CODICI DI SORGENTE CON RITARDO DI CODIFICA CHE MIGLIORANO IL CODICE DI TUNSTALL (*)

di MARIA ALLOCCA (a Trieste) (**)

SOMMARIO. - *Si presenta un modello di codifica di sorgente in cui le parole di codice hanno la medesima lunghezza. Partendo da una osservazione comunicataci da T. Nemetz, si mostra come in tale modello le prestazioni del codice di Tunstall (ottimo quando si richieda l'istantanea codificabilità) possano venire migliorate. Si presenta un algoritmo che, qualora la sorgente verifichi opportune ipotesi, consente di aumentare il rapporto di compressione del codice di Tunstall.*

SUMMARY. - *A model of variable-to-fixed-length source coding is considered. Taking a personal communication by T. Nemetz as a starting point, we show how the performance of a Tunstall code (the best, when the instantaneous coding is required) can be improved in this model. We present an algorithm which allows an increase of the compression ratio of Tunstall codes when the source verifies suitable hypotheses.*

1. Introduzione

La codifica d'una sorgente assegnata, finita, stazionaria e senza memoria, può essere eseguita procedendo in diversi modi. I vincoli di formato sono tra i più immediati: in particolare, imponendo che i messaggi elementari abbiano eguale lunghezza, si ottengono i codici

(*) Pervenuto in Redazione l'11 gennaio 1985.
(**) Indirizzo dell'Autore: Istituto di Matematica dell'Università - Piazzale Europa, 1 - 34100 Trieste.

da blocco a lunghezza variabile (B-LV); richiedendo, invece, che siano le parole di codice ad avere eguale lunghezza, si ottengono i codici *da lunghezza variabile a blocco* (LV-B). Data la simmetria dei vincoli ai quali sono soggetti i codici B-LV e LV-B, è spontaneo supporre le due famiglie dotate di proprietà speculari; per verificare la legittimità di tale ipotesi, cominciamo col riassumere brevemente le principali caratteristiche dei codici LV-B (cfr. anche [1] o [2]).

A tale scopo, introduciamo gli alfabeti $\mathcal{A} = \{a_1, \ldots, a_K\}$ (primario) e $\mathcal{B} = \{b_1, \ldots, b_D\}$ (secondario) e denotiamo con $\mathcal{M}$ e $\mathcal{C}$ l'insieme dei messaggi elementari e, rispettivamente, l'insieme delle parole di codice (dizionario) di φ, generico codice LV-B, avente lunghezza di blocco 1.

La proprietà che caratterizza i codici LV-B ne garantisce l'*univoca decodificabilità,* mentre una richiesta sufficiente ad assicurare la *univoca codificabilità* è la seguente: $\mathcal{M}$ è *completo,* ossia *esauriente* (ogni successione primaria ha almeno un prefisso in $\mathcal{M}$) e *a prefisso* (dati comunque due messaggi elementari, l'uno non è prefisso dell'altro); notiamo che tale richiesta garantisce, in più, l'*istantanea codificabilità* di φ.

Definizione 1.1: Dicesi *estensione* dell'alfabeto di sequenze (finite) $\mathcal{U}$, mediante il suo elemento u *(di estensione,* o *esteso)* l'insieme $\mathcal{U}(u) \triangleq (\mathcal{U} \setminus \{u\}) \bigcup\limits_{a \in \mathcal{A}} \{ua\}$.

Ogni insieme $\mathcal{M}$ completo può essere pensato come il risultato di j estensioni successive (j opportuno numero naturale; $j \in \mathbf{N}$), eseguite a partire da $\mathcal{A}$, di elementi estesi rispettivi $u_1, \ldots, u_j$, pertanto ha cardinalità $T_j \triangleq j \cdot (k-1) + k$.

Un indice ragionevole dell'economicità di φ è il *rapporto di compressione* (r.c.) $E(L)(\mathcal{M}) / l \triangleq$ lunghezza media degli elementi di $\mathcal{M} / l$; fissato $|\mathcal{M}| = T_j (j \in \mathbf{N})$, il valore minimo che l può assumere è $\lceil \log_D T_j \rceil$, mentre un punto di massimo di $E(L)(\mathcal{M})$ è fornito dall'*algoritmo di Tunstall:* per $i = 1, \ldots, j$, u_i è una sequenza di probabilità massima dell'insieme nel quale è prescelta; il *codice di Tunstall,* quindi, è ottimo tra i codici LV-B istantaneamente codificabili (a prefisso).

