Stochastica Vol. VI nº 1 (1982)

INFORMAZIONE RELATIVA IN UNO SPAZIO CON LEGGE D'INDIPENDENZA QUALSIASI.

Carla Poggi

ABSTRACT

The notion of relative measure of information in an abstract information space with generalized independence law is studied. The axiomatic definition is given and the form of dependence on the absolute measures is determined, as a solution of a system of functional equations.

1. Generalita'.

Uno spazio d'informazione [1,2] è una struttura $I=(\Omega,\sigma,\varepsilon,\kappa,*,H)$ in cui:

Ω è un insieme qualunque.

σ è un'algebra di parti di Ω.

ε è una famiglia di partizioni Π_A di insiemi $A\in\sigma$.

κ è una famiglia di classi di elementi di ε tra loro algebraicamente indipendenti (v. [3]).

$*$ è la legge di indipendenza propia dello spazio di informazione considerato (v. [4]).

H: $\varepsilon \to R^+$ è un'applicazione (misura di informazione) che gode delle seguenti proprietà:

(1.1) $H(\{\Omega\})=0$, $H(\{\phi\})=+\infty$. (Si è indicato con $\{A\}$ la partizione impropia di A).

(1.2) $\Pi_A, \Pi_B \in \varepsilon$, $\Pi_B < \Pi_A \rightarrow H(\Pi_A) \geqslant H(\Pi_B)$. ($\Pi_B < \Pi_A$ se $\Pi_A = \{A\}$, $\Pi_B = \{B\}$ e $A \subset B$ oppure se $A=B$ e Π_A è un raffinamento di $_B$).

(1.3) $(\Pi_A, \Pi_B) \in \kappa \rightarrow H(\Pi_A \wedge \Pi_B) = H(\Pi_A) * H(\Pi_B)$.

$(\Pi_A \wedge \Pi_B = \{C_{rs} \neq \phi : C_{rs} = A_r \cap B_s\} = \Pi_{A \cap B}$ (v. [1])).

E' stato dimostrato che ogni legge di indipendenza corrisponde ad una generica successione (finita o no) $\{]\alpha_i, \beta_i[\}$ di intervalli aperti e disgiunti di R^+ (limitati o no) su ognuno dei quali è definita una funzione continua $g_i(x)$ (reale positiva) strettamente crescente e nulla in α_i in modo che per x,y in R^+ sia:

$$(1.4) \qquad x * y = \begin{cases} \bar{g}_i[g_i(x)+g_i(y)] & (x,y) \in]\alpha_i, \beta_i[, \\ \sup(x,y) \text{ altrimenti.} & \end{cases}$$

$\bar{g}_i$ è la funzione (pseudonversa di g_i) definita ponendo

$$\bar{g}_i(x) = \begin{cases} g_i^{-1}(x) & \text{se } x \in [g_i(\alpha_i)=0,\ g_i(\beta_i)] , \\ \beta_i & \text{se } x > g_i(\beta_i) . \end{cases}$$

(Si è posto ovviamente $g_i(\alpha_i) = \lim_{x \to \alpha_i^+} g_i(x)$ e analogamente per $g_i(\beta_i)$).

L'insieme $\Lambda = R^+ - \bar{\Lambda}$ dove $\bar{\Lambda} = \bigcup_i]\alpha_i, \beta_i[$ è costituito dagli elementi $x \in R^+$ (idempotenti) per cui $x*x=x$.

Se $\beta_i = +\infty$, la funzione $g_i(x)$ è ovviamente non limitata.

2. Informazione relativa.

Siano Π' e Π'' due partizioni ($\in \varepsilon$) di un medesimo sottoinsieme $A \in \sigma$.

In questa nota si vuole dare una definizione plausibile di "informazione relativa" di Π' rispetto a Π'', che potremo indicare con $H(\Pi'/\Pi'')$. Essa serà un'aplicazione $K(\Pi',\Pi'')$ a valore in R^+ definita su un sottoinsieme $\bar{\varepsilon}$ di $\varepsilon\times\varepsilon$, ($\bar{\varepsilon}$ è costituito dalle coppie di elementi di ε che sono entrambi partizioni di un medesimo insieme di σ), cui sembra ragionevole imporre di soddisfare le seguenti proprietà:

a) La restrizione di K al sottoinsieme di $\bar{\varepsilon}$ i cui elementi sono del tipo $(\Pi_A,\{A\})$ $(A\in\sigma)$ determina un'applicazione K' di ε in R^+, se si pone $K'(\Pi_A)=K(\Pi_A,\{A\})$, che deve essere ancora una misura di informazione (informazione condizionale) con la stessa legge di indipendenza di I.

b) L'informazione relativa di una partizione di A rispetto ad un'altra più fine o uguale ad essa deve essere nulla.

