Prof. FRANCESCO A.N. PALMIERI

[Appunti per il corso di "TRASMISSIONE ED ELABORAZIONE NUMERICA DEI SEGNALI"

SUN AA. 2014-15

Capitolo 7

Codifica di canale

In questo capitolo viene introdotta la codifica di canale. Vengono presentate le relazioni esistenti tra le quantità informazionali, quali la capacità e l'ambiguità, con la probabilità di errore. Viene inoltre presentato il secondo teorema di Shannon che costituisce uno dei capisaldi della teoria dell'informazione, secondo il quale viene mostrato come, mediante opportuna codifica, sia possibile usare un canale con probabilità di errore asintoticamente nulle, purchè il tasso della sorgente sia al di sotto della capacità di canale. Esempi di semplici codici di canale sono presentati e discussi.

7.1 Introduzione

Abbiamo visto nel capitolo scorso come il canale discreto sia uno schema di riferimento generale per modellare il meccanismo aleatorio di trasformazione di simboli da un alfabeto all'altro. Anche se le applicazioni del modello non si limitano alle comunicazioni, ma vanno dalla soluzione di problemi di diagnostica alla analisi di sistemi, in questo capitolo focalizziamo la nostra attenzione sul modello di comunicazione punto-punto. Nel nostro studio si suppone che l'obiettivo sia la trasmissione affidabile di un insieme di simboli da una sorgente ad un destinatario. L'analisi tende a modellare in maniera equivalente tutti i fenomeni di degradazione che un ipotetico segnale reale subisce in un trasporto non completamente affidabile. Ad esempio la sequenza di simboli, dopo una opportuna modulazione, viene trasmessa su cavo, o su un canale di progazione elettromagnetica libera, o su un supporto magnetico o ottico, ecc. Il ricevitore riconverte i segnali ricevuti in una sequenza di simboli che non sono esattamente quelli trasmessi. Il modello semplificato che esaminiamo in questo capitolo, assume che le degradazioni

siano indipendenti ad ogni *uso del canale* (canale senza memoria)[1].

Successivamente la formulazione verrà generalizzata ad un canale a blocchi

La probabilità di errore media (per simbolo o per blocco) (che in seguito è semplicemente chiamata probabilità di errore) costituisce il parametro di prestazioni ultimo per qualificare la comunicazione. Le prestazioni ottenute in un sistema possono essere considerate soddisfacenti a seconda della applicazione. Ci sono situazioni in cui la probabilità di errore deve essere garantita a ordini di grandezza di 10^{-7}, a volte anche di 10^{-12} e oltre, così come in altri casi sono accettabili valori inferiori di affidabilità. Quindi se le prestazioni ottenute dovessero risultare insoddisfacenti, sembrerebbe che l'unica alternativa sia dotarsi di un canale migliore. Migliori canali possono essere ottenuti aumentando la potenza in trasmissione, o magari agendo su altri parametri negli stadi di modulazione, che si tradurrebbero in una matrice di canale con parametri più favorevoli. Quindi sembrerebbe che il sistema cosi com'è non sia suscettibile di alcuna miglioria. Ma è proprio vero che se si dispone di un canale discreto rumoroso non è possibile utilizzarlo per una comunicazione affidabile? L'idea che il rumore non possa essere combattuto se non con aumenti di potenza è oramai largamante superata. Si pensi alle comunicazioni interplanetarie che operano in condizioni di rapporto segnale/rumore molto inferiore a uno. La soluzione a questa problematica è fornita in maniera elegante dagli strumenti della teoria della informazione che predice quanta informazione può essere trasmessa su un canale con errori. E' una opportuna *codifica* a far sì che una sorgente, a tasso inferiore alla capacità, possa convogliare la sua informazione in un formato che si adatta alle condizioni del canale. L'idea chiave è che introducendo "ridondanza" ovvero riducendo il flusso informativo netto si possono ottenere comunque trasmissioni affidabili.

A questa problematica cerchiamo di dare una risposta in questo capitolo mediante gli strumenti della teoria dell'informazione. Le quantità descrittive del canale quali la mutua informazione, la capacità e la ambiguità forniscono utili indicazioni sulle prestazioni finali quali la probabilità di errore, in relazione alla codifica utilizzata. Impareremo a mettere in relazione i vari parametri del sistema quali la cadenza della sorgente, la cardinalità dei vari alfabeti e il rapporto di codifica.

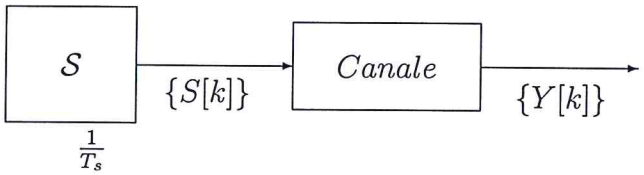


Figura 7.1: Canale senza codifica

7.2 Trasmissione senza codifica

Cominciamo con un primo schema di riferimento riportato in Figura 7.1 dove una sorgente discreta emette un simbolo $S[k]$ ogni T_s secondi prelevandolo da un alfabeto $\mathcal{S} = \{a_1, ..., a_d\}$ di cardinalità d con distribuzione $\Pi_s = \{p_1, ..., p_d\}$. Il canale ad ogni simbolo di ingresso associa un simbolo $Y[k]$ appartenente all'alfabeto di uscita $\mathcal{Y} = \{y_1, ..., y_d\}$ della stessa cardinalità. Il canale viene usato una volta per ogni istante k in maniera indipendente. L'indicazione dell'indice temporale k viene omessa in seguito per semplicità in quanto ad ogni istante il modello prevede una diversa realizzazione della variabile aleatoria di ingresso e del comportamento aleatorio del canale. La matrice