Una condizione sufficiente a garantire per φ l'univoca codificabilità, ma più mite di quella riportata in precedenza, è: $\mathcal{M}$ è *esauriente* ed è assegnata una *regola decisionale,* che consente d'associare ad ogni successione primaria un'unica sua possibile segmentazione in messaggi elementari. Come vedremo, i codici LV-B che soddisfano questa richiesta (ai quali, nel seguito, restringeremo la nostra attenzione) sono in generale affetti da ritardo di codifica.

A questo punto, riesaminiamo la relazione che intercorre tra codici LV-B e B-LV. E' noto che per ogni codice Ψ, di tipo B-LV, è possibile individuare il codice Ψ_1, di Huffman, corrispondente (tale che Ψ e Ψ_1 abbiano eguale lunghezza di blocco), che migliora Ψ, anche se non necessariamente in senso stretto, ed è istantaneamente decodificabile: in questo senso, adoperare un codice B-LV affetto da ritardo di decodifica non è mai vantaggioso. In base all'ipotizzata specularità fra codici B-LV e LV-B, dovrebbe sussistere la proprietà simmetrica della precedente: nessun codice LV-B che non sia a prefisso può essere conveniente. La scoperta d'un codice LV-B che, sebbene affetto da ritardo di codifica, è migliore del codice di Tunstall corrispondente ([4]) è, quindi, sufficiente ad esigere un esame più accurato della situazione. A tale scopo, giova introdurre un modello di codice LV-B, che formalizza la possibilità di scindere ogni operazione di codifica in due fasi (si veda la sez. 3):

a) [PRECODIFICA] segmentazione delle uscite della sorgente in una successione univocamente definita di messaggi elementari, ottenibile applicando una regola decisionale prefissata;

b) traduzione progressiva della successione dei messaggi elementari in una successione di parole di codice, mediante un ulteriore operatore di codifica.

Ipotizzando, come abbiamo fatto sinora, che un codice sia tanto più conveniente, quanto maggiore è, in media, il rapporto fra lunghezza della sequenza primaria codificabile e lunghezza della sua parola di codice, è ragionevole assumere come indice parziale dell'economicità di φ il r.c. j-esimo $E(L_j)(\varphi)/j \cdot l \triangleq$ lunghezza media della sequenza primaria codificabile in φ con j parole del dizionario$/j \cdot l$ $(\forall j \in \mathbf{N})$.

Definizione 1.2: Per ogni coppia di codici LV-B, φ_1 e φ_2, di lunghezze di blocco rispettive l_1 e l_2, diciamo φ_1 *j-migliore* di φ_2, se $E(L_j)(\varphi_1)/j \cdot l_1 \geqslant E(L_j)(\varphi_2)/j \cdot l_2$; diciamo φ_1 *ω-migliore* di φ_2, se $\liminf_{j\to+\infty} E(L_j)(\varphi_1)/j \cdot l_1 \geqslant \liminf_{j\to+\infty} E(L_j)(\varphi_2)/j \cdot l_2$.

(Il problema dell'esistenza di $\lim_{j\to+\infty} E(L_j)(\varphi)/j$ è stato risolto sotto ipotesi particolari relative a φ (cfr. [5])).

Nell'ambito dello schema accennato, mostreremo tra poco che il codice dell'esempio citato è j-migliore, per ogni $j \in \mathbf{N}$, e ω-migliore del codice di Tunstall corrispondente. Questo risultato è sorprendente; tuttavia la dimostrazione non è generalizzabile, in quanto sfrutta la possibilità di utilizzare i metodi standard di trattazione delle catene di Markov, offerta dalla particolare forma di φ. Vale, comunque, la seguente proposizione: per ogni codice di Tunstall fissato, può essere individuato un codice (LV-B; nel seguito, lo sottintenderemo) che ne è 1-migliore, purché una lettera dell'alfabeto primario sia sufficientemente probabile; presenteremo un algoritmo che fornisce uan dimostrazione costruttiva dell'assunto precedente.

Nella sez. 2 introduciamo alcune notazioni; nella sez. 3 proponiamo un modello di codice, mentre la sez. 4 è dedicata alla descrizione di alcune sottofamiglie notevoli di codici. L'esempio e l'algoritmo citati sono discussi nelle sez. 5 e 6.

