FRANCESCO A.N. PALMIERI

APPUNTI SUI CANALI DISCRETI

Capitolo 6

Canali discreti senza memoria

In questo capitolo viene introdotta la formalizzazione matematica dei canali discreti senza memoria, la mutua informazione e la capacità di canale. La discussione è condotta con particolare riferimento a esempi tipici, incluso il canale binario simmetrico e il canale esteso.

6.1 Il canale discreto

In un sistema di comunicazione, anche molto complesso, e che può quindi includere stadi intermedi di modulazione, tratte elettromagnetiche, codifiche e altro, è utile modellare il problema della comunicazione tra i due terminali estremi in maniera globale equivalente. L'obiettivo di riferimento è riprodurre a destinazione, nella maniera più affidabile possibile, una sequenza di simboli emessi da una sorgente discreta. Le degradazioni dovute al canale, genericamente associate al *rumore* presente nel collegamento, tipicamente causeranno degli *errori*, ovvero non tutti i simboli trasmessi potranno essere riprodotti fedelmente a destinazione. Sarà responsabilità del progettista tenere in conto tali effetti e controllare le prestazioni mediante un controllo della potenza in trasmissione e mediante la introduzione di opportune codifiche.

La figura 6.1 mostra lo schema generale di riferimento di un *canale tempo-discreto* dove all'ingresso e all'uscita sono presenti rispettivamente le sequenze $\{X[k]\}$ e $\{Y[k]\}$. I simboli delle due sequenze sono prelevati rispettivamente dall'*alfabeto dei simboli di ingresso* $\mathcal{X} = \{x_1, x_2, ..., x_n\}$, e dall'*alfabeto dei simboli di uscita* $\mathcal{Y} = \{y_1, y_2, ..., y_m\}$. Si noti che in generale gli alfabeti di ingresso e di uscita possono essere diversi, e quindi avere anche cardinalità diverse. Adotteremo qui la formalizzazione generale anche se il

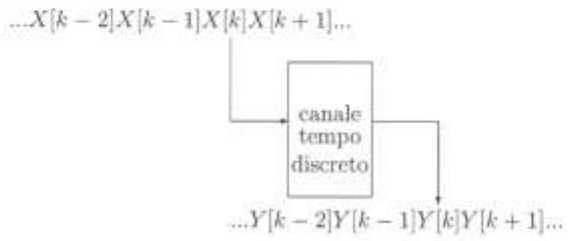


Figura 6.1: Il canale tempo-discreto

caso più tipico, come vedremo in dettaglio in seguito, è quello del *canale rumoroso* in cui i due alfabeti $\mathcal{X}$ e $\mathcal{Y}$ coincidono e il canale è preposto alla riproduzione affidabile in uscita dei simboli di ingresso.

Se il sistema è tale che il simbolo di uscita $Y[k]$ dipende solo dal simbolo di ingresso $X[k]$, il canale è detto *senza memoria*. In tal caso il meccanismo di associazione opera simbolo-a-simbolo. Ogni volta che il sistema opera una associazione su un elemento della sequenza, di dice che il canale è stato *usato* una volta. Se la sequenza di ingresso è anch'essa senza memoria, ovvero è a simboli indipendenti e con distribuzione stazionaria, non è necessario portare in conto l'indice temporale k dei vari *usi del canale*, ma è sufficiente concentrarsi su un sistema con ingresso e uscita aleatoria X e Y rispettivamente. In questo capitolo ci soffermeremo solo su questo tipo di modello rimandando il lettore, per i canali e le sorgenti *con memoria*, a successivi approfondimenti, o alla vasta letteratura sull'argomento.

