Le variabili casuali

Conferenza tenuta al Seminario Matematico dell' Università di Torino l' 11 maggio 1931

1. Il matematico che si accinge per la prima volta allo studio del calcolo delle probabilità deve, ancor oggi, superare una sensazione di disagio più viva di quello che solitamente si prova intraprendendo lo studio di un nuovo capitolo della Scienza non ancora pervenuto ad un completo assetto logico. Invero, nel calcolo delle probabilità, assai più che in altri campi, alcuni ragionamenti e i relativi risultati danno un' impressione di soggettività che non può certo riuscir gradita a chi, dalla propria educazione mentale, è portato sopratutto ad apprezzare l' obiettività dei risultati ed il loro carattere di necessarie conseguenze di alcune premesse.

Si potrebbe anzi credere, non riflettendo a quanto dirò da qui a un momento, che questa certa soggettività del calcolo delle probabilità sia una sua inevitabile caratteristica: retaggio delle difficoltà che notoriamente si incontrano quando si cerca di dare una definizione generale del concetto di *probabilità* di un evento; concetto che, secondo alcuni [1], verso la cui opinione io propendo, sarebbe di sua stessa natura essenzialmente soggettivo.

Fortunatamente però il calcolo delle probabilità, o almeno la parte più essenzialmente matematica di esso, può svolgersi indipendentemente da ogni discussione più o meno filosofica sul concetto di probabilità; così come, per esempio, la geometria

[1] V. p. es. De Finetti: *Fondamenti logici del ragionamento probabilistico*, [« Boll. Unione Mat. It., » 9 (1930)].

proiettiva può svolgersi indipendentemente da ogni considerazione sul significato fisico degli enti che essa chiama « rette », « punti », ecc., e la trigonometria può svolgersi senza preoccuparsi dei metodi con cui poi la geodesia o l'astronomia determineranno effettivamente quei lati e quegli angoli che figurano come dati nei suoi problemi. E invero, concependo il calcolo delle probabilità come la scienza delle relazioni fra più ***probabilità*** le une date e le altre da determinarsi, le prime potranno astrattamente considerarsi come certi numeri reali, compresi fra 0 ed 1, ***associati*** a certi avvenimenti, con la sola condizione dal punto di vista matematico che, nel combinarli fra loro, vengano rispettati i noti principî delle probabilità composte e delle probabilità totali. Nessun inconveniente vi sarà allora a concepire, ad esempio, tali probabilità come ***grado di fiducia*** di un determinato individuo nell'avverarsi di certi avvenimenti; salvo il fatto che, se tali dati non avranno proprio altra giustificazione all'infuori di quella di tradurre numericamente certe opinioni di un certo sig. Tizio, sarà giustificato il dubbio ***a priori*** sull'interesse pratico che potranno avere i risultati ottenuti a partire da siffatte basi!

2. Alla suaccennata concezione del calcolo delle probabilità e alla importante chiarificazione concettuale che ne deriva, ha, in questi ultimi tempi, potentemente contribuito il concetto di ***variabile casuale*** che, nonostante la sua grande semplicità e spontaneità, solo di recente si è affermato ed è stato generalmente adottato. Invero, definendo una variabile casuale (continua o discreta) come una variabile reale tale che a ciascuno dei valori x (in numero finito o infinito) di cui essa è suscettibile, sia ***associato*** un numero reale p_x, compreso fra 0 ed 1, con la condizione che sia $\Sigma p_x = 1$; il calcolo delle probabilità potrà definirsi come la ***teoria delle variabili casuali*** o, più particolarmente, come quella parte delle matematiche che insegna a dedurre dalle ***leggi di probabilità*** (cioè dalle leggi di distribuzione dei p_x) date ***a priori*** di certe variabili casuali, quelle di altre variabili casuali dipendenti dalle prime, o a quelle comunque collegate.

Non è però inopportuno, prima di procedere oltre, soffermarci un momento ad illustrare con qualche esempio questo concetto, tanto importante, di variabile casuale.

Anzitutto un esempio classico di variabile casuale suscettibile di un numero finito di valori: il «*punto*» che si può fare gettando un dado cubico. Evidentemente abbiamo qui una variabile casuale suscettibile soltanto dei sei valori 1, 2, 3, 4, 5 e 6 a ciascuno dei quali è *conveniente* associare, se il dado è perfetto o lo si ritiene tale, la probabilità costante $p = \frac{1}{6}$. Altro esempio di un genere un po' diverso: il «*punto*» che una certa squadra potrà fare in un determinato incontro di *football*. Questa è una variabile casuale suscettibile dei tre valori 0, 1, 2, con probabilità che i «competenti» sapranno spesso assegnare senza grandi incertezze, pur non potendo disconoscere il carattere soggettivo della loro previsione.

Molto più interessanti sono però le variabili casuali *continue* per le quali, nel caso più semplice di una sola dimensione, la *legge di probabilità* è rappresentata da una funzione reale, non negativa, $p(x)$, definita nell'intervallo (a, b) in cui oscilla la variabile, e soddisfacente alla condizione

$$\int_a^b p(x)dx = 1. \tag{1}$$

In tal caso la probabilità che la variabile sia compresa in un certo intervallo (α, β) è espressa dall'integrale

$$\int_\alpha^\beta p(x)dx \tag{2}$$

e, in particolare, quella che essa sia compresa fra x ed $x + dx$, è espressa (a meno d'infinitesimi d'ordine superiore rispetto a dx) dal prodotto

$$p(x)dx.$$

Per esempio, appartengono a questa categoria le cosiddette variabili casuali *normali*, che s' incontrano in molteplici, vitali questioni di calcolo delle probabilità, aventi per legge di probabilità la *legge di* GAUSS:

$$p(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}, \tag{3}$$

dove h è una costante positiva: la cosiddetta *precisione*. La curva rappresentatrice della (3) ha il noto aspetto *a campana*, come si vede nella figura 1ª (relativa al caso di $h=1$).

