono sempre trovare motivi per rifiutare quel tanto di idealizzazione che inevitabilmente vi interviene. Potremmo benissimo rifiutare l' esempio di "un numero reale scelto a caso tra 0 e 1" (ossia: distribuzione uniforme su di un intervallo), dicendo che in pratica non si può che conoscerne le prime n cifre decimali, per cui si hanno soltanto 10^n

casi possibili. Oppure, se si ammette tale esempio, perchè non ammettere del pari quello identico in cui, come casi possibili, si ammettono solo i numeri r a z i o n a l i tra 0 e 1?

Un esempio più preciso, in quanto riferentesi a un criterio preciso di accettabilità o meno a seconda che una costruzione sia indipendente o dipendente dall' assioma della scelta, quello della suddivisione di un segmento (o, meglio, di una circonferenza) in un' infinità numerabile di insiemi sovrapponibili. E' ben noto come una tale scomposizione si realizzi usando l' assioma della scelta (8); ma uno può dire che respinge tale metodo, e solo se si riuscisse a dare un metodo costruttivo diverso accetterebbe (se gli sembrerà conforme al buon senso (9) immaginare che a insiemi sovrapponibili si possano anche in tal caso attribuire probabilità uguali) che tale esempio provi la possibilità di un' infinità numerabile di eventi escludentisi e ugualmente probabili (10).

Comunque, quali conclusioni potrebbero trarsi dal rifiutare certi tipi di esempi? Sarebbe lecito, volendo fare i rigoristi, considerare privi di senso, o almeno privi d' interesse, i problemi in cui interviene una circostanza del genere, p. es. quelli in cui si porrebbe l' alternativa di poter applicare o meno, ad una infinità di casi, il teorema delle probabilità totali. E nessuno potrebbe obiettare nulla contro chi volesse non occuparsi di enunciati del genere.

Ma se invece ci si vuole occupare di problemi del genere, e quindi la alternativa si pone, quali ne sono i termini? Una delle due alternative consiste in un' affermazione categorica: certamente la proprietà additiva completa sussiste: non solo è lecito considerare problemi del genere, ma, in ogni esempio del genere, dalla conoscenza delle probabilità degli addendi è lecito dedurre che la probabilità della somma logica è data dalla somma della serie (e non

---

8) Sulla circonferenza unitaria (o sul segmento $(0, 2\pi)$, considerando coincidenti ascisse che differiscono di $2k\pi$) assegnamo dapprima agli insiemi $I_0, I_1, I_2, \dots, I_n, \dots$ i punti di ascissa $0, 1, 2, \dots, n, \dots$; fra i punti residui scegliamone uno, x_1, ed attribuiamo ad I_n il punto $x_1 + n$ $(n = 0, 1, 2, \dots)$; scegliamo poi analogamente $x_2 \dots$ e così via transfinitamente fino ad aver esaurito tutti i punti. Ammettendo la bene-ordinabilità del continuo, basterebbe dire: ad ogni passo, scegliere "il primo" dei numeri residui. V. p. es. (per applicazione analoga) G. Vitali e G. Sansone, *Moderna teoria delle funzioni di variabile reale,* Parte I, p. 59.

9) Su tale punto, v. dubbi e precisazioni in seguito, n. 6.

10) A proposito dell' "assioma della scelta", vorrei precisare in che senso ritengo che una dimostrazione basata su di esso dovrebbe avere un certo valore anche per coloro che vi si oppongono. Se un certo teorema T è dimostrato usando l' assioma della scelta, una dimostrazione che T è falso sarebbe anche una dimostrazione che l' assioma della scelta è falso. Quindi è praticamente certo che nessuna dimostrazione del contrario di T potrà esser trovata (se non da chi avesse la ventura, dato e per me non concesso che ciò sia possibile, di trovare una dimostrazione dell' assurdità dell' assioma della scelta).

maggiore). L' altra alternativa non è categorica, afferma molto meno, e precisamente non nega che la probabilità della somma logica possa risultare uguale alla somma della serie (come vuole l' alternativa detta), ma non esclude che possa anche riuscire maggiore.