La figura 6.2 mostra schematicamente il modello del canale. Si tratta di un grafo orientato in cui ad ogni ramo è associata una probabilità condizionata. Più formalmente, si definisce *canale di informazione discreto senza memoria*, l'insieme di un alfabeto d'ingresso $\mathcal{X} = \{x_1, x_2, ..., x_n\}$, di un alfabeto di uscita $\mathcal{Y} = \{y_1, y_2, ..., y_m\}$ e di un insieme di probabilità condizionate $P(y_i|x_j) = Pr\{Y = y_i|X = x_j\}$, $i = 1, ..., m$, $j = 1, ..., n$, che rappresentano le probabilità con le quali i simboli di ingresso si traducono in simboli di uscita. Le probabilità che caratterizzano il canale sono contenute

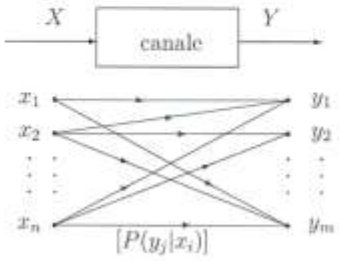


Figura 6.2: Il canale di informazione discreto

in maniera compatta nella *matrice di canale*

$$\mathbf{P}_c = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) & \dots & P(y_m|x_1) \\ P(y_1|x_2) & P(y_2|x_2) & \dots & P(y_m|x_2) \\ \dots & \dots & \dots & \dots \\ P(y_1|x_n) & P(y_2|x_n) & \dots & P(y_m|x_n) \end{bmatrix}$$
$$= \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1m} \\ P_{21} & P_{22} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nm} \end{bmatrix}. \tag{6.1}$$

Ricordiamo come in generale la dimensione dell'alfabeto di uscita possa essere diversa da quella dell'alfabeto di ingresso, anche se il caso più tipico è quello in cui i due alfabeti hanno la stessa cardinalità e il canale è preposto al trasporto affidabile dei simboli. Vedremo alcuni esempi che chiariranno meglio il modello.

La matrice di canale $\mathbf{P}_c$ è tale che la somma degli elementi di ogni riga è pari a uno [1]

$$\sum_{j=1}^{m} P_{ij} = 1 \quad i = 1, ..., n, \tag{6.2}$$

ovvero in forma matriciale

$$\mathbf{P}_c \mathbf{e}_m = \mathbf{e}_n, \tag{6.3}$$

[1] Una matrice di probabilità con tale proprietà è anche detta *matrice di Markov* o *matrice stocastica*.

polin. 4

dove $\mathbf{e}_n = \underbrace{(1, 1, ..., 1)}_{n}^T$ è il vettore unitario di lunghezza n. La distribuzione di probabilità dei simboli d'uscita

$$\Pi_Y = \{P(y_1), P(y_2), ..., P(y_m)\}, \tag{6.4}$$

si valuta dalla conoscenza delle probabilità dei simboli d'ingresso, dette *probabilità a priori*

$$\Pi_X = \{P(x_1), P(x_2), ..., P(x_n)\}, \tag{6.5}$$

mediante la legge della probabilità totale

$$P(y_j) = \sum_{i=1}^{n} P(y_j|x_i)P(x_i), \quad j = 1, ..., m. \tag{6.6}$$

Definendo i vettori delle probabilità a priori e di uscita come

$$\boldsymbol{\pi}_x = [P(x_1), ..., P(x_n)]^T \ , \ \boldsymbol{\pi}_y = [P(y_1), ..., P(y_m)]^T \ , \tag{6.7}$$

è possibile esprimere la relazione tra la distribuzione degli ingressi e quella delle uscite in forma matriciale compatta

$$\boldsymbol{\pi}_y = \mathbf{P}_c^T \boldsymbol{\pi}_x. \tag{6.8}$$

Esempio 6.1 Si consideri il canale avente alfabeto di ingresso $\mathcal{X} = \{x_1, x_2, x_3\}$, alfabeto di uscita $\mathcal{Y} = \{y_1, y_2, y_3\}$ e una matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \tag{6.9}$$

E' qui mostrato un segmento di una possibile realizzazione di ingresso e uscita in cui si è assunto che le probabilità a priori dei simboli di ingresso siano uguali $\Pi_X = \{1/3, 1/3, 1/3\}$.

$$\begin{array}{lcccccccccccccccc} \{X[k]\} = ... & x_1 & x_2 & x_1 & x_3 & x_2 & x_1 & x_2 & x_1 & x_3 & x_3 & x_1 & x_3 & x_2 & x_3 & x_2 ... \\ \{Y[k]\} = ... & y_1 & y_2 & y_2 & y_1 & y_1 & y_1 & y_2 & y_3 & y_3 & y_3 & y_1 & y_2 & y_3 & y_3 & y_2 ... \end{array} \tag{6.10}$$