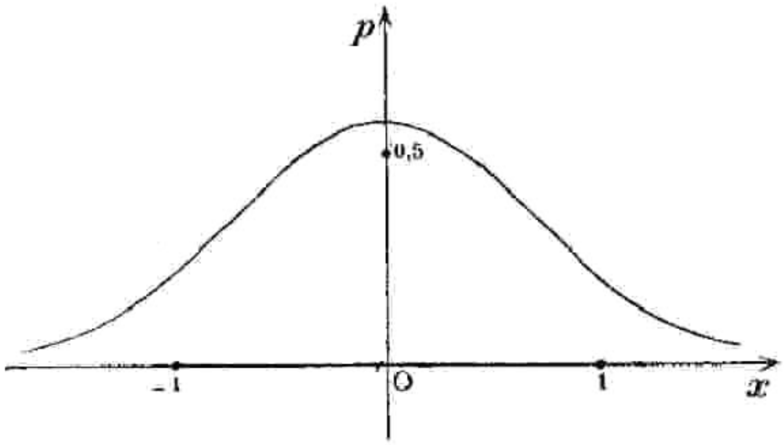

Fig. 1.

Un altro esempio di variabile casuale avente legge di probabilità suscettibile di rappresentazione analitica semplice, è offerto dalla somma

$$y = x_1 + x_2 + x_3$$

di tre termini su cui non si sappia altro all'infuori che sono compresi nell'intervallo (0, 1), cioè, più precisamente, dalla somma di tre variabili casuali oscillanti in (0, 1) con la legge di probabilità costante $p=1$. Invero si trova facilmente, o direttamente o come caso particolare di un problema più generale cui si accennerà più innanzi, che la curva di probabilità di y, variabile casuale oscillante nell'intervallo (0, 3), è formata da tre archi raccordati di parabola ad asse verticale (fig. 2), rappresentati (in coordinate y, p) dalle equazioni:

$$p = \frac{1}{2}y^2 \ (0 \leq y \leq 1), \quad p = \frac{3}{4} - \frac{1}{4}(2y-3)^2 \ (1 \leq y \leq 2),$$

$$p = \frac{1}{2}(3-y)^2 \ (2 \leq y \leq 3).$$

Un esempio semplice di variabile casuale di tutt'altro genere, cioè di carattere statistico, è offerto dall'età x a cui morirà un individuo generico di un certo gruppo di popolazione. Rappresentando graficamente la funzione $p(x)$ si ottiene una curva del genere di quella della fig. 3 e cioè con un minimo principale intorno all'età di 12 anni, un minimo secondario fra i 30 e i 40 anni, ed un massimo molto accentuato intorno all'età di 85 anni.

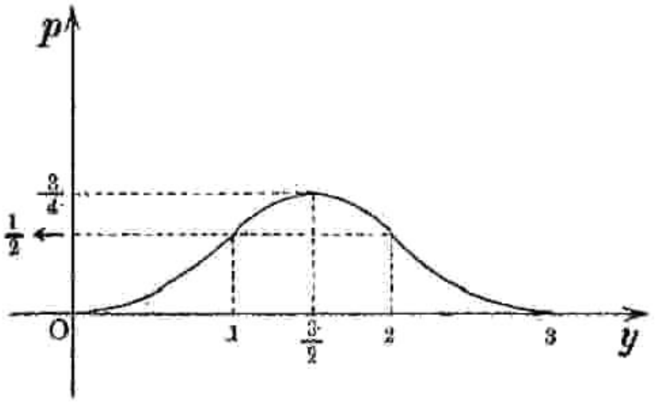

Fig. 2.

Altri semplici esempi di variabili casuali di tipo statistico, sono offerti dall'onere che un istituto di assicurazioni assume assicurando per una certa somma un individuo di una certa età, dalla statura o dal perimetro toracico dei coscritti di un certo paese, ecc.

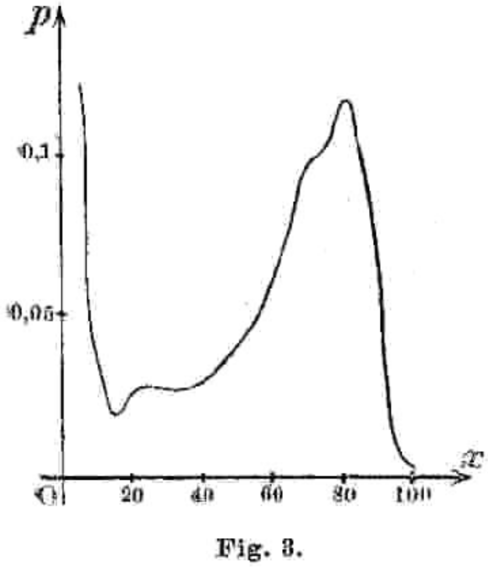

Fig. 3.

Notiamo finalmente, perchè la cosa conduce a non spregevoli semplificazioni formali, che qualsiasi variabile casuale può sempre immaginarsi definita fra $-\infty$ e $+\infty$, pensando che la sua curva di probabilità abbia ordinate sempre nulle in quegli eventuali intervalli a cui la variabile non può mai appartenere.

3. Una delle più comode e più usate rappresentazioni della legge di probabilità di una variabile casuale x, è la rappresentazione geometrica della funzione $p(x)$, di cui si è fatto uso negli esempi precedenti. È però pure molto utile, e concettualmente interessante, la rappresentazione ***statica*** della legge di probabilità, ottenuta pensando ad una distribuzione di materia sull'asse x con ***densità lineare*** misurata da $p(x)$. Invero, la probabilità

$$\int_{\alpha}^{\beta} p(x)dx$$

che la variabile sia compresa fra α e β, potrà allora interpetrarsi come la *massa* del segmento α, β, ecc.... Inoltre si ha il vantaggio che tale rappresentazione si presta anche bene nel caso delle variabili casuali discrete, per cui, invece di una massa uniformemente distribuita nell'asse x, si avranno più masse p_1, p_2, ... concentrate in certi punti x_1, x_2, ...

Per di più la rappresentazione statica suggerisce spontaneamente di considerare, accanto alla funzione $p(x)$, la funzione non decrescente

$$P(x) = \int_{-\infty}^{x} p(x) dx, \tag{4}$$

esprimente la massa a sinistra di x e la probabilità che la variabile sia minore di x, che prende il nome di *funzione di ripartizione* e gode della proprietà che il suo incremento finito $P(\beta) - P(\alpha)$ esprime la probabilità che la variabile che si considera cada nell'intervallo (α, β).

