GIORGIO VITTADINI

UN INDICE NORMALIZZATO PER LA MISURA DELLA DISUGUAGLIANZA FRA LE COMPONENTI DI UNA VARIABILE CASUALE BIDIMENSIONALE

(Estratto dagli "Annali della Facoltà di Scienze Politiche dell'Università di Milano", III - 1983)

MARZORATI EDITORE MILANO - 1983

GIORGIO VITTADINI
Collaboratore ad attività di ricerca presso l'Istituto di Scienze Statistiche e Matematiche « M. Boldrini »

UN INDICE NORMALIZZATO PER LA MISURA DELLA DISUGUAGLIANZA FRA LE COMPONENTI DI UNA VARIABILE CASUALE BIDIMENSIONALE

1. Una recente rilettura di quanto è stato scritto sulla Dissomiglianza — concetto introdotto da C. Gini nel lontano 1914 e ripreso in seguito da altri autori fra cui, in particolare, T. Salvemini e G. Landenna (1) — ha suggerito l'idea di proporre un semplice indice per misurare la « disuguaglianza » che caratterizza le componenti di una variabile casuale bidimensionale (X,Y).

La presentazione di tale indice verrà fatta nelle righe che seguono dopo un breve richiamo sulla classe di Fréchet delle funzioni di ripartizione bidimensionale F(x,y) aventi prefissate marginali $G_1(x)$ e $G_2(y)$.

2. Sia (X,Y) una variabile casuale bidimensionale le cui componenti assumono la stessa serie di modalità $t_1, t_2, \dots, t_k$, caso al quale ci si può sempre ricondurre aggiungendo alle determinazioni della X gli eventuali valori assunti solo dalla Y e viceversa, ed attribuendo probabilità nulla ai nuovi valori.

Siano inoltre:

$$F(t_i,t_j) = \Pr.\ (X \leqslant t_i, Y \leqslant t_j),$$

$$G_1(t_i) = \Pr.\ (X \leqslant t_i),$$

$$G_2(t_j) = \Pr.\ (X \leqslant t_j),$$

le funzioni di ripartizione rispettivamente di (X,Y), di X e di Y e $f(t_i,t_j)$, $g_1(t_i)$ e $g_2(t_j)$ le corrispondenti funzioni di probabilità. È evidente che la predetta $F(t_i,t_j)$ può sempre essere intesa come una funzione della classe di Fréchet (2) generata associando fra loro, in tutti i modi possibili, le componenti X ed Y di (X,Y) aventi prefissate funzioni di ripartizione $G_1(t_i)$ e $G_2(t_j)$, classe alla quale, come è noto, appartengono le due funzioni $F^0(t_i,t_j)$ e $F^1(t_i,t_j)$ che in corrispondenza di ogni coppia (t_i,t_j) assumono rispettivamente:

(1) $F^0(t_i,t_j)$ il valore minore fra $G_1(t_i)$ e $G_2(t_j)$,

(2) $F^1(t_i,t_j)$ il valore maggiore fra 0 e $G_1(t_i)+G_2(t_j) - 1$.

3. Ciò premesso, si definiscano « uguali » le componenti di (X,Y) se, al variare delle prove, assumono determinazioni la cui differenza è sempre nulla.

Ne discende che se X e Y sono « uguali » nel senso suddetto si verifica necessariamente che:

$$
\begin{aligned}
&F(t_i,t_i) = G_1(t_i) = G_2(t_i), && (i = 1, 2, ..., k)\\
(3)\qquad &f(t_i,t_i) = g_1(t_i) = g_2(t_i), && (i = 1, 2, ..., k)\\
&f(t_i,t_j) = 0 && (i \neq j)
\end{aligned}
$$

e, di conseguenza, che il valor medio:

$$(4)\qquad d_{xy} = \sum_{i=i}^{k} \sum_{j=i}^{k} | t_i - t_j | \, f(t_i,t_j)$$

risulta nullo; cioè: $d_{xy} = 0$.

Se le componenti di (X,Y) non sono fra loro « uguali », vale a dire sono « disuguali », non valgono allora le condizioni (3) e si ha di conseguenza: $d_{xy} > 0$.

Stante il fatto che è possibile dimostrare (3), che il valor medio (4) soddisfa alle condizioni:

a) di essere simmetrico rispetto alle due variabili casuali X ad Y e sempre non negativo,

b) di annullarsi se, e solo se, X ed Y sono « uguali »,

c) di soddisfare alla relazione triangolare della distanza,

viene naturale assumere la (4) medesima come indice per la misura della « disuguaglianza » fra le due componenti di (X,Y).

