BRUNO DE FINETTI

A proposito di "correlazione"

SOCIETÀ AN. TIPOGRAFICA EMILIANA
FERRARA 1937-XV

1. — Come spesso avviene, anche nel caso della « correlazione » molte discussioni non hanno origine che da una confusione di concetti. Da qualche tempo si sente dimostrare e ripetere in una quantità di lavori che è necessario distinguere quel concetto di « correlazione » che si estrinseca nel « coefficiente di correlazione » r (di Bravais), e quello corrispondente al concetto di « dipendenza stocastica » del calcolo delle probabilità. Se così diffuso è il bisogno di chiarire questo punto, vorrà dire certamente che anche la confusione era abbastanza diffusa, probabilmente a cagione del fatto che i due concetti sono equivalenti nel caso delle distribuzioni *gaussiane*, e dell'abitudine ingiustificata e dannosa di considerare la distribuzione gaussiana in modo troppo esclusivo come se dovesse rappresentare la regola in quasi tutti i casi presentati dal calcolo della probabilità e dalla statistica, e se ogni diversa forma di distribuzione costituisse un caso eccezionale o irregolare (anche lo stesso nome di « legge normale » può contribuire a tale impressione, e sarebbe forse perciò preferibile abbandonarlo).

Non vedrei alcun motivo di aggiungere parola per chiarire la distinzione fra i due concetti — che mai avrebbero dovuto potersi confondere e che (almeno per quanto mi consta) sono sempre stati tenuti ben distinti dai cultori di calcolo delle probabilità — se non mi sembrasse che, scoprendo che il coefficiente di correlazione non ha quel significato che per un incomprensibile equivoco gli era stato da taluni attribuito, si pensi di aver con ciò svuotato il coefficiente di correlazione di *ogni* significato. Sarebbe come pensare che le trebbiatrici sono inutili perchè abbiamo dimostrato ad un tale che le scambiava per mulini che esse non servono a macinare il grano!

Perciò penso non sia del tutto inutile esporre brevemente il significato del coefficiente di correlazione secondo il calcolo delle probabilità, tanto più che forse qualche considerazione potrà risultare nuova, almeno nella forma e nella luce in cui viene posta secondo il punto di vista della teoria dei numeri aleatori. Inoltre toccherò in fine la questione della terminologia più adatta per evitare strascichi dell'equivoco accennato e per ovviare ad altri inconvenienti che vedremo: la questione è però talmente intricata che, senza poter presentare proposte soddisfacenti, mi limiterò a chiarire quali esigenze si dovrebbe cercar di rispettare nel prendere una deliberazione in merito.

2. — Dato un numero aleatorio X, ne indicheremo, come è d'uso, con $M(X)$ la «speranza matematica» e con $\sigma(X) = \sqrt{;M(X - M(X))^2}$ lo « scostamento quadratico medio ».

È noto che la speranza matematica gode della proprietà additiva: $M(X + Y) = M(X) + M(Y)$, e ciò basta a risolvere tutti i problemi in cui si hanno soltanto combinazioni lineari di numeri aleatori e basta la conoscenza della speranza matematica. Se invece $Z == f(X, Y)$ è funzione non lineare di X e Y, oppure se, tale funzione essendo lineare, interessa una conoscenza della distribuzione di probabilità di Z più precisa di quella che si riduce alla determinazione della speranza matematica, occorrono naturalmente altri elementi. Il primo è appunto il coefficiente di correlazione, che basta, insieme ad M e σ, per tutti i problemi di « secondo grado », in cui si tratti cioè di determinare la speranza matematica di una qualunque funzione di secondo grado di dati numeri aleatori X_1 X_2,...,X_n. Fra i problemi di secondo grado ve ne sono due di interesse pratico immediato, e che bastano a mostrare l'enorme importanza del coefficiente di correlazione di cui esigono l'introduzione: sono i problemi consistenti nella determinazione della speranza matematica del *prodotto* $X.Y$ di due numeri aleatori X e Y e dello scostamento quadratico medio della loro somma $X + Y$ (ovvia generalizzazione dei due problemi, è il calcolo della speranza matematica del prodotto di due combinazioni lineari di dati numeri aleatori, e in particolare del quadrato di una tale combinazione lineare).

Per affrontare la questione in generale, consideriamo una generica funzione di secondo grado di n dati numeri aleatori X_1, X_2,...,X_n, e sia

$$Z = \Sigma_{ij}\, a_{ij}\, X_i X_j$$

(possiamo supporre, senza diminuire la generalità, $a_{ij} = a_{ji}$).

Indicando con $\bar{X}_i = M(X_i)$ la speranza matematica di X_i, possiamo scrivere

$$Z = \Sigma_{ij}\, a_{ij}\, [\bar{X}_i + (X_i - \bar{X}_i)]\, [\bar{X}_j + (X_j - \bar{X}_j)] = \Sigma_{ij}\, a_{ij}\, \bar{X}_i \bar{X}_j + $$
$$+ 2\, \Sigma_{ij}\, a_{ij}\, \bar{X}_i\, (X_j - \bar{X}_j) + \Sigma_{ij}\, a_{ij}\, (X_i - \bar{X}_i)\, (X_j - \bar{X}_j)$$

e la speranza matematica di Z risulta

$$M(Z) = \Sigma_{ij}\, a_{ij}\, \bar{X}_i \bar{X}_j + 2\, \Sigma_{ij}\, a_{ij}\, \bar{X}_i\, M(X_j - \bar{X}_j) + \Sigma_{ij}\, a_{ij}\, M[(X_i - \bar{X}_i)(X_j - \bar{X}_j)] =$$
$$= \Sigma_{ij}\, a_{ij}\, \bar{X}_i \bar{X}_j + \Sigma_{ij}\, a_{ij}\, M[(X_i - \bar{X}_i)(X_j - \bar{X}_j)]$$

perchè $M(X_j - \bar{X}_j) = M(X_j) - \bar{X}_j = \bar{X}_j - \bar{X}_j = 0$.