Anzi vi è spesso vantaggio a sostituire addirittura alla considerazione della funzione $p(x)$ quella della funzione $P(x)$, perchè, così facendo, il concetto di variabile casuale assume una maggiore generalità giungendo a comprendere casi in cui la probabilità che x sia compresa nell'intervallo (α, β) subisce delle discontinuità di prima specie, e quindi esiste bensì la funzione $P(x)$ ma non così la $p(x)$, in quanto la prima non è dapertutto derivabile.

Molto ovvia è l'estensione di queste e delle precedenti considerazioni alle variabili casuali *a due o più dimensioni*; noi non ci soffermeremo perciò a lungo su ciò. Per esempio, una variabile casuale *a due dimensioni*, sarà per noi un punto P mobile nel piano (xy), cui sia associata una funzione sommabile, non negativa $p(x, y)$ il cui integrale doppio esteso a tutto il piano valga 1, mentre lo stesso integrale esteso invece ad un'area σ esprimerà, per definizione, la *probabilità* che il punto P cada nell'area σ. In particolare, il punto di caduta su di un piano orizzontale del proietto sparato da una certa bocca da fuoco, è una variabile casuale a due dimensioni la cui legge di probabilità, supposto scelti convenientemente gli assi cartesiani, può ritenersi espressa da una formula del tipo,

$$p(x, y) = \frac{hk}{\pi} e^{-h^2x^2 - k^2y^2},$$

dove h e k sono due costanti, legate in modo semplice alle cosidette *strisce del 50°/₀ dei colpi* della balistica. Trattasi cioè di una variabile casuale *normale* a due dimensioni.

4. Il calcolo delle probabilità, il cui interesse applicativo è ormai fuori discussione, presenta il più vivo interesse anche pel più puro matematico. Invero, oltre alle questioni concettuali spesso assai difficili che presentano alcuni suoi capitoli, p. es. la teoria degli errori, anche il problema concettualmente elementare e apparentemente facile di dedurre la legge di probabilità di una variabile casuale y legata ad altre date $x_1, x_2, \dots, x_n$ da una data relazione funzionale

$$y = \varphi(x_1, x_2, \dots, x_n), \tag{5}$$

può presentare difficoltà tecniche elevate e richiedere l'impiego di mezzi analitici non elementari. Per esempio, già nel caso semplice in cui la funzione φ si riduca alla somma di x_1, $x_2, \dots, x_n$ e queste variabili oscillino tutte in un medesimo intervallo (a, b), avendo ivi *legge di probabilità costante:*

$$p = \frac{1}{b-a};$$

la determinazione della densità di probabilità $p(x)$ di y non è per niente immediata, nè il risultato finale è elementarissimo. Precisamente, come può immediatamente trarsi dalle formule contenute in un mio lavoro di tre anni fa [1], detto s il massimo intero contenuto in $\dfrac{y-na}{b-a}$, si ha

$$p(y) = \frac{1}{b-a} \, \frac{1}{(n-1)!} \, \overset{n-1}{\underset{s}{\mathfrak{B}}}\left(\frac{y-na}{b-a} - s\right), \quad (na \leq y \leq nb) \tag{6}$$

dove $\overset{n-1}{\underset{s}{\mathfrak{B}}}$ denota una funzione (polinomio di grado $n-1$) che si presenta nella teoria dei numeri di Bernoulli ed ha l'e-

---

[1] *Una quistione di probabilità*, (Atti 1° Congresso Naz. di Scienza delle Assicurazioni. Torino, 1928); v. pure la Nota: *Sul numero delle partizioni di un intero dato*, [Boll. Unione Mat. It., 7 (1928)] e la Memoria: *Su di una variabile casuale connessa con un notevole tipo di partizioni di un numero intero*, [Gior. Ist. Ital. Attuari, 2 (1931)].

spressione seguente

$$\overset{n-1}{\underset{s}{\mathcal{B}}}(x) = \sum_{r=0}^{s} (-1)^r \binom{n}{r} (x + s - r)^{n-1} .$$

Un caso particolare della (6) ($n = 3$, $a = 0$, $b = 1$) ci ha fornito uno degli esempi di variabile casuale indicati nel § 2.

Le difficoltà suaccennate possono talvolta evitarsi rinunciando a determinare esplicitamente la legge di probabilità della variabile casuale x che interessa, e accontentandosi invece di calcolare alcuni numeri, a quella collegati, che siano atti a dare un'idea sintetica dell'andamento generale della funzione $p(x)$, e possano determinarsi senza bisogno di conoscere effettivamente l'espressione esplicita di questa funzione.

Fra questi *numeri* due hanno particolarissima importanza e cioè: 1°) il *valor medio* $M[x]$ della variabile, definito nel caso discontinuo dalla formula

$$M[x] = \Sigma x_j p_j \tag{7}$$

e in quello continuo dall'altra

$$M[x] = \int_{-\infty}^{+\infty} x p(x) dx , \tag{7'}$$

che dà un'idea del punto intorno a cui più si addensano i valori della variabile essendo l'ascissa del *baricentro* della sua massa di probabilità; 2°) lo *scarto quadratico medio* $\mu[x]$, che dà un'idea dell'ampiezza delle oscillazioni di x intorno al suo valor medio, e si definisce per mezzo della formula

$$(\mu[x])^2 = \int_{-\infty}^{+\infty} (x - M[x])^2 p(x) dx , \tag{8}$$

cioè come radice quadrata aritmetica del valor medio del quadrato dello *scarto* $x - M[x]$ della variabile. Entrambi questi elementi possono talvolta determinarsi senza conoscere esplicitamente la funzione $p(x)$; in particolare, se x è una combinazione lineare di più variabili casuali $x_1, x_2, \ldots, x_n$ *indipendenti*, cioè se si ha

