SULLE PROBABILITÀ NUMERABILI E GEOMETRICHE.

In: « *Rendiconti del R. Istituto Lombardo di Scienze e Lettere* », 1928, vol. LXI, fasc. 16-20, pp. 817-824.

REALE ISTITUTO LOMBARDO DI SCIENZE E LETTERE
Estratto dai *Rendiconti* — Vol. LXI - Fasc. XVI-XX — 1928.

SULLE PROBABILITÀ NUMERABILI E GEOMETRICHE.

Nota del dott. BRUNO DE FINETTI

Lo studio delle probabilità quando il numero dei casi possibili è infinito ha assunto recentemente un notevole sviluppo, ed ebbe anche interessanti applicazioni. Se non che, così come finora è stato trattato, lo si può riguardare, più che altro, come una traduzione nel linguaggio del calcolo delle probabilità di risultati raggiunti in campi diversi (principalmente nella teoria della misura), e ciò giustifica il dubbio che la validità di tali metodi sia tutt'altro che generale. E che essi diano quindi soltanto degli esempi di problemi di probabilità numerabili o geometriche, interessanti quanto si vuole, ma sempre esempi, e non una teoria delle probabilità numerabili o geometriche.

Affrontando direttamente la questione dal punto di vista del calcolo delle probabilità si vedrebbe che il fatto essenziale non è tanto di avere dei casi possibili in numero finito o infinito (e in tal caso costituenti una classe numerabile, continua, ...) ma piuttosto quello di considerare delle probabilità nulle. Lasciando per ora da parte lo studio teorico del problema generale, di cui intendo esporre quanto prima i risultati, darò qui degli esempi critici che, mettendo in rilievo l'insufficenza dei metodi usuali, mostrino chiaramente la necessità di nuovi indirizzi.

Probabilità numerabili.

Se i casi possibili costituiscono una classe numerabile

$$a_1, a_2, \ldots\ldots, a_n, \ldots\ldots$$

si considera usualmente (1) l'ipotesi in cui si possono attribuire

(1) V. ad es. A. LOMNICKI, *Nouveaux fondements de la théorie des probabilités.* Fundam. math., T. 4, 1923, p. 34-71.

loro dei « pesi » $p_1, p_2, \ldots, p_n, \ldots$ così che la probabilità di un evento sia il limite del rapporto fra il peso dei casi favorevoli e dei casi possibili nei primi n casi, al crescere di n. È utile notare il significato concreto di tali « pesi » secondo il calcolo delle probabilità. Anche se due eventi a_1, a_2 hanno probabilità nulla ha senso parlare del *rapporto* delle loro probabilità: se infatti consideriamo le probabilità P_1, P_2 di a_1, a_2 *subordinatamente* all'ipotesi che si verifichi o l'uno o l'altro di essi si ha certamente $P_1 + P_2 = 1$ (se $a_1\, a_2$ non sono incompatibili si ha, più in generale, $P_1 + P_2 \geqq 1$) e potremo dire che le loro probabilità stanno fra loro come $P_1 : P_2$. Prendendo ad arbitrio la probabilità di un evento dato come unità, potremo così definire il « peso » di tutti gli altri eventi. Se in particolare si prende $= 1$ la probabilità di un evento certo, il « peso » coincide colla probabilità. Il rapporto di due probabilità (e quindi il « peso ») può però risultare nullo o infinito (nell'esempio precedente può essere $P_1 = 0$ o $P_2 = 0$); se non è così le due probabilità si diranno *commensurabili* o dello stesso *ordine*.

Ora, per poter caratterizzare le probabilità d'una classe d'eventi mediante « pesi », come il metodo esposto per le classi numerabili esige, è necessario che tutte le probabilità siano commensurabili, e già questa sarebbe una restrizione fortissima. Se si ammettesse anche di considerare le serie di pesi in parte nulli (con che si rinuncerebbe a determinare la probabilità dei casi ad essi corrispondenti) non a tutte le classi numerabili di eventi potrebbe corrispondere una tale serie di pesi, come mostra l'esempio 1.