2. Notazioni

Sia $X = \{x_1, \dots, x_K\}$ un *alfabeto*, ovvero un insieme finito, non vuoto. Identifichiamo la generica distribuzione di probabilità (d.p.) P su X col corrispondente vettore di probabilità.

Denotiamo con X^-, X^+ e X^∞, rispettivamente:

— l'insieme delle sequenze (finite), non nulle, di lettere;

— l'estensione di X^- con λ, sequenza nulla (di lunghezza 0);

— l'insieme delle successioni («sequenze infinite») di lettere.

La legge di composizione $\circ$, così definita su $X^+ : \forall x_1, x_2 \in X^+$, $x_1 \circ x_2$ è l'elemento di X^+ che si ottiene scrivendo le lettere di x_2 di seguito alle lettere di x_1, dota X^- e X^+ della struttura di monoide e, rispettivamente, di semigruppo (monoide fornito d'unità).

Per indicare che x è prefisso di y scriviamo: $x < y$; se $x < y$, $y - x$ denota y, privato del prefisso x.

Poniamo, inoltre: $|x| = n, \forall x \in X^n, \forall n \in \mathbf{N}$;

$|x| = \chi_0$ (1), $\forall x \in X^\infty$ ($|x|$ è la *lunghezza* di x).

Sia $\mathfrak{M}$ un alfabeto di sequenze non nulle; diciamo $\mathfrak{M}$:

— *esauriente*, se ogni $x \in X^\infty$ ha prefisso in $\mathfrak{M}$;

— *a prefisso*, se, dati comunque m_i e $m_j \in \mathfrak{M}$, $m_i \nless m_j$ (m_i non è prefisso di m_j);

— *completo*, se gode di entrambe le proprietà precedenti;

— *separabile*, se ogni $x \in X^\infty$ è separabile (rispetto a $\mathfrak{M}$), ossia segmentabile in almeno una successione di elementi di $\mathfrak{M}$;

esaurienza e separabilità sono, evidentemente, proprietà equivalenti.

Con la notazione $\{\mathcal{U}_i\}_{i \geqslant 0}$ indichiamo la generica successione degli insiemi costruibili applicando l'algoritmo di Tunstall, a partire da $\mathcal{U}_0 = X$ (*insiemi di Tunstall*, relativi a X); per ogni $i \geqslant 0$, $|\mathcal{U}_i| = T_i$ ($T_i = (k-1) \cdot i + k$, definito in sez. 1).

(1) χ sta qui a indicare la lettera ebraica «aleph».

3. Un modello di codice (LV-B)

Siano assegnati: gli alfabeti $\mathcal{A} = \{a_1, \ldots, a_K\}$ (primario) e $\mathcal{B} = \{b_1, \ldots, b_D\}$ (secondario); l'insieme dei messaggi elementari $\mathcal{M} = \{m_1, \ldots, m_n\} \subseteq \mathcal{A}^-$, esauriente; il dizionario $\mathcal{C} \subseteq \mathcal{B}^l$, con: $|\mathcal{C}| = n$ e $l = \lceil \log_D n \rceil$ (lunghezza di blocco).

Sulle applicazioni $\tilde{h} : \mathcal{A}^\infty \to \mathcal{M}^\infty$ e $\tilde{\psi} : \mathcal{M} \to \mathcal{C}$ facciamo le seguenti ipotesi:

1) $\forall x \in \mathcal{A}^\infty$, $\tilde{h}(x)$ è una segmentazione di x in messaggi elementari (si ricordi che si è supposto $\mathcal{M}$ esauriente, dunque separabile);

2) $\forall x \in \mathcal{A}^-, \forall x_1, x_2 \in \mathcal{A}^\infty$: (per $i = 1, 2$)
$x < x_i$; $\exists j_i \in \mathbf{N} : x = \tilde{h}(x_i)_1 \circ \ldots \circ \tilde{h}(x_i)_{j_i}$,
risulta: $j = j_1 = j_2$; $\tilde{h}(x_1)_k = \tilde{h}(x_2)_k$ $(1 \leqslant k \leqslant j)$;

3) $\forall x \in \mathcal{A}^\infty$, $\forall k \in \mathbf{N}$, $\exists p \in \mathbf{N}$: noto $(x_1, \ldots, x_p)$, è individuato $(\tilde{h}(x)_1, \ldots, \tilde{h}(x)_k)$;

4) $\tilde{\psi}$ è biiettiva.