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n ,$$

dove $\lambda_1, \lambda_2, \ldots, \lambda_n$ sono dei coeffcienti costanti e $x_1, x_2, \ldots, x_n$ n variabili casuali tali che la eventuale conoscenza del valore di una o più di esse non ha influenza sulle leggi di probabilità delle rimanenti, si hanno le formule importanti:

$$(9)\qquad \begin{cases} M[x] = \lambda_1 M[x_1] + \lambda_2 M[x_2] + \ldots + \lambda_n M[x_n], \\ \mu^2[x] = \lambda_1^2 \mu^2[x_1] + \lambda_2^2 \mu^2[x_2] + \ldots + \lambda_n^2 \mu^2[x_n]. \end{cases}$$

Un altro risultato importante dello stesso genere è che *il valor medio del prodotto di più variabili casuali indipendenti è uguale al prodotto dei valor medi dei singoli fattori.*

Nel caso particolare che x sia una variabile *normale*, cioè seguente la legge di probabilità di GAUSS (3), si hanno poi le formule

$$(10)\qquad M[x] = 0, \quad \mu[x] = \sqrt{M[x^2]} = \frac{1}{\sqrt{2h}}$$

e inoltre si trova che il valor medio del valor assoluto dello scarto $M[|x|]$ è uguale all'inversa di $h\sqrt{\pi}$, ciò che dà luogo alla formula interessante:

$$(11)\qquad \frac{M[x^2]}{(M[|x|])^2} = \frac{\pi}{2},$$

che viene spesso utilizzata per verificare numericamente se una data variabile casuale possa o no ritenersi normale.

Notiamo finalmente che le formule (7') e (8) o, più generalmente, la formula

$$(12)\qquad \int_{-\infty}^{+\infty} \varphi(x) p(x) dx$$

che definisce il valor medio di una funzione qualsiasi φ di x, possono facilmente generalizzarsi al caso in cui, non esistendo la funzione $p(x)$, esiste però la funzione di ripartizione $P(x)$; invero basterà considerare, p. es. in luogo della (12), la formula

$$(12')\qquad \int_{-\infty}^{+\infty} \varphi(x) dP(x),$$

dove l'integrale è preso nel senso di STIELTJES. Quest'osservazione valga anche per il seguito.

5. Fra le funzioni $\varphi(x)$ di cui la (12) definisce il valor medio, hanno prevalente importanza le successive potenze intere positive della variabile data, i cui valor medi soglion chiamarsi, per ovvio motivo, *momenti* dei vari ordini della variabile causale x, e denotarsi con le lettere m_1, m_2,...; si pone cioè, in generale,

$$(13) \qquad m_r = \int_{-\infty}^{+\infty} x^r p(x)dx, \qquad (r = 1, 2, ...),$$

aggiungendo, per simmetria, la notazione

$$m_0 = \int_{-\infty}^{+\infty} p(x)dx = 1.$$

Il momento del prim'ordine m_1 coincide ovviamente col valor medio $M[x]$ della variabile.

La considerazione di questi momenti dei vari ordini dà luogo ad una questione assai importante, non solo pel calcolo delle probabilità, ma anche dal punto di vista puramente analitico e cioè: Data una funzione $p(x)$, integrabile fra $-\infty$ e $+\infty$ assieme ai suoi prodotti per le successive potenze di x, restano univocamente determinati i suoi momenti di tutti gli ordini m_0, m_1, m_2,...; *vale la reciproca?* Data cioè una successione di numeri reali m_0, m_1, m_2,..., esiste sempre ed è unica una funzione non negativa $p(x)$ che ammetta quei numeri come momenti?

Questa non facile questione, che va sotto il nome di *problema dei momenti*, ha dato luogo a numerose ricerche di cui si trova una buona esposizione nel noto trattato di CASTELNUOVO [1]. Il risultato forse più importante ottenuto in questo ordine d'idee è che *la condizione necessaria e sufficiente affinchè il problema dei momenti ammetta almeno una soluzione, è che gli infiniti determinanti:*

$$\Delta_0 = m_0, \quad \Delta_1 = \begin{vmatrix} m_0 & m_1 \\ m_1 & m_2 \end{vmatrix}, \quad \Delta_2 = \begin{vmatrix} m_0 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{vmatrix}, \ldots$$

---

(1) *Calcolo delle probabilità* (2ª Ed., Bologna, Zanichelli, 1925-28), vol. II, pagg. 154-188 (« Appendice », art. II, nn. 5-16); v. pure l'articolo del medesimo A. nel « Giornale dell'Istituto Ital. Attuari », I, fasc. 2° (1930) e BOREL: *Traité du Calcul des Probabilitées etc.*, t. 1, fasc. 1, chap. VI. (Paris, Gauthier-Villars, 1925).

siano tutti positivi (Teor. di HAMBURGER). Inoltre la soluzione è unica se si verifica una certa condizione necessaria e sufficiente supplementare, ch'è certo soddisfatta se, al crescere di n, il rapporto

$$\frac{\sqrt[2n]{m_{2n}}}{n}$$

si mantiene finito (CARLEMAN).

In particolare la condizione di unicità sussiste pei momenti di una variabile casuale normale, che, supposto per semplicità $h=1$, hanno le espressioni seguenti:

$$(14) \quad m_0 = 1; \quad m_{2k} = \frac{1.3.5\ldots(2k-1)}{2^k}, \; m_{2k+1}=0, \; (k=0,1,2,\ldots),$$

ciò che, unitamente ad altre considerazioni, conduce al seguente teorema assai importante per la teoria degli errori:

Se i successivi momenti di una variabile $\mathrm{x_n}$, *dipendente da un indice* n, *tendono, al crescere di* n, *verso i corrispondenti momenti spettanti alla variabile normale, cioè ai valori dati dalle (14); allora la legge di probabilità di* $\mathrm{x_n}$ *tende uniformemente a quella normale; nel senso che, detta* $\mathrm{P_n(x)}$ *la funzione di ripartizione di* $\mathrm{x_n}$ *e* $\mathrm{T(x)}$ *quella della variabile normale, dato un numero positivo* ε, *piccolo a piacere, è sempre possibile trovare un intero* $\mathrm{n_0}$ *tale che per ogni* $\mathrm{n > n_0}$ *e qualunque sia* x, *si abbia*

$$|P_n(x) - T(x)| < \varepsilon.$$

6. In alcune questioni di calcolo delle probabilità, si è rilevato assai utile sostituire alla considerazione della funzione $p(x)$ quella di una sua trasformata di LAPLACE, e cioè della funzione analitica

$$(15) \qquad \varphi(t) = \int_{-\infty}^{+\infty} e^{ixt} p(x)\,dx,$$

che dicesi (con POINCARÉ) *funzione caratteristica* della variabile casuale considerata [1].

