SUI FONDAMENTI DELLA TEORIA DELL'INFORMAZIONE(*)

di GIUSEPPE LONGO (a Trieste)(**)

Alla Memoria di Ugo Barbuti

SOMMARIO.- *Questa memoria è costituita da un certo numero di osservazioni relative ad alcune questioni fondamentali sulla Teoria dell'Informazione. Tali osservazioni riguardano sia la teoria classica (di Shannon) e la portata di alcune ipotesi implicite o esplicite in essa contenute, sia il problema, in un certo senso preliminare, della possibilità di un ampliamento o generalizzazione della teoria stessa.*

SUMMARY.- *This paper consists of a number of remarks concerning some fundamental problems of Information Theory. These remarks concern both the classical (Shannon's) theory and the significance of some implicit or explicit hypotheses contained therein, and the somewhat preliminary problem of the possibility of an extension or generalization of the theory itself.*

1.- La grande semplicità del modello su cui è basata la teoria di Shannon (semplicità che a volte può dar l'impressione di un'eccessiva semplificazione) è ottenuta riducendo il problema della comunicazione al suo livello più basso, e trascurando i due livelli superiori. Cioè, con riferimento ad una classificazione proposta da Weaver [1], e largamente accettata, nell'atto comunicativo si possono distinguere essenzialmente tre livelli:

Livello A (tecnico), che riguarda la questione con quanta *precisione* i messaggi generati dalla sorgente possano essere trasmessi all'utente.

Livello B (semantico), che riguarda la questione con quanta accuratezza il *significato* che la sorgente annette ai messaggi possa essere trasferito all'utente.

Livello C (pragmatico), che riguarda la questione con quanta efficacia il *comportamento* dell'utente possa essere influenzato dai messaggi della sorgente.

Pur essendovi, in qualche misura, un rapporto di subordinazione da A verso C, le relazioni fra questi tre livelli sono piuttosto complicate [2]. E' proprio rinunciando a comprendere nella teoria gli aspetti semantici e

(*) Pervenuto in Redazione il 4 aprile 1990.

(**) Indirizzo dell'Autore: Dipartimento di Elettrotecnica, Elettronica ed Informatica, Università degli Studi, Facoltà d'Ingegneria, Via Valerio 10, 34100 Trieste (Italy).

pragmatici della comunicazione che Shannon ha potuto dare una formalizzazione relativamente semplice dell'aspetto tecnico; è tuttavia necessario tener presente questo aspetto riduttivo della teoria per non essere indotti, ad esempio, a tentarne applicazioni fuori degli ambiti squisitamente tecnici.

2.- Nel caso più semplice, una sorgente d'informazione (finita, stazionaria e senza memoria) si può rappresentare con uno schema del tipo seguente:

$$\mathcal{S} = \begin{pmatrix} a_1 & a_2 \cdots & a_n \\ \\ p_1 & p_2 \cdots & p_n \end{pmatrix}$$

dove gli a_i sono i messaggi e $\{p_i\}$ è una distribuzione di probabilità sui messaggi stessi, considerati come eventi elementari. Se si accetta questa descrizione, apparentemente innocua, di una sorgente d'informazione, si è con ciò stesso accettato tutto un insieme di ipotesi non ovvie:

(i) l'atto elementare con cui la sorgente genera l'informazione è l'indicazione di uno e uno solo degli eventi (messaggi) a_i fra gli n a priori possibili; con ciò l'atto di generazione dell'informazione si compie e si esaurisce: le differenze che generano l'informazione sono tutte e sole quelle riflesse negli indici i ($1 \leq i \leq n$) dei messaggi. Un messaggio non è dunque costituito da parti differenziate, ma è un blocco unico.

(ii) Ciò presuppone che nell'universo iniziale (e potenzialmente assai ricco) dei messaggi si sia introdotta una relazione di equivalenza, le cui classi siano poi state identificate con i messaggi a_i.

(iii) Questa relazione di equivalenza è basata non solo e non tanto sui criteri e sulle esigenze della sorgente, quanto e soprattutto su quelli dell'*utente* dell'informazione, che è sempre necessario considerare accanto alla sorgente per non fornire del processo di generazione dell'informazione un quadro irrealistico.

(iv) La distribuzione di probabilità che fa parte della descrizione della sorgente è nota a priori, ma come si sia giunti a questa conoscenza non si sa; mentre da un punto di vista pratico la difficoltà si può superare con metodi basati sul principio del massimo dell'entropia [3] oppure con metodi numerici che operano sulle frequenze relative osservate [4], [5], dal punto di vista concettuale il problema non si risolve ed è anzi un'osservazione di questo tipo che costituisce la giustificazione di principio di un tentativo di costruire una teoria dell'informazione non basata sulla probabilità (si veda più avanti).