Si noti come i simboli di ingresso $\{x_1, x_2, x_3\}$ all'incirca nella metà dei casi si trasformano in $\{y_1, y_2, y_3\}$ rispettivamente. Nei rimanenti casi ognuno si tramuta in uno degli altri due con uguale probabilità. Si tratta ovviamente solo di un breve segmento di una possibile realizzazione.

Le probabilità $P(y_j|x_i)$, dette *probabilità dirette* del canale, caratterizzano quindi il meccanismo aleatorio con il quale i simboli di ingresso si trasformano nei simboli di uscita. Analogamente, è possibile descrivere, come i simboli in uscita provengano in probabilità dai simboli di ingresso, ovvero, usando il teorema di Bayes, ottenere le relazioni "inverse" del canale

$$P(x_i|y_j) = \frac{P(y_j|x_i)P(x_i)}{\sum_{l=1}^{n} P(y_j|x_l)P(x_l)}, \quad \forall \ i, j. \tag{6.11}$$

Tali probabilità sono dette *probabilità a posteriori* (o *probabilità inverse*) del canale. Ogni $P(x_i|y_j)$ rappresenta la probabilità che, osservato il simbolo y_j in uscita, in ingresso ci sia stato x_i. Da quì il nome "a posteriori" visto che esse descrivono la probabilità dei simboli di sorgente "dopo che" un simbolo in uscita è stato osservato. Si noti che comunque le probabilità a posteriori dipendono anche dalle probabilità a priori. Le probabilità a posteriori possono essere anch'esse organizzate in una matrice

$$\begin{aligned}
\mathbf{P}_p &= \begin{bmatrix} P(x_1|y_1) & P(x_1|y_2) & \cdots & P(x_1|y_m) \\ P(x_2|y_1) & P(x_2|y_2) & \cdots & P(x_2|y_m) \\ \vdots & \vdots & \cdots & \vdots \\ P(x_n|y_1) & P(x_n|y_2) & \cdots & P(x_n|y_m) \end{bmatrix} \\
&= \begin{bmatrix} \frac{P(y_1|x_1)P(x_1)}{P(y_1)} & \frac{P(y_2|x_1)P(x_1)}{P(y_2)} & \cdots & \frac{P(y_m|x_1)P(x_1)}{P(y_m)} \\ \frac{P(y_1|x_2)P(x_2)}{P(y_1)} & \frac{P(y_2|x_2)P(x_2)}{P(y_2)} & \cdots & \frac{P(y_m|x_2)P(x_2)}{P(y_m)} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{P(y_1|x_n)P(x_n)}{P(y_1)} & \frac{P(y_2|x_n)P(x_n)}{P(y_2)} & \cdots & \frac{P(y_m|x_n)P(x_n)}{P(y_m)} \end{bmatrix}.
\end{aligned} \tag{6.12}$$

La generica colonna j-esima della matrice rappresenta tutte le probabilità a posteriori degli n simboli di ingresso corrispondenti alla osservazione di y_j e chiaramente sommano a uno. La matrice delle probabilità a posteriori può essere espressa in forma compatta nella relazione matriciale

$$\mathbf{P}_p = diag(\boldsymbol{\pi}_x)\mathbf{P}_c(diag(\boldsymbol{\pi}_y))^{-1}. \tag{6.13}$$

Si noti inoltre come le probabilità condizionate del canale siano diverse dalle probabilità congiunte

$$P(y_j, x_i) = P(x_i, y_j) = P(y_j|x_i)P(x_i) = P(x_i|y_j)P(y_j), \quad \forall \ i, j, \tag{6.14}$$

che rappresentano la probabilità che all'ingresso e all'uscita del canale ci siano *simultaneamente* x_i e y_j.