Dovrà dunque aversi per la a)

(2.1) $\quad \Pi''_A < \Pi'_A \rightarrow H(\Pi'_A/\{A\}) > H(\Pi''_A/\{A\})$.

(2.2) $\quad (\Pi_A,\Pi_B)\in\kappa \rightarrow H(\Pi_A\wedge\Pi_B/\{A\cap B\}) = H(\Pi_A/\{A\}) * H(\Pi_B/\{B\})$.

e per la b)

(2.3) $\quad \Pi'' < \Pi' \quad H(\Pi''/\Pi') = 0$.

Nel seguito indendiamo caratterizzare le misure di informazione relativa $H(\Pi'/\Pi'')$ che dipendono da Π' e Π'' tramite le misure di informazione assoluta $H(\Pi'\wedge\Pi'')$ e $H(\Pi'')$ (vedi [4]):

$$H(\Pi'/\Pi'') = F\{H(\Pi'\wedge\Pi''),H(\Pi'')\}$$

dove la funzione $F(x,y)$ abbia carattere universale. Le (2.1), (2.2) e (2.3) con tale ipotesi si traducono allora nel seguente sistema di equazioni funzionali nell'incognita $F(x,y)$

(2.4) $\quad x_1 < x_2 \rightarrow F(x_1,y) \leqslant F(x_2,y) \qquad \forall\ x_1,x_2 \geqslant y$

(2.5) $\quad F(x_1 * x_2, y_1 * y_2) = F(x_1,y_1) * F(x_2,y_2) \quad \forall\ x_1 \geqslant y_1, x_2 \geqslant y_2$

(2.6) $\quad F(x,x) = 0.$

3. La funzione $F(x,y)$.

Poniamo

$$L = \{(x,y): y<x\} \quad L \subset (R^+)^2, \quad \bar{L} = \{(x,y): y=x\}$$

$$a = \sup\{x : x \in \Lambda\} \quad (a \leqslant +\infty).$$

Poichè $H(\Pi' \wedge \Pi'') \geqslant H(\Pi'')$ la funzione $F(x,y)$ è definita sull'insieme $L \cup \bar{L}$.

<u>Lemma 1.</u> a) Se Λ è limitato

$$(3.1) \qquad \begin{aligned} F(x,y) &= 0 && \forall (x,y) \in \{[0,a]^2\} \cap L \\ F(x,y) &= F(x,a) && \forall (x,y)\ \{[a,+\infty[\times [0,a]\} \cap L \end{aligned}$$

b) Se Λ è non limitato, allora $F(x,y) = 0 \quad \forall (x,y) \in (R^+)^2 \cap L.$

<u>Dimostrazione.</u> Poichè $F(x,y) = F(x,y) * 0$ e $F(a,a) = 0$, se $x>y$ per la (2.5) si ha: $F(x,y) = F(x,y) * F(a,a) = F(x*a, y*a)$ e quindi

$$\begin{aligned} F(x,y) &= F(a,a) = 0 && \text{se } (x,y) \in [0,a]^2 \cap L \\ F(x,y) &= F(x,a) && \text{se } (x,y) \in \{[a,+\infty[\times [0,a]\} \cap L \end{aligned}$$

In modo del tutto analogo si prova b): introducendo in corrispondenza ad (x,y) un elemento $\lambda \in \Lambda$ per cui sia $\lambda \geqslant x > y$ si trova $F(x,y) = F(\lambda,\lambda) = 0$.

In virtù della (2.6) e del lemma 1 per determinare la funzione $F(x,y)$ possiamo limitarci a studiarla sul sottoinsieme di $(R^+)^2$ in cui $a \leqslant y < x$.