3.- Proprio la necessità, messa ora in luce, di considerare le coppie sorgente- utente, e non le sorgenti isolate, sembra indicare che, accanto ad elementi "oggettivi", nella descrizione delle sorgenti d'informazione debbano essere introdotti anche elementi "soggettivi" i quali per ogni messaggio generato dalla sorgente indichino o misurino l'*interesse* dell'utente, o l'*utilità* che l'utente annette al messaggio in vista di un suo scopo o fine [6]. L'inserimento di questi parametri soggettivi rende la teoria certamente più flessibile e più adatta a rappresentare un modello per certe situazioni reali. Osserviamo a questo proposito che

(i) La sorgente d'informazione arricchita coi parametri "utilità" è forse un modello migliore per un processo di comunicazione che coinvolge il livello A (tecnico) e il livello C (pragmatico); sembra invece difficile riuscire a formalizzare in questo modo il processo al livello B (semantico), e ciò sembra ancora una volta indicare quanto gli aspetti semantici della comunicazione siano difficilmente formalizzabili.

(ii) I parametri supplementari, qui sopra chiamati "utilità", possono ricevere un'interpretazione affatto diversa, cioè di "costo unitario". Con ciò la sorgente arricchita con questi parametri si presta alla verifica di certi "teoremi di codifica", la cui interpretazione in termini economici è abbastanza interessante. Vale forse la pena illustrare brevemente un esempio, cui premettiamo lo schema della sorgente "arricchita":

$$\mathfrak{S} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ p_1 & p_2 & \dots & p_n \\ u_1 & u_2 & \dots & u_n \end{pmatrix}$$

dove gli a_i sono al solito i messaggi, e $\{p_i\}$ è una distribuzione di probabilità sugli a_i. Gli u_i sono numeri non negativi (che possono essere interpretati come utilità, o come costi, o altrimenti) che rappresentano un "valore" soggettivamente attribuito agli a_i dal particolare utente cui la sorgente è collegata.

Ora si supponga che, in base alla distribuzione $\{p_i\}$ la sorgente $\mathfrak{S}$ emetta una lettera a_i che viene codificata (prima di essere inviata al canale che la deve trasmettere all'utente) in una parola di codice w_i, di lunghezza n_i. Ricevendo l'informazione relativa ad a_i, l'utente deve pagare il conto, che gli è inviato da un supervisore (che può essere la società che fornisce il servizio di canale) che calcola il costo c_i di a_i in base all'utilità (o qualità) u_i di a_i per l'utente e alla lunghezza (o quantità) n_i della corrispondente parola di codice w_i. Dunque

$$c_i = c_i\,(u_i, n_i) \tag{3.1}$$

e l'utente cercherà di minimizzare il costo medio

$$\bar{c} = \sum_{i=1}^{n} p_i c_i (u_i, n_i) \tag{3.2}$$

scegliendo opportunamente il codice, cioè la legge con cui w_i è associata ad a_i.

Si veda la figura 1.

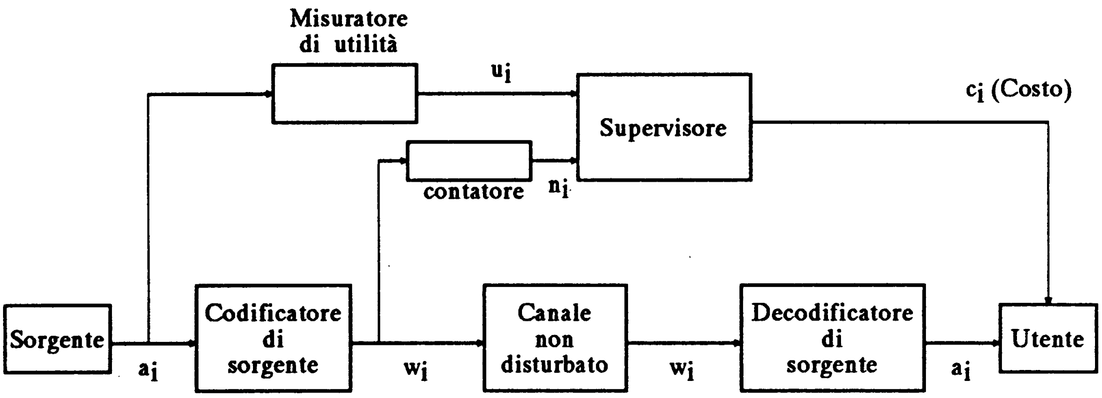

Fig. 1 - Schema della trasmissione di informazione utile.