E' interessante notare come le probabilità a posteriori possano essere utilizzate da un ricevitore che osservi l'uscita Y per ottenere informazioni sull'ingresso X. Infatti, se un ricevitore in uscita osserva il simbolo y_j, dalla conoscenza della struttura del canale e delle probabilità a priori, può calcolarsi le n probabilità a posteriori e avere quindi una idea di quale potrebbe essere stato il simbolo in ingresso ad aver generato quel simbolo in uscita. Confrontando i valori delle probabilità a posteriori, il ricevitore può "decidere" di scegliere il simbolo di ingresso $\hat{x}$ a cui corrisponde il valore massimo. Ovvero se in uscita si è osservato y_j

$$\hat{x} = \arg \max_{i=1,...,n} P(x_i|y_j). \tag{6.15}$$

Tale criterio è detto *criterio MAP* (Maximum A Posteriori) o *criterio a massima probabilità a posteriori* . Il criterio MAP verrà approfondito in seguito. [2]

6.1.1 La probabilità di errore

La matrice di canale e le probabilità a priori, consentono di valutare le probabilità di qualunque evento collegato al canale. Approfondiamo il caso tipico di canale quadrato ($n = m$) e valutiamo l'efficacia con la quale il canale trasferisce i simboli dell'alfabeto $\mathcal{X}$, nell'alfabeto $\mathcal{Y}$ nella corrispondenza $x_i \leftrightarrow y_i$, $i = 1, ..., m$, che consideriamo come quella relativa al "corretto trasporto."[3] La probabilità che un simbolo sia ricevuto erroneamente o correttamente è facilmente calcolabile. Più precisamente, si definisce *probabilità di errore condizionata al simbolo i-esimo*

$$P(e_i) \stackrel{def}{=} Pr\{\text{errore} \mid \text{trasmesso } x_i\} = \sum_{j=1; j\neq i}^{n} P(y_j|x_i), \tag{6.16}$$

[2]Credo sia interessante far notare al lettore come il problema della decisione sul simbolo di ingresso più probabile, una volta che si sia osservata una particolare uscita, sia un problema di *diagnostica*. Infatti se attribuiamo ai simboli di ingresso il significato di possibili "cause," e ai simboli di uscita quello di "sintomi," lo scopo della decisione è, dalla conoscenza dei meccanismi più o meno aleatori con cui le cause generano dei sintomi, risalire alla causa, o alle cause, più probabili. Si noti come anche le probabilità a priori delle varie cause giochino un ruolo cruciale. Una diagnosi medica si basa sempre, in maniera più o meno consapevole, su questo meccanismo di decisione.

[3]Si noti come la corrispondenza, purché sia biunivoca, è arbitraria. Infatti si potrebbero adottare delle associazioni diverse da definire come "favorevoli." Si noti però che una qualunque diversa associazione sarebbe equivalente ad una rietichettatura, o permutazione dei simboli. Pertanto la associazione $x_i \leftrightarrow y_i$, $i = 1, ..., m$ non comporta perdita di generalità e sarà sempre considerata come quella di riferimento.

ovvero la probabilità con la quale il simbolo i-esimo viene confuso con altri simboli dopo la trasmissione sul canale. La *probabilità di errore media*, ovvero la probabilità di errore mediata su tutti i simboli, è invece

$$P(e) = \sum_{i=1}^{n} P(e_i)P(x_i) = \sum_{i=1}^{n} \sum_{j=1 j\neq i}^{n} P(y_j|x_i)P(x_i), \tag{6.17}$$

e rappresenta una misura globale di affidabilità del canale.

E' generalmente più semplice esprimere la probabilità di errore in funzione della probabilità dell'evento complemento, ovvero in termini di *probabilità di corretta ricezione*

$$P(c_i) \stackrel{def}{=} Pr\{\text{corretta ricezione} \mid \text{trasmesso } x_i\} = P(y_i|x_i), \tag{6.18}$$

ovvero

$$P(e) = 1 - P(c) = 1 - \sum_{i=1}^{n} P(c|x_i)P(x_i) = 1 - \sum_{i=1}^{n} P(y_i|x_i)P(x_i). \tag{6.19}$$

Le relazioni precedenti si scrivono in forma matriciale come

$$P(e) = 1 - P(c) = 1 - diag(\mathbf{P}_c)^T \boldsymbol{\pi}_x, \tag{6.20}$$

dove $diag(\mathbf{P}_c)$ è il vettore colonna formato dagli n elementi della diagonale di $\mathbf{P}_c$. [4]

Esaminiamo ora alcuni tipici esempi di canale.