Es. 1. — I casi possibili siano tutti ugualmente probabili, e costituiscano una classe numerabile $a_1, a_2, \ldots, a_n, \ldots$ in cui le probabilità si possano calcolare secondo la regola usuale del passaggio al limite; costruiamo allora la classe numerabile $A_1, A_2, \ldots, A_h, \ldots$ dove A_1 è la somma dei casi $a_1\, a_4\, a_9 \ldots$ corrispondenti agli indici quadrati (a_n tali che $E(\sqrt{n}) > E(\sqrt{n-1})$, $E =$ « entier »); i casi rimanenti $a_2\, a_3\, a_5\, a_6\, a_7\, a_8\, a_{10} \ldots$ indichiamoli nell'ordine $a_1^{(1)}\, a_2^{(1)}\, a_3^{(1)} \ldots$ e formiamo A_2 come somma degli $a_n^{(1)}$ tali che $E(n^{\frac{2}{3}}) > E\left((n-1)^{\frac{2}{3}}\right)$. In generale indichiamo $a_1^{(h)}\, a_2^{(h)} \ldots a_n^{(h)} \ldots$ i casi non contenuti in $A_1, A_2, \ldots, A_h$, e disposti nello stesso ordine iniziale, e formiamo A_{h+1} come somma di quelli per cui $E\left[n^{\frac{h}{h+1}}\right] > E\left[(n-1)^{\frac{h}{h+1}}\right]$.

Tutti i casi possibili a_n compaiono in uno (e manifestamente in uno solo) degli A_h. Infatti $a_1^{(h)}$, il primo dei casi non contenuti in $A_1 \ldots A_h$, appartiene sempre a A_{h+1}, ed è quindi ovvio che a_n deve figurare certamente in uno degli A_h con $h \leq n$. Dimostriamo ora che ciascuna delle probabilità degli $A_1 \ldots, A_h \ldots$ ha ordine maggiore di tutte le precedenti. In $a_1^{(h)} \ldots a_n^{(h)}$ A_{h+1} ha $E(n^{\frac{h}{h+1}})$ elementi; ne rimangono $n - E(n^{\frac{h}{h+1}})$, di cui A_{h+2} ne conterrà $E\left\{\left[n - E\left(n^{\frac{h}{h+1}}\right)\right]^{\frac{h+1}{h+2}}\right\}$. E si ha facilmente

$$\lim_{n=\infty} \frac{E(n^{\frac{h}{h+1}})}{E\left\{\left[n - E\left(n^{\frac{h}{h+1}}\right)\right]^{\frac{h+1}{h+2}}\right\}} = \left[\lim_{n=\infty} \frac{n^{-\left(\frac{1}{h+1}\right)^2}}{1 - n^{-\frac{1}{h+1}}}\right]^{\frac{h+1}{h+2}} = 0 \qquad \text{c. v. d.}$$

Abbiamo quindi un esempio in cui *i casi possibili* ($A_1, A_2, \ldots$ $\ldots, A_h, \ldots$) *costituiscono una classe numerabile, le loro probabilità hanno tutte ordini distinti, fra cui non ne esiste uno maggiore di tutti gli altri.* E quindi se a un caso si attribuisce il peso 1, i pesi precedenti sono 0 e i successivi ∞ (1).

Ciò ha del resto ben poca importanza per la nostra critica di fronte a un risultato ben più decisivo dimostrato dall'es. 2 (su cui in particolare vorrei richiamare l'attenzione): *anche nel caso in cui le infinite probabilità siano tra loro commen-*

(1) Procedendo analogamente, ma sostituendo alla

$$E(n^{\frac{h}{h+1}}) > E\left((n-1)^{\frac{h}{h+1}}\right)$$

la condizione

$$E(n^{\frac{1}{h+1}}) > E\left((n-1)^{\frac{1}{h+1}}\right),$$

si avrebbe un esempio in cui i casi possibili ($A_1, A_2, \ldots A_h \ldots$) costituiscono una classe numerabile, e hanno probabilità d'ordini tutti distinti (ciascuno minore dei precedenti): il primo ha probabilità d'ordine maggiore di tutti gli altri, ma pure ha probabilità nulla.

surabili il metodo del passaggio al limite può non essere applicabile.