Ora (a prescindere dagli altri argomenti sviluppati nella nota (1), per i quali la mia opinione è anche più netta), non vedo comunque come si possa considerare ragione sufficiente per rigettare un atteggiamento *dubitativo* e sostenere recisamente un atteggiamento *affermativo* il fatto di non trovare abbastanza sodisfacenti gli esempi addotti (a puro e non indispensabile scopo illustrativo). Accettando di ragionare su questa base, si dovrebbe dire (se è lecito spiegarsi con un paragone paradossale) che fino al momento in cui è stato costruito in modo "sodisfacente" un esempio di numero trascendente era lecito considerare provato che tutti i numeri reali fossero algebrici.

Ed anche qualora le due alternative non presentassero questa dissimmetria, di essere l' una dubitativa e l' altra restrittiva, per cui nel dubbio va ovviamente preferita la prima, si può dire che in generale le argomentazioni tendenti a scoraggiare dalla considerazione di certi problemi non danno una preferenza all' una o all' altra alternativa qualora invece tali problemi si vogliano considerare; il dedurne che una delle alternative va scartata appare pertanto ispirato al preconcetto che tale alternativa sia "la meno naturale".

Ed infine, anche se fosse dimostrato che non è possibile costruire esempi sodisfacenti (nel senso che dovrebbe essere precisato) nell' ambito dei problemi usuali ove i "casi possibili" costituiscono un insieme a un numero finito di dimensioni, come si potrebbe giovarsi di tale circostanza per assumere valida l' additività completa nei casi ove la s' invoca per dimostrare ad es. teoremi sul "limite della frequenza", quando una rappresentazione "naturale" di essi richiederebbe uno spazio a infinite dimensioni?

Queste argomentazioni, che hanno qui necessariamente un carattere un po' vago ed astratto com'è vaga la nozione di "concepibilità" cui si riferiscono, troveranno riscontro perfettamente anche con riferimento alle considerazioni alquanto più precise consentite dai due ulteriori aspetti più precisi che abbiamo nominato e che ora passeremo ad esaminare.

4. Conseguenze della "indistinguibilità"

L' indistinguibilità di eventi derivante dall' impossibilità di misure esatte è a volte addotta come ulteriore motivo di rifiuto per esempi che pur non fossero ritenuti rifiutabili quanto a metodi usati nella costruzione. Così ad es., anche se si supponesse trovato un metodo indipendente dall' assioma della scelta per definire la suddivisione della circonferenza in insiemi $I_0, I_1, I_2, \ldots$ $\ldots, I_n, \ldots$ disgiunti e sovrapponibili quali precedentemente considerati nella nota in calce (7), chi volesse valersi di quest' altro argomento per sollevare obiezioni potrebbe dire che l' esempio non è utilizzabile ai fini del calcolo delle probabilità in quanto siffatti insiemi sono ovunque densi cosicchè ogni punto x può appartenere a uno qualunque di essi se lo si conosce con una esattezza comunque grande ma non perfetta.

Più precisamente, supponiamo che "i nostri mezzi fisici non permetteranno di poter fare più distinzione" (11) tra valori che differiscano tra loro per meno di un certo valore ε; in tal caso, l'evento consistente nel fatto che il valore di un certo numero aleatorio X appartenga ad un insieme I sarà decidibile se il valore assunto, x, *nonchè tutti i punti dell'intervallo* $(x-\varepsilon, x+\varepsilon)$, appartengono ad I, risp. al complementare di I; se invece in detto intervallo cadono tanto punti di I che del suo complementare, l' evento è, coi mezzi supposti, indecidibile. Se, per ogni insieme I e per ogni ε positivo, conveniamo di indicare con $I(+\varepsilon)$ l' insieme di tutti i punti che distano non più di ε da I, e con $I(-\varepsilon)$ l' insieme di tutti i punti che distano non meno di ε dal complementare di I, ragionando in tal modo si dovrebbe dire che della probabilità di I è lecito soltanto affermare che è compresa fra quella (teorica) di $I(-\varepsilon)$ e di $I(+\varepsilon)$; la differenza di tali due probabilità è la probabilità di valori per cui l' appartenenza ad I è indecidibile, e si può addirittura pensare che in tutti questi casi la decisione sia presa a sorte o ad libitum, p. es. sempre in favore di I o sempre nel senso opposto, realizzando così per la probabilità risp. il confine superiore o inferiore predetti.

