Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Giuseppe Longo, Franco Buttazzoni

Un applicazione dell'energia informazionale alle sorgenti stazionarie

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **45** (1968), n.5, p. 311–316.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1968_8_45_5_311_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1968.

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

RÉSUMÉ. — On demontre la possibilité de construire une théorie de l'information à partir de l'énergie informationnelle de M. Onicescu. On retrouve un résultat classique, c'est-à-dire qu'on peut employer un dictionnaire très restreint, pourvu qu'on fasse assez longues les successions de symbols qui transportent l'information.

§ 1. INTRODUZIONE. – Si consideri una sorgente di informazione Q come un processo aleatorio discreto definito matematicamente dalle sue proprietà statistiche [1]. Il generico membro di questo processo è costituito da una successione doppiamente infinita:

$$\xi = \cdots x_{-n}\, x_{-n+1} \cdots x_{-1}\, x_0\, x_1 \cdots x_{n-1}\, x_n \cdots \tag{1.1}$$

in cui le « lettere » x sono tratte da un insieme finito A di a elementi $(a > 1)$ detto l'alfabeto della sorgente Q. Sia S la totalità delle successioni del tipo (1.1).

Tra i sottoinsiemi di S sono particolarmente interessanti quelli del tipo seguente: fissate k posizioni (istanti) nella $\xi : t_n , t_{n+1} , \cdots , t_{n+k-1}$ e fissate k lettere non necessariamente distinte $\alpha_{t_n} , \alpha_{t_{n+1}} , \cdots , \alpha_{t_{n+k-1}}$, allora l'evento: la sorgente emette la lettera α_{t_i} all'istante $t_i\, (n \leq i < n + k)$ è costituito da tutte le successioni ξ per cui $x_{t_i} = \alpha_{t_i}$. Tale insieme sarà detto « cilindro sottile di ordine k » [2], e sarà indicato con ${}_nO^{(k)}$. Per n e k fissati, si dice « cilindro di ordine k », ${}_n\omega^{(k)}$, ogni riunione finita di cilindri sottili di ordine k:

$${}_n\omega^{(k)} = \bigcup_{i \in I} {}_nO_i^{(k)}, \tag{1.2}$$

dove I è una famiglia finita di indici, ovvero:

$${}_n\omega^{(k)} = \{ \xi \mid (\alpha_{t_n}\, \alpha_{t_{n+1}} \cdots \alpha_{t_{n+k-1}}) \in E \} , \tag{1.3}$$

essendo E un sottoinsieme del prodotto cartesiano $A \times A \times \cdots \times A = A^k$.

Sia Ω il minimo campo che contiene tutti i cilindri di ordine finito, e sia P una misura di probabilità su Ω. Così la sorgente Q resta completamente caratterizzata dalla terna $\{S , \Omega , P\}$. È noto che per caratterizzare statisticamente la sorgente è sufficiente assegnare la P su tutti i cilindri sottili di ordine finito [1].

(*) Lavoro eseguito nell'ambito degli Istituti di Elettrotecnica e di Meccanica dell'Università di Trieste.

(**) Nella seduta del 19 novembre 1968.

§ 2. STAZIONARIETÀ DELLA SORGENTE. – Per ogni valore di n e di k, i cilindri sottili di ordine k, ${}_nO^{(k)}$, formano una partizione di S di ordine a^k. Assoggettiamo ora ogni $\xi \in S$ alla seguente trasformazione di traslazione T:

$$\xi = (\cdots x_{-n} \cdots x_{-1}\, x_0\, x_1 \cdots x_n \cdots) \xrightarrow{T} T\xi = (\cdots y_{-n} \cdots y_{-1}\, y_0\, y_1 \cdots y_n \cdots), \tag{2.1}$$

dove:

$$y_j = x_{j-1} \quad \text{per ogni } j.$$

La T induce in modo naturale una corrispondenza Φ tra i cilindri sottili: ad un cilindro sottile ${}_nO_i^{(k)}$ corrisponde un cilindro sottile, che indicheremo con $\Phi\,{}_nO_i^{(k)}$, in generale diverso dal primo (1). La partizione $\{{}_nO_i^{(k)}\}$ $(i = 1, 2, \cdots, a^k)$ di S si trasforma di conseguenza nella partizione $\{\Phi\,{}_nO_i^{(k)}\}$ $(i = 1, 2, \cdots, a^k)$ di S.