---

[1] L'introduzione, molto opportuna, dell'immaginario nella (15) è dovuta al LÉWY. [Cfr. il suo bel *Calcul des Probabilitées*. (Paris, Gauthier-Villars, 1925), 2ème Partie, Ch. II].

Dalla (15), posto

$$\varphi(t) = \varphi_1(t) + i\varphi_2(t)$$

segue subito che

$$(16) \qquad \begin{cases} \varphi_1(t) = \displaystyle\int_{-\infty}^{+\infty} p(x) \cos tx \, dx \\ \varphi_2(t) = \displaystyle\int_{-\infty}^{+\infty} p(x) \sin tx \, dx ; \end{cases}$$

inoltre, dal teorema d'inversione di FOURIER, si ha la formula inversa:

$$(17) \qquad p(x) = \frac{1}{2\pi} \int_{-\infty}^{*+\infty} e^{-ixt} \varphi(t) dt ,$$

che serve a calcolare la $p(x)$ nota la funzione caratteristica e in cui l'asterisco sta a denotare che dell'integrale improprio bisogna considerare il *valor principale nel senso di* CAUCHY, cioè bisogna prendere il limite per $z \to \infty$ dell'integrale esteso fra $-z$ e $+z$. Questa piccola complicazione può però evitarsi separando il reale dall'immaginario, con che si ha la formula

$$(17') \qquad p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [\varphi_1(t) \cos tx + \varphi_2(t) \sin tx] dt ,$$

dove l'integrale è preso nel senso ordinario.

Per esempio, nel caso della variabile normale, si ha

$$(18) \qquad \varphi(t) = e^{-\frac{t^2}{4h^2}} .$$

L'importanza della considerazione della funzione caratteristica apparisce subito riprendendo in esame la somma di più variabili casuali indipendenti. Invero, mentre la determinazione diretta della legge di probabilità di una sifatta somma, presenta, come abbiamo visto nel § 4, non lievi difficoltà; se invece ci si riferisce alle funzioni caratteristiche, si ha il teorema semplicissimo che:

La funzione caratteristica della somma di più variabili casuali indipendenti, è il prodotto delle funzioni caratteristiche dei singoli addendi.

Donde, in virtù della (18), segue immediatamente che:

La somma di più variabili casuali normali di precisioni $h_1, h_2, \ldots, h_n$, *è una nuova variabile casuale normale la cui precisione* h *è data dalla formula:*

$$\frac{1}{h^2} = \frac{1}{h_1^2} + \frac{1}{h_2^2} + \ldots + \frac{1}{h_n^2} \tag{19}$$

Quest'ultimo risultato suol talora enunciarsi dicendo che *la legge di* GAUSS *è stabile.* Non bisogna però dimenticare che anche altre leggi di probabilità, p. es. la cosiddetta «*legge di* CAUCHY», godono di un'analoga proprietà.

Sussiste inoltre un *teorema di convergenza,* analoga a quello enunciato pei momenti, ossia si ha che:

Affinchè una legge di probabilità dipendente da un parametro (p. es. da un indice n*) tenda alla legge di* GAUSS, *è necessario e sufficiente che il logaritmo della sua funzione caratteristica tenda a* $-t^2/4h^2$, *e ciò uniformemente in ogni intervallo finito.*

7. Una semplicissima proprietà delle variabili casuali ci condurrà subito al teorema forse più interessante, certo il più celebre del calcolo delle probabilità: la cosiddetta «*legge dei grandi numeri*» o teorema di BERNOULLI. Intendiamo alludere al fatto che, con un semplicissimo spezzamento in parti dell'integrale (o della somma) con cui si definisce lo scarto quadratico medio μ di una variabile casuale x, si dimostra immediatamente che *la probabilità che il valore di* x *sia compreso nell'intervallo* $(M[x] - \alpha\mu, M[x] + \alpha\mu)$, *dove* α *è un qualsiasi numero reale maggiore di* 1, *è certo non inferiore ad* $1 - 1/\alpha^2$. (Teorema di BIENAYMÉ-TCHEBYCHEFF).

Invero, appoggiandosi sulla precedente disuguaglianza e sulla non meno semplice osservazione che la frequenza f con cui un certo evento di probabilità costante p (p. es. sorteggio di una pallina bianca da un'urna di composizione costante) si è presentato in una serie di n prove (cioè il rapporto ν/n, dove ν è il numero delle volte in cui l'evento si è verificato), è una variabile casuale ad n valori con valor medio p e scarto quadratico medio $\sqrt{p(1-p)/n}$; si vede subito che la probabilità della disuguaglianza

$$|f - p| \leq \varepsilon,$$

essendo ε un numero positivo arbitrariamente prefissato, è certo non inferiore a

$$1 - \left[\frac{\varepsilon}{\sqrt{p(1-p)/n}}\right]^{-2} = 1 - \frac{p(1-p)}{n\varepsilon^2},$$

e quindi tende ad 1 per $n \to \infty$. Si ha cioè che:

Se f *è la frequenza con cui un certo avvenimento di probabilità costante* p *si è presentato in* n *prove, la probabilità che il valor assoluto della differenza* f — p *sia minore di un numero positivo* ε, *arbitrariamente prefissato, tende alla certezza al crescere di* n. (Teor. di BERNOULLI *o legge dei grandi numeri*).

Introducendo la nozione, dovuta a CANTELLI [1] di *tendenza al limite nel senso del calcolo delle probabilità*, stabilendo cioè la definizione seguente:

Diremo che la variabile casuale x_n, *dipendente da un indice n, tende al limite* a *per* $n \to \infty$ *«nel senso del calcolo delle probabilità» (o, come anche si dice, «stocasticamente») e scriveremo:*

$$\operatorname*{Lim}_{n \to \infty} x_n = a, \qquad \text{(con « L » maiuscola)}$$

allorchè, dato un numero positivo ε *arbitrariamente piccolo, la probabilità della disuguaglianza*

$$|x_n - a| \leq \varepsilon,$$

tende all'unità al crescere di n;

la legge dei grandi numeri può, più semplicemente ed espressivamente, enunciarsi nel modo seguente:

La frequenza in n *prove di un evento di probabilità costante* p, *tende stocasticamente a* p *per* $n \to \infty$.