Qualora il supervisore scelga come funzione costo il prodotto $u_i\, n_i$, cioè

$$c_i (u_i, n_i) = u_i\, n_i \tag{3.3}$$

il costo medio diviene

$$\bar{c} = \sum_{i=1}^{n} p_i\, u_i\, n_i \tag{3.4}$$

e minimizzare $\bar{c}$ equivale a minimizzare la quantità

$$\frac{\sum_{i=1}^{n} p_i\, u_i\, n_i}{\sum_{I=1}^{n} p_i\, u_i} \tag{3.5}$$

che è la lunghezza media delle parole del codice rispetto alla nuova distribuzione di probabilità

$$\left\{ \frac{p_i\, u_i}{\sum_{i=1}^{n} p_i\, u_i} \right\}. \tag{3.6}$$

Si consideri ora il seguente esempio: una società costruisce una certa macchina complicata che è soggetta ai guasti $a_1, a_2, \ldots, a_n$ con probabilità $p_1, p_2, \ldots, p_n$. Ovviamente non tutti i guasti hanno la stessa importanza: alcuni limitano solo di poco la funzione della macchina, mentre altri sono più gravi e possono addirittura farla arrestare. Misuriamo allora con u_1, $u_2, \ldots, u_n$ la "gravità" dei guasti (ad esempio gli u_i possono essere un indice di riduzione nella funzionalità della macchina). Una volta avvenuto un guasto a_i, è necessario individuarlo e ripararlo, il che richiede un tempo t_i, che tiene il posto della lunghezza n_i della parola di codice nell'esempio della sorgente. Il costo di un guasto è allora fornito dalla perdita di funzionalità moltiplicata per la durata in cui si ha tale perdita, cioè

$$c_i = u_i\, t_i$$

e in media ancora

$$\bar{c} = \sum_{i=1}^{n} p_i\, u_i\, t_i.$$

In base a considerazioni di questo tipo si può anche dare un'interpretazione in termini economici di un teorema di codifica, che non si può tuttavia presentare qui e per il quale rimandiamo il lettore alla bibliografia [2], [7], [8].

4.- Vogliamo ora passare al secondo ordine di osservazioni, cioè quelle relative non più alla teoria di Shannon e alle sue varianti, nelle quali è assegnata sempre esplicitamente una distribuzione di probabilità sugli eventi- messaggi, bensì alla teoria proposta da Kampé de Fériet e Forte. In questa proposta la probabilità è dichiaratamente ignorata e si prende come punto di partenza per costruire la Teoria dell'Informazione una famiglia di assiomi considerati caratteristici per l'informazione stessa, la quale viene dunque considerata come concetto o ente primitivo.

Un'analisi abbastanza accurata della proposta citata, limitatamente all'informazione di evento, è stata compiuta in [9] e a questo lavoro si rimanda il lettore desideroso di maggiori particolari. Ci limiteremo qui ad alcuni punti fondamentali, tra i quali daremo particolare rilievo al concetto di indipendenza.

J. Kampé de Fériet e B. Forte sono stati giustificati a intraprendere la loro ricerca [10] dall'osservazione che per assegnare ad una famiglia di eventi (o ad un evento) una distribuzione di probabilità è necessario partire da una certa informazione relativa a quegli eventi. Da ciò questi autori concludono che l'informazione precede, da un punto di vista logico, la probabilità e quindi ha senso tentare di costruire una teoria assiomatica della prima che prescinda dalla seconda.

In realtà la precedente osservazione va integrata con quella che la ripetuta esecuzione dell'esperienza i cui risultati sono gli eventi d'interesse può influire, e in generale influisce, sullo stato d'informazione dell'osservatore. La cosa diviene più chiara qualora si rifletta che il termine "probabilità" è in questo discorso impiegato in due accezioni piuttosto diverse: probabilità a priori e probabilità a posteriori [11], e che se l'informazione è base per la prima, essa è a sua volta determinata dalla seconda.

5.- E' importante osservare che, nel caso classico, la probabilità e l'informazione sono, dal punto di vista logico, su un piano paritetico, essendo l'una funzione invertibile dell'altra. Tuttavia, da punto di vista psicologico la probabilità prevale sull'informazione, e ciò, io stimo, per motivi storici: un tentativo di formalizzazione del concetto di probabilità è assai più antico, e tale concetto è assai più familiare anche al profano che non quello d'informazione.