Posto: $S^- \triangleq \{x \in \mathcal{A}^- \mid \exists \tilde{x} \in \mathcal{A}^\infty, \exists j \in \mathbf{N} : x = \tilde{h}(\tilde{x})_1 \circ \ldots \circ \tilde{h}(\tilde{x})_j\}$, definiamo la *regola decisionale* (o *di preferenza*) $h : \mathcal{A}^\infty \cup S^- \to \mathcal{M}^\infty \cup \mathcal{M}^-$ e l'applicazione $\psi : \mathcal{M}^\infty \cup \mathcal{M}^- \to \mathcal{C}^\infty \cup \mathcal{C}^-$, come segue:

$$h|_{\mathcal{A}^\infty} = \tilde{h}; \ \forall x \in S^-, \ h(x) = (\tilde{h}(\tilde{x})_1, \ldots, \tilde{h}(\tilde{x})_j),$$

con riferimento alle notazioni introdotte nella definizione di S^-; $\forall m = (m_{i_1}, m_{i_2}, \ldots) \in \mathcal{M}^\infty \cup \mathcal{M}^-$, $\psi(m) = (\tilde{\psi}(m_{i_1}), \tilde{\psi}(m_{i_2}), \ldots)$.

Definizione 3.1 - Un codice φ è la composizione di due applicazioni, h e ψ, soddisfacenti le proprietà precedentemente indicate: $\varphi = \psi \circ h$. Le richieste fatte possono essere così giustificate: 1), 2) e 4) garantiscono per φ l'univoca codificabilità, mentre 1) e 4) ne assicurano l'univoca decodificabilità; 3) esclude che φ sia impossibile da costruire, perché affetto da ritardo di codifica infinito.

Il ritardo di codifica, tuttavia, può risultare arbitrariamente elevato, come accade per $\varphi = \psi \circ h$, così definito:

Siano: $\mathcal{A} = \{0, 1\}$; $\mathcal{M} = \mathcal{C}\{0, 00, 1\}$; $\psi \equiv$ identità; $r \in \mathbf{N}$, fissato; segmentiamo ogni $x \in \mathcal{A}^\infty$, scegliendo al passo i-esimo:

— $0 \ldots 0$ $(i \cdot r$ zeri), se possibile;

— $0 \ldots 01$, in caso contrario;

segmentiamo, poi, ogni sequenza di zeri (risp. $0 \ldots 1$) ottenuta in:

— $|00|\ldots|00|$ (risp. $|00|\ldots|1|$), se essa contiene un numero pari di 0;

— $|0|\ldots|00|$ (risp. $|0|\ldots|1|$), in caso contrario;

viene, così, assegnata un'applicazione $\tilde{h}: \mathcal{A}^\infty \rightarrow \mathcal{M}^\infty$, che individua una regola decisionale h. In base a h, la successione di zeri, in particolare, è segmentata inizialmente in:

$$\underbrace{0\ldots0}_{r}\ \underbrace{0\ldots0}_{2r}\ \underbrace{0\ldots0}_{3r}\ \ldots,$$

quindi il primo messaggio elementare delle i-esima sequenza risultante è individuato con un ritardo di (almeno) $(i \cdot r - 2)$ lettere primarie, dove i può essere arbitrariamente elevato.

4. I codici α_j e α_ω

Assegnata su $\mathcal{A}$ la d.p. $P = (p_1, \ldots, p_k)$, in relazione a $\varphi = \psi \circ h$ (di lunghezza di blocco 1), possiamo introdurre le seguenti variabili aleatorie:

X, uscite della sorgente (a valori in $\mathcal{A}^\infty$);

$M = \{M_i\}_{i \geqslant 1}$, uscite della sorgente dei messaggi, ottenute segmentando le uscite della sorgente, in base alla regola di preferenza h (a valori in $\mathcal{M}^\infty$);

$$Y_j = h^{-1}(M_j);\ L_j = |Y_1 \circ \ldots \circ Y_j| = \sum_1^j{}_i |Y_i| \quad (\forall j \in \mathbf{N}).$$

Il r.c. j-esimo $E(L_j)(\varphi)/j \cdot l = \sum_1^j{}_i E[|Y_i|]/j \cdot l$ dipende, tramite il numeratore, esclusivamente da $\mathcal{M}$ e h, quindi è sufficiente, per poter valutare la convenienza di un codice, conoscerne l'operatore di precodifica. Fortunatamente, inoltre, le regole decisionali più utili sono molto semplici; in particolare, fissato $j \in \mathbf{N}$ e $\mathcal{M}$, è facile individuare $h(j)$, punto di massimo di $E(L_j)(\mathcal{M}, h)/j$:

Siano $\tilde{\mathcal{M}}^j = \{\tilde{m}_1, \ldots, \tilde{m}_T\}$ e $h^*: \tilde{\mathcal{M}}^j \rightarrow \mathcal{M}^j$ l'insieme (esauriente) delle sequenze primarie, ottenibili componendo in $\circ j$ messaggi elementari e, rispettivamente, un'applicazione che associa a ciascun $\tilde{m}_i$ una sua possibile segmentazione in j sequenze di $\mathcal{M}$. $\forall x \in \mathcal{A}^\infty$, diciamo $h'(x)$ la suddivisione di x che si ottiene scegliendo, ad ogni passo, $\tilde{m}_i$ di lunghezza massima possibile; designamo, inoltre, con $h''(x)$ l'elemento di $\mathcal{M}^\infty$, risultante dalla sostituzione di ogni termine di $h'(x)$ con la sua immagine in h^*. La regola decisionale h'', individuata dall'applicazione h'' definita, è $h(j)$; infatti, detta h la generica regola di preferenza, si ha:

$$E(L_j)(\mathfrak{M}, h) = \sum_1^T{}_t \, |\tilde{m}_t| \cdot \text{prob}\{x \,|\, h(x)_1 \circ \ldots \circ h(x)_j = \tilde{m}_t\} \leqslant$$
$$\leqslant \sum_1^T{}_i \sum_1^T{}_t \, |\tilde{m}_i| \cdot \text{prob}\{x \,|\, h''(x)_1 \circ \ldots \circ h''(x)_j = \tilde{m}_i;$$
$$h(x)_1 \circ \ldots \circ h(x)_j = \tilde{m}_t\} = E(L_j)(\mathfrak{M}, h'').$$

Nella medesima ipotesi ($\mathfrak{M}$ fissato) non è agevole, invece, determinare $h(\omega)$, punto di massimo di $\liminf_{j \to +\infty} E(L_j)(\mathfrak{M}, h)/j$, in quanto l'espressione precedente non ha una forma semplice, a meno che non valgano ipotesi particolari su $\{Y_i\}_{i \geqslant 1}$ (cfr. [5]).

Definizione 4.1 - Diciamo *codici* $\alpha_j(\alpha_\omega)$ i codici ottenibili fissando $h = h(j)$ (risp. $h = h(\omega)$).

Per giustificare la nostra impostazione, occorre verificare che le famiglie dei codici α_j e α_h possono essere distinte, per $j \neq h$. A tale scopo, è sufficiente considerare il seguente esempio:

Siano: $\mathfrak{A} = \{0, 1\}$; $P = (\alpha, \beta)$; $\mathfrak{M} = \{0, 01, 010, 1\}$; $\mathfrak{B} = \mathfrak{C} = \{a, b, c, d\}$; $\varphi_1(\varphi_2)$ un codice α_1 (risp. α_2), relativo a $\mathfrak{M}$ e $\mathfrak{C}$. Con semplici calcoli, risulta: $E(L_2)(\varphi_1) - E(L_2)(\varphi_2) = -\alpha^3\beta^2 \lneqq 0$; pertanto, φ_2 non può coincidere con φ_1, giacché ne è strettamente 2-migliore.

5. Un esempio notevole

Con la terminologia adottata, possiamo dire che l'esempio preannunciato nella sez. 1 fornisce un codice α_1, j-migliore $(\forall j \in \mathbf{N})$ e ω-migliore del codice di Tunstall corrispondente.

Siano dati: $\mathfrak{A} = \{0, 1, 2\}$; $P = (\alpha, \beta, \gamma)$, con $\alpha \gneqq \beta \geqslant \gamma$; $\mathfrak{B} = \{a, b, c, d, e\}$; $\mathfrak{M} = \{000, 00, 0, 1, 2\}$. Per definizione, risulta $\mathfrak{U}_1 = \{00, 01, 02, 1, 2\}$. $\mathfrak{M}$ e $\mathfrak{U}_1$ hanno la stessa cardinalità, quindi è uguale, e pari a 1, la lunghezza di blocco minima sufficiente a codificare ciascuno dei due insiemi: sia $\mathfrak{C} = \{a, b, c, d, e\}$.