Molte delle proprietà dei limiti ordinari valgono anche pei limiti stocastici.

Notiamo che il teorema di BERNOULLI ha ricevute diverse estensioni di cui la più notevole è, a mio giudizio, la dimostrazione, dovuta al CANTELLI [2], che, non solo la probabilità

---

(1) *La tendenza ad un limite nel senso del calcolo delle probabilità*, [« Rend. Circ. Matem. », Palermo, 61 (1916)].

(2) *Sulla probabilità come limite della frequenza*, [Rend. Lincei (5), 26¹ (1° sem. 1917), pagg. 39-45].

della disuguaglianza $|f^{(n)} - p| \leq \varepsilon$, dove $f^{(n)}$ è la frequenza in n prove tende, per $n \to \infty$, alla certezza; ma che *lo stesso succede anche per la probabilità che detta frequenza tenda, nel senso ordinario della parola, alla probabilità* p, *oscillando infinite volte intorno ad essa,* ciò che costituisce la cosiddetta *« legge forte dei grandi numeri ».*

Molto recentemente è stato poi dimostrato, dal KINTCHINE [1] e ritrovato, indipendentemente da lui, dal LÉVY [2], che, detto ν il numero delle volte in cui, in n prove, l'evento di probabilità costante p si è verificato, *la differenza* $|\nu - np|$ *è, stocasticamente parlando, asintoticamente uguale a*

$$\sqrt{2p(1-p)n \log \log n}\,;$$

nel senso che la probabilità d'infinite realizzazioni della disuguaglianza

$$\nu - np > c\sqrt{2p(1-p)n \log \log n}\,, \tag{20}$$

è zero se $c > 1$ *e uno se* $c < 1$.

8. Lo schema delle prove ripetute cui si riferisce il teorema di BERNOULLI, ci offre un'altra questione d'interesse non minore di quella di cui ci siamo occupati nel § precedente, e cioè; *qual'è la legge di probabilità della differenza* $\nu - pn$? e, secondariamente: *questa legge, tende essa ad un limite* (in senso da precisarsi opportunamente) *al crescere indefinito di* n?

Alla prima di queste due domande si risponde assai agevolmente; propriamente, con considerazioni del tutto elementari, si vede subito che la differenza $\nu - np$ è una variabile casuale (ad n valori), di valor medio nullo e scarto quadratico medio uguale a $\sqrt{np(1-p)}$, tale che la probabilità di un certo suo valore $\bar{\nu} - np$ è data da

$$p_{\bar{\nu}} = \binom{n}{\bar{\nu}} p^{\bar{\nu}} (1-p)^{n-\bar{\nu}}. \tag{21}$$

---

(1) *Ein Satz der Wahrscheinlichkeitsrechnung,* [« Fundamenta Math. », 6 (1924), pagg. 9-20].

(2) *Sulla legge forte dei grandi numeri,* [« Giorn. Ist. It. Attuari », 2 (1931), pagg. 1-21].

Meno facile ma anche più interessante è invece rispondere alla seconda domanda, e cioè determinare la distribuzione asintotica, per $n \to \infty$, delle probabilità (21).

A tale scopo bisogna sostituire al coefficiente binomiale che figura nella (21) la sua espressione mediante fattoriali, e a questi le loro espressioni fornite dalla nota formula di Stirling; si giunge così, con un calcolo un po' lungo ma privo di difficoltà concettuali, al risultato non meno semplice che importante, che la probabilità di uno scarto $\bar{\nu} - np = x$ è *asintoticamente* uguale a

$$\frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-\frac{x^2}{2np(1-p)}},$$

cioè che *detto scarto tende a comportarsi come una variabile normale di precisione*

$$h = \frac{1}{\sqrt{2np(1-p)}}. \tag{22}$$

Più esattamente si ha il seguente teorema il cui enunciato precisa il senso da attribuirsi alla frase «tende a comportarsi» di cui sopra:

Se, in una serie di n *prove, un certo evento ha probabilità costante* p *e si pone* $\mu^{(n)} = \sqrt{\mathrm{np}(1-\mathrm{p})}$, *la probabilità che esso si verifichi un numero di volte compreso fra* $\mathrm{np} + \alpha\mu^{(n)}$ *e* $\mathrm{np} + \beta\mu^{(n)}$, *dove* α *e* β *sono due numeri reali qualsiasi, indipendenti da* n; *tende nel senso ordinario della parola, al limite*

$$\frac{1}{\sqrt{\pi}}\int_\alpha^\beta e^{-x^2}dx$$

per n *crescente indefinitamente.*

Questa proposizione, la cui importanza teorica e pratica è superfluo soffermarsi ad illustrare, suole citarsi sotto il nome di *primo teorema limite del calcolo delle probabilità*, e basterebbe da sola a porre in una posizione specialissima la legge di probabilità di Gauss, la cui comparsa in questo problema delle prove ripetute deve interpretarsi come una proprietà asintotica dei coefficienti binomiali. Propriamente essa è da mettere in relazione col fatto che una spezzata i cui vertici rappresentino le probabilità (21), a condizione di ridurre convenientemente la scala delle ordinate e delle ascisse, tende, al crescere di n, alla «curva degli errori», cioè al diagramma della legge di GAUSS.

9. Ma perchè, a proposito, il diagramma della legge di GAUSS suol chiamarsi la «curva degli errori»?

La ragione di questa denominazione risiede in una proprietà della legge di GAUSS, anche più importante di quella posta in luce nel § precedente; proprietà di cui tutti hanno sentito più o meno vagamente parlare, ma di cui non molti hanno penetrato lo spirito. Intendo alludere al fatto che *gli errori accidentali d'osservazione si distribuiscono sensibilmente secondo la legge di* GAUSS, ossia, in altre parole, che *l'entità* x *dell'errore in una determinata osservazione, può considerarsi come una variabile casuale normale, cioè seguente la legge di probabilità* (3), *dove* h *è una costante conveniente* (*« precisione »* della serie d'osservazioni considerata).