Da un punto di vista formale, nel caso classico, dato il supporto $(\Omega, \mathcal{A})$, dove Ω è un insieme arbitrario e $\mathcal{A}$ un'algebra di parti di Ω (gli eventi d'interesse), le due funzioni

$$J : \mathcal{A} \to \overline{R}^{+} \tag{5.1}$$

$$P : \mathcal{A} \to [0, 1] \tag{5.2}$$

si dicono rispettivamente un'*informazione* e una *probabilità* se soddisfano i seguenti postulati:

J1) $J(\Omega) = 0, J(\emptyset) = +\infty$
J2) $A \subset B \Rightarrow J(A) \geq J(B)$
J3) A e B indipendenti $\Rightarrow J(A \cap B) = J(A) + J(B)$
J4) A e B incompatibili, cioè $A \cap B = \emptyset \Rightarrow J(A \cup B) =$

$$= -\frac{1}{k} \log (e^{-kJ(A)} + e^{-kJ(B)})$$

per l'informazione, e i seguenti

P1) $P(\Omega) = 1, P(\emptyset) = 0$
P2) $A \subset B \Rightarrow P(A) \leq P(B)$
P3) A e B indipendenti $\Rightarrow P(A \cap B) = P(A) P(B)$
P4) A e B incompatibili, cioè $A \cap B = \emptyset \Rightarrow P(A \cup B) =$
$= P(A) + P(B)$

per la probabilità.

Lo schema classico di deduzione dell'informazione dalla probabilità è il seguente:

Data una probabilità P che soddisfi P1), P2) e P3) si ricava un'informazione J che deve soddisfare J1), J2) e J3). La soluzione, sotto ipotesi piuttosto ampie è $J = \log \frac{1}{P}$ e quindi J4) è la traduzione di P4), proprietà che resta ai margini di questo schema di deduzione.

Si osservi che l'indipendenza di cui si parla in P3) e l'indipendenza di cui si parla in J3) sono la stessa cosa, in questo schema di deduzione.

Viceversa, lo schema di deduzione della probabilità dall'informazione è il seguente: data un'informazione J che soddisfi J1), J2) e J3) si ricava una probabilità P che soddisfa P1), P2) e P3). La soluzione, sotto ipotesi poco restrittive è $P = e^{-J}$ e quindi P4) è la traduzione della proprietà J4), che viene richiesta alla J proprio in vista di ottenere per la P la P4), che è evidentemente una proprietà comoda.

A proposito di questo doppio schema deduttivo, peraltro assai elementare, se si prescinde dalle eventuali difficoltà tecniche che s'incontrano nella soluzione delle equazioni funzionali cui si riduce il problema, si deve fare la seguente osservazione:

L'assioma J4), dall'apparenza piuttosto astrusa, può essere trovato "naturale" (cioè appunto avente natura assiomatica) soltanto da chi è già a conoscenza dell'assioma P4), la cui natura è assai più immediata. Una tale assenza di naturalità si può peraltro rilevare anche in P3), che non discende dalla natura mensurale della probabilità, ma è piuttosto una scelta che caratterizza le probabilità fra le altre misure. In questo caso è J3) l'assioma "naturale", e P3) ne è in fondo la traduzione. Mi rendo conto che un tal punto di vista non è facilmente accettabile per i teorici della probabilità, ma sono persuaso che tali difficoltà siano di natura puramente psicologica.

6.- Consideriamo ora il concetto di *indipendenza*, il quale, come abbiamo accennato, è uno dei punti di fondamentale importanza in tutta la questione.

Nella probabilità la caratterizzazione degli insiemi indipendenti avviene sulla base dell'assioma P3); in altre parole P3) è in fondo una definizione dell'indipendenza. Il carattere di questa relazione è dunque metrico, e non strutturale, o algebrico, com'è ad esempio quello dell'incompatibilità, oggetto dell'assioma P4). Poiché d'altra parte è abbastanza naturale ammettere che l'indipendenza (così come l'incompatibilità) debba essere una relazione primaria, strutturale, tra gli eventi, e non sovrastrutturale, sembra naturale richiedere per l'indipendenza (come lo si richiede per l'incompatibilità) che essa sia assegnata a priori tra gli eventi e che ogni sovrastruttura metrica, ad esempio la probabilità, sovrapposta alla struttura algebrica degli eventi, debba necessariamente conformarsi a tale relazione preesistente.

Ciò naturalmente non si deve richiedere soltanto alla probabilità: cioè, così come gli insiemi "indipendenti", cioè "strutturalmente" o "algebricamente indipendenti" (M-indipendenti), debbono anche risultare "stocasticamente indipendenti" (P-indipendenti) rispetto ad una qualunque probabilità si voglia sovraimporre alla struttura (e ciò sarà appunto il criterio di accettabilità per una probabilità), allo stesso modo gli insiemi M-indipendenti debbono essere anche "informazionalmente indipendenti" (J-indipendenti).

E addirittura, essendovi un'unica nozione di indipendenza, quella algebrica, tutte le altre dovranno essere traduzioni in termini sovrastrutturali di quella.

Questo punto di vista rende più simmetrico l'insieme dei postulati P1-P4 e J1- J4 e anche la doppia deduzione $P \rightleftarrows J$.