Designamo con φ e ξ un codice α_1, relativo a $\mathfrak{M}$ e $\mathfrak{C}$ e, rispettivamente, un codice di Tunstall, relativo a $\mathfrak{U}_1$ e $\mathfrak{C}$. $\{Y_i\}_{i \geqslant 1}$, segmentazione aleatoria delle uscite della sorgente nei messaggi di $\mathfrak{M}$, ottenibile applicando $h(1)$, è una catena di Markov. Conglobando 1 e 2 in una «super-lettera», di probabilità complessiva $(1 - \alpha)$ (il che non comporta variazione alcuna nel calcolo degli $E(L_j)(\varphi)$), l'evoluzione della catena è descritta dalla d.p. iniziale:

$\Pi = (\text{prob}\{Y_1 = 000\}, \text{prob}\{Y_1 = 00\}, \text{prob}\{Y_1 = 0\}, \text{prob}\{Y_1 = 1,2\}) =$
$= (\alpha^3, \alpha^2(1-\alpha), \alpha(1-\alpha), 1-\alpha)$ e dalla matrice di transizione R:

	000	00	0	1,2
000	α^3	$\alpha^2(1-\alpha)$	$\alpha(1-\alpha)$	$1-\alpha$
00	0	0	0	1
0	0	0	0	1
1,2	α^3	$\alpha^2(1-\alpha)$	$\alpha(1-\alpha)$	$1-\alpha$

Definiamo per ricorrenza le seguenti quattro successioni:

$(a_0 = b_0 = c_0 = 0;\ d_0 = 1);\ a_1 = \alpha^3;\ b_1 = \alpha^2(1-\alpha);$
$c_1 = \alpha(1-\alpha);\ d_1 = 1-\alpha;$

$(\forall n \geqslant 2)$ per $x = a, b, c,\ x_n = x_1(a_{n-1} + d_{n-1});$
$d_n = d_1(a_{n-1} + d_{n-1}) + b_{n-1} + c_{n-1}.$

E' facile provare, procedendo per induzione su n, che $\forall n \geqslant 1$:

$$(^\circ) \quad a_n + d_n = \sum_{i=0}^{n} [-(b_1 + c_1)]^i;$$

$$(^{\circ\circ}) \quad R^n = \begin{pmatrix} a_n & b_n & c_n & d_n \\ a_{n-1} & b_{n-1} & c_{n-1} & d_{n-1} \\ a_{n-1} & b_{n-1} & c_{n-1} & d_{n-1} \\ a_n & b_n & c_n & d_n \end{pmatrix};$$

notiamo che $(^{\circ\circ})$ implica: $a_n + b_n + c_n + d_n = 1$.

Utilizzando $(^\circ)$ e $(^{\circ\circ})$, dal calcolo diretto risulta, qualunque sia $n \in \mathbf{N}$:

$$\begin{aligned} \text{prob}\{Y_n = 000\} &= (\Pi * R^{n-1})_1 = a_n; \\ \text{prob}\{Y_n = 00\ \} &= (\Pi * R^{n-1})_2 = b_n; \\ \text{prob}\{Y_n = 0\ \ \} &= (\Pi * R^{n-1})_3 = c_n; \\ \text{prob}\{Y_n = 1,2\} &= (\Pi * R^{n-1})_4 = d_n; \end{aligned}$$

pertanto: $E[|Y_n|] = 3a_n + 2b_n + c_n + d_n = 2a_n + b_n + 1$ e finalmente:

$$E(L_n)(\varphi) = \sum_{i=1}^{n} E[|Y_i|] = n + \sum_{i=1}^{n} (2a_i + b_i).$$

Sviluppiamo la sommatoria:

$$\sum_{i=1}^{n} (2a_i + b_i) = (2a_1 + b_1) \sum_{i=1}^{n} \sum_{j=0}^{i-1} [-(b_1 + c_1)]^j =$$

$$= \frac{2a_1 + b_1}{1 + b_1 + c_1}\{n - \sum_{1}^{n}{}_i [-(b_1 + c_1)]^i\} =$$

$$= \frac{2a_1 + b_1}{1 + b_1 + c_1}\{n + (b_1 + c_1)\sum_{0}^{n-1}{}_i [-(b_1 + c_1)]^i\} \geqslant$$

$$\geqslant \frac{2a_1 + b_1}{1 + b_1 + c_1} \cdot n = \frac{\alpha^2(\alpha + 1)}{1 + \alpha(1 - \alpha^2)} \cdot n.$$