Si vede così che potremo determinare in ogni caso $M(Z)$ conoscendo, oltre gli $\bar{X}_i = M(X_i)$, i termini $M[(X_i - \bar{X}_i)(X_j - \bar{X}_j)]$.

Per $i = j$ è per definizione $M(X_i - \bar{X}_i)^2 = \sigma^2(X_i)$; per $i \neq j$ la conoscenza degli $M(X_i)$ e $\sigma(X_i)$ non basta invece a determinare $M[(X_i - \bar{X}_i)(X_j - \bar{X}_j)$ ma dà però una limitazione dalla quale risulta spontanea l'introduzione del « coefficiente di correlazione ».

3. — Essendo infatti X ed Y due numeri aleatori consideriamo la loro combinazione lineare $Z_t = X + tY$, ove t è un qualunque numero reale; avremo

$$\sigma^2(Z_t) = M[(X - \bar{X}) + t(Y - \bar{Y})]^2 = M(X - \bar{X})^2 + 2\, t\, M[(X - \bar{X})\, (Y - \bar{Y})] +$$
$$+ t^2\, M(Y - \bar{Y})^2 = \sigma^2(X) + 2\, t\, M[(X - \bar{X})\, (Y - \bar{Y})] + t^2\, \sigma^2(Y).$$

Considerando t come un parametro, $\sigma^2(Z_t)$ risulta dunque funzione di secondo grado di t; d'altronde $\sigma^2(Z_t)$ è per sua natura positivo o (o al più nullo), e quindi dev'essere

$$|\, M[(X - \bar{X})\, (Y - \bar{Y})]\, | < \sigma(X).\sigma(Y),$$

che è la limitazione annunciata; vedremo tosto che questa limitazione è la più precisa possibile, nel senso che $M[(X - \bar{X})\, (Y - \bar{Y})]$ può effettivamente assumere tutti i valori compresi fra $\pm\, \sigma(X).\sigma(Y)$ (estremi inclusi). Appare così spontanea la convenienza di introdurre il « coefficiente di correlazione », definito appunto da

$$r(X,Y) = \frac{M[(X-\overline{X})(Y-\overline{Y})]}{\sigma(X)\ \sigma(Y)}$$

che risulta compreso sempre fra ± 1, e che non dipende da eventuali coefficienti positivi di proporzionalità da cui possano venire affetti X e Y (cioè, se $a > 0$ e $b > 0$, $r(aX, bY) = r(X, Y)$; più in generale $r(aX, bY) = \pm\, r(X, Y)$ segno +o – a seconda che a e b hanno segni uguali od opposti).

Per la speranza matematica del prodotto di due numeri aleatori X e Y si ha allora in generale la formula

$$M(X.Y) = M(X).M(Y) + r(X, Y).\sigma(X).\sigma(Y)$$

(da cui si apprende subito che $M(X.Y)$ è ad ogni modo compreso tra $M(X).M(Y) \pm \sigma(X).\sigma(Y)$),

mentre per lo scostamento quadratico medio della somma si ha

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2\, r(X, Y)\ \sigma(X)\ \sigma(Y)$$

(da cui $\sigma(X + Y)$ risulta compreso ad ogni modo tra $|\sigma(X) - \sigma(Y)|$ e $\sigma(X) + \sigma(Y)$).

A seconda del segno di $r(X, Y)$ si dirà che X e Y sono correlati positivamente o negativamente: essi sono adunque correlati *positivamente* quando $M(X.Y) > M(X).M(Y)$ ossia $\sigma^2(X + Y) > \sigma^2(X) + \sigma^2(Y)$; *negativamente* quando all'opposto $M(X.Y) < M(X).M(Y)$ ossia $\sigma^2(X + Y) < \sigma^2(X) + \sigma^2(Y)$.

Se $r = 0$, ossia $M(X.Y) = M(X).M(Y)$, ossia $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$, i due numeri aleatori si dicono *non-correlati*. Sono in particolare non correlati numeri aleatori stocasticamente indipendenti, perchè è noto e si vede facilmente che, nel caso dell'indipendenza, $M(X.Y) = M(X).M(Y)$; non vale però naturalmente l'affermazione reciproca.

Per dimostrare che r può effettivamente assumere tutti i valori tra –1 e + 1 basta considerare due numeri aleatori indipendenti X e X' che, senza diminuire la generalità, potremo supporre abbiano $M(X) = M(X') = 0$, $\sigma(X) = \sigma(X') = 1$, e definire poi, scelto r compreso fra –1 e + 1, il numero aleatorio Y come segue:

$$Y = rX + \sqrt{1-r^2}\, X'.$$

Si ha

$$\sigma^2(Y) = r^2 + (1-r^2) = 1, \quad r(X,Y) = M(X.Y) = M(rX^2 + \sqrt{1-r^2}\, XX') =$$
$$= rM(X^2) + \sqrt{1-r^2}\, M(XX') = r, \quad \text{c.d.d.}$$

4. — I casi estremi $r(X, Y) = \pm 1$ significano invece che X e Y sono linearmente dipendenti. L'equazione $\sigma^2(Z_t) = 0$ ha allora infatti la radice $t = \pm$ **Errore.** $= k$ (segno + o – a seconda che $r = +1$ o $r = -1$).
Ma perchè un numero aleatorio abbia scostamento quadratico medio nullo è necessario che sia nulla la probabilità di uno scarto dalla speranza matematica superiore ad un ε comunque piccolo, ossia, se si accetta il principio esteso delle probabilità totali (cioè: esteso alle classi numerabili), il numero aleatorio ha probabilità uguale ad uno di risultare uguale alla speranza matematica. Nel nostro caso ne consegue che, a meno di casi di probabilità nulla, Z_t = Z_t;¯, ossia X – X;¯ = k(Y – Y;¯).