<u>Lemma 2.</u> Se $a \leqslant y < x$, $F(x,y)$ è funzione della sola differenza $g(x)-g(y)$ ovvero può porsi

$$F(x,y) = f[g(x)-g(y)] \tag{3.2}$$

dove $g(x)$ è la funzione che individua l'operazione $*$ in $[a,+\infty[$.

<u>Dimostrazione.</u> Se $a \leqslant y < x$ allora $x*y = g^{-1}[g(x)+g(y)]$. Se $z>a$ $F(x,y) = F(x,y)*F(z,z) = F(x*z, y*z) = F\{g^{-1}[g(x)+g(z)], g^{-1}[g(y)+g(z)]\}$. Ponendo allora $g(x)=\xi$, $g(y)=\eta$, $g(z)=\zeta$ risulta

$$F\{g^{-1}(\xi+\zeta), g^{-1}(\eta+\zeta)\} = F\{g^{-1}(\xi), g^{-1}(\eta)\}$$

Per la funzione $G(\xi,\eta) = F\{g^{-1}(\xi), g^{-1}(\eta)\}$ deve dunque aversi per ogni valore positivo di ξ,η,ζ:

$$G\{\xi+\zeta,\ \eta+\zeta\} = G(\xi,\eta) \text{ da cui segue subito:}$$

$G(\xi,\eta) = f(\xi-\eta)$ ovvero $F(x,y) = f[g(x)-g(y)]$.

Nel seguente paragrafo caratterizzeremo la funzione $f(u)$ per completare lo studio di $F(x,y)$.

4. La funzione f(u).

<u>Lemma 3.</u> $f(u)$: $]0,+\infty[\to R^+$ gode delle seguenti propietà:

$$u<v \quad . \quad f(u) \leqslant f(v) \tag{4.1}$$

(4.2) $$f(u+v) = f(u)*f(v)$$

<u>Dimostrazione.</u> La (4.1) è conseguenza immediata della (2.4). Per dimostrare la (4.2) osserviamo che se $a \leqslant y_1 < x_1$ e $a \leqslant y_2 < x_2$ si ha per le (2.5) e (3.2):

$$f[g(x_1*x_2)-g(y_1*y_2)] = f[g(x_1)-g(y_1)]* f[g(x_2)-g(y_2)]$$

ovvero ricordando l'espressione di x_1*x_2 e di y_1*y_2 (v. (1.5)):

(4.3) $$f[g(x_1)+g(x_2)-g(y_1)-g(y_2)]=f[g(x_1)-g(y_1)]*f[g(x_2)-g(y_2)]$$

Se poniamo quindi $u=g(x_1)-g(y_1)$ e $v=g(x_2)-g(y_2)$ per le (4.3) è:

$$f(u+v)=f(u) * f(v)$$

(Per la continuità di $g(x)$ infatti è sempre possibile variando x_1, y_1, x_2, y_2 far variare u e v in tutto R^+).

Siano

$f_o = \lim_{u \to 0^+} f(u)$ $(f_o \geqslant 0)$

$\beta = \inf \{\lambda \in \Lambda : \lambda > f_o\}$ $(\beta \geqslant f_o)$

$p = \inf \{u > 0 : f(u) \geqslant \beta\}$ $(0 \leqslant p \leqslant +\infty)$

<u>Lemma 4.</u> Nell'intervallo aperto $]p,+\infty[$ si ha $f(u)=\text{cost.}= \beta$.

<u>Dimostrazione.</u> In base alle posizioni fatte si ha per qualinque $u>p$: $f(u) \geqslant \beta$.

Se d'altra parte per qualche $u>p$ fosse strettamente $f(u)>\beta$ si potrebbe trovare $\lambda \in \Lambda : f(u) > \lambda > f_o$ (Λ è chiuso, quindi $\beta \in \Lambda$; se $\beta = f_o$ allora β è per Λ punto di accumulazione). Posto allora:

$\varepsilon=\lambda-f_o>0$ e preso $n\in N$ abbastanza grande che risulti $f(u/n)-f_o<\varepsilon$, ovvero $f(u/n)<\lambda$ (si ricordi la definizione di f_o) si avrebbe per la (4.2):

$$f(u)=f(nu/n)=f(u/n)*f(u/n)*f(u/n)*\ldots\leqslant\lambda$$

che è assurdo essendosi supposto $f(u)>\lambda$.