6.1.2 Il canale binario

Questo esempio, che costituisce di gran lunga il canale più importante nella teoria delle comunicazioni, è definito mediante i due alfabeti binari

$$\mathcal{X} = \{0,1\}\ , \qquad \mathcal{Y} = \{0,1\}, \tag{6.21}$$

e la matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} (1-p_0) & p_0 \\ p_1 & (1-p_1) \end{bmatrix}. \tag{6.22}$$

Il grafo relativo è rappresentato in figura 6.3. Si noti come la matrice di

[4] La nostra convenzione sulla funzione $diag(.)$ è che, se essa è applicata a un vettore, restituisce la matrice diagonale con gli elementi del vettore nella diagonale principale. Se la funzione è applicata a una matrice (quadrata), essa restituisce un vettore colonna costituito dagli elementi della diagonale principale della matrice.

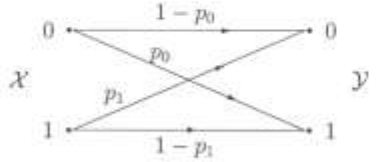


Figura 6.3: Il canale binario

canale dipenda solo dai due parametri, p_0 e p_1, che rappresentano proprio le probabilità di errore condizionate dai simboli "0" e "1": $P(e_0) = p_0$, $P(e_1) = p_1$. Quindi, in generale, la probabilità che un simbolo venga confuso con l'altro può essere diversa per i due simboli. Il calcolo della probabilità di errore media richiede la conoscenza delle probabilità a priori dei simboli in ingresso

$$\boldsymbol{\pi}_x^T = [P(0), P(1)] = [P(0), (1 - P(0))], \tag{6.23}$$

ed è pertanto

$$\begin{aligned} P(e) &= 1 - P(c) = 1 - (P(c_0)P(0) + P(c_1)P(1)) \\ &= 1 - (1 - p_0)P(0) - (1 - p_1)(1 - P(0)) \\ &= p_0 P(0) + p_1(1 - P(0)). \end{aligned} \tag{6.24}$$

Le probabilità dei simboli d'uscita sono

$$\begin{aligned} P\{Y = 0\} &= P\{Y = 0|X = 0\}P(0) + P\{Y = 0|X = 1\}P(1) \\ &= (1 - p_0)P(0) + p_1(1 - P(0)), \\ P\{Y = 1\} &= P\{Y = 1|X = 0\}P(0) + P\{Y = 1|X = 1\}P(1) \\ &= p_0 P(0) + (1 - p_1)(1 - P(0)). \end{aligned} \tag{6.25}$$

Le probabilità a posteriori sono

$$\begin{aligned} P\{X = 0|Y = 0\} &= \frac{P\{Y = 0|X = 0\}P(0)}{P\{Y = 0\}} \\ &= \frac{(1 - p_0)P(0)}{P\{Y = 0\}}, \\ P\{X = 1|Y = 0\} &= \frac{P\{Y = 0|X = 1\}P(1)}{P\{Y = 0\}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{p_1(1-P(0))}{P\{Y=0\}},\\
P\{X=0|Y=1\} &= \frac{P\{Y=1|X=0\}P(0)}{P\{Y=1\}}\\
&= \frac{p_0 P(0)}{P\{Y=1\}},\\
P\{X=1|Y=1\} &= \frac{P\{Y=1|X=1\}P(1)}{P\{Y=1\}}\\
&= \frac{(1-p_1)(1-P(0))}{P\{Y=1\}}.
\end{aligned}
\tag{6.26}
$$