Supponiamo ora che la P sia invariante rispetto alla T, nel senso che:

$$P({}_nO_i^{(k)}) = P(\Phi\,{}_nO_i^{(k)}) \qquad (i = 1, 2, \cdots, a^k) \tag{2.2}$$

per ogni valore finito di k. Diremo allora che la sorgente $Q = \{S, \Omega, P\}$ è stazionaria. È chiaro che agli effetti della misura di probabilità P, la (2.2) equivale alla possibilità di svincolarsi, nella definizione dei cilindri sottili, dal valore di n. In altre parole, in $S, TS, T^2S, \cdots$, i cilindri sottili corrispondenti ad una stessa k-upla di lettere $(\alpha_1\alpha_2 \cdots \alpha_k)$ sono tra loro diversi, ma – nel caso di sorgente stazionaria – hanno tutti la stessa probabilità.

§ 3. L'ENERGIA INFORMAZIONALE DI UN CAMPO FINITO. – D'ora in poi supporremo che la sorgente sia stazionaria. Fissato un valore di k, gli a^k cilindri sottili corrispondenti generano un campo finito $\Omega^{(k)}$ di $2^{a^k} - 1$ elementi (cilindri di ordine k) sul quale la probabilità P, definita su Ω, induce una probabilità, che chiameremo ancora P, la quale ci permette di calcolare la quantità media di informazione di $\Omega^{(k)}$. Tale quantità si può esprimere ad esempio mediante l'entropia di Shannon [3]:

$$\mathcal{H}(\Omega^{(k)}) = -\sum_{j=1}^{a^k} P(O_j^{(k)}) \lg_2 P(O_j^{(k)}), \tag{3.1}$$

oppure mediante l'energia informazionale di Onicescu [4]:

$$\mathcal{E}(\Omega^{(k)}) = \sum_{j=1}^{a^k} P(O_j^{(k)})\, P(O_j^{(k)}). \tag{3.2}$$

Assegnati due campi di probabilità finiti, A e B, sia $A \times B$ il campo finito costituito dal prodotto cartesiano di A e B, in cui si introduca una misura

(1) Se degli ${}_nO_i^{(k)}$ facevano parte le successioni aventi le lettere $\alpha_n\, \alpha_{n+1} \cdots \alpha_{n+k-1}$ ordinatamente nei posti $t_n\, t_{n+1} \cdots t_{n+k-1}$, del cilindro $\Phi\,{}_nO_i^{(k)}$ fanno parte le successioni aventi le stesse lettere nei posti $t_{n-1}\, t_n \cdots t_{n+k-2}$.

di probabilità definita al modo seguente:

$$\text{(3.3)} \qquad \mathrm{P}(a_j\, b_k) = \mathrm{P}(a_j)\, \mathrm{P}_{a_j}(b_k) \qquad (a_j \in \mathrm{A}\,,\ b_k \in \mathrm{B}),$$

essendo $\mathrm{P}_{a_j}(b_k)$ la probabilità di b_k condizionata da a_j.