Lemma 5. Se $p>0$ (e quindi $f_o<\beta$) allora

$$0<u<p \rightarrow f(u)=h^{-1}[u.h(\beta)/p]$$

dove $h(x)$ è la funzione che definisce l'operazione $*$ in $]f_o,\beta[$.

Dimostrazione. $\forall u,v$ con $0<u,v;u+v<p$ si ha per la (4.2)
$f(u+v)=f(u)*f(v)=h^{-1}\{h[f(u)]+h[f(v)]\}$ ovvero

$$h.f(u+v)=h.f(u) + h.f(v)$$

La funzione $h.f(u)$ deve dunque soddisfare l'equazione di Cauchy nel dominio $u+v<p$ $(u,v>0)$; poichè essa è monotona deve essere in $]0,p[$:

$$h.f(u)=ku \text{ ovvero } f(u)=h^{-1}(ku) \text{ con } k \text{ costante}$$

In base al lemma 4 e alla monotonia è $\lim\limits_{u\to p^-} h^{-1}(ku)\leqslant\beta$ e quindi $k\leqslant h(\beta)/p$.

Se però fosse strettamente $k<h(\beta)/p$ ovvero $p<h(\beta)/k$, presi u e v minori di p con $p<u+v<h(\beta)/k$ si avrebbe:

$$f(u)*f(v)=h^{-1}\{h[f(u)]+h[f(v)]\}=h^{-1}[k(u+v]<\beta$$

mentre per la (4.2) e per il lemma 4 deve essere

$$f(u)*f(v)=f(u+v)=\beta$$

Osservazioni.

1) $f_o \epsilon \Lambda$: in caso contrario infatti f_o sabebbe interno ad un intervallo $]\alpha,\beta[\subset \bar{\Lambda}$ con $\alpha,\beta \epsilon \Lambda$ e risulterebbe di conseguenza $\lim_{u\to 0^+} f(u)=\alpha\neq f_o$.

2) In base ai lemmi n.4 e n.5 e alla (4.1) si riconosce immediatamente che deve porsi $f(p)=\beta$. La funzione $f(u)$ è così determinata per ogni $u>0$.

5. Conclusione.

Quanto provato nei §§3 e 4 può sintetizzarsi nel seguente teorema che determina ogni informazione relativa del tipo $H(\Pi'/\Pi'')=F[H(\Pi'\wedge\Pi''),H(\Pi'')]$.

Teorema. Dato uno spazio di incertezza $I=(\Omega,\sigma,\varepsilon,\kappa,*,H)$ ogni funzione $F(x,y)$ atta a determinare l'informazione relativa di esperienza è del tipo:

$$F(x,y) = \begin{cases} 0 & \forall(x,y)\,\epsilon([0,a]^2\cap L)\cup\bar{L} \\ f[g(x)-g(\sup(y,a))] & \forall(x,y)\,\epsilon L-[0,a]^2 \end{cases}$$

dove $f(u)$ può essere una qualunque delle funzioni di R^+ in R^+ definite scegliendo ad arbitrio $\lambda\epsilon\Lambda$ e ponendo:

$f(u)=\bar{h}(ku)$ con $k\geqslant 0$ se $\beta\epsilon\Lambda$ con $\beta\neq\lambda$: $]\lambda,\beta[\subset\bar{\Lambda}$

$f(u)=\lambda=$cost. se λ è per Λ punto di accumulazione.

($h(x)$ è la funzione che definisce $*$ in $]\lambda,\beta[$ ed $\bar{h}$ è la sua pseudoinversa).

Osservazioni.

3) Qualunque sia dunque lo spazio I, scegliendo $\lambda=0$ e ponendo eventualmente $k=0$ il sistema (2.4)÷(2.6) è sempre soddisfatto dalla funzione identicamente nulla.

4) La funzione $f(u)$ è sempre una funzione continua in $]0,+\infty[$.
La funzione $F(x,y)$ è invece quasi continua potendo presentare discontinuità per $x=y$.
Tali discontinuità possono essere eliminate scegliendo $\lambda=0$ e determinando $f(u)$ di conseguenza; in tal caso si potrà ittenere per la funzione $F(x,y)$ una espressione non identicamente nulla solo se 0 è per Λ punto isolato.