Quando le probabilità di errore condizionate sono uguali ($p_0 = p_1 = p$), il canale è detto *canale binario simmetrico*, o BSC (dall'inglese *Binary Symmetric Channel*). La probabilità di errore media diventa semplicemente

$$P(e) = pP(0) + p(1-P(0)) = p, \tag{6.27}$$

ed è indipendente dalle probabilità a priori. Le probabilità dei simboli di uscita diventano

$$
\begin{aligned}
P\{Y=0\} &= (1-p)P(0) + p(1-P(0))\\
&= P(0)(1-2p) + p,\\
P\{Y=1\} &= pP(0) + (1-p)(1-P(0))\\
&= 1-p-P(0)(1-2p).
\end{aligned}
\tag{6.28}
$$

Analogamente le probabilità a posteriori

$$
\begin{aligned}
P\{X=0|Y=0\} &= \frac{(1-p)P(0)}{P(0)(1-2p)+p},\\
P\{X=1|Y=0\} &= \frac{p(1-P(0))}{P(0)(1-2P)+p},\\
P\{X=0|Y=1\} &= \frac{pP(0)}{1-p-P(0)(1-2p)},\\
P\{X=1|Y=1\} &= \frac{(1-p)(1-P(0))}{1-p-P(0)(1-2p)}.
\end{aligned}
\tag{6.29}
$$

E' anche interessante notare che nel BSC, se le probabilità a priori sono uguali, $P(0) = P(1) = \frac{1}{2}$, abbiamo

$$P\{Y=0\} = P\{Y=1\} = \frac{1}{2},$$

$$\begin{aligned} P\{X=0|Y=0\} &= 1-p, \\ P\{X=1|Y=0\} &= p, \\ P\{X=0|Y=1\} &= p, \\ P\{X=1|Y=1\} &= 1-p. \end{aligned} \tag{6.30}$$

Per maggiore chiarezza esaminiamo quanche esempio numerico di canale binario.

Esempio 6.2 Supponiamo che un canale binario sia simmetrico con $p = 0.2$. Le probabilità a priori siano $P(0) = P(1) = 0.5$. Un segmento tipico dei flussi binari all'ingresso e all'uscita sono qui riportati:

...0101110100110010010101010011010010110101001010110100101001 10...

...0001010110111010011101010111010010010101101011111011010110100...

Si noti come nella sequenza di ingresso i simboli siano divisi in percentuali *quasi* uguali (si ricordi che tratta solo di una realizzazione della sequenza aleatoria) e come circa il 20% sia in errore, indipendentemente dal simbolo. Le probabilità delle uscite sono da equazione (6.28): $P\{Y=0\} = P\{Y=1\} = 0.5$ e le probabilità a posteriori da (6.29)

$$\begin{aligned} P\{X=0|Y=0\} &= 0.8, \\ P\{X=1|Y=0\} &= 0.2, \\ P\{X=0|Y=1\} &= 0.2, \\ P\{X=1|Y=1\} &= 0.8. \end{aligned} \tag{6.31}$$

La probabilità di errore media è da (6.27): $P(e) = p = 0.2$. Le probabilità a priori calcolate mostrano che la migliore strategia per un ricevitore MAP che osservi l'uscita per risalire all'ingresso, sarebbe quella di lasciare i bit inalterati. Infatti dalle probabilità a posteriori, se si è ricevuto uno "0" il simbolo più probabile ad essere stato trasmesso è ancora lo zero. Lo stesso vale per il simbolo "1". Ovviamente dopo la decisione MAP $P(e)$ non cambia.

Esempio 6.3 Consideriamo ora lo stesso canale dell'esempio precedente con $p = 0.2$, ma con simboli d'ingresso non equiprobabili: $P(0) = 0.1$ e quindi $P(1) = 0.9$. Sono qui mostrati due segmenti tipici di una sequenza ingresso-uscita

...1101110111111111111011111111101111111101111111111111111111101111...

...1100110110111011010111101111110111010111111101111011011011101...