Vorrei però precisare anche qui la modificazione che occorre se si ritiene (come è mia opinione) di dover abbandonare il principio esteso delle probabilità totali. Allora può essere $\sigma(X) = 0$ e quindi Prob. {|X – X;⏋ > ε} = 0 per ogni $\varepsilon > 0$, senza che sia necessariamente Prob. {|X – X;⏋ > 0} = 0; se ad es. $X = 1/n$, con n « intero scelto ad arbitrio » (nel senso che ogni intero determinato ha probabilità nulla) è X;¯ = $\sigma(X) = 0$, ma Prob. $\{X > 0\} = 1$.

Per evitare di dover ripetere continuamente una precisazione del genere, e per non usare una forma a rigore inesatta, come si avrebbe affermando nel caso precedente che è senz'altro X – X;¯ = k (Y – Y;¯), ritengo opportuno introdurre il segno $X \doteq Y$, da leggere « X coincide con Y », per indicare che la disuguaglianza $|X - Y| > \varepsilon$ ha probabilità nulla qualunque sia $\varepsilon > 0$, ossia che $M(|X - Y|) = 0$. Possiamo allora esprimere altrettanto correttamente quanto brevemente la conclusione trovata dicendo che se $r(X, Y) = \pm 1$ risulta X – X;¯ = k(Y – Y;¯), ove k è positivo o negativo insieme con r. E l'enunciato sarà esatto sia per chi accetta che per chi non accetta il principio delle probabilità totali, salvo la diversa interpretazione della definizione di « coincidenza ».

5. — Non mi sto a dilungare sul significato che ha il «coefficiente di correlazione » come indice statistico di « concordanza », significato spesso lumeggiato, ed anche recentemente su queste stesse pagine dal Pietra ([1]); la formula $M(X.Y)$ riassume del resto tutto l'aspetto mate-

---

[1] Cfr. questo *Supplemento*, A. II, n. 2-3, 1936 (*L'ostracismo al coefficiente di correlazione?*).

matico di tale significato.

Piuttosto può essere utile un cenno su di una possibile interpretazione geometrica, atta a rendere intuitive le relazioni fra speranze matematiche, scostamenti quadratici medi e coefficienti di correlazione. Poichè di numeri aleatori possiamo considerare delle combinazioni lineari, potremo interpretarli come vettori di uno « spazio astratto ». Considerando $\sigma(X)$ come *modulo* del vettore X, e quindi $\sigma(Y-X)$ come *distanza* $d(X, Y)$ dei due vettori X e Y, veniamo a definire uno spazio distanziale, o spazio « D » nel senso di Fréchet, purchè si considerino rappresentati da un medesimo vettore tutti i numeri aleatori la cui differenza coincida con un numero fisso. Infatti X ed $X+a$ hanno « distanza » nulla, perchè $(X+a)-X=a$, $\sigma(a)=0$, e inversamente $\sigma(Y-X)=0$ significa $Y-X=a$. In ogni altro caso è $\sigma(Y-X)>0$, e vale (come sostanzialmente, si e visto nel n. 3) la disuguaglianza triangolare

$$|\sigma(Y)-\sigma(X)| \leq \sigma(Y-X) \leq \sigma(Y)+\sigma(X) .$$

Di più, lo spazio S così definito risulta uno s*pazio astratto metrico*, o spazio D_M,[2] in quanto che $\sigma(X)\,\sigma(Y)\,r(X, Y)$ vi si può interpretare come prodotto interno essendo funzione lineare omogenea simmetrica di X e Y che si riduce al quadrato del modulo per $Y=X$. I1 coefficiente di correlazione figura quindi come coseno dell' « angolo » fra i vettori X e Y, angolo che si può univocamente definire ponendo $r(X, Y)=\cos\alpha(X, Y)$, $0 \leq \alpha(X, Y) \leq \pi$ (ossia α compreso tra 0° e 180°).

La formula già vista per $\sigma(X+Y)$ diviene allora

$$\sigma^2(X+Y)=\sigma^2(X)+\sigma^2(Y)+2\sigma(X).\sigma(Y)\cos\alpha(X,Y),$$

e si vede, ricordando il teorema di Carnot, che $\sigma(X+Y)$ è il terzo lato di un triangolo quando gli altri due sono $\sigma(X)$ e $\sigma(Y)$ e racchiudono un angolo $\alpha(X, Y)$. Nel caso della non-correlazione (o, in particolare, dell'indipendenza) ci si riduce al teorema di Pitagora: $\sigma(X+Y)$ è l'ipotenusa del triangolo rettangolo di cui $\sigma(X)$ e $\sigma(Y)$ sono i cateti.

La formula per $M(XY)$ dice analogamente che il termine correttivo da aggiungere a $M(X).M(Y)$ è il prodotto interno dei vettori rappresentativi di X e Y, cioè $\sigma(X)\sigma(Y)\cos\alpha(X, Y)$.

---

[2] Cfr. la mia nota *Spazi. astratti metrici* (D_M), « Atti Accad. Pontificia », A. LXXXIII, sess. VI, 1930.

La non-correlazione significa *ortogonalità*; la correlazione positiva o negativa significa invece che α è rispettivamente *acuto* od *ottuso*; i casi estremi $r = \pm 1$ dicono che $\alpha = 0$ risp. $\alpha = \pi$, ossia che i due vettori (e quindi, a meno d'un numero fisso additivo, i numeri aleatori) non differiscono che per una costante moltiplicativa, rispettivamente positiva o negativa..

Molte proprietà note degli spazi metrici potranno suggerire imagini adatte per studiare e risolvere vari problemi. Ogni numero aleatorio Y si può ad es. scomporre nella somma di uno proporzionale ad X (componente parallelo) e di uno non correlato ad X (componente ortogonale). Più generalmente, dati n numeri aleatori $X_1, X_2, \ldots, X_n$ linearmente indipendenti (per cui cioè non esistano dei coefficienti $a_0, a_1, a_2, \ldots, a_n$ tali che necessariamente $a_0 = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$), si potrà esprimerli come combinazioni lineari di n numeri aleatori non correlati a scostamento quadratico medio unitario $Y_1\ Y_2 \ldots Y_n$ (sistema unitario ortogonale) con $M(Y_i) = 0$, $\sigma(Y_i) = 1$, $r(Y_i, Y_j) = 0$ $(i \neq j)$, e per σ varrà la solita espressione del modulo d'un vettore.

