Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Giuseppe Longo, Franco Buttazzoni

Un’interpretazione dell’intensità dell’energia informazionale

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **44** (1968), n.4, p. 520–526.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1968_8_44_4_520_0>

Teoria dell'informazione. — *Un'interpretazione dell'intensità dell'energia informazionale.* Nota di GIUSEPPE LONGO e FRANCO BUTTAZZONI, presentata (*) dal Socio B. FINZI.

RÉSUMÉ. — On considère l'énergie informationnelle de M. Onicescu et on montre que l'intensité de l'énergie informationnelle constitue pour une distribution continue une mesure de son écart de la distribution uniforme.

§ 1. – INTRODUZIONE.

È stata recentemente proposta [1] come misura dell'informazione associata ad uno schema finito A costituito da n eventi A_i, di probabilità p_i:

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}, \tag{1}$$

dove

$$\sum_{i=1}^{n} p_i = 1, \quad \text{e} \quad p_i \geq 0 \qquad (i = 1, \cdots, n), \tag{2}$$

la quantità

$$E_n = \sum_{1}^{n} {}_i\, p_i^2, \tag{3}$$

detta « energia informazionale dello schema A ». Tale quantità – che è il valor medio della variabile aleatoria semplice che assume i valori p_i con probabilità p_i – assume i valori 1 corrispondentemente allo stato di massima determinazione (tutte le p_i sono nulle tranne una che vale 1), ed $1/n$ corrispondentemente allo stato di massima indeterminazione (tutte le p_i valgono $1/n$).

L'energia informazionale di Onicescu può servire [2], sia pure con certe limitazioni [3], come base per una Teoria dell'Informazione. Recentemente [4] l'energia informazionale è stata utilizzata anche in connessione con problemi diversi.

§ 2. – CASO DISCRETO NUMERABILE.

Prendendo in esame il caso di uno schema discreto ma infinito:

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}, \tag{4}$$

(*) Nella seduta del 20 aprile 1968.

con $\sum_{i=1}^{\infty} p_i = 1$ e $p_i \geq 0$ $(i = 1, 2, \cdots)$, è naturale definire la relativa energia informazionale mediante la serie:

$$E_\infty = \sum_{i=1}^{\infty} p_i^2. \tag{5}$$

Poiché $p_i^2 \leq p_i$ e $\sum_{i=1}^{\infty} p_i = 1$, la serie (5) è sempre convergente, e anzi:

$$E_\infty \leq 1.$$

§ 3. – CASO GENERALE.

Sia Ω un insieme arbitrario, $\mathfrak{A}$ una σ–algebra di sottinsiemi di Ω, cioè una famiglia di sottoinsiemi, chiusa rispetto alla riunione numerabile e alla differenza, alla quale appartenga Ω stesso. Sia poi μ una misura elementare definita su $\mathfrak{A}$, cioè una funzione d'insieme reale non negativa numerabilmente additiva.

Consideriamo ora un'altra misura φ, definita anch'essa su $\mathfrak{A}$ e tale che $\varphi(\Omega) = 1$. L'insieme $(\Omega, \mathfrak{A}, \varphi)$ si dirà uno «spazio di probabilità». Eventualmente sottraendo alla φ una componente singolare (in particolare discreta), si può supporre che la φ sia assolutamente continua rispetto alla μ.

Per il teorema di Radon–Nikodym, per ogni $A \in \mathfrak{A}$ si può scrivere allora:

$$\varphi(A) = \int_A g(x)\, \mu(dx), \tag{6}$$

dove la $g(x)$ è μ – sommabile su ogni $A \in \mathfrak{A}$ e in particolare su Ω. Pertanto esiste una successione di funzioni elementari $\{f_n(x)\}$ tale che $f_n(x) \uparrow g(x)$ per ogni $x \in \Omega$. Per ogni i, si ha:

$$f_i(x) = \sum_{1}^{\infty}{}_k \, h_{ik}\, \chi(A_{i,k}), \tag{7}$$

dove $\chi(A)$ è la funzione caratteristica di A; h_{ik} sono costanti reali positive distinte fra loro; $A_{i,k} \in \mathfrak{A}$; $\bigcup_{k=1}^{\infty} A_{i,k} = \Omega$ per ogni i.