Qual'è però la portata precisa di quest'asserzione? In particolare: siamo di fronte ad un fatto di carattere sperimentale, oppure la circostanza che gli errori seguono la legge di GAUSS è matematicamente deducibile da ipotesi più semplici?

Storicamente parlando non c'è dubbio che la «legge degli errori» ha avuta un'origine teorica; però, essendo generalmente noto che il ragionamento con cui GAUSS la dedusse dal cosiddetto *postulato della media* (anche a prescindere dalle non infondate critiche rivolte a questo stesso postulato) dà luogo ad obiezioni che ne infirmano il valore; così molti, quasi per reazione, sono ancor'oggi propensi a credere che la legge di GAUSS non abbia che un valore sperimentale, al punto da essere indotti a chiedersi se non sia il caso di cercare empiricamente qualche altra formula che meglio rappresenti i dati sperimentali. E dico «ancor'oggi» perchè, se una simile concezione poteva giustificarsi qualche tempo fà, essa non mi sembra più difendibile al giorno d'oggi in cui, un gruppo di ricerche oltremodo interessanti e profonde, iniziatesi con TCHEBYCHEFF e continuate con LIAPOUNOFF, CANTELLI, LÉROY, ecc., ha condotto ad una rigorosa *dimostrazione* della legge di GAUSS, a partire da ipotesi accettabilissime e di alto grado di generalità.

L'interesse concettuale di queste nuove ricerche risiede specialmente nel fatto che esse risalgono ad una concezione dell'origine stessa dell'errore, la cui prima idea può già riscontrarsi in LAPLACE, secondo cui l'errore osservato x sarebbe la somma di un grandissimo numero di *errori elementari*, dovuti a cause ignote o mal note, indipendenti e tali che ciascuna di esse, se fosse sola ad agire, produrrebbe un effetto di un ordine di grandezza assai minore di quello di x. In altre parole: l'errore x si considera come la somma di un grandissimo numero di variabili casuali indipendenti $x_1, x_2, \ldots, x_n$ sulle cui singole leggi di probabilità poco o nulla si sa, all'infuori dell'accennata relazione fra gli ordini di grandezza delle x_k e quello di x. Ora la cosa estremamente interessante e, a prima vista, piuttosto sorprendente, è che, sotto ipotesi estremamente poco restrittive sulle ignote leggi di probabilità delle x_k, si può dimostrare rigorosamente che, per $n \to \infty$, *la variabile somma* x *tende asintoticamente a divenire una variabile normale*, cioè a seguire la legge di GAUSS, ciò che costituisce quel che suol chiamarsi il *secondo teorema limite del calcolo delle probabilità.*

Più esattamente, per citare soltanto uno dei risultati più semplici ed espressivi ottenuti in questo ordine d'idee, sussiste il seguente teorema (di LINDEBERG):

Se $x_1, x_2, \ldots, x_n$ *sono* n *variabili casuali indipendenti di valor medio nullo, assumenti valori aritmeticamente non superiori ad un certo* l, *e se si indica con* P(x) *la funzione di ripartizione della variabile casuale somma:*

$$x = x_1 + x_2 + \ldots + x_n ,$$

e con μ *il suo scarto quadratico medio; fissato ad arbitrio un numero positivo* ε *è sempre possibile determinare un altro numero positivo* η *tale che dalla disuguaglianza*

$$\frac{l}{\mu} < \eta$$

segua l'altra

$$\left| P(x) - \frac{1}{\sqrt{2\pi}\mu} \int_{-\infty}^{x} e^{-\frac{x^2}{2\mu^2}} dx \right| < \varepsilon. \tag{23}$$

In altre parola: *se il rapporto* $1/\mu$ *è sufficientemente piccolo, la* x *è sensibilmente una variabile normale di precisione* $h = 1/\sqrt{2}\mu$.

Quanto ai metodi con cui si perviene a questo e altri consimili risultati, non possiamo che rimandare specialmente ai già citati trattati del CASTELNUOVO e del LÉVY. In essi ha, com'è da aspettarsi, una parte notevole la considerazione della funzione caratteristica e i teoremi di convergenza cui si è parlato in fine dei §§ 5 e 6.

Nel giudicare della portata del secondo teorema limite per la teoria degli errori, non deve però perdersi di vista che esso vale solo sotto certe ipotesi, sia pure molto poco restrittive, sulle leggi di probabilità degli errori elementari che non è detto debbano esser *sempre* soddisfatte. Per esempio, se questi seguissero la cosiddetta *legge di* CAUCHY, cioè fossero delle variabili casuali governate dalla legge di probabilità

$$p(x) = \frac{1}{\pi}\frac{a}{a^2 + x^2}, \tag{24}$$

dove a è una costante, anche l'errore totale seguirebbe, com'è stato posto in chiara luce dal LÉVY, la legge di CAUCHY invece che quella di GAUSS.

Giova però avvertire che se uno strumento di misura desse errori distribuiti secondo la legge di CAUCHY, o altre consimili leggi *eccezionali* (come le chiama il LÉVY), accadrebbe il fatto paradossale che, ripetendo n volte una stessa misura, la *precisione* della media aritmetica dei risultati, invece di crescere al crescere di n, resterebbe stazionaria o, peggio, diminuirebbe! In tali condizioni, l'unica cosa sensata sarebbe perciò di gettare tra i ferri vecchi lo strumento adoperato.

10. Posto in luce, coi, due teoremi limiti, lo speciale carattere e la speciale importanza della legge di probabilità di GAUSS, torniamo ora alle variabili casuali generiche per riprendere brevemente in esame il problema fondamentale, cui si è già accennato nel § 4, di determinare la legge di probabilità, e per essa la funzione di ripartizione $Q(y)$, di una variabile casuale y, *funzione* di più variabili casuali $x_1, x_2, \dots, x_n$ di date funzioni di ripartizione $P_1(x_1)$, $P_2(x_2), \dots, P_n(x_n)$, cioè legata a questa da una relazione *deterministica* del tipo

$$y = \varphi(x_1, x_2, \dots, x_n),$$

in virtù della quale il valore di y resti *determinato* quando sono fissati quelli di $x_1, x_2, \dots, x_n$.