$$\mathbf{P}_c = [Pr\{Y = y_i | S = a_j\}]_{i,j=1,...,d} = [P(y_i|a_j)]_{i,j=1,...,d}\,, \tag{7.1}$$

contiene tutte le probabilità di transizione. Se il canale è preposto al trasporto dei simboli $\{a_1, ..., a_d\}$ nei simboli $\{y_1, ..., y_d\}$ nella corrispondenza $a_j \rightarrow y_j, \quad j = 1, ..., d$, la probabilità di errore media è

$$P(e) = \sum_{i=1}^{d} \sum_{j=1;j\neq i}^{d} P(a_i, y_j), \tag{7.2}$$

che può anche essere riscritta come

$$P(e) = \sum_{i=1}^{d} \sum_{j=1;j\neq i}^{d} P(y_j|a_i)P(a_i) = 1 - \sum_{i=1}^{d} P(y_i|a_i)P(a_i), \tag{7.3}$$

[1] Canali con memoria possono essere studiati con modellistica di vario tipo. Gli strumenti di analisi sono analoghi e costituiscono generalizzazioni del caso senza memoria qui discusso. Tali estensioni vanno oltre gli scopi introduttivi di queste note e saranno trattati altrove.

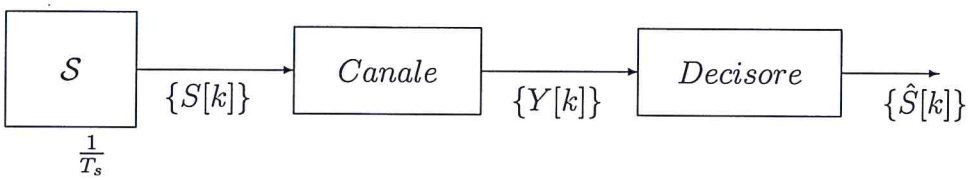


Figura 7.2: Canale senza codifica con decisore

o in forma matriciale

$$P(e) = 1 = diag(\mathbf{P}_c)^T \boldsymbol{\pi}_s. \tag{7.4}$$

La domanda che si pone è: la nostra conoscenza della struttura del canale e delle probabilità dell'ingresso può consentire una migliore inferenza sui simboli di sorgente una volta osservate le uscite?

Per esaminare questo problema è necessario che si preveda all'uscita un *ricevitore*, o un *dispositivo di decisione*, che dalla conoscenza della struttura del problema stimi il simbolo di sorgente senza necessariamente accettare l'osservazione così com'è.

La figura 7.2 mostra lo schema di comunicazione senza codifica a cui è stato aggiunto lo stadio di decisione. Lo schema è più generale di quello di figura 7.1 in quanto assumiamo che l'alfabeto $\mathcal{Y}$ di uscita del canale possa avere una cardinalità diversa da quella di $\mathcal{S}$. Quindi il canale ad ogni uso trasforma un simbolo $S \in \mathcal{S} = \{a_1, ...a_d\}$ in un simbolo $Y \in \mathcal{Y} = \{y_1, ..., y_m\}$ mediante delle probabilità di transizione contenute nella matrice di canale $\mathbf{P}_c$ di dimensioni $d \times m$. Le probabilità dell'ingresso sono $\Pi_S = \{p_1, ..., p_d\}$. Il dispositivo di decisione è una funzione deterministica $D(.)$ che per ogni simbolo osservato Y, "decide" per un simbolo $\hat{S} \in \mathcal{S}$

$$\hat{S} = D(Y) \in \mathcal{S} = \{a_1, ..., a_d\}. \tag{7.5}$$

Il lettore noti che si tratta di un problema di "diagnostica." Le "cause" (simboli di ingresso) si trasformano secondo un meccanismo aleatorio in "sintomi" (osservazioni in uscita al canale) e il decisore, dall'osservazione dei sintomi e dalla conoscenza della struttura probabilistica del sistema, deve fare del suo meglio per risalire alle cause.

Esempio 7.1 Consideriamo un canale binario simmetrico con probabilità di errore $p_e = 0.3$. In tale sistema il 30% dei simboli in uscita è errato e gli errori sono equamente distribuiti tra i due simboli. L'alfabeto di ingresso è $\mathcal{S} = \{a_1, a_2\}$ con probabilità a priori $\Pi_S = \{0.2, 0.8\}$. All'uscita del canale l'alfabeto è $\mathcal{Y} = \{y_1, y_2\}$ e la probabilità di errore media

$$P(e) = p_e Pr\{S = a_1\} + p_e Pr\{S = a_2\} = p_e = 0.3. \tag{7.6}$$

Se invece di prendere i simboli così come sono restituiti dal canale, aggiungessimo un dispositivo di decisione che segue la regola

$$D(y_1) = a_2; \quad D(y_2) = a_2, \tag{7.7}$$

ovvero aggiungessimo un ricevitore che decide sempre per il simbolo a_2 (Figura 7.3), la probabilità di errore sarebbe

$$\begin{aligned} P(e) &= 1 - P(c) = 1 - (Pr\{c|S = a_1\}Pr\{S = a_1\} + Pr\{c|S = a_2\}Pr\{S = a_2\}) \\ &= 1 - (0 \cdot Pr\{S = a_1\} + 1 \cdot Pr\{S = a_2\}) = 0.2. \end{aligned} \tag{7.8}$$