Facendo tendere ε a zero, la restrizione si attenua ma non scompare: l' appartenza di x ad I sarà decidibile quando esso sia un punto *interno* ad I, o risp. *esterno* ad I, mentre sarà indecidibile se x è un punto della *frontiera* di I (se, cioè, ogni suo intorno per quanto piccolo contiene sia punti di I che del complementare di I). Sarà per i *punti frontiera* che, in tale accezione, corrispondente all' ammissione di disporre di mezzi fisici atti a raggiungere qualunque grado di esattezza si prefissi (errore inferiore a un ε comunque piccolo ma prefissato), sussiste il dubbio per la decisione, che si potrà pensare presa a sorte, o ad libitum. I confini inferiore e superiore per la probabilità sono in tal caso risp. la misura interna ed esterna secondo Jordan-Peano (o l' analogo rispetto ad altra distribuzione non uniforme).

(11) Così si esprime il Cantelli in (3), pag. 446.

Tale conclusione è proprio la stessa cui conduce l'applicazione della additività semplice: essa consente di dare, per la probabilità di I, soltanto questa limitazione ricavabile dal confronto con segmenti in numero finito, e non quelle più precise che richiedono successioni di segmenti e l'applicazione ad essi (in numero infinito) della proprietà additiva. E' l'ammissione dell'additività completa che obbliga invece, ad es., a dire che è nulla la probabilità che un valore scelto a caso in un intervallo sia razionale, mentre in base all'additività semplice nulla si può dire (la probabilità può avere un valore qualunque, tra 0 e 1 inclusi).

Qualunque valore si voglia dare all'argomento, esso può dunque invocarsi se mai per inibire la considerazione di certi problemi o per farli considerare con un certo grado di imprecisione, ma non per criticare esclusivamente o di preferenza le considerazioni che postulano l'additività completa oppure quelle che ritengono viceversa che non sussistano ragioni valide che obblighino a postularla.

Quanto a mia opinione personale, non ritengo che la circostanza della "indistinguibilità" abbia a giocare nel caso della probabilità un ruolo speciale, influendo sull'impostazione concettuale. Se l'impostazione concettuale si riferisce sempre a problemi alquanto idealizzati, è ovvio che anche nel calcolo delle probabilità accada di considerare in teoria eventi la cui decidibilità richiederebbe un'esattezza assoluta (p. es. distinguere se un valore è razionale o irrazionale); se tale esattezza è in pratica irraggiungibile, si interpreteranno in pratica i risultati entro i limiti in cui ciò sembrerà aver senso, allo stesso modo che in ogni altro campo, senza alterare l'impostazione teorica. Tutti sappiamo, in geometria applicata, che due rette quasi parallele hanno in comune "molti punti", ossia un punto determinato assai imprecisamente, ma guai se una siffatta circostanza si volesse introdurre come una premessa nella costruzione degli assiomi per la geometria in senso matematico.

5. La "verificabilità" come "frequenza".

Ultimo argomento, quello riguardante la "verificabilità" della "probabilità" come "frequenza" nei casi in discussione, in cui non dovesse sussistere la proprietà additiva completa. Mi sembra anzi che, nel pensiero di Cantelli (nota cit. (3), esclusi i punti già sviluppati in (2) e discussi nella nota (1) precedente), queste osservazioni occupino la posizione preminente.

Nel discutere tale punto sorvolerò naturalmente tutte le questioni critiche sul significato delle relazioni fra probabilità e frequenza; faccio tale premessa solo per potermi esprimere nella forma usuale magari talvolta in modo un po' semplicistico senza che chi conosce mie precedenti precisazioni ritenga le abbia abbandonate. Anche come terminologia, mi atterrò a quella della nota del Cantelli parlando di eventi di cui si possono fare ripetute "prove" (mentre nel mio modo di esprimere ogni "prova" sarebbe un "evento"). In ciò non v'è nulla che nuoccia, agli effetti che interessano.