§ 4. L'ENERGIA INFORMAZIONALE UNITARIA E IL SUO LIMITE. – Posto ora $\mathcal{E}(\Omega^{(kn)}) = \mathcal{E}_{kn}$, consideriamo la successione $\{\mathcal{E}_{kn}\}$ $(n = 1\,,2\,,\cdots)$, con $\Omega^{(kn)} = \Omega^{(k)} \times \cdots \times \Omega^{(k)} = [\Omega^{(k)}]^n$. Essa risulta monotona non crescente. Infatti, conformemente alla definizione (3.2), si ha:

$$\text{(4.1)} \qquad \mathcal{E}_{kn} = \sum_{j_1,\cdots,j_n=1}^{a^k} \mathrm{P}^2(\mathrm{O}_{j_1}^{(k)}, \mathrm{O}_{j_2}^{(k)}, \cdots, \mathrm{O}_{j_n}^{(k)}),$$

dove si è generalizzata in modo ovvio la definizione (3.3). Poiché ogni elemento di $\Omega^{(k(n+1))}$ si ottiene associando un elemento di $\Omega^{(kn)}$ ad un elemento di $\Omega^{(k)}$, si ha:

$$\text{(4.2)} \qquad \mathrm{P}(\mathrm{O}_{j_1}^{(k)}, \cdots, \mathrm{O}_{j_n}^{(k)}) = \sum_{j_{n+1}=1}^{a^k} \mathrm{P}(\mathrm{O}_{j_1}^{(k)}, \mathrm{O}_{j_2}^{(k)}, \cdots, \mathrm{O}_{j_n}^{(k)}, \mathrm{O}_{j_{n+1}}^{(k)}),$$

e, sostituendo nella (4.1), si ottiene subito:

$$\mathcal{E}_{kn} = \sum_{j_1,\cdots,j_n=1}^{a^k} \left\{ \sum_{j_{n+1}=1}^{a^k} \mathrm{P}(\mathrm{O}_{j_1}^{(k)}, \mathrm{O}_{j_2}^{(k)}, \cdots, \mathrm{O}_{j_n}^{(k)}, \mathrm{O}_{j_{n+1}}^{(k)}) \right\}^2 \geq$$

$$\geq \sum_{j_1,j_2,\cdots,j_{n+1}=1}^{a^k} \mathrm{P}^2(\mathrm{O}_{j_1}^{(k)}, \mathrm{O}_{j_2}^{(k)}, \cdots, \mathrm{O}_{j_n}^{(k)}, \mathrm{O}_{j_{n+1}}^{(k)}) = \mathcal{E}_{k(n+1)}.$$

Tenendo presente che $0 \leq \mathcal{E}_{kn} \leq 1$, dalla monotonia della successione $\{\mathcal{E}_{kn}\}$, segue che essa ammette un limite, $\mathcal{E}_{(k)}$, con:

$$0 \leq \mathcal{E}_{(k)} \leq 1.$$

Definiamo ora l'energia informazionale unitaria, E_k, del campo finito $\Omega^{(k)}$ come la radice k-esima dell'energia informazionale corrispondente (3.2):

$$\text{(4.3)} \qquad \mathrm{E}_k = \sqrt[k]{\mathcal{E}_k}.$$

Essa rappresenta in un certo senso la quantità media di energia informazionale spettante alla generica lettera dei cilindri di ordine k.

Anche la successione $\{\mathrm{E}_{kn}\}$ ammette limite. Infatti, in primo luogo, dalla definizione (4.3) segue:

$$0 \leq \mathrm{E}_{kn} \leq 1,$$

e quindi essa avrà almeno un punto di accumulazione. Vogliamo dimostrare che ne ha uno solo, il quale sarà il suo limite. Se infatti ne avesse due, e'_k ed e''_k, con $e''_k > e'_k$, dalla $\{\mathrm{E}_{kn}\}$ si potrebbero estrarre due sottosuccessioni disgiunte, $\{E_{kl_i}\}$ ($i \in N^+$) ed $\{E_{km_j}\}$ ($j \in N^+$), tali che:

$$(4.4) \qquad \begin{cases} \{E_{kl_i}\} \to e''_k \\ \{E_{km_j}\} \to e'_k . \end{cases}$$