Il decisore adottato migliora le prestazioni del sistema rispetto a un sistema che lascia i simboli invariati! Questo esempio ci lascia in prima battuta certamente un pò perplessi. Comunque se guardiamo attentamente al problema non è troppo difficile convincersi che a causa della dissimmetria delle probabilità dell'ingresso conviene che la decisione sia sempre polarizzata sul simbolo a_2. Quindi in generale una associazione sull'uscita del canale che "non si fidi" dell'osservazione potrebbe portare ad una probabilità di errore migliore. Il problema del progetto del decisore va quindi affrontato in maniera sistematica, come vedremo in seguito.

La funzione del decisore può essere schematizzata come un canale deterministico, trattandosi di una funzione fissa che associa senza ambigutà un simbolo di uscita al simbolo osservato. La matrice di canale $\mathbf{D}$ di dimensioni $d \times m$ rappresenta la regola di decisione. Ricordiamo che il canale deterministico, già discusso nel capitolo precedente, ha un solo elemento diverso da zero e pari a uno in ogni riga. La figura 7.3 mostra lo schema equivalente dell'esempio 7.1. Lo schema generale è mostrato in figura 7.4. Enfatizziamo la maggiore generalità dello schema con il decisore rispetto a quello senza in quanto quest'ultimo è ottenuto banalmente per $m = d$ e per $\mathbf{D}$ uguale alla matrice identità.

7.3 Ricevitore MAP

E' naturale ora chiedersi quale sia la regola di decisione per cui risulta minima la probabilità di errore.

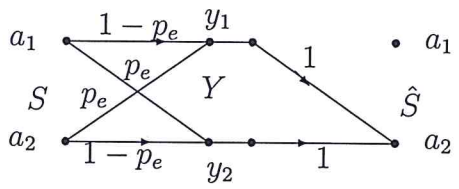


Figura 7.3: Lo schema equivalente dell'esempio 7.1

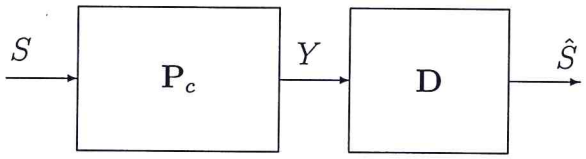


Figura 7.4: La cascata del canale e del decisore

Teorema 7.1 *La regola di decisione D_0 che minimizza*

$$P(e) = Pr\{\hat{S} \neq S\} = \sum_{i=1}^{n} \sum_{j=1;j\neq i}^{n} Pr\{\hat{S} = a_i, S = a_j\}, \tag{7.9}$$

è quella che alla osservazione di $Y = y_\ell$ associa

$$\hat{s} = D_0(y_\ell) = a_i, \quad se \;\; Pr\{S = a_i | Y = y_\ell\} \geq Pr\{S = a_j | Y = y_\ell\}, \quad \forall \; j \neq i. \tag{7.10}$$

La regola di decisione è nota come *Regola a Massima Probabilità a Posteriori* (nella terminologia anglosassone MAP è l'acronimo di "Maximum A Posteriori"). L'ipotesi a posteriori più probabile è quindi la migliore scelta. A questa regola eravamo già arrivati nel capitolo precedente ipotizzando che tale scelta fosse la più sensata. Il teorema appena enunciato ne sancisce la ottimalità rispetto agli errori.

La matrice $\mathbf{D}_0$ del decisore avrà un uno nella i-esima posizione della riga j-esima se la matrice delle probabilità a posteriori $\mathbf{P}_p$ ha il suo massimo in corrispondenza dell'i-esimo elemento della colonna j-esima. Possono verificarsi delle situazioni di parità per cui la scelta tra le varie ipotesi equiprobabili (a posteriori) è equivalente, ovvero produce la stessa probabilità di errore.

Prova: Riscriviamo la probabilità di errore come

$$P(e) = \sum_{j=1}^{m} Pr\{e|Y = y_j\} Pr\{Y = y_j\}. \tag{7.11}$$

La minimizzazione di $P(e)$ si ottiene minimizzando ogni termine della sommatoria in quanto ognuno di essi dipende dalla regola che si adotta in conseguenza della osservazione $Y = y_j$. Inoltre la regola di decisione non influenza $Pr\{Y = y_j\}$ e pertanto deve minimizzare separatamente le probabilità di errore condizionate alle osservazioni dei vari simboli di $\mathcal{Y}$, ovvero $D(y_j)$ deve minimizzare $Pr\{e|Y = y_j\}$. Se la scelta una volta osservato il simbolo y_j è $D(y_j) = a_i$

$$Pr\{e|Y = y_j\} = 1 - Pr\{c|Y = y_j\} = 1 - Pr\{S = a_i|Y = y_j\}. \tag{7.12}$$

Quindi la migliore associazione è $D_0(y_j) = a_*$ tale che

$$Pr\{S = a_*|Y = y_j\} \geq Pr\{S = a_i|Y = y_j\}, \quad \forall \ i \neq *. \tag{7.13}$$