Abbiansi dunque gli eventi $E_1, E_2, \dots, E_h, \dots$, costituenti un'infinità numerabile e completa di eventi incompatibili (di modo cioè che in ogni prova deve verificarsene uno e uno solo); le prove siano indipendenti. Siano le loro probabilità $p_1, p_2, \dots, p_h, \dots$; prescindendo dal postulare l'additività completa si hanno tre casi possibili: 1) solo un numero finito delle p_h sono positive, e la loro somma vale 1 (le altre, come la loro somma, sono nulle), 2) le p_h positive sono infinite, e la somma della serie vale 1 (le altre eventualità, anche complessivamente, hanno probabilità nulla), 3) la somma della serie delle p_h è minore di 1 (e non importa distinguere se le p_h sono tutte nulle, o se v e n'è un numero finito, o un'infinità, o tutte meno un numero finito, o tutte, che sono invece positive); se si ammette l'additività completa possono sussistere soltanto le prime due ipotesi e la terza cade: si tratta appunto di vedere per quali ragioni questo terzo caso dovrebbe rigettarsi in base alle considerazioni che discutiamo.

Supponiamo ora di fare delle "prove", diciamo N prove; otterremo, per i nostri eventi, delle frequenze $f_1, f_2, \dots, f_h, \dots$; per la legge dei grandi numeri ciascuna delle f_h sarà quasi certamente vicinissima a p_h purchè N sia sufficientemente grande. In base a ciò - questo è il punto da discutere - si potrà escludere per le p_h un andamento che non possa essere realizzato per le f_h.

Cosa possiamo dire sul comportamento delle frequenze? ovviamente che sono multipli di $1/N$, sono tutte nulle tranne al più N, e la loro somma vale 1. E' chiaro che tali conclusioni non si possono trasportare tali e quali alle p_h (ciò sarebbe assurdo già per il fatto che N è arbitrario); ma si potrà escludere che ve ne sia più di un numero finito positive? e se si ammette che la somma si possa trasformare in una serie, si dovrà tuttavia conservare la conclusione che la somma debba dare l'unità? e perchè?

Evidentemente un ragionamento per superficiale analogia fra due casi tanto radicalmente diversi come la somma di un numero finito di termini e la somma di una serie non ha alcun valore. E a parte ciò, se qualche conclusione si vuol trarre non si può fare a meno di esaminare un po' da vicino fino a qual punto si possa pretendere che probabilità e frequenza praticamente coincidano.

Già dal punto di vista meramente aritmetico, dato che la frequenza non varia che per multipli di $1/N$, è ovvio che essa non può riprodurre con buona approssimazione la probabilità che se $1/N$ è piccolo rispetto alla probabilità stessa, ossia N grande rispetto ad $1/p$; dal punto di vista probabilistico, poichè lo scarto tra probabilità e frequenza ha valore quadratico medio $\sqrt{pq/N}$ ($\simeq \sqrt{p/N}$, se p è piccolo, come qui interessa), volendo che esso sia piccolo rispetto a p (per darne una valutazione sodisfacente) occorrerà che N sia tanto grande rispetto a $1/p$ che anche $\sqrt{Np}$ risulti grande.

Ma ciò è praticamente impossibile già nel caso banale di un numero finito di casi possibili, purchè molto numeroso: ogni tentativo del genere di verificare che le 40 ! permutazioni di un mazzo di 40 carte da gioco hanno uguale probabilità ($= 1/40!$), dato che il numero N delle prove non può che essere piccolissimo in confronto a 40 !, farebbe concludere (a dimenticare che era assurdo pretendere un "accordo" migliore) che la equiprobabilità va rigettata, e tutti i casi sono impossibili (probabilità nulla) tranne N che hanno probabilità molto più grande ($1/N$).

In questo caso (e in ogni caso con un numero finito di eventualità) si può rimediare (almeno idealmente) prendendo N tanto grande da render praticamente certo che nessuna delle frequenze si scosti percentualmente dalla rispettiva frequenza per più di un prefissato margine di tolleranza; ma se si ha un' infinità di casi con probabilità positiva, e quindi necessariamente tendente a zero, non si può trovare un N per cui t u t t a la successione delle frequenze rappresenti con buona approssimazione t u t t a la successione delle probabilità. L' impossibilità di verifica è una circostanza intrinseca ineliminabile, e non può esser sfruttata per negare valore ad uno "schema" del genere (e del resto, semmai, dovrebbe inficiare tanto il caso in cui l' additività completa sussiste che quello in cui non sussiste); occorre ammettere (nè vedo come ciò possa non essere conforme al "buon senso" di qualunque cultore di calcolo delle probabilità secondo qualunque tendenza) che per nessun N determinato si può trasportare al comportamento globale delle p_h una conclusione tratta dall' esame del comportamento globale delle f_h; soltanto per i casi E_h aventi una probabilità p_h sufficientemente grande in confronto ad $1/N$ può attendersi (dopo N prove) una sufficiente concordanza tra probabilità e frequenza (e ciò non conta affatto con riguardo alle proprietà asintotiche, implicanti tutte le p_h).

