Riv. Mat. Univ. Parma (2) 8 (1967), 189-195

L. ARLOTTI e N. PINTACUDA (*)

Probabilità e Informazione. (**)

Introduzione.

Nel presente lavoro ci riferiremo costantemente per quel che concerne i concetti e la terminologia agli assiomi della teoria della informazione senza probabilità, quale è esposta nei lavori di J. KAMPÉ DE FÉRIET e B. FORTE [1].

Da tali assiomi rimangono definite su una σ-algebra $\mathscr{F}$: una funzione J, che ha tutte le proprietà intuitive di una informazione, e viene caratterizzata essenzialmente tramite la relazione funzionale

$$J(A \cup B) = F\big(J(A),\ J(B)\big) \qquad \text{per} \qquad A \cap B = \varnothing;$$

e una funzione P che può considerarsi una probabilità generalizzata su $\mathscr{F}$, comportandosi su $\mathscr{F}$ come una probabilità, eccetto per la circostanza che in generale

$$P(A \cup B) \neq P(A) + P(B) \qquad \text{per} \qquad A \cap B = \varnothing$$

(diremo che la funzione P è, in generale, *non additiva*).

È ovvio far rilevare che il sistema assiomatico è autocompatibile, e che vi rientra il caso di uno spazio di probabilità di KOLMOGOROV finito, nel quale si assuma, in accordo con la formula di HARTLEY,

$$J(A) = -\log P(A)\,,$$

risultando in tale caso

$$J(A \cup B) = -\log\left[e^{-J(A)} + e^{-J(B)}\right].$$

(*) Indirizzi: L. ARLOTTI, Seminario Matematico, Università, Ferrara, Italia. N. PINTACUDA, Istituto Matematico, Università, Pavia, Italia.

(**) Ricevuto: 7-III-1967.

1. - Probabilità e informazione.

Un esame delle proprietà formali degli assiomi di KAMPÉ DE FÉRIET e FORTE mette in luce una loro simmetria intrinseca: è facile vedere che la trasformazione $P \to \exp(-hP)$ con $h > 0$ lascia invariato il sistema assiomatico nel suo complesso, scambiando ordinatamente gli assiomi per la funzione P con i corrispondenti assiomi per la funzione J.

Si può pertanto asserire a priori, e una verifica diretta lo conferma, che se $J\colon \mathscr{F} \to \overline{R}^+$ è un'applicazione che soddisfa agli assiomi per la J, la scrittura

$$P(X) = \exp\{-h\, J(X)\}, \qquad h > 0\,,$$

definisce un'applicazione $P\colon \mathscr{F} \to [0, 1]$ che rende soddisfatti gli assiomi per la P. E viceversa, se $P\colon \mathscr{F} \to [0, 1]$ è una tale funzione, la scrittura

$$J(X) = -k \log P(X), \qquad k > 0\,,$$

definisce una $J\colon \mathscr{F} \to \overline{R}^+$ che soddisfa agli assiomi per J.

Non è detto invece che, date due applicazioni $J\colon \mathscr{F} \to \overline{R}^+$, $P\colon \mathscr{F} \to [0, 1]$ verificanti gli assiomi, interceda fra esse una relazione del tipo [2]

$$P(X) = \exp\{-h\, J(X)\}, \qquad h > 0.$$

In effetti si mostra facilmente che ciò non accade, se su una σ-algebra $\mathscr{F}$ munita di una probabilità P definiamo una J caratterizzata dalla proprietà

$$J(A \cup B) = \inf[J(A),\ J(B)], \qquad A \cap B = \emptyset\,.$$

La teoria della informazione senza probabilità presenta a questo punto un duplice quesito:

1) determinare sotto quali condizioni le funzioni P e J risultano legate da una relazione $P = \exp(-h\, J)$, $h > 0$;

2) determinare, in un'algebra $\mathscr{F}$ munita di probabilità (additiva) P, quali sono le funzioni J che soddisfano agli assiomi, una volta che si assuma come indipendenza degli elementi di $\mathscr{F}$ l'indipendenza *stocastica* (brevemente: quali sono le funzioni di informazione « compatibili con la probabilità »).

Il presente lavoro dà una risposta a questi due quesiti.

2. - Struttura dell'informazione compatibile con una probabilità.

Diremo che due elementi $X \in \mathscr{F}$, $Y \in \mathscr{F}$ sono P-indipendenti, se $P(X \cap Y) = P(X)\,P(Y)$, e che sono J-indipendenti se $J(X \cap Y) = J(X) + J(Y)$.