$\triangle$

Esempio 7.2 Consideriamo un canale binario con cancellazione. L'alfabeto di ingresso sia $\mathcal{S} = \{0, 1\}$ e quello di uscita $\mathcal{Y} = \{0, 1, *\}$. La distribuzione dei simboli di ingresso sia $\Pi_S = \{0.3, 0.7\}$ e la matrice di canale

$$\mathbf{P}_c = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.7 & 0.2 \end{pmatrix}. \tag{7.14}$$

La matrice di canale corrisponde ad un probabilità di cancellazione $p_c = 0.2$ e una probabilità di errore $p_e = 0.1$ per entrambi i simboli. Valutando la matrice delle probabilità a posteriori si ottiene

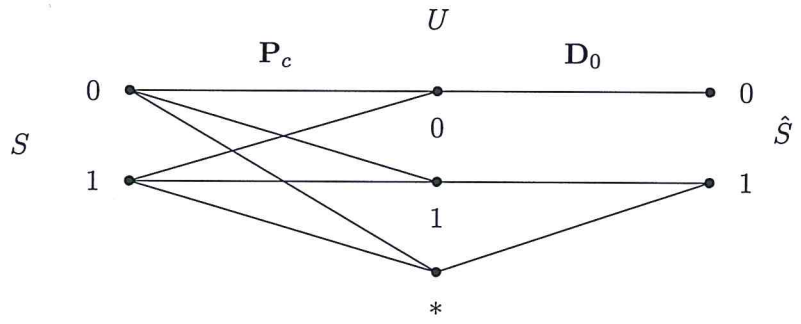
$$\mathbf{P}_p = \begin{pmatrix} 0.7500 & 0.0577 & 0.3000 \\ 0.2500 & 0.9423 & 0.7000 \end{pmatrix}. \tag{7.15}$$

Il ricevitore che minimizza la probabilità di errore, che realizza la regola MAP, si ottiene guardando alle colonne di $\mathbf{P}_p$ ed è

$$D_0(0) = 0; \quad D_0(1) = 1; \quad D(*) = 1. \tag{7.16}$$

La matrice $\mathbf{D}_0$ è quindi

$$\mathbf{D}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \tag{7.17}$$



La figura mostra la cascata di canale e decisore che hanno matrice complessiva

$$\mathbf{P}_t = \mathbf{P}_c\mathbf{D}_0 = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{pmatrix}. \tag{7.18}$$

Si tratta di un canale binario non simmetrico con probabilità di errore

$$P(e) = 1 - diag(\mathbf{P}_t)^T \boldsymbol{\pi}_S = 0.16, \tag{7.19}$$

dove $\pi_s = (0.3 \;\; 0.7)^T$.

7.4 Ricevitore ML

La regola MAP può essere riscritta usando il teorema di Bayes come

$$\frac{Pr\{Y = y_\ell | S = a_i\} Pr\{S = a_i\}}{Pr\{Y = y_\ell\}} \geq \frac{Pr\{Y = y_\ell | S = a_j\} Pr\{S = a_j\}}{Pr\{Y = y_\ell\}}, \quad \forall\ j \neq i. \tag{7.20}$$

Notiamo che se i vari simboli di $\mathcal{S}$ sono equiprobabili, la disuguaglianza diventa

$$Pr\{Y = y_\ell | S = a_i\} \geq Pr\{Y = y_\ell | S = a_j\}, \quad \forall\ j \neq i. \tag{7.21}$$

Quindi, uno volta che si è osservato il simbolo y_ℓ, la migliore scelta ai fini della minimizzazione della probabilità di errore è il simbolo a_i a cui corrisponde la $Pr\{Y = y_l | S = a_i\}$ più grande. Il criterio è noto come *Regola a Massima Verosimiglianza*, o regola MV (nella terminologia anglosassone è la *regola ML*, che è l'acronimo di "Maximum Likelihood"). La funzione

$$Pr\{Y = y_\ell | S = a_j\}, \quad \forall\ \ell, i, \tag{7.22}$$

è la *funzione di Verosimiglianza* (Likelihood Function).

Il criterio ML è molto usato in pratica quando non si ha conoscenza delle probabilità dei simboli di ingresso che pertanto vengono assunti avere probabilità uniformi. Si noti che la applicazione del criterio ML è molto più semplice in quanto esso non richiede il calcolo delle probabilità a posteriori: per valutare la ipotesi a massima verosimiglianza basta infatti guardare alle colonne della matrice di canale $\mathbf{P}_c$.