Alla luce delle precedenti considerazioni sull'indipendenza, si considerino ora gli assiomi di Kampé de Fériet e Forte per l'informazione di evento [10]:

Dato il supporto $(\Omega, \mathcal{A})$, si definisce *informazione di evento* un'applicazione $J : \mathcal{A} \to \overline{R}^+$ tale che

I $\quad J(\Omega) = 0, J(\emptyset) = +\infty$
II $\quad A \subset B \Rightarrow J(A) \geq J(B)$
III $\quad A$ e B indipendenti $\Rightarrow J(A \cap B) = J(A) + J(B)$
IV $\quad A$ e B incompatibili (cioè $A \cap B = \emptyset$) $\Rightarrow J(A \cup B) = f(J(A), J(B))$.

E' immediato osservare che non c'è alcun riferimento ad una misura di probabilità, e altrettanto immediato osservare che la nozione di "indipendenza" non è affatto esplicitata. Inoltre, l'assioma IV è introdotto in certo modo "dopo" gli altri, per evitare che la teoria sia troppo sterile.

Naturalmente, prima di affrontare un esame della teoria assiomatica basata sui postulati I-IV, ci si potrebbe a buon diritto chiedere perché mai, visto che la teoria di Shannon va così bene per affrontare i problemi relativi alla comunicazione (livello tecnico), e visto che in nessuna delle applicazioni della teoria la presenza della probabilità costituisce un ostacolo, perché mai ci si debba sforzare di eliminare la probabilità dalla teoria dell'informazione.

La risposta a questa domanda potrebbe essere che è importante vedere fino a che punto la probabilità e l'informazione siano tra loro legate, ovvero fino a che punto l'informazione sia un concetto indipendente dal concetto di probabilità, o, in altri termini, fino a che punto l'informazione probabilistica (di Shannon) sia un'informazione "naturale". E' in certo modo consolante il risultato, ottenuto da Kampé de Fériet e Forte dopo aver percorso una strada abbastanza lunga, che l'informazione di Shannon, dunque l'informazione probabilistica, è l'unica che goda di proprietà abbastanza "normali" e che quindi abbia il diritto di essere chiamata "informazione" nei numerosi sensi che questo termine ha acquisito. Resta tuttavia dubbio che proprio questo fosse l'intento originario degli autori sopra citati.

Dal punto di vista qui adottato, viceversa, secondo il quale probabilità e informazione sono soltanto sovrastrutture, in rapporto paritetico, di una struttura algebrica sottostante, non si vede perché, dopo aver introdotto sul supporto un'informazione, non vi si possa introdurre anche una probabilità e indagare sui loro rapporti. Tali rapporti dipenderanno in larga parte dal concetto di indipendenza, che diviene dunque il fulcro delle considerazioni che seguono.

7.- Supponiamo che lo spazio di probabilità $(\Omega, \mathcal{A}, P)$ sia "interpretato", costituisca cioè il modello per una situazione reale. Se il modello è adeguato, cioè se lo studio del modello porta a risultati di portata e precisione sufficienti rispetto alla situazione reale, tutti gli elementi rilevanti della situazione sono contenuti e rappresentati nell'algebra $\mathcal{A}$. L'attribuzione di un valore $P(A)$ ad un qualunque elemento $A \in \mathcal{A}$ deve rispettare e non può modificare la struttura di $\mathcal{A}$. La stessa prescrizione deve valere per tutte le altre funzioni di insieme, e in particolare per l'informazione.

Una conseguenza di tale prescrizione, come si è visto, è che esiste un solo tipo di indipendenza, quella strutturale, o insiemistica, e che questa indipendenza dev'essere riflessa ed espressa da qualunque sovrastruttura.

Il concetto di *indipendenza insiemistica* (o *algebrica*, o *qualitativa* o *strutturale*), indicata per brevità con *M- indipendenza*, è precisato dalla seguente

DEFINIZIONE 1. Data una famiglia $\{X_i\}_{i \in E}$ di sottoinsiemi di un insieme Ω, essa si dice M-indipendente se per qualunque $E' \subset E$ risulta

$$\bigcap_{i \in E'} X_i^{\varepsilon_i} \neq \emptyset$$

dove $\varepsilon_i = 1$ oppure 0 e $X_i^1 = X_i$, $X_i^0 = X_i^c$ (complemento di X_i in Ω).

I rapporti fra la M-indipendenza e l'indipendenza stocastica sono illustrati dai due teoremi seguenti.

Sia al solito $(\Omega, \mathcal{A})$ un supporto e P una probabilità su di esso.