Se $P = \exp(-h\,J)$, chiaramente le nozioni di P-indipendenza e di J-indipendenza coincidono.

È chiaro allora che: condizioni necessarie perchè sussista la relazione $P = \exp(-h\,J)$ fra una P e una J sono che

1°) P risulti funzione di J,

2°) gli elementi di $\mathscr{F}$ P-indipendenti siano J-indipendenti, e viceversa.

Ci proponiamo quindi di formulare delle condizioni sufficienti; e faremo vedere che se P, oltre a verificare le due condizioni necessarie, è una funzione additiva su $\mathscr{F}$, e se inoltre la funzione $P = P(J) = \psi(J)$ è definita in un intervallo ed ivi monotona e provvista di derivata prima misurabile secondo LEBESGUE, allora:

1°) la relazione funzionale fra P e J è del tipo

$$P = \exp(-h\,J), \qquad h > 0\,.$$

2°) la struttura di J è data dalla «regola di composizione»

$$J(A \cup B) = -\frac{1}{\alpha}\log[e^{-\alpha\,J(A)} + e^{-\alpha\,J(B)}]\,, \qquad A \cap B = \varnothing,\ \alpha > 0\,.$$

3. - Dimostrazione.

Posto $J(A \cup B) = F\big(J(A),\ J(B)\big)$ per $A \cap B = \varnothing$, cominciamo col provare il seguente:

L e m m a: $F(x,\ y) = \psi^{-1}[\psi(x) + \psi(y)]$.

P r o v a. Poichè la funzione $\psi\big(J(X)\big)$ per $X \in \mathscr{F}$ è additiva su $\mathscr{F}$, per $A \cap B = \varnothing$, deve aversi

$$\psi\big(J(A \cup B)\big) = \psi\big(J(A)\big) + \psi\big(J(B)\big);$$

d'altronde

$$\psi\big(J(A \cup B) = \psi\big(F\big(J(A),\ J(B)\big)\big);$$

uguagliando e ponendo $x = J(A)$, $y = J(B)$

$$\psi(F(x, y)) = \psi(x) + \psi(y).$$

Ma poichè ψ, in quanto monotona, è invertibile

$$F(x, y) = \psi^{-1}[\psi(x) + \psi(y)]. \tag{1}$$

L'aver determinato la forma della $F(x, y)$ semplifica il compito di imporre che la funzione $J(X) = \psi^{-1}(P(X))$ soddisfi agli assiomi per l'informazione. Per la sua stessa forma, la funzione $F(x, y)$ è — come facilmente si verifica — simmetrica, associativa e semicontinua inferiormente; inoltre le condizioni $F(x, y) \geqslant 0$, $F(x, y) \leqslant x$, e $\psi(+\infty) = 0$, che si traggono dagli assiomi, sono compatibili se e solo se $\psi(x)$ è una funzione positiva decrescente, nulla all'infinito.

Rimane dunque da imporre la sola condizione che, se $F(x, x^c) = 0$, sia

$$F(x + y, x + y^c) = x. \tag{2}$$

Si ha, dalla (1),

$$\psi^{-1}(\psi(x) + \psi(x^c)) = 0, \qquad \psi(x) + \psi(x^c) = \psi(0),$$

$$\psi(x^c) = \psi(0) - \psi(x), \qquad x^c = \psi^{-1}[\psi(0) - \psi(x)]$$

e, sostituendo nella (2),

$$\psi^{-1}\left\{\psi(x + y) + \psi(x + \psi^{-1}[\psi(0) - \psi(y)])\right\} = x,$$

$$\psi(x + y) = \psi(x) - \psi\left\{x + \psi^{-1}[\psi(0) - \psi(y)]\right\}. \tag{3}$$

Gli assiomi dell'informazione per J si traducono dunque nell'equazione funzionale (3) per la ψ; di quest'equazione cercheremo le soluzioni positive, decrescenti, nulle all'infinito e provviste di derivata prima misurabile. In dette ipotesi è lecito derivare rispetto ad x la (3):

$$\psi'(x + y) = \psi'(x) - \psi'\left\{x + \psi^{-1}[\psi(0) - \psi(y)]\right\},$$

ossia

$$\psi'\left\{x + \psi^{-1}[\psi(0) - \psi(y)]\right\} = \psi'(x) - \psi'(x + y) \tag{4}$$