Esempio 7.2 (cont.): Riprendiamo l'esempio del canale con cancellazione. Poiché la regola MV si basa solo sulla matrice $\mathbf{P}_c$, la decisione è

$$D_{MV}(0) = 0; \quad D_{MV}(1) = 1; \quad D_{MV}(*) = 0. \tag{7.23}$$

La decisione conseguente la ricezione del simbolo $*$, potrebbe anche essere $D_{MV}(*) = 0$ in quanto le probabilità condizionate (verosimiglianze) sono le stesse. Le prestazioni del ricevitore MV possono essere valutate assumendo probabilità a priori con distribuzione uniforme, o probabilità a priori note, che però non sono state usate nel progetto del ricevitore. Assumendo di prendere come ricevitore MV quello di equazioni (7.23), la matrice complessiva del canale è

$$\mathbf{P}_{tMV} = \mathbf{P}_c\mathbf{D}_{MV} = \begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix}. \tag{7.24}$$

La probabilitá di errore del ricevitore ML per probabilità a priori uguali è

$$P(e) = 1 - diag(\mathbf{P}_{tMV})^T \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = 0.2. \tag{7.25}$$

Questa è anche la probabilità di errore minima del decisore MAP quando i simboli di sorgente sono equiprobabili. Se invece si usa il ricevitore ML con le probabilità $\Pi_S = \{0.3, 0.7\}$ si ha

$$P(e) = 1 - diag(\mathbf{P}_{tMV})^T \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} = 0.24. \tag{7.26}$$

Si noti come le prestazioni ottenute in quest'ultimo caso siano peggiori di quelle ottenute per il ricevitore MAP. Questo è una ovvia conseguenza del non avere tenuto conto delle probabilità a priori nel progetto del ricevitore.

Esempio 7.3 Si consideri un canale avente simboli di ingresso $\mathcal{S} = \{A, B, C\}$, simboli d'uscita $\mathcal{Y} = \{0, 1, *, \#\}$ e matrice di canale

$$\mathbf{P}_c = \begin{pmatrix} 0.8 & 0.1 & 0.05 & 0.05 \\ 0.7 & 0.2 & 0.05 & 0.05 \\ 0.9 & 0.05 & 0.02 & 0.03 \end{pmatrix}. \tag{7.27}$$

La distribuzione dei simboli d'ingresso è $\Pi_S = \{0.1, 0.1, 0.8\}$. Si confrontino il ricevitore MAP e il ricevitore MV per distribuzione dei simboli d'ingresso Π_S e uniforme. **Soluzione:** La matrice delle probabilità a posteriori valutata numericamente è

$$\mathbf{P}_p = \begin{pmatrix} 0.0920 & 0.1429 & 0.1923 & 0.1471 \\ 0.0805 & 0.2857 & 0.1923 & 0.1471 \\ 0.8276 & 0.5714 & 0.6154 & 0.7059 \end{pmatrix}. \tag{7.28}$$

Il ricevitore MAP, che si ottiene guardando alle colonne di $\mathbf{P}_p$, è

$$\mathbf{D}_0 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \tag{7.29}$$

Il ricevitore MV, che si ottiene guardando alle colonne di $\mathbf{P}_c$, è

$$\mathbf{D}_{MV} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \tag{7.30}$$

La decisione corrispondente alle ultime due colonne di $\mathbf{P}_c$ è arbitraria tra i simboli A e B e ne è stata scelta una a caso. La matrice totale nel caso del ricevitore MAP è

$$\mathbf{P}_{t0} = \mathbf{P}_c\mathbf{D}_0 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \tag{7.31}$$

Mentre nel caso del ricevitore MV abbiamo

$$\mathbf{P}_{tMV} = \mathbf{P}_c\mathbf{D}_{MV} = \begin{pmatrix} 0 & 0.1000 & 0.9000 \\ 0 & 0.2000 & 0.8000 \\ 0 & 0.0500 & 0.9500 \end{pmatrix}. \tag{7.32}$$

La probabilità di errore per i ricevitori MAP e MV, se i simboli d'ingresso sono distribuiti come $\Pi_S = \{0.1, 0.1, 0.8\}$ sono rispettivamente

$$P(e) = 1 - diag(\mathbf{P}_{t0})^T \pi_S = 0.2, \tag{7.33}$$

$$P(e) = 1 - diag(\mathbf{P}_{tMV})^T \pi_S = 0.22\ . \tag{7.34}$$

Il ricevitore MAP in questo caso ha le prestazioni migliori essendo l'ottimo. Altrimenti se le probabilità all'ingresso sono uniformi abbiamo rispettivamente

$$P(e) = 1 - diag(\mathbf{P}_{t0})^T \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} = 0.6667, \tag{7.35}$$

$$P(e) = 1 - diag(\mathbf{P}_{tMV})^T \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} = 0.6167. \tag{7.36}$$

In questo caso il ricevitore MV, che coincide con il ricevitore MAP per probabilità d'ingresso uniformi, è il migliore.

7.5 La disuguaglianza di Fano

In un canale, o nella cascata di un canale e di un decodificatore, la probabilità di errore media caratterizza globalmente l'affidabilità del sistema. La domanda naturale è se esista una relazione con le quantità informazionali quali l'ambiguità, la mutua informazione e la capacità. In effetti una relazione deve esserci in quanto in un canale $X \rightarrow Y$ l'ambiguità $\mathcal{H}(X|Y)$ misura l'incertezza residua sull'ingresso X, una volta che si sia osservata l'uscita Y. Quindi se l'uscita Y ci fornissse sufficiente informazione su X, l'ambiguità dovrebbe essere nulla e analogamente dovremmo essere in grado di risalire a X con probabilità di errore nulla. Le relazioni tra queste quantità sono di importanza cruciale per comprendere i limiti teorici alla affidabilità di un trasporto e di una decodifica. La disuguaglianza di Fano fornisce una prima utile relazione tra la probabilità di errore e l'ambiguità. Il risultato è presentato in generale con riferimento ad un canale discreto $X \rightarrow Y$ che può essere il solo canale rumoroso, o la cascata di un canale e di un decisore (anche non ottimo), o la cascata di un codificatore, di un canale e di un decisore. Approfondiremo l'interpretazione e l'utilizzo del teorema in seguito.