TEOREMA 1. *Se $\mathcal{K} \subset \mathcal{A}$ è una famiglia P-indipendente, $\mathcal{K}$ è anche M-indipendente.*

Quindi l'indipendenza stocastica implica l'indipendenza qualitativa, e ciò è in accordo col punto di vista qui adottato. L'implicazione inversa è un po' più recondita, ed è predicata dal seguente importante teorema, dovuto a Marczewski [12], [13]:

TEOREMA 2. *Sia $\mathcal{K}$ una famiglia M-indipendente, e γ (E) una funzione definita per $E \in \mathcal{K}$ e avente valori nell'intervallo* [0, 1]. *Allora esiste una e una sola misura di probabilità P sull'algebra $\mathcal{A}$ $(\mathcal{K})$ generata da $\mathcal{K}$ tale che*

1) $P(E) = \gamma(E)$ *per ogni* $E \in \mathcal{K}$
2) *gli insiemi di $\mathcal{K}$ sono P-indipendenti.*

Un corollario di questo teorema [9] è che $\mathcal{K}$ è in $\mathcal{A}$ $(\mathcal{K})$ una famiglia massimale di insiemi P-indipendenti, qualunque sia la funzione γ di partenza.

C'è da osservare che, naturalmente, la famiglia $\mathcal{K}$ non è P- indipendente rispetto ad una qualunque probabilità P che si possa assegnare sull'algebra $\mathcal{A}$ $(\mathcal{K})$. In effetti, assegnata una P arbitraria su $\mathcal{A}$ $(\mathcal{K})$, resta in quest'algebra identificata la classe Φ_P delle famiglie P-indipendenti, la quale è contenuta nella classe Φ_M delle famiglie M-indipendenti di $\mathcal{A}$ $(\mathcal{K})$.

Finora sono state quindi introdotte due nozioni di indipendenza, quella algebrica e quella stocastica, e i loro rapporti sono stati schematicamente illustrati. Se a questo punto si vuole introdurre la nozione di informazione in modo che essa risulti indipendente dalla probabilità, è necessario introdurre una terza nozione di indipendenza, quella, appunto, informazionale, la J-indipendenza. E' inoltre necessario investigare quali siano i rapporti tra questa indipendenza e le prime due.

Fin dal principio si può dire che la J-indipendenza si situa accanto alla P-indipendenza, in quanto si tratta di un'indipendenza di natura metrica, o quantitativa, a differenza della M-indipendenza, la cui natura è algebrica, o qualitativa. Tuttavia, Kampé de Fériet e Forte non dedicano la loro attenzione ad approfondire il problema della definizione della J-indipendenza e quello, immediatamente successivo, dei suoi rapporti con la M-indipendenza e con la P-indipendenza. Sembra anzi che, a proposito di questa nozione, si possa rilevare in [10] una certa tortuosità, se non addirittura contraddizione. Dapprima, infatti, per decidere se due eventi A e B dell'algebra $\mathcal{A}$ sono J-indipendenti si afferma che si deve osservare "la natura stessa degli eventi rappresentati dai punti di Ω" e, se questa indagine lo autorizza, si decide che A e B sono J-indipendenti, e si richiede che $J(A \cap B) = J(A) + J(B)$.

Ora si osservi che

1) i valori di $J(A)$ e di $J(B)$ (così come quelli dell'informazione di ogni insieme di una qualunque famiglia J- indipendente) non sono vincolati da alcuna condizione, ed appaiono perciò completamente arbitrari (ciò è richiesto [9], [10] per poter attribuire un carattere di "universalità" alla legge di composizione f che compare nel postulato IV del paragrafo (6));
2) non è assegnato alcun criterio preciso su cui basare la decisione sulla J-indipendenza di A e B; dunque questa decisione ha carattere non formale, essendo fondata sulla "natura degli eventi".

A questo punto, quindi, sembrerebbe che, con un criterio di tipo empirico adottato di volta in volta, si potesse, analogamente a ciò che si fa nel calcolo delle Probabilità, definire la J-indipendenza per più di due insiemi, per giungere alla nozione di famiglia di insiemi J- indipendenti; e, poi, si potrebbe affrontare il problema dei rapporti tra J-indipendenza ed M-indipendenza.

Invece, abbastanza inopinatamente, si "precisa la nozione di J- indipendenza mediante quella di M-indipendenza: tra le famiglie M-indipendenti di $\mathcal{A}$ se ne sceglie una, $\mathcal{K}$, da considerare anche come J-indipendente".

Di conseguenza si pongono diverse questioni: in primo luogo, con quale criterio viene scelta $\mathcal{K}$ come famiglia J-indipendente nella classe Φ_M

delle famiglie M-indipendenti di $\mathcal{A}$? Inoltre: mentre con la definizione di J-indipendenza basata sull'additività dell'informazione le famiglie J-indipendenti formano una classe Φ_J ben individuata, con questa seconda definizione si sceglie una famiglia $\mathcal{K}$ che, a priori, potrebbe anche non appartenere a Φ_J.