Si noti come solo circa il 10% dei bit dell'ingresso sia pari al simbolo "0", come circa il 20% di tutti bit sia in errore e come gli errori siano indipendenti dal

simbolo. La probabilità di errore è infatti da (6.27): $P(e) = 0.2$. Da equazioni (6.28) e (6.29) abbiamo le probabilità delle uscite, $P\{Y = 0\} = 0.26$, $P\{Y = 1\} = 0.74$, e le probabilità a posteriori

$$\begin{aligned} P\{X = 0|Y = 0\} &= 0.3077, \\ P\{X = 1|Y = 0\} &= 0.6923, \\ P\{X = 0|Y = 1\} &= 0.0270, \\ P\{X = 1|Y = 1\} &= 0.9730. \end{aligned} \tag{6.32}$$

E' interessante notare come, dalle probabilità a posteriori, un ricevitore che decida con la regola MAP associerebbe a qualunque sequenza di uscita una sequenza di tutti "1". Infatti quando l'uscita è "0", $P\{X = 1|Y = 0\} > P\{X = 0|Y = 0\}$, e quindi conviene "decidere" per "1". Diversamente, se si è osservato "1", $P\{X = 1|Y = 1\} > P\{X = 0|Y = 1\}$, e pertanto conviene lasciare il simbolo inalterato. La probabilità di errore a valle del decisore MAP diventa

$$P(e)_{MAP} = Pr\{\text{trasmesso } 0\} = Pr\{X = 0\} = 0.1, \tag{6.33}$$

che è migliore di $P(e)$. La struttura di questo esempio è evidentemente fortemente influenzata dalla dissimmetria delle probabilità a priori.

Esempio 6.4 Costruiamo ora un esempio con un canale binario non simmetrico con matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} 0.8 & 0.2 \\ 0.01 & 0.99 \end{bmatrix}. \tag{6.34}$$

Si assuma che i simboli di ingresso siano equiprobabili, ovvero $P(0) = P(1) = 0.5$. La struttura della matrice di canale indica che il bit "0" verrà invertito con una probabilità di 0.2, mentre il bit "1" solo con una probabilità di 0.01. Un segmento tipico del flusso binario all'ingresso e all'uscita è

...0101110100110010010101010011010010110101001010110100101001l0...

...0111110100110011010101010111011010110101001011110100111001l0...

Si noti come i due simboli siano distribuiti quasi equamente al 50% nell'ingresso, come sia stato invertito circa il 20% degli "0" e solo l'1% degli "1" (in effetti in questo segmento nessun "1" è stato invertito). La probabilità media di errore, da equazione (6.24), è: $P(e) = 0.1050$. Le probabilità dei simboli di uscita, da (6.25), sono $P\{Y = 0\} = 0.4050$ e $P\{Y = 1\} = 0.5950$. Le probabilità a posteriori da (6.26) sono

$$\begin{aligned} P\{X = 0|Y = 0\} &= 0.9877, \\ P\{X = 1|Y = 0\} &= 0.0123, \\ P\{X = 0|Y = 1\} &= 0.1681, \\ P\{X = 1|Y = 1\} &= 0.8319. \end{aligned} \tag{6.35}$$

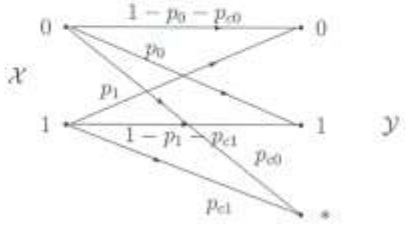


Figura 6.4: Il canale binario con cancellazione

Da questi risultati si evince che un ricevitore MAP lascerebbe i simboli di uscita inalterati, in quanto se si osserva "0" in uscita, il simbolo di ingresso più probabile è ancora "0". Analogamente se si osserva "1" in ingresso è ancora "1" ad essere il più probabile. Quindi $P(e)_{MAP} = P(e)$.