L'espressione (7) induce naturalmente la partizione di Ω data da $\{A_{i,k}\}$, cui è associata l'energia informazionale:

$$E_{i\infty} = \sum_{1}^{\infty}{}_k \, \varphi^2(A_{i,k}) \text{ [1]}. \tag{8}$$

Per quanto detto nel § 2, l'espressione (8) ha sempre significato.

(1) Potrà verificarsi ovviamente il caso di partizioni finite.

La successione di partizioni $\{\{A_{i,k}\}\}$ (al variare di i) indotta dalla successione $\{f_i\}$ può sempre considerarsi monotona, nel senso che la partizione $\{A_{i+1,k}\}$ è più fine della partizione $\{A_{i,k}\}$ per ogni i. Infatti se così non fosse, basterebbe considerare in luogo della partizione $\{A_{i+1,k}\}$ la partizione prodotto $\{A_{i,k}\} \otimes \{A_{i+1;k}\}$ i cui elementi sono le intersezioni degli elementi della $\{A_{i,k}\}$ e della $\{A_{i+1,k}\}$.

Ne segue che la successione $\{E_{i\infty}\}$ delle energie corrispondenti alla successione delle partizioni $\{\{A_{i,k}\}\}$ è monotona non crescente; infatti, essendo, per ogni j:

$$A_{i,j} = \bigcup_{k_j} A_{i+1,k_j},$$

per l'additività della φ segue che:

$$\varphi^2(A_{i,j}) = \left(\sum_{k_j} \varphi(A_{i+1,k_j})\right)^2 \geq \sum_{k_j} \varphi^2(A_{i+1,k_j}),$$

da cui l'asserto, sommando su ambo i membri rispetto a j.

Si ha anzi $\lim\limits_{i\to\infty} E_{i\infty} = 0$. Infatti, poiché

$$\lim_{i\to\infty} \sup_k \mu(A_{i,k}) = 0,$$

essendo

$$E_{i\infty} = \sum_{k=1}^{\infty} \varphi^2(A_{i,k}) \leq \sup_k \varphi(A_{i,k}) \sum_{k=1}^{\infty} \varphi(A_{i,k}) = \sup_k \varphi(A_{i,k}),$$

la tesi segue dall'assoluta continuità di φ.

Dunque, se nel caso di uno schema continuo si volesse mantenere la definizione di energia informazionale data nel caso discreto, il modo più naturale di farlo – quello di un passaggio al limite attraverso una successione di schemi discreti – porterebbe ad un'energia informazionale nulla.

§ 4. – L'INTENSITÀ DELL'ENERGIA INFORMAZIONALE.

Conviene pertanto seguire un'altra via, per la cui trattazione ci limiteremo al caso, più interessante in pratica, in cui Ω è l'intervallo $|x| \leq K$ della retta reale e μ è l'ordinaria misura (lunghezza).

Seguendo uno schema già introdotto per l'entropia da Rényi [5], consideriamo la variabile aleatoria reale ξ, tale che $|\xi| \leq K$, alla quale sia associata una funzione di ripartizione monotona non decrescente assolutamente continua $f(x)$; dunque esiste una funzione densità $d(x)$ tale che, per ogni $|x| \leq K$ sia:

$$f(x) = \int_{-\infty}^{x} d(t)\, dt = \int_{-K}^{x} d(t)\, dt. \tag{9}$$

Alla funzione $f(x)$ si può associare la funzione d'insieme $\varphi(\mathrm{A})$ (A ora appartiene alla famiglia dei boreliani sulla retta reale ed è contenuto in $|x| \leq \mathrm{K}$), così definita:

$$\varphi(\mathrm{A}) = \underset{\mathrm{A}}{\mathrm{var}}\, f(x)\ ^{(2)}.$$

Consideriamo ora una successione di variabili aleatorie «semplici» $\{\xi^{(n)}\}$, così definite:

$$\xi^{(n)} = \frac{[n\xi]}{n}\,; \tag{10}$$

dove $[x]$ è la parte intera del numero x. Le v. a. rappresentate dalla (10) si possono anche definire mediante l'espressione:

$$\xi^{(n)} = \frac{k}{n} \qquad \text{per} \quad \frac{k}{n} \leq \xi < \frac{k+1}{n}$$

$$(k = -[n\mathrm{K}]\,, -[n\mathrm{K}] + 1\,, \cdots - 1\,, 0\,, 1\,, \cdots, [n\mathrm{K}] - 1)\,.$$