Preso ora $l_i < m_j$, si ha per la monotonia della $\{\mathcal{E}_{kn}\}$:

$$\mathcal{E}_{kl_i} \geq \mathcal{E}_{km_j} .$$

Ora, qualunque sia $\varepsilon > 0$, poiché la $\{\mathcal{E}_{kn}\}$ converge, pur di prendere l_i ed m_j ($l_i < m_j$) sufficientemente grandi, si ha anche:

$$(4.5) \qquad \mathcal{E}_{kl_i} < \mathcal{E}_{km_j} + \varepsilon ,$$

e, estraendo da ambo i membri della (4.5) la radice di ordine kl_i, anche:

$$(4.6) \qquad E_{kl_i} < \sqrt[kl_i]{\mathcal{E}_{km_j} + \varepsilon} .$$

Se ε è sufficientemente piccolo, tale che almeno da un certo m_j in poi risulti $\mathcal{E}_{km_j} + \varepsilon < 1$, dalla (4.6) segue:

$$(4.7) \qquad E_{kl_i} < \sqrt[kl_i]{\mathcal{E}_{km_j} + \varepsilon} < \sqrt[km_j]{\mathcal{E}_{km_j} + \varepsilon} .$$

Facendo ora tendere l_i ed m_j all'infinito, con la condizione $m_j > l_i$, dalle (4.4) e (4.7) si ricava:

$$(4.8) \qquad e''_k \leq \lim_{m_j \to \infty} \sqrt[km_j]{\mathcal{E}_{km_j} + \varepsilon} ,$$

e, poiché il limite che compare a secondo membro della (4.8) si discosta da e'_k di una quantità arbitraria, anche:

$$e''_k \leq e'_k ,$$

contro l'ipotesi. Dunque esiste un numero e_k tale che:

$$(4.9) \qquad \lim_{n \to \infty} E_{kn} = e_k .$$

§ 5. Esistenza di un vocabolario per la trasmissione dell'informazione. – Dimostreremo ora che:

$$(5.1) \qquad \lim_{n \to \infty} | \mathcal{E}_{kn} - e_n^{kn} | = 0.$$

Dalla (4.9) segue che, per n sufficientemente grande, si ha:

$$(5.2) \qquad e_k - \varepsilon < E_{kn} < e_k + \varepsilon$$

per ogni $\varepsilon > 0$. Elevando tutti i membri della (5.2) alla potenza kn, si ha anche:

$$e_k^{kn} - (\varepsilon e_k^{kn-1} - \cdots \pm \varepsilon^{kn}) < \mathcal{E}_{kn} < e_k^{kn} + (\varepsilon e_k^{kn-1} + \cdots + \varepsilon^{kn}),$$

ovvero, posto $\varepsilon e_k^{kn-1} - \cdots \pm \varepsilon^{kn} = \varepsilon_1$ e $\varepsilon e_k^{kn-1} + \cdots + \varepsilon^{kn} = \varepsilon_2$, e detto ε' il maggiore tra ε_1 e ε_2, anche:

$$e_k^{kn} - \varepsilon' < \mathcal{E}_{kn} < e_k^{kn} + \varepsilon',$$

cioè la (5.1), essendo ε' una quantità arbitraria.

Se $0 \leq e_k < 1$, poiché $\lim\limits_{n\to\infty} e_k^{kn} = 0$, la (5.1) implica anche:

$$\lim_{n\to\infty} \mathcal{E}_{kn} = \mathcal{E}_{(k)} = 0. \tag{5.3}$$

Le considerazioni precedenti valgono in particolare anche se $k = 1$, nel qual caso l'energia informazionale $\mathcal{E}_n$, definita dalla:

$$\mathcal{E}_n = \sum_{j_1, j_2, \cdots, j_n = 1}^{a} \mathrm{P}^2 (\mathrm{O}_{j_1}^{(1)}, \mathrm{O}_{j_2}^{(1)}, \cdots, \mathrm{O}_{j_n}^{(1)}),$$

può venir considerata come il valor medio di una variabile aleatoria ${}_n y$ definita in $\Omega^{(n)}$, che assume i valori ${}_n y_{j_1 \cdots j_n} = \mathrm{P}(\omega_{j_1}^{(1)}, \cdots, \omega_{j_n}^{(1)})$:

$$\mathcal{E}_n = \sum_{j_1, \cdots, j_n = 1}^{a} \mathrm{P}(\mathrm{O}_{j_1}^{(1)}, \cdots, \mathrm{O}_{j_n}^{(1)}) \cdot {}_n y_{j_1 \cdots j_n} = \mathfrak{M}({}_n y). \tag{5.4}$$