Teorema 7.2 (Disuguaglianza di Fano) *Dato un canale discreto senza memoria, con alfabeti di ingresso e uscita* $\mathcal{X} = \{x_1, ..., x_n\}$ *e* $\mathcal{Y} = \{y_1, ..., y_n\}$ *aventi la stessa cardinalità* $|\mathcal{X}| = |\mathcal{Y}| = n$*, avendo definito la probabilità di errore come*

$$P(e) = \sum_{i=1}^{n} \sum_{j=1; j \neq i}^{n} P(x_i, y_j), \tag{7.37}$$

vale la seguente disuguaglianza

$$\mathcal{H}(X|Y) \leq \mathcal{H}(e) + P(e) \log_2(n-1) \tag{7.38}$$

dove $\mathcal{H}(e)$ *è l'entropia binaria*

$$\mathcal{H}(e) = -P(e) \log_2 P(e) - (1 - P(e)) \log_2 (1 - P(e)). \tag{7.39}$$

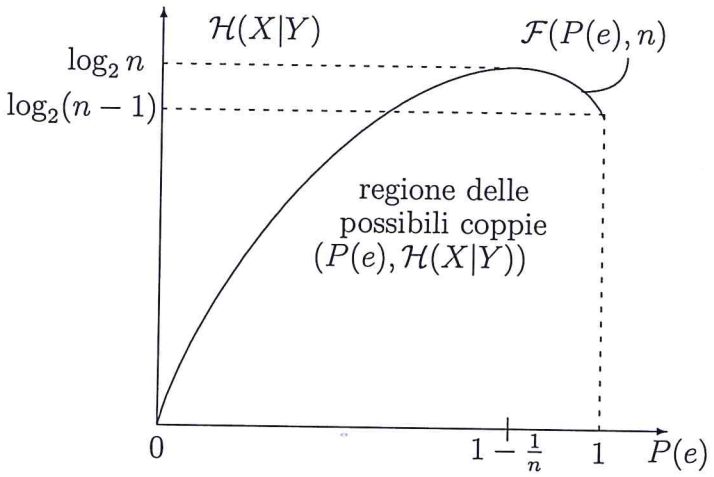


Figura 7.5: La regione delle possibili coppie $(P(e), \mathcal{H}(X|Y))$ della disuguaglianza di Fano.

Il teorema fornisce una relazione generale per un qualunque canale discreto, tra $P(e)$ e $\mathcal{H}(X|Y)$. La funzione maggiorante

$$\mathcal{F}(P(e), n) = \mathcal{H}(e) + P(e)\log_2(n-1), \tag{7.40}$$

è mostrata in Figura 7.5 dove è evidenziata la regione delle possibili coppie $(P(e), \mathcal{H}(X|Y))$.

Il risultato del teorema è suscettibile di una interessante interpretazione: il limite superiore alla incertezza residua su X, una volta osservato Y, è la entropia collegata all'errore in uscita più un temine $P(e)\log_2(n-1)$, che è l'informazione collegata alla posizione dell'errore sull'uscita.

Prova: La ambiguità, la probabilità di errore, la probabilità di corretta ricezione e la entropia binaria $\mathcal{H}(e)$ possono essere scritte rispettivamente come

$$\begin{aligned}
\mathcal{H}(X|Y) &= \sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i,y_j)\log_2\frac{1}{P(x_i|y_j)} \\
&= \sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n} P(x_i,y_j)\log_2\frac{1}{P(x_i|y_j)} + \sum_{i=1}^{n} P(x_i,y_i)\log_2\frac{1}{P(x_i|y_i)}; \\
P(e) &= \sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n} P(x_i,y_j); \\
P(c) &= 1-P(e) = \sum_{i=1}^{n} P(x_i,y_i);
\end{aligned}$$

$$\begin{aligned}\mathcal{H}(e) &= -P(e)\log_2 P(e) - (1-P(e))\log_2(1-P(e)) \\ &= -\sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n} P(x_i,y_j)\log_2 P(e) - \sum_{i=1}^{n} P(x_i,y_i)\log_2(1-P(e)).\end{aligned}$$

La disuguaglianza si ottiene combinando linearmente, e usando la disuguaglianza $\log_2 x \leq (\log_2 e)(x-1)$