Osserviamo anzitutto che se il metodo è applicabile agli eventi $a_1\, a_2 \,.....\, a_n \,.....$ disposti nell' ordine in cui sono scritti, non lo è in generale quando essi subiscono una permutazione. Dire che il metodo non è applicabile significa asserire che gli eventi non possono essere ordinati in modo che sussistano le proprietà richieste. La constatazione che esse non sussistono per un dato ordinamento della successione non significherebbe nulla. Se però la serie dei pesi è convergente la dipendenza dall' ordine scompare, e il metodo è applicabile per qualunque ordinamento o per nessuno a seconda che lo è o non lo è per un ordinamento particolare. Di ciò appunto ci gioveremo per costruire un esempio facile in cui il metodo risulta non applicabile.

Es. 2. — I casi possibili siano $a_1\, a_2 \,.....\, a_n \,.....$ Se si verifica un' ipotesi, α, di probabilità $\frac{1}{2}$, gli $a_1 \,.....\, a_n \,.....$ sono ugualmente probabili (probabilità $= 0$) e la probabilità di una loro somma si calcola col metodo usuale del limite. Nel caso contrario (ipotesi « non α », di probabilità $\frac{1}{2}$) essi hanno le probabilità $p_1\, p_2 \,.....\, p_n \,.....$ la cui serie (convergente) ha per somma 1, e la probabilità di ogni loro somma è la somma (eventualmente serie) delle probabilità. È chiaro che la probabilità di a_n sarà $\left(\frac{1}{2} \times 0\right) + \left(\frac{1}{2} \times p_n\right) = \frac{1}{2}\, p_n$. I « pesi » rimangono dunque inalterati (una costante moltiplicativa è inessenziale), e il metodo del limite (anche cambiando d'ordine) non potrà che dare gli stessi risultati relativi all' ipotesi « non α ». Ma ciò è assurdo, perchè la probabilità di una somma di a_n sarà $p = \frac{p' + p''}{2}$ (p' = probabilità nell' ipotesi α, p'' nell' ipotesi « non α »), e non può essere sempre $p' = p''$ dato che, ad esempio, per le classi finite è sempre $p' = 0$, mentre ciò non si può avere sempre per p''.

È invece facile dimostrare che il metodo usuale è sempre valido in una classe numerabile se le probabilità costituiscono una serie convergente, e la sua somma è $= 1$.

Probabilità nel continuo.

Passiamo al caso delle probabilità geometriche o nel continuo.

Gli esempi seguenti mostreranno quanta maggiore varietà di casi si presenti nella teoria delle probabilità di quanti non ne abbia condotti a prevedere la teoria della misura.

Dal fatto che intervalli uguali abbiano probabilità uguali non scende che la probabilità di un aggregato sia proporzionale alla sua misura secondo Lebesgue (e quindi: *dal fatto che la probabilità in un generico intervallo* (α, β) *sia data da* $\int_\alpha^\beta \varphi(x)\,dx$ *non scende che per ogni aggregato sia data dall'integrale di Lebesgue*).

Es. 3. — Abbiasi una classe numerabile di casi possibili $a_1\, a_2 \ldots\ldots a_n \ldots\ldots$, ugualmente probabili, e in cui le probabilità si calcolino col passaggio al limite (negli n del tipo $2^m - 1$). Consideriamo una variabile t che in corrispondenza ad $a_1\, a_2 \ldots$ $\ldots a_n \ldots\ldots$ assuma ordinatamente i valori

$$\frac{1}{2};\ \frac{1}{4},\ \frac{3}{4};\ \frac{1}{8},\ \frac{3}{8},\ \frac{5}{8},\ \frac{7}{8};\ \ldots;\ \frac{1}{2^m},\ \frac{3}{2^m},\ \ldots,\ \frac{2^m-1}{2^m};\ldots$$

La probabilità che sia $\alpha < t < \beta$ $(0 \leqq \alpha \leqq \beta \leqq 1)$ è $\beta - \alpha$: la probabilità di un intervallo è uguale alla sua lunghezza. Pure, l'aggregato dei numeri razionali, che ha misura nulla, ha probabilità uguale ad 1.