Se $\alpha(X, Y)$ è l'angolo di due numeri aleatori X e Y, un terzo numero aleatorio Z non potrà formare con X e con Y due angoli $\alpha(X, Z)$ ed $\alpha(Y, Z)$ qualunqui, ma si avrà (ovviamente, se si pensa all'imagine geometrica) $\alpha(X, Y) \leq \alpha(X, Z) + \alpha(Y, Z) \leq \pi - \alpha(X, Y)$; il caso estremo $\alpha(X, Z) + \alpha(Y, Z) = \alpha(X, Y)$ si ha se e solo se $Z = aX + bY$, $a > 0$, $b > 0$ (vettore complanare, compreso nell'angolo concavo fra i due vettori), e l'altro $\alpha(X, Z) + \alpha(Y, Z) = \pi - \alpha(X, Y)$ se e solo se $Z = -(aX + bY)$ $a > 0$, $b > 0$ (condizione predetta pel vettore opposto). Ciò mette in luce che vi sono delle limitazioni fra i gradi di correlazione due a due di diversi numeri aleatori. In particolare, se tre numeri aleatori sono tutti tra loro ugualmente correlati, non potendo essi formare due a due un angolo, superiore a **Errore.** (ossia a 120°), il coefficiente di correlazione non può essere minore di $-1/2$.

Analoghe limitazioni sussistono per quattro o più numeri aleatori, e può interessare in particolare di estendere la precedente ricerca al caso di più numeri aleatori ugualmente correlati due a due. Abbiansi dunque n numeri aleatori $X_1, X_2 \ldots X_n$ ugualmente correlati; per comodità li supporremo a speranza matematica nulla e scostamento quadratico medio σ uguale (ciò altera i moduli ma non gli angoli dei vettori!). Dimostriamo che, sotto questa ipotesi, condizione necessaria e sufficiente perché il coefficiente di correlazione raggiunga il minimo valore comune possibile è che $\Sigma X_i = 0$. Poniamo infatti

Y = **Errore.** ΣX_i, $X_i = X_i - Y$: se non fosse $Y = 0$, avremmo così ottenuto
una nuova ennupla Y_1, Y_2 ... Y_n di numeri aleatori ugualmente correlati e con coefficiente di correlazione più piccolo. Dimostriamolo. Si vede anzitutto che Y risulta non-correlato con ciascuno degli Y_i; proviamolo per Y_1:

$$n^2 YY_1 = \left(\sum_{i=1}^{n} X_i\right)\left(nX_1 - \sum_{i=1}^{n} X_i\right) = \left(X_1 + \sum_{i=2}^{n} X_i\right)\left((n-1)X_1 - \sum_{i=2}^{n} X_i\right) =$$

$$= (n-1)X_1^2 + (n-2)\sum_{i=2}^{n} X_1 X_i - \sum_{i=2}^{n} X_i^2 - \sum_{ij=2}^{n}{}' X_i X_j$$

(ove Σ' significa la somma estesa ai termini $i \neq j$);

$$n^2 M(YY_1) = (n-1)\sigma^2 + (n-2)(n-1)\sigma^2 r - (n-1)\sigma^2 - (n-1)(n-2)\sigma^2 r = 0,$$

il che, essendo $M(Y) = M(Y_i) = 0$, significa $r(Y, Y_1) = 0$.

Perciò $M(Y_i\, Y_j) = M(X_i\, X_j) - \sigma^2(Y)$
(infatti $X_i\ X_j = (Y_i + Y)\ (Y_j + Y) = Y_i\ Y_j + Y(Y_i + Y_j) + Y^2$, e il secondo termine ha $M = 0$), e in particolare per $i = j$

$$\sigma^2(Y_i) = M(Y_i^2) = M(X_i^2) - \sigma^2(Y) = \sigma^2(X_i) - \sigma^2(Y) = \sigma^2 - \sigma_0^2$$

ove si scriva $\sigma_0 = \sigma(Y)$, mentre per $i \neq j$, $M(Y_i\, Y_j) = \sigma^2 r - \sigma_0^2$.
Abbiamo tosto allora

$$r(Y_i Y_j) = \frac{M(Y_i Y_j)}{\sigma(Y_i)\sigma(Y_j)} = \frac{\sigma^2 r - \sigma_0^2}{\sigma^2 - \sigma_0^2} = r - (1-r)\frac{\sigma_0^2}{\sigma^2 - \sigma_0^2} < r \quad \text{c.d.d.}$$

Sia dunque $\Sigma\, X_i = 0$; avremo $\sigma^2(\Sigma\, X_i) = 0$,

ossia $\quad M(\Sigma X_i)^2 = M(\Sigma_i X_i^2 + \Sigma'_{ij}\, X_i X_j) = n\sigma^2 + n(n-1)\sigma^2 r = 0,$

da cui $$r = -\frac{1}{n-1}:$$

otteniamo con ciò insieme il cercato minimo di r e la prova che la condizione $\Sigma X_i = 0$ è anche sufficiente.

Tutto questo ragionamento assume un aspetto intuitivo se se ne pensa l'interpretazione vettoriale, ove significa che n vettori di S_n, per formare due a due un medesimo angolo quanto più grande possibile, debbono avere direzione e senso dei raggi congiungenti il centro O di un simplesso equilatero ai suoi vertici; l'angolo α di due vettori così costruiti è dato da cos α = –**Errore.**. Se abbiamo un insieme infinito di numeri aleatori $X_1 X_2 ... X_n$ ... (o un insieme che si può rendere numeroso a piacere, ad es. moltiplicando delle «prove » di un fenomeno), essi non possono essere quindi in particolare ugualmente correlati se non essendo o non-correlati, o correlati positivamente; questa conclusione può essere interessante perchè mostra una differenza intrinseca profonda tra le possibilità di esplicarsi di una correlazione positiva e di una correlazione negativa. Più che per l'interesse che può avere la questione in sè, la trattazione del presente problema si proponeva però di dare un esempio dell'utilità dell'interpretazione vettoriale nel suggerire e rendere intuitivi procedimenti e calcoli. Osserviamo ancora a tale proposito che anche il semplice procedimento per mostrare che $r(X, Y)$ può assumere tutti i valori tra -1 e $+1$ era suggerito dall'ovvia considerazione geometrica che, dati due vettori unitari ortogonali u e v, $u \cos \alpha + v \operatorname{sen} \alpha$ è un vettore unitario formante con u un angolo α.