D'altro canto, $\forall n \in \mathbf{N}, E(L_n)(\xi) = n(1 + \alpha)$, pertanto:

$$E(L_n)(\varphi) - E(L_n)(\xi) \geqslant \frac{\alpha^2(\alpha + 1)}{1 + \alpha(1 - \alpha^2)} \cdot n - n\alpha = n\alpha \cdot \frac{\alpha^3 + \alpha^2 - 1}{1 + \alpha(1 - \alpha^2)} > 0,$$

se $\alpha \geqslant 0.755$: *φ è strettamente n-migliore di ξ, qualunque sia $n \in \mathbf{N}$, se $\alpha \geqslant 0.755$.*

Il r.c. n-esimo, valutato in φ, può essere espresso come:

$$(r \mid \Pi_* \sum_{0}^{n-1}{}_i \frac{R^i}{n}),$$

con $r = (3, 2, 1, 1)$. Al divergere di n a $+\infty$, la matrice limite di R^n (limite eseguito elemento per elemento) ha le righe coincidenti con:

$$\left(\frac{a_1}{1 + b_1 + c_1}, \frac{b_1}{1 + b_1 + c_1}, \frac{c_1}{1 + b_1 + c_1}, \frac{1 - a_1}{1 + b_1 + c_1}\right);$$

da ciò segue:

$$\lim_{n \to +\infty} \frac{E(L_n)(\varphi)}{n} = \frac{2a_1 + 2b_1 + c_1 + 1}{1 + b_1 + c_1} = \frac{\alpha^2 + \alpha + 1}{1 + \alpha(1 - \alpha^2)}.$$

Essendo, ovviamente: $\lim_{n \to +\infty} \frac{E(L_n)(\xi)}{n} = 1 + \alpha$, risulta:

$$\lim_{n \to +\infty} \frac{E(L_n)(\varphi)}{n} - \lim_{n \to +\infty} \frac{E(L_n)(\xi)}{n} = \frac{\alpha(\alpha^3 + \alpha^2 - 1)}{1 + \alpha(1 - \alpha^2)} > 0, \text{ se}$$

$\alpha \geqslant 0.755$: *φ è strettamente ω-migliore di ξ, se $\alpha \geqslant 0.755$.*

Dal punto di vista della trattazione matematica, il prezzo che viene pagato per l'incremento dei rapporti di compressione è un aumento di complessità della sorgente dei messaggi, giacché $\{Y_i\}_{i \geqslant 1}$ non è più composta da variabili indipendenti ed equidistribuite. Beninteso, esistono casi particolari fortunati, come quello dell'esempio analizzato, in cui i legami di dipendenza delle Y_i sono abbastanza semplici.

6. Un algoritmo competitivo con quello di Tunstall

Siano: $\mathcal{A} = \{0, 1, \ldots, (k - 1\}$; $P = (p_0, \ldots, p_{k-1})$; senza perdita di generalità, possiamo supporre: $p_0 \geqslant p_1 \geqslant \ldots \geqslant p_{k-1}$. Fissata la lunghezza di blocco $1 \geqslant \log_D T_1$, definiamo: $r \triangleq \max\{n \mid T_n \leqslant D^l\}$ (r consente di individuare la cardinalità massima d'un insieme completo codificabile utilizzando la lunghezza di blocco l ed è certamente diverso da 0).

$\mathcal{U}_t$ è estensione di $\mathcal{U}_{t-1}$, mediante un suo elemento di probabilità massima, che designamo con u_t. $\forall n \in \mathbf{N}$ sia:

$$\mathcal{M}_{r,n} \triangleq \mathcal{U}_{r-1} \bigcup_{a=0}^{k-2} \{u_r \circ \underbrace{0 \circ \ldots \circ 0}_{n-1} \circ a\};$$

poiché $|\mathcal{M}_{r,n}| = |\mathcal{U}_r|$, l è la minima lunghezza di blocco sufficiente a codificare $\mathcal{M}_{r,n}$. Si verifica facilmente che:
$E(L_1)(\mathcal{M}_{r,n}, h(1)) = E(L_1)(\mathcal{U}_{r-1}, h(1)) + P^{|u_r|}(u_r) \cdot n \cdot p_0^{n-1}(1-p_{k-1})$;
$f(x) = x \cdot p_0^{x-1}$ ha per unico punto di massimo $\tilde{x} = -1/\log p_0$, quindi $E(L_1)(\mathcal{M}_{r,n}, h(1))$ assume valore massimo in $\mathcal{M}_r = \mathcal{M}_{r,\tilde{n}}$, con

$$\tilde{n} \in \left\{\left\lfloor -\frac{1}{\log p_0}\right\rfloor, \left\lceil -\frac{1}{\log p_0}\right\rceil\right\}.$$