Il fatto però che non tutti i numeri da 0 a 1 sono valori possibili per t può lasciare dei dubbi sul valore probativo di tale esempio. Si può evitare quest'obiezione in modo semplice considerando una seconda variabile casuale t_p che, in un'ipotesi A di probabilità p, segue la solita legge uniforme (probabilità uguale alla misura secondo Lebesgue), e nell'ipotesi contraria, « non A », di probabilità $1 - p$, coincide colla t precedente. La probabilità che sia $\alpha < t_p < \beta$ è, così nella 1ª che nella 2ª ipotesi, $\beta - \alpha$, e quindi è essa stessa uguale a

$$p(\beta - \alpha) + (1 - p)(\beta - \alpha) = \beta - \alpha.$$

La probabilità che t_p sia razionale è, nelle due ipotesi, 0 o 1, e quindi è uguale a $p \times 0 + (1 - p) \times 1 = 1 - p$. Abbiamo quindi una variabile casuale t_p, i cui valori possibili sono tutti i numeri tra 0 e 1, la probabilità d'appartenere a un intervallo è uguale alla lunghezza, ma la probabilità d'appartenere a un aggregato non è uguale alla misura di Lebesgue. In particolare può supporsi che A abbia probabilità nulla ($p = 0$) pur non essendo impossibile. Allora si ha una variabile casuale t_0 che, pur potendo assumere tutti i valori di (0, 1), e non dando luogo quindi all'obiezione accennata, segue la stessa legge di probabilità di t.

L'esempio precedente mostra anche che *l'uguale densità di probabilità negli intervalli non implica l'uguale probabilità dei punti* (1). Potrebbe tuttavia sussistere il dubbio che se la probabilità dei punti non fosse discontinua, come lo è nell'esempio precedente, per l'ipotesi dell'omogeneità dovesse essere costante. Daremo quindi un nuovo esempio in cui la probabilità del *punto* t $(0 < t < 1)$ è proporzionale alla funzione continua $\varphi(t)$ mentre la *densità* di probabilità è costante.

Es. 4. — Consideriamo, in coordinate cartesiane, la figura piana compresa tra l'asse delle ascisse, la curva d'equazione parametica

$$x(t) = \int_0^t \frac{dt}{\varphi(t)} \qquad y(t) = \varphi(t)$$

e le ordinate corrispondenti a $t = 0$, $t = 1$. I punti interni ad essa siano i casi possibili; porzioni d'area uguale, segmenti di lunghezza uguale, abbiano uguale probabilità. Allora risulta immediatamente che la probabilità del segmento corrispondente al parametro t $(x = x(t),\ 0 \leqq y \leqq y(t))$ è proporzionale a $\varphi(t)$, quella della striscia $\alpha < t < \beta$ $(0 \leqq \alpha \leqq \beta \leqq 1)$ è

$$\int_\alpha^\beta \varphi(t) \frac{dx}{dt} dt = \int_\alpha^\beta \varphi(t) \frac{dt}{\varphi(t)} = \int_\alpha^\beta dt = \beta - \alpha.$$

È ovvio ora come inversamente si possa avere una densità di probabilità variabile da punto a punto, mentre i singoli punti hanno tutti eguale probabilità.

---

(1) È erronea quindi — se presa alla lettera — l'espressione « tutti i valori sono ugualmente probabili » per significare che la densità di probabilità è uniforme. Non si potranno ad esempio caratterizzare le tre ipotesi del paradosso di Bertrand dicendo che si considerano ugualmente probabili rispettivamente tutti gli angoli, tutte le distanze, tutti i centri delle corde.

Vale la pena invece di mostrare, col seguente esempio, come, pur conservando omogenea la densità, i punti di un intervallo possano aver tutti probabilità di ordine minore (non commensurabili) con quelli di un altro.