6. — Anzichè rappresentare con un medesimo vettore tutti i numeri aleatori la cui differenza coincide con una costante, si può benissimo adottare una rappresentazione in cui siano rappresentati da un medesimo vettore soltanto i numeri aleatori *coincidenti*. Se, come *distanza* $d'(X, Y)$ dei numeri aleatori X e Y, anzichè $\sigma(X-Y)$, si considera $\sqrt{M(X-Y)^2}$, si ha infatti $d'(X, Y) = 0$ se e solo se $X = Y$. Non si ha però una rappresentazione totalmente nuova : poichè

$$M(X^2) = [M(X)]^2 + [\sigma(X)]^2$$

si vede subito che ogni vettore del nuovo spazio S' si può scomporre in due componenti che risultano ortogonali fra loro : il primo rappresenta un numero fisso (cioè $\bar{X}$), l'altro un numero aleatorio a speranza matematica nulla. In altre parole, scrivendo $X = \bar{X} + (X - \bar{X})$ i due componenti $\bar{X}$ e $X - \bar{X}$ risultano ortogonali. Il primo è un puro numero fisso, ossia il componente secondo l' « asse dei numeri reali», e il « modulo » ne è il « modulo » o « valore assoluto » nel senso usuale ($|\bar{X}|$); per il secondo si ha (per definizione) $M(X - \bar{X})^2 = \sigma^2(X)$, e quindi l'iperpiano ortogonale all'asse dei numeri reali è lo spazio S considerato nel paragrafo precedente; lo si può pertanto considerare come proiezione dello spazio S' (X e Y si rappresentano sul medesimo vettore di S se $X - Y = \alpha$, ossia $X - Y$ parallelo all'asse dei numeri reali lungo il quale si effettua la proiezione).

L'osservazione del Pietra ([3]) che la relazione tra $\sqrt{M(X^2)}$, $M(X)$ e $\sigma(X)$ è traducibile graficamente nel teorema di Pitagora assume così un significato geometrico preciso in S': la media quadratica $\sqrt{M(X^2)}$ è il modulo del vettore che rappresenta X, mentre $M(X)$ e $\sigma(X)$ sono i moduli dei due componenti ortogonali $\bar{X}$ e $X - \bar{X}$.

7. — Può essere interessante aggiungere qualche osservazione relativa a dei casi particolari.

Il caso più semplice, quello degli *eventi*, presenta la particolarità notevole che, per esso, la *non-correlazione* significa *indipendenza* stocastica. Siano infatti E' ed E'' due eventi, $E = E'E''$ il coverificarsi di essi, e indichiamone le probabilità con $p' = P(E')$, $p'' = P(E'')$, $p = P(E)$, e gli scostamenti quadratici medi con $\sigma' = \sqrt{p'q'}$, $\sigma'' = \sqrt{p''q''}$ (al solito indicando $q' = 1 - p'$, $q'' = 1 - p''$). È allora $p = p'p'' + r\,\sigma'\sigma''$, ed è $p' = p\,p''$ (indipendenza stocastica) se e solo se $r = 0$. L'equivalenza fra i due concetti di non-correlazione e indipendenza sussiste in generale quando i due numeri aleatori X e Y considerati hanno ciascuno soltanto due valori possibili, x_1 e x_2, y_1 e y_2 (a rigore: se « coincidono » con dei numeri aleatori X' e Y' godenti di tale proprietà: $X \doteq X'$, $Y \doteq Y'$). In ogni altro caso (lo proveremo tra poco) il concetto di indipendenza, è effettivamente più ristretto di quello di non-correlazione.

Nel caso degli eventi si può indicare una formula per r che ne mostra sotto un altro aspetto il significato. La probabilità $P(E'/E'')$ di E' subordinatamente ad E' venga indicata con $p'(1 + \rho)$: allora $\rho = 0$ nel caso di indipendenza, $\rho > 0$ risp. $\rho < 0$ nel caso di correlazione positiva o negativa, ed è anche $P(E''/E') = p''(1 + \rho)$ (tale fatto costituisce sostanzialmente il teorema di Bayes). Possiamo allora scrivere

$$p = p'\,p''(1+\rho) = p'\,p'' + r\,\sigma'\sigma''$$

da cui

[3] Cfr. questo *Supplemento*, A. II, n. 2-3, 1936 (*Il teorema di Pitagora e la Statistica*).

$$r\,\sigma'\sigma'' = \rho\, p'\, p''$$

D'altronde, posto (indicandosi con « $\bar{E}$ » la negazione di « E »),

$$q' = P(\bar{E}'), \quad q'' = P(\bar{E}''), \quad q = P(\bar{E}'\bar{E}''), \quad q = q'q''(1+\tau),$$

è anche

$$q = q'q''(1+\tau) = q'q'' + r\,\sigma'\sigma''$$

e quindi

$$r\,\sigma'\sigma'' = \tau\, q'q''$$

Moltiplicando membro a membro le due espressioni

$$r^2\,\sigma'^2\sigma''^2 = \rho\,\tau\, p'q'p''q'', \quad \text{ma} \quad \sigma'^2 = p'q', \quad \sigma'' = p''q'',$$

e finalmente $$r^2 = \rho\,\tau$$

(e si vede subito che il segno di r è quello di ρ e di τ, che hanno necessariamente il medesimo segno). Il coefficiente di correlazione esprime quindi la media geometrica fra il coefficiente di aumento o diminuzione della probabilità che un evento si verifichi verificandosi l'altro o non si verifichi non verificandosi l'altro.