Si dirà: sta bene; ma facciamo crescere N, e considerando N crescente indefinitamente si giungerà alla conclusione voluta. Ma ciò non accade: il ragionamento che si pretendeva fare per le frequenze su un numero N grande quanto si voglia ma determinato di prove, sfuma e si dissolve variando N. Nulla vieta infatti p. es. che tutte le f_h (h fisso, N tendente a infinito)

tendano a zero; ciò avviene senz' altro, ad es., se si suppone che nessuno dei casi possibili E_h si verifichi più di una volta (o più di un numero finito di volte) in una successione indefinita di prove (e non vedo come tale possibilità di comportamento non debba essere accolta quale conforme al "buon senso" di chicchessia).

Anche il criterio della frequenza, secondo le considerazioni svolte, sembra quindi atto, se mai, ad escludere ogni problema in cui figurano probabilità piccole (non "verificabili" con un numero di prove materialmente effettuabili), ma non, qualora non si accolga siffatto divieto, a creare discriminazioni a favore dell' una o dell' altra alternativa, pro o contro l' additività completa (12).

6. Il paradosso di Hausdorff

Il paradosso di Hausdorff ha un significato puramente geometrico: è possibile ripartire la superficie di una sfera in tre insiemi A, B, C, disgiunti; mediante rotazione, ciascuno di essi può esser portato a sovrapporsi a qualunque altro, oppure alla somma degli altri due. Da ciò scende che sulla superficie sferica la condizione che una funzione d' insieme risulti invariante per rotazione è incompatibile con l' additività anche semplice (mentre il risultato classico del Vitali per il caso più semplice del segmento, o meglio circonferenza, è che ivi l' incompatibilità sussiste solo con l'additività completa).

Tale risultato è paradossale nel senso in cui lo sono molte proprietà indubitabili relative ad insiemi infiniti che a prima vista appaiono sorpren-

12) Nello stesso modo si potrebbe cbiarire anche dal punto di viasta dell' interpretazione basata sulle frequenze la validità delle argomentazioni svolte in (1), n. 26, a proposito del paradosso del Lévy. Aggiunto tale osservazione perchè non si ritenga (come è avvenuto) che quelle argomentazioni avessero soltanto l' intenzione di superare il paradosso da un punto di vista "astratto"; conformemente al punto di vista generale esposto nel n. 2, esse vanno intese in ogni senso interpretativo.

Poichè il paradosso deriva (come rilevato nel punto cit.) dall' apparente "evidenza" della proprietà *conglomerativa* (cfr. (1), nn. 26, 27, 28, 30, 31), mostriamo come si possa costruire un esempio in cui essa non valga per la "frequenza-limite". Dalla successione dei numeri naturali si estragga una prima sottosuccessione prendendo il primo numero composto e alternativamente uno sì e uno no dei numeri primi; poi una seconda sottosuccessione operando la stessa scelta sui numeri residui, e così via. Si ottiene una scomposizione dell' insieme dei numeri naturali in una infinità numerabile di sottoinsiemi numerabili: la frequenza-limite dei numeri primi vale "uno" in ciascun sottoinsieme, e tuttavia vale "zero" nella successione totale. Ammettendo una valutazione di probabilità basata su tale frequenzadenti; ciò non può che condurre, come dice il Lévy (13), a pensare che "le continu à deux dimensions est plus complexe encore que nous le pensions" e a ripetersi con Pascal che "notre imagination se lassera plus tôt de concevoir que la nature de fournir" (14). Volerne trarre un argomento contro l'assioma della scelta (o contro le proprietà geometriche delle rotazioni, o contro un qualunque altro punto ci fosse antipatico nell'insieme dei concetti impiegati per la dimostrazione) non è che un segno di forte ripugnanza verso il risultato (così all'epoca dell'introduzione degli irrazionali un assertore della loro "inconcepibilità" avrebbe potuto rifiutare l'argomento della diagonale del quadrato dicendo che esso anzichè dimostrarne l'irrazionalità confuta il pregiudizio che un numero non possa essere simultaneamente pari e dispari).