$$\begin{aligned}&\mathcal{H}(X|Y) - P(e)\log_2(n-1) - \mathcal{H}(e) \\ &= \sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n} P(x_i,y_j)\log_2\frac{P(e)}{(n-1)P(x_i|y_j)} + \sum_{i=1}^{n} P(x_i,y_i)\log_2\frac{1-P(e)}{P(x_i|y_i)} \\ &\leq \log_2 e\sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n} P(x_i,y_j)\left(\frac{P(e)}{(n-1)P(x_i|y_j)} - 1\right) \\ &\quad + \log_2 e\sum_{i=1}^{n} P(x_i,y_i)\left(\frac{1-P(e)}{P(x_i|y_i)} - 1\right) \\ &= (\log_2 e)\frac{P(e)}{n-1}\sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n}\frac{P(x_i,y_j)}{P(x_i|y_j)} - (\log_2 e)\sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n} P(x_i,y_j) \\ &\quad + (\log_2 e)(1-P(e))\sum_{i=1}^{n}\frac{P(x_i,y_i)}{P(x_i|y_j)} - (\log_2 e)\sum_{i=1}^{n} P(x_i,y_i) \\ &= (\log_2 e)\frac{P(e)}{n-1}\sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n} P(y_j) - (\log_2 e)P(e) \\ &\quad + (\log_2 e)(1-P(e))\sum_{i=1}^{n} P(y_i) - (\log_2 e)(1-P(e)) \\ &= (\log_2 e)\frac{P(e)}{n-1}\sum_{i=1}^{n}(1-P(y_i)) - (\log_2 e)P(e) \\ &\quad + (\log_2 e)(1-P(e)) - (\log_2 e)(1-P(e)) \\ &= (\log_2 e)\frac{P(e)}{n-1}(n-1) - (\log_2 e)P(e) = 0. \qquad \square \qquad (7.41)\end{aligned}$$

Manipolando la disuguaglianza di Fano otteniamo delle utili osservazioni. Poiché $\mathcal{H}(X|Y) = \mathcal{H}(X) - \mathcal{I}(X;Y)$ e $\mathcal{I}(X;Y) \leq C$, abbiamo che

$$\mathcal{H}(X) - C \leq \mathcal{H}(X|Y) \leq \mathcal{H}(e) + P(e)\log_2(n-1), \qquad (7.42)$$

ovvero

$$\mathcal{H}(X) \leq C + \mathcal{F}(P(e), n). \qquad (7.43)$$

La formula esprime un limite superiore alla entropia dell'ingresso al canale in funzione della capacità e della probabilità di errore. La figura 7.6 mostra

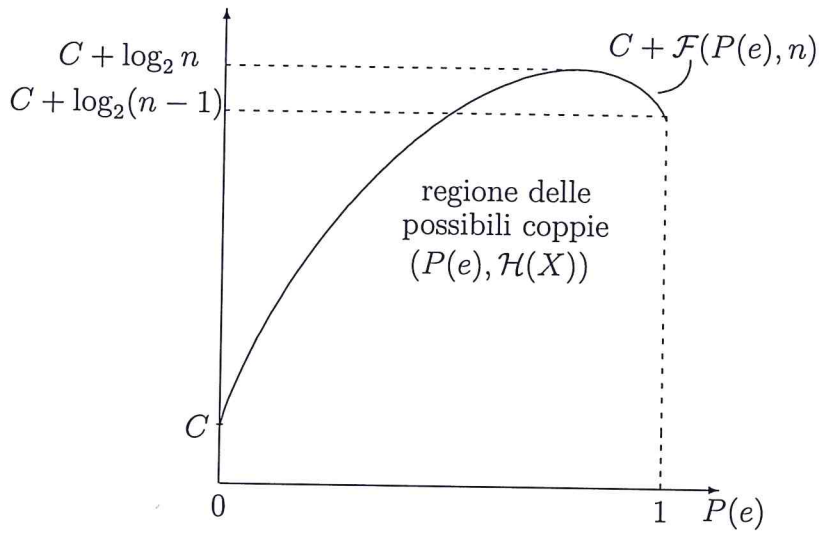


Figura 7.6: La regione delle possibili coppie $(P(e), \mathcal{H}(X))$.

graficamente il risultato con l'indicazione della regione delle possibili coppie $(P(e), \mathcal{H}(X))$. Pertanto *solo la condizione* $\mathcal{H}(X) < C$, *è compatibile con una la probabilità di errore pari a zero.* Ovvero l'informazione presente all'ingresso non può eccedere la capacità del canale se si vuole probabilità di errore nulla. Per ottenere probabilità di errore nulla, o meglio, come vedremo in seguito, asintoticamente nulla, sarà necessario operare un opportuna codifica sul segnale d'ingresso. Comunque la relazione ottenuta rivela che è impossibile andare al di sotto di una probabilità minima descritta dalla maggiorazione se il flusso informativo all'ingresso non è al di sotto della capacità di canale. Il risultato ottenuto conferisce alla definizione di capacità data nel capitolo precedente un significato cruciale. L'analogia fluidodinamica del sistema di comunicazione è molto accattivante e conferma come la teoria dell'informazione rivesta un ruolo centrale nell'analisi e nel progetto dei sistemi di comunicazione. Il risultato è una versione semplificata del cosiddetto *teorema inverso della codifica* che sarà presentato più in generale in seguito.

Esempio 7.4 Consideriamo un canale binario simmetrico con probabilità di errore $p = 0.1$. La capacità del canale è $C = 1 - \mathcal{H}(p) = 0.5310$ bit. Poiché $n = 2$

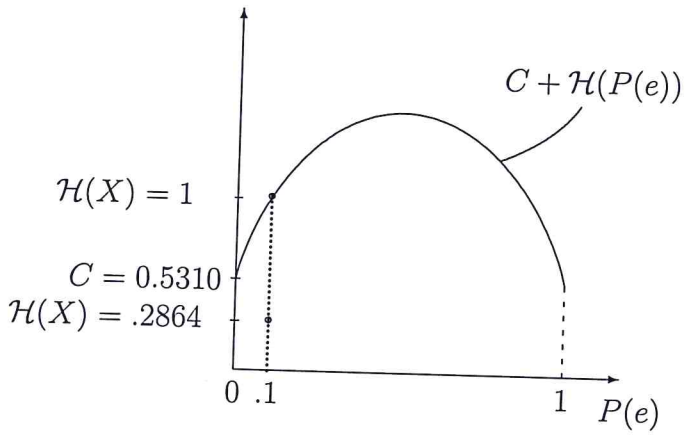


Figura 7.7: Le coppie $(\mathcal{H}(X), P(e))$ per l'esempio 7.4.