Es. 5. — Consideriamo il rettangolo

$$0 \leqq x \leqq 1, \quad 0 \leqq y \leqq \frac{1}{2},$$

in cui porzioni d'area uguale, segmenti di lunghezza uguale, abbiano uguale probabilità, e tutti i punti siano casi ugualmente probabili. La funzione $t(x, y)$ sia, per $x \leqq \frac{1}{2}$, la metà del parametro della curva di Peano che riempie il corrispondente quadrato; per $x > \frac{1}{2}$ sia $t(x, y) = x$. La probabilità di t ha densità costante in (0, 1), ma per ogni valore $t > \frac{1}{2}$ (corrispondente a un segmento di lunghezza $\left.\frac{1}{2}\right)$ è infinitamente maggiore che pei valori $t \leqq \frac{1}{2}$ (corrispondenti ai singoli punti del I° quadrato).

Es. 6. — Un esempio ancor più paradossale si può costruire analogamente partendo da un cubo e facendo corrispondere ai t di $\left(0, \frac{1}{2}\right)$ i segmenti di una faccia, ai t di $\left(\frac{1}{2}, 1\right)$ i singoli punti rimanenti in modo che la densità di probabilità sia costante entro ciascuna delle due parti. Si ha allora che un punto della prima parte è infinitamente più probabile che un punto della seconda, mentre un intervallo comunque piccolo della seconda parte ha probabilità infinitamente maggiore anche di tutta la prima parte.

Per completare in un altro senso le conclusioni dell'Es. 3, osserviamo che ovviamente: *dal fatto che la probabilità di un intervallo sia uguale (in data unità di misura) alla sua lunghezza, scende che la probabilità di un aggregato è compresa fra la sua misura interna ed esterna* (nel senso di « longore supero » e « infero » di Peano, Form. Math.), e, più in generale *dal fatto che la probabilità in un generico intervallo* (α, β)

è data da $\int_\alpha^\beta \varphi(\mathrm{x})\,\mathrm{dx}$, *scende che per ogni aggregato è compresa fra* $\underline{\int\varphi}$ *e* $\overline{\int\varphi}$ (ove $\underline{\int\varphi}$ indica il limite superiore della somma d'integrali estesa a una qualunque classe finita di intervalli contenuta nell'aggregato, e $\overline{\int\varphi}$ il limite inferiore della somma d'integrali estesa a una qualunque classe finita di intervalli contenente l'aggregato).

Il fatto interessante che si tratta di mettere in luce è ora che questa limitazione *non può essere ristretta*, cioè che la probabilità può effettivamente assumere il massimo o il minimo così stabiliti. Daremo la dimostrazione completa per il caso omogeneo ($\varphi =$ costante) per comodità di esposizione, osservando che il caso generale non presenta nessuna nuova difficoltà.

Es. 7. — Nell'intervallo (0, 1) consideriamo un aggregato u di misura esterna L; dimostriamo che è possibile definire una variabile casuale x ($0 < x < 1$) che ha in ogni intervallo probabilità uguale alla lunghezza, e in u probabilità uguale alla misura esterna L.

Definiamo la variabile t come nell'Es. 3, e poniamo $x = t$ se t non è un punto limite di u; se invece t è un punto limite di u poniamo x uguale a uno degli u che differiscono da t per meno di $\frac{1}{2^{\mathrm{m}}}$, ove 2^{m} è il denominatore ridotto di t (la scelta di x si potrà fare secondo una regola fissa per prescindere dal principio di Zermelo). Anche x, come t, ha probabilità omogenea in ogni intervallo; dimostriamo di più che ha in u probabilità uguale alla misura esterna L. Infatti la probabilità che x appartenga ad u è uguale alla probabilità che t appartenga alla figura limite di u; la figura limite di u ha misura esterna e interna uguali ad L, e quindi la probabilità cercata è L, c. v. d. Per ovviare alla stessa obiezione che, al § 3, ci fece introdurre t_{p} e t_0, potremo qui considerare e definire in modo perfettamente analogo x_{p} e x_0.

La seconda parte del teorema è immediato corollario della prima. Se infatti il procedimento si applica all'aggregato complementare di u, si ottiene una variabile casuale la cui probabilità di appartenere ad u è uguale alla misura esterna, perchè la misura esterna dell'aggregato complementare è il complemento della misura interna.

Roma, 18 settembre 1928 (A. VI).
(Istit. Centr. di Statistica).