Ma cosa c'entra la probabilità? C'entra, secondo Borel, per sostenere, che una funzione additiva invariante per rotazione d e v e esistere, perchè tale è la probabilità che un punto scelto a caso sulla sfera appartenga al dato insieme. Egli ritiene cioè che la frase "scegliere un punto a caso" abbia essa stessa un significato tanto intuitivo e preciso da servire di per sè a garantire l'esistenza e univocità di una estensione "naturale" a tutti gli insiemi di una nozione di "misura"; il Lévy ci informa che il Borel ha sempre nutrito tale convincimento per riguardo alla stessa nozione di "misura", e ciò può far pensare che il riferimento alla probabilità sia piuttosto un travisamento di questa convinzione originaria che un convincimento proprio della sua concezione della probabilità. E' caratteristico della mentalità del Borel questo credere nella esistenza di enti matematici "naturali", indipendentemente da più o meno artificiose "definizioni", ed è proprio a tale fede che sono dovuti i brillanti risultati che raggiunse in tanti campi, a cominciare da quello delle funzioni analitiche ove appunto si pose, fin dalla sua tesi (1894), la questione di dare un significato "naturale" al prolungamento

limite, per essa si avrebbe: probabilità di estrarre un numero primo, nulla ($p = 0$); tuttavia, subordinatamente a ciascuna delle ipotesi possibili (di estrazione di un numero appartenente alla prima, seconda, . . . , n-esima, . . . sottosuccessione), probabilità uno di estrarre un numero primo.

Osservo del resto che tale problema, dal punto di vista della teoria di von Mises, era stato già completamente studiato dal Wald in *Die Widerspruchsfreiheit des Kollektivbegriffes* (cit. sub. (17) nella (1), senza però rilevare tale parte del contenuto, sfuggitami perchè non avevo recentemente riletto quel lavoro).

13) Nella sua recensione del libro di Borel, cit. (4), immediatamente precedente alla nota cit. (6).

14) Quasi una parafrasi, del resto, della celebre battuta dell'Amleto di Shakespeare.

di una funzione (15). Certo la fede è utile perchè indica una via e incita a seguirla, e può darsi che corrisponda a un'intuizione verace; non è però sufficiente ad assicurare tale veracità, che solo la scoperta di dimostrazioni e la conseguente introduzione di definizioni, più o meno artificiose che appaiano, assicura. E può darsi che uno scienziato, per quanto geniale, insieme ad una fiducia in intuizioni valide ne senta altrettanta in altre o non valide, o almeno richiedenti qualche ritocco.

Comunque, ciò che importa in questioni del genere è l'occasione che esse danno di chiarire sotto aspetti particolari la situazione psicologica di ciascuno degli autori di fronte a date argomentazioni: è con tale osservazione che il Lévy giustifica il suo intervento in opposizione alle idee del Borel, e a maggior ragione devo richiamarmi a detta osservazione per giustificarmi nell'interloquire nel dialogo fra di essi.

Secondo me, la frase "scegliere a caso" non significa nulla, o meglio, può significare tutto quel che si vuole. Può addirittura avere solo il senso di "in modo non predeterminato", e cioè semplicemente "scegliere" senza riferimento a probabilità, oppure con "distribuzione di probabilità qualsivoglia"; nel senso in cui qui rammentata la locuzione significa invece "secondo una distribuzione di probabilità che appaia particolarmente naturale", ma, o ciò si lascia nel vago, oppure si precisano in postulati le proprietà volute (p. es. additività completa, invarianza per traslazione, estendibilità ad ogni insieme, o a un dato corpo d'insiemi, ecc.), e allora è una questione di analisi il vedere se distribuzioni sodisfacenti alla condizioni desiderate esistono o no, e nel caso affermativo se ce n'è una univocamente determinata o parecchie fra cui la frase "scelto a caso" intesa come si suppone precisato non fornisce criteri discriminativi. La risposta dell'analisi potrà apparire più o meno strana, oppure strana a qualcuno e ovvia ed altri, ma ciò nulla sposta.