Otteniamo così, al variare di n, una successione di variabili aleatorie $\{{}_n y\}$, il cui valor medio $\mathfrak{M}({}_n y)$, che per grandi valori di n si confonde con e_1^n, tende a zero:

$$\lim_{n\to\infty} \mathfrak{M}({}_n y) = 0. \tag{5.5}$$

Per il teorema di Cebicev [5], dalla convergenza in media segue la convergenza in probabilità; cioè, assegnati due numeri reali positivi, ε ed η, con $\eta < 1$, esiste un intero $n(\varepsilon, \eta)$ tale che per ogni intero $n > n(\varepsilon, \eta)$, la probabilità che la v.a. ${}_n y$ si discosti dal suo valor medio per meno di ε è maggiore di $1 - \eta$. Nel nostro caso (tenendo presente la (5.1)):

$$\text{Prob.}\ (|\,{}_n y - e_1^n\,| < \varepsilon) > 1 - \eta. \tag{5.6}$$

Dalla (5.6) si vede che gli elementi di $\Omega^{(n)}$ per cui la v.a. ${}_n y$ assume valori arbitrariamente vicini ed e_1^n (che diremo elementi della prima classe) hanno una probabilità complessiva che differisce da 1 di una quantità arbitraria η, e tutti gli altri hanno una probabilità complessiva pari a η. Detto allora ω il generico elemento della prima classe, si ha:

$$\mathrm{P}(\omega) \simeq e_1^n \qquad (n > n(\varepsilon, \eta)),$$

e quindi il numero N di elementi della prima classe soddisferà alla limitazione:

$$\mathrm{N} \leq e_1^{-n} = 2^{n \lg \frac{1}{e_1}} \text{ (2)}. \tag{5.7}$$

(2) I logaritmi sono considerati in base 2.

Siccome il numero totale di elementi di $\Omega^{(n)}$ è $2^{\alpha^n} - 1 \simeq 2^{\alpha^n}$, il rapporto tra il numero N degli elementi della prima classe ed il numero totale è circa:

$$2^{n \lg \frac{1}{e_1} - \alpha^n} . \tag{5.8}$$

Ora, se il limite e_1 è positivo, pur di scegliere un n sufficientemente grande, si avrà:

$$e_1 > 2^{-\frac{\alpha^n}{n}}$$

e quindi il rapporto (5.8) si potrà rendere molto minore di 1. Quindi, pur di trasmettere l'informazione con successioni sufficientemente lunghe (n abbastanza grande), il numero delle sequenze interessanti per il trasporto dell'informazione sarà solo una piccola frazione del numero totale di sequenze: il vocabolario sarà dunque abbastanza ristretto da risultare maneggevole.

Conclusioni analoghe si possono trarre anche a partire dall'entropia di Shannon, anziché dall'energia informazionale di Onicescu [1].

Gli autori sono grati al prof. Octav Onicescu che ha loro indicato il tema di questa Nota.

Bibliografia.

[1] A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover, New York 1957.

[2] P. Billingsley, *Ergodic Theory and Information*, John Wiley & Sons, New York 1965.

[3] C. E. Shannon, *A Mathematical Theory of Communication*, « Bell System Tech. J. », 27, 379–423, 623–656 (1948).

[4] O. Onicescu, *L'Energie Informationnelle*, « C. R. Acad. Sc. Paris », t. 263, p. 841–842 (nov. 1966).

[5] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press 1946.