A ciascuna variabile aleatoria $\xi^{(n)}$ corrisponde una funzione di ripartizione discreta («a scalino») del tipo:

$$f^{(n)}(x) = \int_{-\mathrm{K}}^{x} \sum_{k,\,-\infty}^{+\infty} \left\{ \left[f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right] \delta\left(t - \frac{k+1}{n}\right) \right\} \mathrm{d}t = \tag{11}$$

$$\sum_{k,\,-[n\mathrm{K}]}^{[nk]-1} \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right\} = f\left(\frac{[nx]}{n}\right) - f\left(\frac{-[n\mathrm{K}]}{n}\right).$$

Ovviamente $f^{(n)}(x) \to f(x)$.

Si può associare alla variabile aleatoria $\xi^{(n)}$ l'energia informazionale:

$$\mathrm{E}^{(n)} = \sum_{k,\,-[n\mathrm{K}]}^{[n\mathrm{K}]-1} \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right\}^2. \tag{12}$$

Si vede immediatamente che:

$$\lim_{n \to \infty} \mathrm{E}^{(n)} = 0. \tag{13}$$

Il modo più semplice di persuadersene è quello di considerare la sottosuccessione $\{\mathrm{E}^{(k\bar{n})}\}$ per un qualunque intero $\bar{n}$; la successione delle corrispon-

(2) Nel caso più generale, alla $f(x)$ data dalla (9) si deve aggiungere una componente discreta data da:

$$f_{\mathrm{D}}(x) = \int_{-\mathrm{K}}^{x} \sum_i p_i\, \delta(t - t_i)\, \mathrm{d}t,$$

dove i t_i sono i punti (atomi) in cui $\varphi(t_i) = p_i \neq 0$ e $\delta(x)$ è la funzione di Dirac.

denti partizioni è strettamente monotona, nel senso che ogni partizione è più fine della precedente. Allora, maggiorando la (12), si ha:

$$(14) \qquad 0 \leq \mathrm{E}^{(n)} \leq \sup_k \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right\} \sum_{-[n\mathrm{K}]}^{[n\mathrm{K}]-1}{}_k \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right\} =$$

$$= \sup_k \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right\} \cdot \left\{ f\left(\frac{[n\mathrm{K}]}{n}\right) - f\left(\frac{-[n\mathrm{K}]}{n}\right)\right\} .$$

Il primo fattore dell'ultimo membro della (14) tende a zero per la supposta continuità di $f(x)$, mentre il secondo resta limitato (tende anzi ad 1).

Siccome nessuna ipotesi, oltre a quella della continuità, è stata fatta sulla $f(x)$, il risultato espresso dalla (13) è valido in generale. Dunque l'energia informazionale, definita per le distribuzioni continue come limite della successione di energie relative ad una successione di distribuzioni elementari che tenda alla distribuzione data, non caratterizza affatto la ripartizione. Infatti qualunque essa sia, la sua energia è sempre nulla.

Osserviamo ora che per n fissato si hanno $2[n\mathrm{K}]$ atomi per la $\xi^{(n)}$. Il minimo valore per l'energia di una variabile aleatoria discreta avente $2[n\mathrm{K}]$ valori è dato da $\frac{1}{2[n\mathrm{K}]}$ e corrisponde al caso equiprobabile. Quindi:

$$\mathrm{E}^{(n)} \geq \frac{1}{2[n\mathrm{K}]} ,$$

ovvero:

$$(15) \qquad \mathrm{E}^{(n)} \cdot 2[n\mathrm{K}] \geq 1 .$$

Ovviamente, anche $\frac{1}{2[n\mathrm{K}]}$ è infinitesima per $n \to \infty$, e con un passaggio al limite la (15) ci assicura che $\mathrm{E}^{(n)}$ è un infinitesimo di ordine non superiore a $\frac{1}{2[n\mathrm{K}]}$.