Supposto di riferirci alla distribuzione in cui, per eventi corrispondenti a segmenti (o rettangoli, prismi, . . . nel caso di due, tre, . . . dimensioni) la probabilità è proporzionale alla misura (lunghezza, area, volume, . . .), non penso affatto che debba esisterne un prolungamento privilegiato, "naturale", a tutti gli insiemi, o a parte di essi. Il solo prolungamento obbligato è quello agli insiemi misurabili secondo Jordan-Peano (in quanto per essi la valutazione della probabilità discende dall'additività semplice, che è fuori

15) E. B o r e l, *Sur quelques points de la théorie des fonctions*, (Thèse) - Annales de l'Ecole Normale, 3^{e} S., t. 12, 1895, riprodotta in *SELECTA* (Jubilé scientifique de M. E. B.), Gauthier - Villars, Paris, 1940, pp. 3-48; il criterio informatore qui accennato è chiaramente ed esplicitamente esposto nella "Introduzione" (prime tre pagine della memoria).

discussione); tuttavia sono disposto ad attenuare ancora tale conclusione dicendo che la probabilità ha un valore praticamente significativo solo per quegli insiemi I per cui il rapporto fra le misure di $I(+\varepsilon)$ ed $I(-\varepsilon)$ non supera $1+\theta$ (ove ε e θ sono i margini d'intederminazione risp. per le distanze e per le probabilità, ed $I(\pm\varepsilon)$ ha lo stesso significato che nel n. 4), nel qual caso vanno esclusi certi insiemi misurabili secondo J.-P. (p. es. quello ottenuto dividendo l'intervallo in un numero finito di parti uguali minori di ε, e prendendone una sì e una no) e compresi altri non misurabili, magari neppure secondo Lebesgue (p. es. un segmento abbastanza grande più un insieme non misurabile contenuto in altro segmento abbastanza piccolo).

Il fatto che l'estensione al corpo di Borel sodisfacente all'additività completa sia univoca, e sodisfi anche l'invarianza per traslazione, dice che un modo possibile di valutare le probabilità per tali insiemi consiste nell'assumere la probabilità uguale alla misura di Lebesgue; non dice però che questo sia il solo modo possibile (anche se quelle proprietà possono apparire, per certi problemi e per certi individui, ragioni di preferenza). Nulla vieta ad es. che la probabilità dei razionali sia valutata ad 1 anzichè a 0: basta ad es. pensare che i casi possibili siano le frazioni $m/2^n$ con m dispari minore di 2^n, e che la probabilità di ogni insieme I sia data dal limite per $n \to \infty$ di $f(n)/2^{n+1}$ ove $f(n) =$ numero dei termini con denominatore $\leq 2^n$ appartenenti ad I; questa distribuzione coincide con quella uniforme per ogni intervallo (e, quindi, per ogni insieme misurabile J.-P.) eppure ammette solo valori razionali. Più ancora interessa notare: non dico soltanto che la probabilità d'insiemi non misurabili J.-P. può essere indifferentemente valutata o in base alla misura di Lebesgue o anche differentemente, ma che si può far benissimo a meno di valutarla se ciò non interessa o se non ci si vuole pronunciare in merito.

Allo stesso modo passando alla totalità degli insiemi (anche non misurabili-L.): non dico che esista un modo (privilegiato) di estendere la misura, e la probabilità, ad essi (come vorrebbe il Borel); non posso pretendere a priori che ciò possa farsi rispettando l'additività completa, o l'invarianza per traslazione, o altre proprietà che potrei desiderare sodisfatte (sarà l'analisi a dirmi che la prima condizione è impossibile, la seconda possibile qui ma non più - come s'è visto - passando sulla sfera, ecc.; e potrò essere più o meno stupito che sia così, ma non rimane che prenderne atto). Non dico neppure che si debba, in qualche modo, fare l'estensione; nessuno è obbligato a considerare dei problemi che non lo interessano. Dico però che chi voglia attribuire una probabilità ad ogni insieme deve avere il diritto di farlo, in un qualunque modo sodisfacente l'additività semplice (e sarà libero di far sì che valga l'additività completa su certi corpi d'insiemi, o l'invarianza per traslazione, o che altro vorrà, ammesso che le proprietà da lui desiderate non siano impossibili o incompatibili) (16).