Un'altra osservazione che ne scende, è che, *assegnate* le *probabilità* p' e p'', il coefficiente di correlazione r non può più variare tra -1 e $+1$, ma soltanto fra $-\sqrt{p'p''/q'q''}$ e $+\sqrt{p'q''/q'p''}$ (supposto $p' < p''$, $p' + p'' \leq 1$; negli altri casi sarebbero da scambiare p' con p'' risp. i p coi q). Il limite inferiore è -1 soltanto se $p' + p'' = 1$, quello superiore è $+1$ soltanto se $p' = p''$ e quindi solo per $p' = p'' = \frac{1}{2}$ i due limiti sono ± 1. Ma è possibile che i due limiti, chiamiamoli r_1 ed r_2, siano comunque prossimi allo zero (basta prendere molto piccola p'); più in generale, scelti ad arbitrio r_1 ed r_2, purchè sia, come necessario, $-1 \leq r_1 < 0 < r_2 \leq 1$, è possibile determinare p' e p'' di modo che

$$r_1 = -\sqrt{p'p''/q'q''}, \quad r_2 = +\sqrt{p'q''/q'p''}$$

(la soluzione è univoca, e precisamente $p' = -\,r_1 r_2 / (1 - r_1 r_2)$, $p'' = -\,r_2 / (r_2 - r_1)$).

8 — Le conclusioni raggiunte pel caso particolare degli « eventi » mostrano che, in generale, *quando siano assegnate le leggi di probabilità* di due numeri aleatori X ed Y, non è più possibile che $r(X, Y)$ assuma un qualunque valore tra -1 e $+1$; è del resto ovvio, poichè allora deve aversi $X = a\,Y + b$ ($a > 0$) risp. $X = a\,Y + b$ ($a < 0$), che i due limiti sono raggiungibili solo se le due funzioni di ripartizione Φ_1 e Φ_2 sono simili ($\Phi_1(\xi) = \Phi_2(a\xi + b)$, $a > 0$), risp. anti-simili ($\Phi_1(\xi) = = 1 - \Phi_2(a\xi + b)$, $a < 0$), risp. (perchè lo siano entrambi) simili e simmetriche ($\Phi_1(\xi) = \Phi_2(a\xi + b) = 1 - \Phi_2(-a\xi + b)$).

In generale, i due valori estremi r_1 ed r_2 si ottengono considerando i due casi estremi in cui X ed Y siano funzioni decrescenti o crescenti l'uno dell'altro; in tale ultimo caso, se X assume il valore ξ tale che $\Phi_1(\xi) = t$, Y assume il valore η tale che $\Phi_2(\eta) = t$; nel caso opposto, a ξ tale che $\Phi_1(\xi) = t$ corrisponde η tale che $\Phi_2(\eta) = 1 - t$. Indicando con $\xi(t)$ ed $\eta(t)$ i valori per cui $\Phi_1(\xi) = t$, risp. $\Phi_2(\eta) = t$ (funzioni inverse), il coefficiente di correlazione risulta nei due casi estremi considerati

$$r = \int_0^1 \xi(t)\eta(t)\,dt \qquad\qquad r = \int_0^1 \xi(t)\eta(1-t)\,dt$$

pur di aver « normalizzato » X e Y così che la speranza matematica ne sia nulla e lo scostamento quadratico medio ne sia unitario ([4]). Per dimostrare che i due casi considerati forniscono effettivamente i valori estremi r_2 ed r_1, osserviamo anzitutto che, se nella distribuzione di probabilità sul piano ξ, η si avessero due probabilità uguali p concentrate in due punti ξ_1, η_1 e ξ_2, η_2 con $\xi_2 > \xi_1$ ma $\eta_2 < \eta_1$, r aumenterebbe del termine $p(\xi_2 - \xi_1)\,(\eta_1 - \eta_2)$ spostando le due probabilità in ξ_2, η_1 e ξ_1, η_2, ciò che non altera le distribuzioni Φ_1 e Φ_2. Salvo qualche ulteriore precisazione e termini correttivi che si possono render trascurabili, il medesimo ragionamento sussiste per probabilità non concentrate- nei due punti ξ_1, η_1 e ξ_2, η_2, ma contenute in loro intorni. Finchè a $\xi(t)$ non si faccia corrispondere $\eta(t)$, si può sempre quindi aumentare r col procedimento d'inversione accennato, e analogamente r si può sempre far diminuire finchè a $\xi(t)$ non si faccia corrispondere $\eta(1 - t)$.

E finiamo completando una dimostrazione lasciata in sospeso:

---

[4] Si noti che $\int \xi\eta\; dt$ può interpretarsi come coseno dell'angolo tra le due funzioni ξ ed η nello spazio funzionale, e quindi il limite superiore di $\cos(X, Y)$ è il coseno di $\xi(t)$ ed $\eta(t)$ nello spazio funzionale. Analoghe considerazioni per il minimo riallacciano la metrica dello spazio dei numeri aleatori e dello spazio funzionale.

che non-correlazione e indipendenza coincidono soltanto se Φ_1 e Φ_2 hanno tutta la probabilità concentrata ciascuna in due soli punti. Fissiamo un valore a (tra 0 ed 1), e facciamo corrispondere a $\xi(t)$

$$\begin{cases} \eta(t) & \text{se} \quad t \leq a \\ \eta(1+a-t) & \text{se} \quad t > a \end{cases}$$

Avremo $$r(a) = \int_0^a \xi(t)\eta(t)dt + \int_a^1 \xi(t)\eta(1+a-t)dt,$$

e al variare di a tra 0 ed 1, $r(a)$ varia con continuità tra gli estremi r_2 ed r_1. Per un certo valore $a = a_0$ sarà $r(a_0) = 0$. Si vede subito che tale distribuzione, non-correlata, non coincide con quella del caso di indipendenza, salvo il caso detto. Si può anche osservare che, per ogni r tra r_1 e r_2, la distribuzione è univocamente determinata per Φ_1 e Φ_2 aventi probabilità concentrata in due soli punti, mentre in ogni altro caso vi sono infinite soluzioni.