In [9] si è cercato di approfondire un pochino il problema, e uno dei risultati cui si è pervenuti è il seguente

TEOREMA 3. *Nel caso shannoniano (cioè quando l'informazione J soddisfa il postulato* J4) *del paragrafo* (5) *in luogo del più generale postulato* IV *del paragrafo* (6)) *c'è equivalenza tra J- indipendenza e P-indipendenza rispetto alla funzione d'insieme* $P = \exp(-kJ)$ $(k > 0)$, *che risulta essere una misura di probabilità. Dunque c'è equivalenza tra la J-indipendenza e la M-indipendenza, nel senso del teorema di Marczewski.*

Non vogliamo qui dilungarci ulteriormente su questi problemi; ci basta averne dato un'idea, anche se sommaria.

8.- Per concludere questa rapida rassegna di osservazioni e critiche intorno ai fondamenti della Teoria dell'Informazione, accenniamo ancora al concetto di "differenza", che ha una parte fondamentale nella teoria, anzi costituisce in ultima analisi il fondamento su cui la teoria stessa è basata.

L' unità d'informazione è costituita dalla minima differenza che causa una differenza [14], e in effetti solo ciò che si vuole, o si può, distinguere dal resto dell'universo del discorso può produrre informazione. Naturalmente le differenze non sono mai assolute, cioè non tutte le differenze causano differenze. In altre parole l'utente dell'informazione può desiderare (o essere in grado) di conservare le differenze che la sorgente ha prodotto, oppure può trascurarle, con un'operazione logica che equivale ad un ingrossamento della relazione di "distinguibilità" tra messaggi elementari. Inoltre, la stessa sorgente può essere, da diversi utenti, considerata in modo più o meno fine; cioè può in diversi utenti produrre differenze più o meno numerose (la limitazione superiore a tale finezza è costituita dal contenuto "in sè" della sorgente, se questa espressione ha qualche significato).

La relazione di "distinguibilità" fra messaggi elementari di una sorgente può essere descritta mediante un *grafo probabilistico* [15], i cui vertici corrispondono ai messaggi, e in cui due messaggi distinguibili (per il particolare utente considerato) sono rappresentati da due vertici collegati da un lato (non orientato). Inoltre la distribuzione di probabilità sull'insieme dei messaggi viene trasferita ai vertici del grafo, che si dice perciò probabilistico.

E' chiaro allora che le sorgenti ordinarie corrispondono a grafi probabilistici completi.

Il punto di vista qui considerato ha rilievo quando i messaggi generati dalla sorgente siano di tipo "analogico": disegni, lettere manoscritte, etc., e quando l'uscita di una tale sorgente debba essere codificata in un alfabeto discreto. In queste circostanze l'utente dell'informazione può, in un primo momento, essere soddisfatto di questa rappresentazione discreta, la quale è ovviamente meno fedele di quanto sarebbe una riproduzione analogica. Si supponga, tuttavia, che in seguito l'interesse dell'utente nell'uscita della sorgente aumenti e che egli ne desideri una rappresentazione più precisa.

Da un canto la natura analogica della sorgente permette di ricercare tale rappresentazione più precisa, ma d'altro canto il risultato sarà ancora una rappresentazione discreta. Il passaggio dalla prima, grossolana rappresentazione alla seconda, più raffinata, può essere formalmente descritto con l'ausilio dei grafi probabilistici sopra menzionati, il cui impiego permette anche di valutare la grossolanità della prima rappresentazione rispetto alla seconda.

Naturalmente sarebbe auspicabile, per ragioni di economia, che nella costruzione della seconda rappresentazione, più fine, si potesse far uso dell'informazione che la prima grossolana descrizione fornisce sull'uscita della sorgente. In altre parole, sarebbe utile trovare una codifica in due passi, di cui il secondo completasse, in qualche senso, il primo. Allora ciascun messaggio della sorgente sarebbe codificato due volte, e sarebbe dunque associato a due parole di codice, la cui giustapposizione genererebbe la parola del codice finale, che dovrebbe essere il meno ridondante possibile (in termini di lunghezza) rispetto all'usuale codifica in un solo passo che si impiega per le sorgenti ordinarie. Il confronto tra il procedimento in due passi e il procedimento ordinario si compie [15] confrontando la somma dell'entropia del grafo corrispondente alla prima grossolana codifica (non tutti i messaggi sono tra loro distinguibili) e dell'entropia del grafo complementare con l'entropia (di Shannon) della distribuzione di probabilità assegnata. La somma delle prime due entropie, come si dimostra, non risulta mai inferiore all'entropia della distribuzione di probabilità, e ciò costituisce una riprova formale della correttezza del punto di vista adottato.