4. Deve ora osservarsi che l'indice (4) assume il suo « minimo » valore (4) se la funzione $F(t_i,t_j)$ che caratterizza (X, Y) coincide con la funzione $F^0(t_i,t_j)$ definita con la (1).

In tal caso, deve ritenersi che le due componenti di (X,Y) presentano fra loro la « minima » disuguaglianza la cui misura è fornita da:

$$(5)\qquad d^0_{xy} = \sum_{i=i}^{k} \sum_{j=i}^{k} | t_i - t_j | \, f^0(t_i,t_j) = \min. d_{xy}$$

dove $f^0(t_i,t_j)$ è la funzione di probabilità associata a $F^0(t_i,t_j)$.

Ed è charo che sarà $d^0_{xy} = 0$ se, e solo se, le componenti di (X, Y) sono fra loro « uguali ».

D'altro canto, l'indice (4) assume il suo « massimo » valore (4), se la funzione $F(t_i,t_j)$ che caratterizza (X,Y) coincide con la funzione $F^1(t_i,t_j)$ definita con la (2).

In tal caso deve ritenersi che le due componenti di (X,Y) presentano fra loro la « massima » disuguaglianza la cui misura è fornita da:

$$(6)\qquad d^1_{xy} = \sum_{i=i}^{k} \sum_{j=i}^{k} | t_i - t_j | \, f^1(t_i,t_j) = \max. d_{xy}$$

dove $f^1(t_i,t_j)$ è la funzione di probabilità associata a $F^1(t_i,t_j)$. Dalle (4), (5) e (6) discende che si può scrivere:

$$(7)\qquad 0 \leqslant d^0_{xy} \leqslant d_{xy} \leqslant d^1_{xy}$$

e tale serie di disuguaglianze consente, in via naturale, di costruire il seguente indice normalizzato per la misura della disuguaglianza fra X ed Y:

$$(8)\qquad D_{xy} = \frac{d_{xy} - d^0_{xy}}{d^1_{xy} - d^0_{xy}}$$

Detto indice, infatti, assume: valore 1, se X ed Y presentano fra loro la « massima » disuguaglianza, nel senso che rendono massimo il valore medio (4); assume valore 0, se X ed Y presentano la « minima » disuguaglianza nel senso che rendono minimo lo stesso valore (4). E, in particolare, si dirà che X ed Y sono fra loro « uguali », se $d_{xy} = 0$ nel qual caso è anche necessariamente $d^0_{xy} = 0$.

BIBLIOGRAFIA

(1) C. Gini, *Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche*, « Atti del R. Ist. Ven. di Scienze, Lettere ed Arti », Tomo LXXIV, 1914-15; *Nuovi contributi alla teoria delle relazioni statistiche*, « Atti del R. Ist. Ven. di Scienze, Lettere ed Arti », Tomo LXXIV, 1914-15; T. Salvemini, *Nuovi procedimenti per il calcolo degli indici di dissomiglianza e di connessione*, « Statistica », 1949; *Dissomiglianza a più dimensioni*, « Atti della Soc. Ital. Stat. », 1951; *L'indice di dissomiglianza tra distribuzioni continue*, « Metron », vol. XVI, 1952; *Su alcuni aspetti della dissomiglianza e della concordanza*, « Riv. Ital. Stat. », vol. VII, 1953; *L'indice quadratico di dissomiglianza tra di-*

stribuzione gaussiane, « Atti della Soc. Ital. Stat. », 1954; G. LANDENNA, *La dissomiglianza*, « Statistica », 1956;
(2) M. FRÉCHET, *Sur les tableaux de correlation dont les marges sont données*, « Annales de l'Université de Lyon », s. III, fasc. 4: « Sciences », 1951;
(3) G. LANDENNA, *op. cit.*, (1).
(4) G. DALL'AGLIO, *Sugli estremi dei momenti delle funzioni di ripartizione doppia*, « Annali della Scuola Normale Superiore », 1956.

SUMMARY: *Index of dissimilarity of two dimensional random variable components.*

In this paper, following the concept of dissimilarity, introduced by C. Gini and closer examined by T. Salvemini and G. Landenna, an index of dissimilarity of two-dimensional random variable components is proposed. This index is introduced by the « class of Fréchet » of the two dimensional distribution function.

SOMMAIRE: *Un indice normalisé pour mesurer la dissimilarité entre les composantes d'une variable aléatoire bidimensionnelle.*

Une récente réflexion sur le concept de dissimilarité, introduit par C. Gini et, pour le suite, par d'autres auteurs (T. Salvemini et G. Landenna) a suggéré de proposer une nouvelle mesure de dissimilarité entre les deux composants de une variable aléatoire bidimensionelle. Cet indice est introduit à partir de la « classe de Fréchet » concernant les variables bidimensionelles.