9. — Nella presente nota ho sempre usato il termine « correlazione » nel significato che si estrinseca nella misura del « coefficiente di correlazione ». Ma, come si è accennato in principio, detto termine è stato usato in più significati distinti, e diverse sono le opinioni sulla terminologia che sarebbe opportuno adottare per evitare il prolungarsi di ogni possibilità di equivoci. Esaminiamo brevemente la questione, le proposte che sono state fatte e quelle che si potrebbero suggerire.

Uno dei significati in cui è stato usato il termine di « correlazione » è, come detto, quello corrispondente alla « dipendenza stocastica » del calcolo delle probabilità; misurare tale grado di « dipendenza » (o meglio, come è preferibile dire sempre quando si tratta di indici, « determinare un numero che dia una certa idea di tale grado di "dipendenza" ») significa definire un indice che assume i valori estremi nei due casi estremi in cui X e Y sono stocasticamente indipendenti oppure sono funzioni l'uno dell'altro, e negli altri casi valori intermedi, più vicini a questo o a quell'estremo a seconda che, in base a un certo criterio (in larga misura arbitrario) si è indotti a considerare la « dipendenza » come più o meno stretta. A differenza del coefficiente di correlazione r, di cui l'elemento essenziale è il *segno*, un indice di dipendenza (se all'indipendenza e alla dipendenza funzionale si fanno corrispondere i valori 0 e 1) non potrà variare che fra 0 e 1.

Per la terminologia, si potrebbe:

- o stabilire di conservare il termine di « correlazione » nel senso di « dipendenza stocastica » in cui fu impropriamente usato, stabilendo naturalmente di abbandonarlo in ogni senso diverso;
- o usare il termine di « dipendenza », magari con la specificazione « stocastica », come si usa nel calcolo delle probabilità;
- o infine introdurre un termine nuovo e diverso.

Quest'ultima soluzione sembra la migliore, perché il termine di « correlazione » si presta meglio nel senso in cui qui stesso lo abbiamo usato, mentre quello di «dipendenza », se non si vuole ripetere sempre « stocastica », può spesso generare ambiguità con la dipendenza nel senso dell'analisi. Tale inconveniente è divenuto sensibile da quando nel calcolo delle probabilità si parla correntemente di numeri aleatori, per i quali si debbono usare entrambe le nozioni di dipendenza, nell'uno e nell'altro senso; inoltre, è poco opportuno che nozioni così disparate siano contraddistinte tra loro solo da un avverbio. Ciò meglio s'addice alla semplice distinzione di diversi casi particolari della stessa nozione, come nelle locuzioni « linearmente dipendente », « algebricamente dipendente », che esprimono modi particolari di « dipendenza » in un unico senso, quello dell'analisi.

Trovare una nuova parola non è però tanto facile, se si vuole da un lato che essa renda intuitivo il significato, come è pur necessario dato che il termine esprime un concetto di uso corrente nella pratica (p. es. assicurativa) e dovrebbe entrare nel linguaggio comune, e dall'altro che goda di molte almeno fra le possibilità grammaticali di cui è ricco il vocabolo « dipendenza » e che sono utilissime se non necessarie nel calcolo delle probabilità. Esso infatti dà luogo, sia per l'affermazione che per la negazione, al nome, verbo, avverbio, e aggettivo usabile in senso transitivo o riflessivo, come risulta dalla tabella

... *X dipendente da Y* ...	... *X indipendente da Y* ...
... *X* e *Y dipendenti* (*tra loro*) ...	... *X* e *Y indipendenti* (*tra loro*) ...
... *dipendenza* ...	... *indipendenza* ...
... *X dipende da Y* ...	... *X non dipende da Y* ...
... *X, dipendentemente da Y,* ...	... *X, indipendentemente da Y,* ...

Il termine di « connessione », proposto da Gini e Pietra, dà l'aggettivo *connesso* utilizzabile in entrambi i modi (*connessi* (tra loro), e *connesso con*), ma negazione .e verbo mi sembrano inutilizzabili (*connettere*, *sconnesso*). Quanto al senso, *connessione* richiama piuttosto qualcosa di rigido, è dire per es. che dei rischi sono connessi non mi pare si attagli bene all'idea.

Si potrebbe proporre « influenza » (*influenzato da*, *influente su*, *influenzantisi*, *influisce su*, *si influenzano*); come senso considererei tale parola come la soluzione ideale, perchè rende impeccabilmente il preciso significato della « dipendenza stocastica », ma ha il grave difetto di non ammettere la negazione: la negazione col « non » (*non-influenzantisi*, ecc.) riuscirebbe alquanto pesante, perchè già ad es. *influenzantisi* è un po' stiracchiato.

Si potrebbe proporre « legame » (*legati* (tra loro), *legato a*) che ha manchevolezze grammaticali analoghe a quelle di « connessione », ma mi sembra un po' più aderente al senso: « legare » è unire in modo più elastico che « connettere », e inoltre, dicendo ad es. che due rischi sono legati tra loro, mi sembra che s'intuisca il significato anche indipendentemente dalla eventuale convenzione d'introdurre tale terminologia nel linguaggio scientifico. I francesi usano già del resto il termine « loi liée » per indicare la distribuzione di probabilità di un numero aleatorio corrispondente a un valore determinato di un altro: se tale numero aleatorio si dice « legato », la locuzione di « distribuzione legata » (traduzione di « loi liée ») risulterebbe di per sè ricollegata al termine esprimente il concetto generale.

Osserviamo ancora che « correlazione » (che abbiamo escluso per ragioni d'altra indole), non si presterebbe grammaticalmente in modo più opportuno.

Esisterà qualche altra parola adatta allo scopo ma più duttile? Meriterebbe di cercarla, prima di prendere una decisione: comunque, dovendo scegliere fra i termini considerati, darei la preferenza a uno dei due ultimi (*influenza* o *legame*).