Non possiamo qui addentrarci in una descrizione particolareggiata del procedimento di codifica in due passi, e rimandiamo il lettore al lavoro

[15]. Desideriamo tuttavia osservare che un tal procedimento può essere interessante nei casi seguenti:

a) Reperimento dell'informazione. Per trovare un oggetto in una classe di oggetti si può compiere la ricerca esaminando dapprima gli oggetti frettolosamente per scartare quelli "molto dissimili" dall'oggetto cercato e poi esaminando più accuratamente gli oggetti "non molto dissimili" da quello cercato, per individuarlo con precisione.

b) Riconoscimento delle forme. Per attribuire un oggetto assegnato ad una classe tra alcune possibili, si può procedere come in a), scartando dapprima le classi di oggetti "molto dissimili" e poi esaminando con più cura le classi "non molto dissimili" per trovare quella giusta.

c) Trasmissione cooperativa. Se una sorgente d'informazione alimenta i canali di due utenti e se le capacità dei due canali sono ambedue minori dell'entropia della sorgente, si può pensare di comprimere il tasso attraverso i due canali in modo che all'uscita dell'uno vi sia una descrizione "complementare" (nel senso sopra precisato) a quella che si ritrova all'uscita dell'altro, sicché i due utenti, cooperando, possano ricostruire perfettamente i messaggi della sorgente [16], [17], [18].

Queste ultime osservazioni si collegano ad un filone di ricerca che, nell'ambito della Teoria dell'Informazione, ha ricevuto recentemente un impulso assai notevole: cioè il problema della comunicazione con sorgenti, canali e utenti multipli.

BIBLIOGRAFIA

[1] C.E. SHANNON and W. WEAVER, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, 1949.

[2] G. LONGO, *Informazione e utilità*, Atti dell'Istituto Veneto di Scienze, Lettere ed Arti, CXXXI, 1973, Cl. Sc. Mat. e Nat. p. 289-323.

[3] E.T. JAYNES, *Prior Probabilities*, IEEE Trans. on Systems Science and Cybernetics, vol. SSC-4, 1968, p. 227-241.

[4] E. PFAFFELHUBER, *Error Estimation for the Determination of Entropy and Information Rate from Relative Frequencies*, Kybernetik, *8* (1971), p. 50-51.

[5] T. NEMETZ, *On the Experimental Determination of the Entropy*, Kybernetik, *10* (1972), p. 137-139.

[6] M. BELIS and S. GUIAŞU, *A Quantitative-Qualitative Measure of Information in Cybernetic Systems*, IEEE Trans. on Information Theory, IT- 14, 1968, p. 593-594.

[7] G. LONGO, *Qualitative-Quantitative Measure of Information*, Springer Verlag, Vienna-New York, CISM Courses and Lectures n. 138.

[8] G. LONGO, *A Coding Theorem for Sources Having Utilities*, SIAM Journal on Applied Mathematics, *30*, 4 (1976), p. 739-748.

[9] G. LONGO, *Probabilità e Informazione*, Atti dell'Istituto Veneto di Scienze, Lettere ed Arti, CXXXII, 1973, Cl. Sc. Mat. e Nat., p. 11- 56.

[10] J. KAMPÉ DE FÉRIET and B. FORTE, *Information et Probabilité*, C.R. Acad. Sc. Paris, *265* (1967), p. 110-114, p. 142-146, p. 359-363.

[11] R. CARNAP, *Logical Foundations of Probability*, Univ. of Chicago Press, 1950.

[12] E. MARCZEWSKI, *Ensembles indépendents et leurs applications à la théorie de la mesure*, Fund. Math., *35* (1948), p. 13-28.

[13] D.A. KAPPOS, *Strukturtheorie der Wahrscheinlichkeitsfelder und - räume*, Springer, 1960, p. 77.

[14] G. BATESON, *Steps to an Ecology of Mind*, Chandler Publishing Company, 1972.

[15] J. KÖRNER and G. LONGO, *Two-step Encoding for Finite Sources*, IEEE Transactions on Information Theory, IT-19, 1973, p. 778-782.

[16] G. LONGO, *Source Encoding: New Trends and Open Problems*, Int. Rep. n. 20, CISM-Udine 1974.

[17] G. LONGO, *Communication Theory: New Trends and Open Problems*, Invited Paper, Proc. 5th Coll. on Microwave Communication, Budapest, 1974.

[18] G. LONGO, *Fuzzy Sets, Graphs and Source Coding*, Proceedings of the NATO Advanced Study Institute on New Directions in Signal Processing in Communication and Control, Darlington, 1974.

Un riferimento bibliografico generale sull'argomento è

[19] G. LONGO, *Teoria dell'informazione*, Boringhieri, 1980.