D'ora in poi sostituiremo $n\mathrm{K}$ a $[n\mathrm{K}]$ – com'è lecito fare – se ciò ci sembrerà opportuno [3].

Supponiamo ora che la funzione di densità $\mathrm{d}(x)$ sia continua nell'intervallo chiuso $|x| \leq \mathrm{K}$. Consideriamo il

$$(16) \qquad \lim_{n\to\infty} \mathrm{E}^{(n)} \cdot 2n\mathrm{K} = \lim_{n\to\infty} 2n\mathrm{K} \cdot \sum_{-n\mathrm{K}}^{n\mathrm{K}-1}{}_k \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right\}^2 =$$

$$= 2\mathrm{K} \lim_{n\to\infty} n \sum_{-n\mathrm{K}}^{n\mathrm{K}-1}{}_k \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right\}^2 ;$$

poniamo ora, per n fissato:

$$(17) \qquad \mathrm{M}^{(n)} = \sup_k \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right\} .$$

(3) Tale sostituzione è legittima poiché $\lim_{n\to\infty} \frac{[n\mathrm{K}]}{n} = \mathrm{K}$ ed è questo rapporto che interesserà particolarmente (cfr. (18)).

Dalla (15) segue allora:

$$(18)\qquad \lim_{n\to\infty} \mathrm{E}^{(n)}\cdot 2\,n\mathrm{K} \leq 2\,\mathrm{K}\cdot \lim_{n\to\infty} n\mathrm{M}^{(n)} \sum_{-n\mathrm{K}}^{n\mathrm{K}-1}{}_{k} \left\{ f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right\} =$$

$$= 2\,\mathrm{K} \lim_{n\to\infty} \{n\cdot \mathrm{M}^{(n)}\} \left\{ f\left(\frac{[n\mathrm{K}]}{n}\right) - f\left(\frac{-[n\mathrm{K}]}{n}\right)\right\}.$$

Poiché la f è continua,

$$\lim_{n\to\infty} f\left(\frac{[n\mathrm{K}]}{n}\right) = f(\mathrm{K}) = 1 \quad \text{e} \quad \lim_{n\to\infty} f\left(\frac{-[n\mathrm{K}]}{n}\right) = f(-\mathrm{K}) = 0$$

(cfr. nota [3]) e dunque:

$$(19)\qquad \lim_{n\to\infty} \mathrm{E}^{(n)}\cdot 2\,n\mathrm{K} \leq 2\,\mathrm{K} \lim_{n\to\infty} n\cdot \mathrm{M}^{(n)}.$$

Ancora per la continuità della f, per ogni valore di k e per n fissato si può scrivere:

$$(20)\qquad f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) = \frac{1}{n}\cdot f'(x_k^{(n)}) = \frac{1}{n}\cdot \mathrm{d}\,(x_k^{(n)}),$$

dove $\frac{k}{n} \leq x_k^{(n)} \leq \frac{k+1}{n}$; e quindi, per la (17):

$$\mathrm{M}^{(n)} = \sup_k \frac{1}{n}\,\mathrm{d}\,(x_k^{(n)}) = \frac{1}{n} \sup_k \mathrm{d}\,(x_k^{(n)}),$$

e, posto $\sup\limits_k \mathrm{d}\,(x_k^{(n)}) = \mathrm{D}^{(n)}$, dalla (19) segue:

$$(21)\qquad \lim_{n\to\infty} \mathrm{E}^{(n)}\cdot 2\,n\mathrm{K} \leq 2\,\mathrm{K} \lim_{n\to\infty} \mathrm{D}^{(n)} = 2\,\mathrm{KD},$$

dove $\mathrm{D} = \lim\limits_{n\to\infty} \mathrm{D}^{(n)}$ è il massimo della funzione $\mathrm{d}\,(x)$ nell'intervallo chiuso $|x| \leq \mathrm{K}$, in cui essa è per ipotesi continua.