10. — L'altro significato è quello per il quale Gini e Pietra hanno proposto il termine di « *concordanza* » (risp. *discordanza*, *indifferenza*). Tale concetto è analogo, ma più generale, di quello di *correlazione* nel senso stretto (corrispondente cioè ad r). È analogo perchè riguarda la tendenza di un numero aleatorio ad assumere preferentemente valori maggiori o minori a seconda del valore maggiore o minore assunto dall'altro. Quindi si ha da distinguere il *senso* (concordanza o discordanza), così come per la correlazione (positiva o negativa). Ma il concetto di concordanza è più generale perchè comprende genericamente tutti gli aspetti sotto cui tale tendenza può essere considerata e studiata (aspetti che si estrinsecano in altri indici che furono proposti — come quelli di *omofilia* — che si potranno introdurre).

Data la posizione a sè che ha il coefficiente di correlazione r nel calcolo delle probabilità (e che ho cercato di lumeggiare), ritengo che sarebbe opportuno seguire la proposta di Gini - Pietra adottando il termine di «concordanza » per il senso generico, ma riservando quello di « correlazione » per il solo significato corrispondente ad r. Altrimenti la frase «correlati positivamente » non avrebbe alcun senso definito, perchè non è detto che un altro indice di concordanza. debba risultare sempre positivo negativo o nullo a seconda che sia positivo negativo o nullo r.

Secondo Fréchet, sarebbe invece opportuno chiamare r « indice di linearità » (egli riserverebbe il termine di « correlazione » al significato di « dipendenza stocastica »). Tale denominazione non si presterebbe però a dedurne l'equivalente di « correlato », « correlato positivamente o negativamente », « non-correlato », locuzioni indispensabili, come s'è visto, nel calcolo delle probabilità. Mi sembra inoltre che essa corrisponda a un'idea un po' limitata riguardo al significato di r, che indicherebbe soltanto la tendenza a una « regressione lineare »; il significato fondamentale di r, quale abbiamo cercato di lumeggiare, è invece del tutto indipendente dal fatto che la distribuzione di probabilità abbia o non abbia tendenza ad addensarsi lungo una determinata linea, e che questa, ove esista, sia una retta o una curva. La sola differenza è che, se tale linea è retta, r può dare un'idea dell'addensamento della distribuzione intorno ad essa, se tale linea non esiste nessun equivoco può sorgere, mentre se esiste ed è curva bisogna avvertire di non credere che r consenta ancora la conclusione che sussisteva nel caso di regressione rettilinea.

11. — Ancora una sola osservazione, per chiarire come un indice di concordanza possa non coincidere *neppure come segno* con r, e insieme per suggerire quello che mi sembra forse il più semplice e intrinsecamente significativo indice di concordanza (e che non mi risulta sia stato considerato). Per esprimerlo nel modo più semplice consideriamo un caso particolare: data la distribuzione dei matrimoni secondo le età dei coniugi, consideriamo la probabilità che, scegliendo a caso e indipendentemente due coppie, la più giovane delle mogli sia quella del più giovane dei mariti. Dette X_1 e Y_1, X_2 e Y_2, l'età dello sposo e della sposa nelle due coppie, abbiamo cioè da considerare la probabilità c che

$$(X_1 - X_2)(Y_1 - Y_2) > 0 \ ,$$

e, se $\varphi(x, y)\, dx\, dy$ è la probabilità che le età dei due sposi siano comprese rispettivamente fra x e $x + dx$, y e $y + dy$, si può esprimere

$$c = \int_C \varphi(x, y)\, \varphi(\xi, \eta)\, dx\, dy\, d\xi\, d\eta$$

ove C è il campo definito da $(x - \xi)\,(y - \eta) > 0$, ossia limitato dai piani $x = \xi$, $y = \eta$. Si può considerare c come indice di concordanza: si avrebbe concordanza, discordanza o indifferenza a seconda che c risulti maggiore, minore od uguale a 1/2; sarebbe $c = 1$, risp. $c = 0$, solo nei casi estremi in cui l'età della sposa fosse rispettivamente funzione crescente o funzione decrescente dell'età dello sposo.

Ora è facile vedere che c non ha necessariamente lo stesso segno di r: basta osservare che c rimane inalterato sostituendo ad X e Y due loro funzioni crescenti $f(X)$, $g(X)$ (infatti è allora $[f(X_1) - f(X_2)]\,[g(Y_1) - g(Y_2)]$ dello stesso segno di $[X_1 - X_2]\,[Y_1 - Y_2]$) mentre $r(X, Y)$ e $r[f(X), g(Y)]$ possono avere benissimo segni differenti ([5])).

Nonostante tale diversità di comportamento fra c e r, esiste fra di essi un'analogia di significato che vale la pena d'esser messa in rilievo. Se, oltre al segno di $(X_1 - X_2)\,(Y_1 - Y_2)$ vogliamo considerarne anche il valore per dare a ogni disuguaglianza un peso tanto maggiore quanto maggiori sono le differenze $X_1 - X_2$ e $Y_1 - Y_2$, all'integrale che esprime c andrebbe sostituito il seguente

$$\int (x - \xi)(y - \eta)\varphi(x, y)\varphi(\xi, \eta)\, dx\, dy\, d\xi\, d\eta =$$
$$= M[X_1 Y_1 + X_2 Y_2 - X_1 Y_2 - X_2 Y_1]$$

ed è facile vedere che tale espressione vale

---

[5] Ad es., se i valori possibili per (X, Y) sono

$(-1,1)$	con probabilità	1/4
$(0, -1)$	" "	1/2
$(a, 1)$	" "	1/4

è $r(X, Y) = 0$ per $a = 1$; variando a (tra 0 e ∞ esclusi), ossia. sostituendo X con la sua funzione crescente $g_a(X) = X +$ **Errore.** $X(X + 1)$,
si ha invece $r > 0$ per $a > 1$, $r < 0$ per $a < 1$.

$$2\sigma(X)\sigma(Y)r(X,Y).$$

Il significato di c può essere considerato pertanto come l'analogo di quello di r, quando si voglia prescindere dalla *grandezza* delle disuguaglianze per osservarne solamente il *senso*.