6. Casi particolari.

Può essere talvolta oportuno fare uso di una misura di informazione relativa secondo la quale per ogni partizione Π di Ω resulti $H(\Pi/\{\Omega\})=H(\Pi)$. Tra le funzioni $F(x,y)$ di cui nel teorema del paragrafo precedente si dovranno perciò individuare quelle che soddisfano l'ulteriore condizione

$$F(x,0) = x \tag{6.1}$$

Si riconosce immediatamente che di funzioni di tal tipo ve n'è una (e una sola) se (e solo se) $\Lambda-\{0\}=\phi$. Qualunque sia $\lambda\in\Lambda$ infatti per la (6.1) e per la (2.5) si ha:

$$\lambda=F(\lambda,0)=F(\lambda,0)*F(\lambda,\lambda)=F(\lambda*\lambda,0*\lambda)=F(\lambda,\lambda) = 0$$

Si viceversa è $\Lambda-\{0\}=\phi$ $(\Lambda=\{0\})$ l'intervallo $]0,+\infty[$ è contenuto in $\bar{\Lambda}$; pertanto $h(x)$ coincide con $g(x)$ e $\bar{h}(x)=g^{-1}(x)$. A norma del teorema del paragrafo precedente si ha quindi:

$$(6.2) \qquad F(x,y) = \begin{cases} 0 & \forall (x,y) \in \bar{L} \\ g^{-1}\{k[g(x)-g(y)]\} & \forall (x,y) \in L \end{cases}$$

che soddisfa la (6.1) se e solo se $k=1$.

Viene individuata così, nel caso particolare trattato, la seguente espressione particolarmente significativa per l'informazione relativa:

$$H(\Pi'/\Pi'') = \begin{cases} 0 & \text{se } \Pi'' < \Pi' \\ g^{-1}\{g[H(\Pi' \wedge \Pi'')]-g[H(\Pi'')]\} & \text{altrimenti} \end{cases}$$

7. Appendice.

Qualora introducendo il concetto di informazione relativa di esperienza si fosse preferito imporre fin dall'inizio la condizione (6.1) accanto alle (2.4)÷(2.6) si sarebbe pervenuti immediatamente alle (6.2) per via diretta, prescindendo cioè dal contenuto dei paragrafi precedenti, come qua di seguito illustreremo sommariamente.

Si prova dapprima (nello stesso modo con cui lo si è fatto nel §6) che dalle (6.1) e (2.5) segue che $\Lambda-\{0\}=\phi$. $(a=0)$. Con il procedimento già illustrato nel lemma 2 si riconosce quindi che e $F(x,y)=f[g(x)-g(y)]$, $\forall (x,y) \in L$.

Ricordando poi che $g(0)=0$ si ricava dalla (6.1) che deve essere $f[g(x)]=x$ ovvero $f=g^{-1}(x)$ e quindi che deve porsi: $F(x,y)=g^{-1}[g(x)-g(y)]$ ovvero

$$H(\Pi'/\Pi'')=g^{-1}\{g[H(\Pi' \wedge \Pi'')]-g[H(\Pi'')]\}$$

ritrovando così direttamente quanto si era già provato per altra via.

Bibliografia.

[1] B. FORTE, N. PINTACUDA, "Sull'informazione associata alle esperienza incomplete" Ann. Mat. Pura e Appl. (IV) vol. LXXX pp. 215-234 (1968).

[2] C. BERTOLUZZA, F. BARBAINI, "Measures d'information totalement composables compatibles avec l'axiome d'independence généralisée". Coll. Int. Du C.N.R.S. N. 276 (Theorie de l'information) Edition du C.N.R.S. Paris 1978 pp. 57-66.

[3] D.A. KAPPOS, "Strukturtheorie der Wahrscheinlichkeitsfelder-und Raume". Berlin, Springer, 1971 p. 71.

[4] J. KAMPÉ DE FERIET: L'indépendence des événements dans la théorie généralisée de l'Information" Journées Lyonnaises Questionnaires, 1975.

Istituto di Matematica Applicata.
Università di Pavia.
Corso Strada Nuova, 65
27100 - PAVIA (Italia).