e per $x = 0$

$$\psi'\{\psi^{-1}[\psi(0) - \psi(y)]\} = \psi'(0) - \psi'(y) . \tag{5}$$

Derivando la (3) rispetto ad y, invece, si ha

$$\psi'(x + y) = \psi'\{x + \psi^{-1}[\psi(0) - \psi(y)]\} \frac{\psi'(y)}{\psi'\{\psi^{-1}[\psi(0) - \psi(y)]\}}$$

e, tenendo conto delle (4), (5),

$$\psi'(x + y) = \psi'(y) \frac{\psi'(x) - \psi'(x + y)}{\psi'(0) - \psi'(y)}$$

ossia

$$\psi'(x + y)[\psi'(0) - \psi'(y) + \psi'(y)] = \psi'(x)\, \psi'(y) ,$$

$$\psi'(x + y)\, \psi'(0) = \psi'(x)\, \psi'(y) .$$

Poniamo ora $G(t) = \psi'(t)$, $H(t) = G(t)/G(0)$; da cui

$$G(t) = H(t)\, G(0), \qquad H(0) = 1.$$

L'equazione funzionale diventa

$$G(x + y)\, G(0) = G(x)\, G(y) ,$$

$$H(x + y)\, G^2(0) = H(x)\, H(y)\, G^2(0) ,$$

$$H(x + y) = H(x)\, H(y) . \tag{6}$$

La (6) è una classica equazione funzionale di CAUCHY, la cui soluzione generale misurabile secondo LEBESGUE è [3]: $H(x) = c\, e^{-\alpha x}$, con c, α costanti. Ne segue $G(x) = \text{cost.}\, e^{-\alpha x}$, da cui

$$\psi'(x) = \text{cost.}\, e^{-\alpha x} ,$$

$$\psi(x) = -\lambda\, e^{-\alpha x} + \text{cost.}, \quad \lambda = \text{cost.} .$$

Poichè ψ dev'essere decrescente e nulla all'infinito, $\alpha > 0$, la costante additiva si annulla; e si avrà:

$$\psi(x) = -\lambda\, e^{-\alpha x}, \qquad \alpha > 0 ,$$

$$\psi^{-1}(t) = -\frac{1}{\alpha} \log\left(-\frac{1}{\lambda}\, t\right) .$$

Sostituendo queste espressioni nella (1) si ottiene per F:

$$F(x,\, y) = -\frac{1}{\alpha} \log\left[-\frac{1}{\lambda}(-\lambda\, e^{-\alpha x} - \lambda\, e^{-\alpha y})\right],$$

(7) $$F(x,\, y) = -\frac{1}{\alpha} \log(e^{-\alpha x} + e^{-\alpha y}) ,$$

che costituisce il risultato annunciato.

4. - Conclusione.

Dato che nella (7) non compare la costante λ, possiamo senz'altro, senza pregiudizio di generalità, porre $\lambda = -1$.

Abbiamo così

$$J = -\frac{1}{\alpha} \log P .$$

Il parametro α corrisponde all'arbitrarietà nella scelta dell'unità di misura dell'informazione, e con una adeguata scelta della base del logaritmo può ridursi all'unità.

Otteniamo così per l'informazione l'espressione classica

$$J = \log\,(1/P) .$$

Risulta con ciò provato nel presente lavoro che, sotto opportune condizioni qualitative sulla relazione funzionale $J = J(P)$, la sola informazione compatibile con una probabilità su $\mathcal{F}$ resta l'informazione di HARTLEY.

5. – Gli autori desiderano esprimere il loro ringraziamento ai Proff. FORTE e KAMPÉ DE FÉRIET per i loro insegnamenti, ai quali è attinto il materiale del presente lavoro.

La prova del Lemma del n. 3 è dovuta alla Dott. ARLOTTI, la successiva dimostrazione è del Dott. PINTACUDA.

Bibliografia.

[1] J. Kampé de Fériet, B. Forte, ... (in corso di stampa).

[2] B. Forte, *Sulle pseudoprobabilità associate all'informazione*, Ann. Univ. Ferrara **12** (1966).

[3] J. Aczel, *Lectures on Functional Equations*, New York 1965.

Sommario.

In connessione con la teoria dell'informazione senza probabilità, viene formulata in questo lavoro una condizione sufficiente per la compatibilità della funzione di informazione con la probabilità definita su una σ-algebra.

Si trova, sotto condizioni piuttosto generali, che l'unica definizione possibile dell'informazione è quella classica di Hartley-Shannon.

Summary.

A sufficient condition is given, in the frame of the theory of information without probability, in order to ensure compatibility between the information function and a probability measure over a σ-algebra.

The classical Hartley-Shannon formula is shown to be, under quite reasonable assumptions, the only possible definition of an information function.

* * *