Per la (15) e per la (21) si ha dunque:

$$(22)\qquad 1 \leq \lim_{n\to\infty} 2\,n\mathrm{K}\cdot \mathrm{E}^{(n)} \leq 2\,\mathrm{K}\cdot\mathrm{D}.$$

Dunque, se esiste il $\lim\limits_{n\to\infty} 2\,n\mathrm{K}\cdot \mathrm{E}^{(n)}$, si può concludere che $\mathrm{E}^{(n)}$ è infinitesima come $1/n$.

Per trovare l'espressione del limite che compare nella (22), sfruttiamo ancora la (20). Si ha dunque:

$$(23)\qquad \lim_{n\to\infty} 2\,n\mathrm{K}\cdot \mathrm{E}^{(n)} = \lim_{n\to\infty} 2\,n\mathrm{K} \sum_{-n\mathrm{K}}^{n\mathrm{K}-1}{}_{k} \left\{\frac{1}{n}\,\mathrm{d}\,(x_k^{(n)})\right\}^2 \leq 2\,\mathrm{K} \lim_{n\to\infty} \sum_{-n\mathrm{K}}^{n\mathrm{K}-1}{}_{k} \frac{1}{n}\,\mathrm{d}^2\,(x_k^{(n)}) =$$

$$= 2\,\mathrm{K} \int_{-\mathrm{K}}^{\mathrm{K}} \mathrm{d}^2\,(x)\,\mathrm{d}x,$$

espressione che ha senso, poiché la $\mathrm{d}(x)$ è limitata nell'intervallo d'integrazione. Si può ovviamente dedurre che:

$$\lim_{n\to\infty} n\mathrm{E}^{(n)} = \int_{-\mathrm{K}}^{\mathrm{K}} \mathrm{d}^2(x)\,\mathrm{d}x\,. \tag{24}$$

§ 5. – Conclusioni.

L'integrale che figura al secondo membro della (24) è stato chiamato [1, 2] «intensità di energia informazionale». Si è dunque visto che il suo significato, a meno della costante 2 K, è quello di fornire una certa misura dello scostamento della distribuzione $f(x)$ da quella uniforme $f(x) = \frac{1}{2\mathrm{K}}x + \frac{1}{2}$ sull'intervallo $|x| \leq \mathrm{K}$. Tale scostamento è stato misurato a partire dalla successione

$$\left\{ \frac{\mathrm{E}^{(n)}}{\frac{1}{2n\mathrm{K}}} \right\}$$

dei rapporti tra le energie informazionali corrispondenti. In generale a funzioni $f(x)$ continue e dotate di derivata prima $\mathrm{d}(x)$ continua nell'intervallo $|x| \leq \mathrm{K}$, corrispondono diversi valori del numero

$$\mathrm{I}_f = 2\,\mathrm{K}\int_{-\mathrm{K}}^{\mathrm{K}} \mathrm{d}^2(x)\,\mathrm{d}x\,,$$

che ne costituisce pertanto una caratterizzazione. Come si è visto, per l'intensità I_f, valgono le limitazioni:

$$1 \leq \mathrm{I}_f \leq 2\,\mathrm{KD},$$

dove $\mathrm{D} = \max_{|x|\leq \mathrm{K}} \mathrm{d}(x)$. Si ha $\mathrm{I}_f = 1$ nel caso uniforme, mentre scostamenti via via crescenti da questo caso corrispondono a valori di I_f sempre maggiori.

Bibliografia.

[1] O. Onicescu, «C. R. Acad. Sc. Paris», t. 263, 841–842 (1966).

[2] O. Onicescu, Seminario presso l'Istituto di Meccanica dell'Università di Trieste (1967).

[3] J. Aczél e Z. Daróczy, «Acta Mathematica Academiae Scientiarum Hungaricae», t. XIV, 95–121, Budapest 1963.

[4] A. Perez, *Sur l'énergie informationnelle de M. O. Onicescu*, «Rev. Roum. Math. Pures et Appl.», t. XII, n. 9, 1341–1347 (1967).

[5] A. Rényi e J. Balatoni, *Über den Begriff der Entropie*. «Arbeiten zur Informationstheorie», I, VEB, Berlin 1961.