In altre parole, il fatto che in certi problemi si possano considerare delle distribuzioni di probabilità particolarmente interessanti, come quella che ad n casi attribuisce uguale probabilità o, nel caso di lunghezze aree volumi, attribuisce probabilità uguali a lunghezze (aree, volumi) uguali, e il fatto che ciò suggerisca di usare la frase "scelta a caso" per indicare che si valutano la probabilità in tale modo, non giustifica una interpretazione, per così dire mitica, di tale frase, che faccia assurgere quasi la scelta" a caso" a nozione intrinsecamente significativa, dalla quale l'esistenza e le prorpietà della corrispondente distribuzione (in ogni possibile campo) potrebbero venir dedotte come logicamente necessarie.

Questo punto di vista mi sembra non in contrasto con quello del Lévy, il quale si limita all'aspetto negativo (lasciando da parte la facoltà di valutare ad arbitrio le probabilità il cui valore non è obbligato, facoltà troppo ispirata alla concezione strettamente soggettiva per pensare di trovarla ivi). Egli ammette cioè che l'estensione probabilità = misura - L. non abbia senso univocamente obbligato se non quando l'insieme considerato abbia frontiera di misura nulla (per dire secondo la nostra precedente conclusione, corrispondente alle considerazioni del Lévy sul procedimento di determinare progressivamente un numero mediante estrazione delle successive cifre decimali).

7. Conclusione

In argomenti del genere, ove non si tratta di deduzioni di date proprietà da date ammissioni, ma di scelta di metodi di impostare e interpre-

16) Allo stesso punto di vista sembra ispirarsi L. J. Savage: riporto il seguente passaggio dal testo (per ora ettografato) del corso che sta svolgendo all'Università di Chicago.

"However convenient countable additivity may be, it, like any other assumption, ought not be listed among the postulates for a concept of personal probability unless it can be agreed that its violation deserves to be called inconsistent. " . . . " "It therefore seems better not to assume countable additivity outright as a postualate, but to recognize it as a special hypothesis yielding a very large class of useful theorems".

A quanto riferisce il Savage, anche B. O. Koopman si è espresso nel medesimo senso, citandone i tre lavori sotto indicati; ciò non mi stupisce data la sostanziale concordanza delle mie vedute con quelle del K. nei lavori che ho letto, dove non ricordo tuttavia fosse discussa l'additività completa.

B. O. Koopman, *The Axioms and Algebra of Intuitive Probability*, "Annals of Mathematics", 41 (1940); *Intuitive Probability and Sequences*, ibid., 42 (1941); *The Bases of Probability*, "Bull. Amer. Math. Soc.", 46 (1940).

tare una teoria, non si tratta tanto di giudicare quale punto di vista sia "giusto" o "sbagliato", quanto di cercar di chiarire i moventi di ciascun punto di vista, di penetrarne l' intima consistenza, di rendersi conto della influenza che l' adozione d' un punto di vista o d' un altro porta su tutto il campo delle possibili applicazioni, sia di natura astratta che di natura pratica.

In simili considerazioni è ovvio che abbia una grande parte il cosiddetto "buon senso", ma anch' esso a sua volta è relativo al punto di vista d' ogni autore, e, per di più, ha anche peculiarità del tutto personali, che non permettono di trarne alcunchè di certamente accetto per tutti. Anche nella nota citata, il Lévy, pur professando di andar quasi sempre d' accordo con Borel in fatto di probabilità, dice a un certo punto: "Mais cette fois mon bon sens refuse de s' accorder avec le sien".

Dato ciò, spero non di aver confutato opinioni altrui e avvalorato le mie proprie, ma di aver chiarito le reciproche posizioni, spiegato i motivi per cui non trovo di poter accettare certi modi di vedere, e quelli per cui altri sono invece necessari in accordo con la visione che sostengo della teoria delle probabilità (17).

Bruno de Finetti

17) Scrivendomi dopo aver letto il manoscritto della presente nota, il Cantelli dà delle precisazioni che non mi sembra modifichino l'interpretazione qui data del suo punto di vista, nè il mio punto di vista rispetto ad esso. Egli si ripromette comunque di ritornare sull'argomento, e sarà certo meglio per il lettore ricorrere a tale espressione diretta del suo pensiero piuttosto che a quella che avessi voluto più estesamente sviluppare parafrasando il contenuto di affrettati scambi epistolari.