abbiamo che la disuguaglianza di Fano fornisce

$$\mathcal{H}(X) \leq C + \mathcal{H}(P(e)). \tag{7.44}$$

Figura 7.7 mostra la regione delle possibili coppie $(P(e), \mathcal{H}(X))$. Se l'ingresso X ha distribuzione uniforme $\Pi_X = \{0.5, 0.5\}$, l'entropia è $\mathcal{H}(X) = 1$ bit e la probabilità di errore è $P(e) = p = 0.1$. Il punto sul piano $(P(e), \mathcal{H}(X))$ è sulla frontiera della regione ed è indicato in figura 7.7. Notiamo che un tale ingresso non potrà mai essere compatibile con una probabilità di errore arbitrariamente piccola in quanto la sua entropia eccede la capacità. Si noti anche come il punto è anche sulla frontiera e quindi la probabilità di errore è anche la minima ottenibile. Se invece la distribuzione dell'ingresso è non uniforme, ad esempio $\Pi_X = \{0.05, 0.95\}$, l'entropia è $\mathcal{H}(X) = 0.2864$ bit e siamo al di sotto della capacità. Il canale così com'è fornisce una probabilità di errore che è ancora $P(e) = p\ 0.05 + p\ 0.95 = 0.1$ (canale uniforme) e la situazione corrisponde al secondo punto indicato in figura 7.7. L'esempio suggerisce che dovrebbe essere possibile ottenere una probabilità di errore migliore, al limite nulla, e fa intuire come sia necessario riorganizzare il canale includendo un opportuno stadio di codifica come vedremo in seguito.

7.5.1 Il teorema di Fano per il canale uniforme

Il canale uniforme è molto comune nelle applicazioni e merita una attenzione particolare a riguardo delle relazioni tra le quantità informazionali e la probabilità di errore. Ricordiamo che un canale uniforme ha matrice di

canale del tipo

$$\mathbf{P}_c = \begin{bmatrix} (1-p_e) & \frac{p_e}{n-1} & \frac{p_e}{n-1} & \cdots & \frac{p_e}{n-1} \\ \frac{p_e}{n-1} & (1-p_e) & \frac{p_e}{n-1} & \cdots & \frac{p_e}{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \frac{p_e}{n-1} & \frac{p_e}{n-1} & \frac{p_e}{n-1} & \cdots & (1-p_e) \end{bmatrix}, \tag{7.45}$$

dove p_e è proprio la probabilità di errore (il BSC è un caso particolare).

Teorema 7.3 *Per un canale uniforme di dimensione n, probabilità di errore p_e e avente ingresso uniforme, si ha*

$$\mathcal{H}(X|Y) = \mathcal{H}(p_e) + p_e \log_2(n-1). \tag{7.46}$$

Quindi in questo caso il teorema di Fano vale con il segno di uguaglianza. L'interpretazione data in precedenza è ancora più calzante: l'ambiguità sull'ingresso è equivalente alla incertezza sull'errore e su dove esso si è verificato.

Prova: Abbiamo già dimostrato nel capitolo precedente come per un canale simmetrico con matrice di canale avente le probabilità $\mathbf{p} = (p_1, p_2, ..., p_{n-1}, 1 - \sum_{i=1}^{n-1} p_i)$ sulla prima riga, la mutua informazione sia

$$\mathcal{I}(X;Y) = \mathcal{H}(Y) - \mathcal{H}(\mathbf{p}). \tag{7.47}$$

Se la distribuzione dei simboli di ingresso è uniforme, per simmetria anche quella delle uscite sarà uniforme e pertanto $\mathcal{H}(Y) = \mathcal{H}(X) = \log_2 n$. Poiché $\mathcal{I}(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y)$, sostituendo si ha

$$\mathcal{H}(X|Y) = \mathcal{H}(\mathbf{p}). \tag{7.48}$$

Nel caso specifico del canale uniforme, $p_1 = (1-p_e)$ e $p_i = p_e/(n-1)$, $i = 2, ..., n$ e quindi

$$\begin{aligned} \mathcal{H}(\mathbf{p}) &= (1-p_e)\log_2 \frac{1}{1-p_e} + (n-1)\frac{p_e}{n-1}\log_2 \frac{n-1}{p_e} && (7.49) \\ &= (1-p_e)\log_2 \frac{1}{1-p_e} + p_e \log_2 \frac{n-1}{p_e} && (7.50) \\ &= \mathcal{H}(p_e) + p_e \log_2(n-1). \ \triangle && (7.51) \end{aligned}$$

Si noti che per ingressi uniformi $\mathcal{H}(X) = \log_2 n$ e $\mathcal{I}(X;Y) = C$ e quindi la formula ci da

$$C = \log_2 n - \mathcal{H}(p_e) - p_e \log_2(n-1), \tag{7.52}$$

che è semplicemente la formula della capacità del canale uniforme già ricavata nel capitolo precedente.