Da ciò segue che: $\forall r \geqslant 1$, $E(L_1)(\mathcal{M}_r, h(1)) - E(L_1)(\mathcal{U}_r, h(1)) = P^{|u_r|}(u_r)\,[p_0^{\tilde{n}-1} \cdot (1-p_{k-1}) \cdot \tilde{n} - 1] \underset{\neq}{\geq} 0$, se $p_0^{\tilde{n}-1} \cdot (1-p_{k-1}) \cdot \tilde{n} \underset{\neq}{\geq} 1$: se la prima lettera di $\mathcal{A}$ ricorre con sufficiente frequenza, l'algoritmo fornisce $\{\mathcal{M}_i\}_{i\geqslant 1}$ *migliore* di $\{\mathcal{U}_i\}_{i\geqslant 1}$, ovvero tale che: *ciascun codice* α_1, *relativo a* $\mathcal{M}_i$, *è* 1-*migliore di ogni codice di Tunstall, relativo a* $\mathcal{U}_i$, *qualunque sia* $i \in \mathbf{N}$.

Esempio - Siano $\mathcal{A} = \{0, 1, 2\}$; $P = (0.7, 0.2, 0.1)$, nel qual caso $\tilde{n} = 3$ e $\tilde{n} \cdot p_0^{\tilde{n}-1}(1-p_{k-1}) = 1.323 \underset{\neq}{\geq} 1$.

— $\mathcal{U}_1 = \{00, 01, 02, 1, 2\}$; per $\mathcal{M}_1 = \{0, 1, 2, 0000, 0001\}$, si ha:
$E(L_1)(\mathcal{M}_1, h(1)) - E(L_1)(\mathcal{U}_1, h(1)) = 0.2261$;

— $\mathcal{U}_2 = \{000, 001, 002, 01, 02, 1, 2\}$;
per $\mathcal{M}_2 = \{00000, 00001, 00, 01, 02, 1, 2\}$, si ha:
$E(L_1)(\mathcal{M}_2, h(1)) - E(L_1)(\mathcal{U}_2, h(1)) = 0.15827$, e così via.

7. Conclusioni

Il maggiore inconveniente nella trattazione matematica dei codici LV-B non a prefisso è che la successione $M = \{\mathcal{M}_i\}_{i\geqslant 1}$ relativa ad ognuno di essi è composta da variabili dipendenti, in maniera complessa, sicché, almeno in generale, non si può fare uso dei comodi strumenti che sono disponibili nel caso di processi stocastici stazionari. Tuttavia, sul piano pratico, si è visto che l'introduzione di un certo ritardo di codifica può essere conveniente ed è particolarmente importante che ciò avvenga per sorgenti «sghembe», visto che l'esperienza empirica indica che i codici LV-B sono particolarmente indicati appunto nel caso di tali sorgenti. Naturalmente, sempre sul piano pratico, rimane aperto il problema di trovare un compromesso ragionevole tra l'esigenza di conseguire un rapporto di compressione molto elevato e quella di contenere i ritardi di codifica entro limiti accettabili.

Si ringrazia il prof. Andrea Sgarro per l'incoraggiamento e per i preziosi consigli che hanno reso possibile questo lavoro.

BIBLIOGRAFIA

[1] G. LONGO, *Teoria dell'Informazione*, Boringhieri (1980).

[2] F. JELINEK e G. LONGO, *Algorithms for Source Coding*, in «Coding and Complexity», a cura di G. Longo, CISM Courses and Lectures, n. 216 (Springer, Vienna e New York 1976).

[3] I. CSISZAR e J. KÖRNER, *Information Theory*, Academic Press (1981).

[4] T. NEMETZ, comunicazione personale.

[5] M. ALLOCCA, *Compressione di messaggi a lunghezza variabile*, Pubblicazioni dell'Istituto di Matematica Applicata, Università degli Studi di Trieste (1984).