Tale problema, tecnicamente difficile in generale, diviene però molto facile per $n = 1$ e specialmente poi se la relazione fra y e x è *biunivoca*, cioè se la funzione $\varphi(x)$ è sempre crescente o sempre decrescente. Invero, basta osservare che, se x_0 e x sono due valori qualunque di x e y_0 e y i corrispondenti valori di y, la condizione necessaria e sufficiente affinchè y sia compresa y_0 e y è che x sia compresa fra x_0 e x, per dedurne che

$$Q(y) - Q(y_0) = \pm [P(x) - P(x_0)],$$

dove il segno $+$ si riferisce al caso di φ *crescente* e il segno $-$ a quella di φ *decresente.* Facciamo ora tendere x_0 a $-\infty$, col che $P(x_0)$ tenderà a *zero* mentre $Q(y_0)$ tenderà a *zero* o ad *uno* secondochè φ è *crescente* o decrescente; avremo così la formula importante

$$(25) \qquad Q(y) = \begin{cases} P(x) & (\text{se } \varphi \text{ è } \textit{crescente}) \\ 1 - P(x) & (\text{se } \varphi \text{ è } \textit{decrescente}), \end{cases}$$

donde, se esistono le densità di probabilità $p(x)$ e $q(y)$, segue subito che

$$(26) \qquad q(x) = \frac{p(x)}{|\varphi'(x)|}.$$

Se invece la relazione fra x e y non è biunivoca, allora bisogna tener conto del fatto che un determinato valore di y può provenire da diversi valori di x e applicare il principio delle probabilità totali. Così p. es., nel caso che sia

$$y = x^2,$$

si trova subito che è

$$Q(y) = \begin{cases} 0 & (y < 0) \\ 1 + P(\sqrt{y}) - P(-\sqrt{y}) & (y > 0) \end{cases}, \quad q(y) = \begin{cases} 0 & (y > 0) \\ \dfrac{p(\sqrt{y}) + p(-\sqrt{y})}{2\sqrt{y}} & (y > 0). \end{cases}$$

Nel caso generale l'unica cosa facile a stabilirsi, ma non molto significativa, è un'espressione di $q(y)$ sotto forma di integrale multiplo; precisamente, dal principio delle probabilità composte segue subito che dev'essere:

$$(27) \qquad q(y)dy = \iint_{D_y} \cdots \int p(x_1)p(x_2) \ldots p(x_n) dx_1 dx_2 \ldots dx_n,$$

dove l'integrale n — plo del secondo membro è esteso al dominio D_y dello spazio $(x_1, x_2, \dots, x_n)$ definito dalla condizione:

$$y \leq \varphi(x_1, x_2, \dots, x_n) \leq y + dy. \tag{28}$$

11. Le relazioni fra variabili casuali considerate nel § precedente non sono però le più generali possibili e nemmeno le più interessanti. Invero, rimanendo per semplicità nel caso di due sole variabili casuali x e y, si può benissimo concepire una relazione fra le loro leggi di probabilità $p(x)$ e $q(y)$, indipendentemente da ogni relazione *deterministica* fra i loro valori, indipendentemente cioè del fatto che esista una funzione φ tale che fissato un valore di x resti in conseguenza determinato un valore $\varphi(x)$ di y.

Per esempio, la maggior parte delle coppie di variabili casuali *dipendenti* forniteci dalla statistica, non sono tali in quanto esista (o si conosca) una relazione deterministica fra i loro valori, ma in quanto la conoscenza del valore dell'una, o dei più probabili valori di questa, ha influenza sulla legge di probabilità dell'altra. Così, in particolare, le stature individuali in un gruppo di padri e nel gruppo dei loro figli, sono due variabili casuali ovviamente non indipendenti, non in quanto esista una formula che, data la statura del padre permetta di calcolare quella del figlio, ma in quanto si presume (e l'osservazione lo conferma) che da padri d'alta statura nasceranno prevalentemente figli d'alta statura e viceversa.

S'induce da quanto sopra che il tipo più generale di dipendenza fra due (o più) variabili casuali, sarà la dipendenza *funzionale* fra le loro leggi di probabilità globalmente considerate, dipendenza che solo in casi particolari degenererà in dipendenza *punto da punto o deterministica* che dir si voglia. Perveniamo così alla seguente definizione molto generale di dipendenza fra due variabili casuali:

Diremo che la variabile casuale y *dipende dalla variabile casuale* x, *quando la curva di probabilità della prima cioè, per fissare le idee, il diagramma della sua densità di probabilità* q(y), *è funzione, nel senso di* VOLTERRA, *della curva di probabilità* p(x) *di* x*: cioè, in simboli, quando si ha:*

$$[q] = \Phi([p]).$$

Questo concetto di dipendenza mi sembra, nonostante la sua grande semplicità, di notevole importanza, specialmente in relazione al cosiddetto ***principio di indeterminazione*** (di HEISENBERG) della fisica odierna. Invero è noto come ragioni profonde hanno indotto di recente ad abbandonare, nello studio del microcosmo atomico, come sperimentalmente e concettualmente impossibile la ricerca di relazioni di tipo deterministico fra le varie grandezze che si considerano, p. es. fra le coordinate di un elettrone in un istante t_0 e quelle in un istante successivo t_1. Tali grandezze vengono invece concepite come altrettante variabili casuali le cui leggi di probabilità sono in parte date e in parte da determinarsi tenendo conto dei legami funzionali fra di esse stabiliti dal fenomeno fisico che si studia. Pertanto, considerando specialmente il caso schematico in cui le grandezze in gioco siano soltanto due: x e y, il fenomeno fisico non si cercherà più di tradurre matematicamente, perchè ciò è impossibile, in una relazione ***deterministica*** del tipo

$$y = \varphi(x) ;$$

bensì in una relazione ***funzionale*** del tipo

$$[q] = \Phi([p])$$

fra le leggi di probabilità delle due variabili casuali x e y.

In altre parole l'ente matematico corrispondente ad una ***legge fisica*** nel microcosmo, non è, in generale, una funzione ordinaria φ, bensì un ***funzionale*** Φ o, come altri dicono, una ***funzione di linea***.

F. TRICOMI