6.1.3 Il canale binario con cancellazione

Questo canale prevede che in ricezione i simboli binari trasmessi possano essere ricevuti erroneamente, o non ricevuti affatto. L'alfabeto d'ingresso è $\mathcal{X} = \{0, 1\}$ e quello di uscita è $\mathcal{Y} = \{0, 1, *\}$, in cui al simbolo "$*$" viene attribuito il significato di "cancellazione". La matrice di canale è

$$\mathbf{P}_c = \begin{bmatrix} 1-p_0-p_{c0} & p_0 & p_{c0} \\ p_1 & 1-p_1-p_{c1} & p_{c1} \end{bmatrix}. \tag{6.36}$$

Le probabilità p_0 e p_1 sono le probabilità di errore condizionate e p_{c0} e p_{c1} le probabilità di cancellazione condizionate ai simboli "0" e "1" ripettivamente. La figura 6.4 mostra il grafo relativo al canale. Il caso più tipico si ha per cancellazioni e per errori simmetrici: $p_{c0} = p_{c1} = p_c$ e $p_0 = p_1 = p$, per cui la matrice di canale diventa

$$\mathbf{P}_c = \begin{bmatrix} 1-p-p_c & p & p_c \\ p & 1-p-p_c & p_c \end{bmatrix}. \tag{6.37}$$

Nello studio del canale con cancellazione, poiché la cancellazione non è un vero e proprio errore, piuttosto che parlare di probabilità di errore media, conviene caratterizzare la probabilità di corretta ricezione

$$P(c) = P(c_0)P(0)+P(c_1)P(1) = (1-p_0-p_{c0})P(0)+(1-p_1-p_{c1})(1-P(0)), \tag{6.38}$$

che nel caso simmetrico diventa

$$P(c) = 1 - p - p_c. \tag{6.39}$$

La probabilità media di cancellazione è

$$P(\text{canc}) = P(\text{canc}_0)P(0)+P(\text{canc}_1)P(1) = p_{c0}P(0)+p_{c1}(1-P(0)), \tag{6.40}$$

che nel caso simmetrico diventa semplicemente

$$P(\text{canc}) = p_c. \tag{6.41}$$

Anche le probabilità dei simboli di uscita e le probabilità a posteriori sono facilmente ottenibili e sono lasciate al lettore per esercizio.

6.1.4 Canale ternario

Esaminiamo ora un esempio di canale i cui alfabeti di ingresso e di uscita hanno cardinalità tre e sono costituiti dagli stessi simboli

$$\mathcal{X} = \{A, B, C\}, \quad \mathcal{Y} = \{A, B, C\}. \tag{6.42}$$

La matrice di canale è di dimensioni 3×3 ed è per definizione

$$\begin{aligned} \mathbf{P}_c &= \begin{bmatrix} P\{Y=A|X=A\} & P\{Y=B|X=A\} & P\{Y=C|X=A\} \\ P\{Y=A|X=B\} & P\{Y=B|X=B\} & P\{Y=C|X=B\} \\ P\{Y=A|X=C\} & P\{Y=B|X=C\} & P\{Y=C|X=C\} \end{bmatrix} \\ &= \begin{bmatrix} P_{AA} & P_{AB} & P_{AC} \\ P_{BA} & P_{BB} & P_{BC} \\ P_{CA} & P_{CB} & P_{CC} \end{bmatrix}. \end{aligned} \tag{6.43}$$

La probabilità di errore segue lo schema generale ed è

$$\begin{aligned} P(e) &= P\{Y \neq X\} = 1 - P\{X = Y\} \\ &= 1 - (P\{X = A, Y = A\} + P\{X = B, Y = B\} + P\{X = C, Y = C\}) \\ &= 1 - (P\{Y = A|X = A\}P(A) + P\{Y = B|X = B\}P(B) \\ &+ P\{Y = C|X = C\}P(C)) \\ &= 1 - (P_{AA}P(A) + P_{BB}P(B) + P_{CC}P(C)) \\ &= 1 - diag(\mathbf{P}_c)^T \boldsymbol{\pi}_x, \end{aligned} \tag{6.44}$$

dove $\boldsymbol{\pi}_x = [P(A)P(B)P(C)]^T$ è il vettore delle probabilità a priori. Anche il calcolo delle probabilità delle uscite e delle probabilità a posteriori seguono lo schema generale e sono lasciate al lettore per esercizio.

Con riferimento a un caso numerico con probabilità a priori uniformi, $P(A) = P(B) = P(C) = 1/3$, e con matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}, \tag{6.45}$$

è qui mostrato un segmento di una possibile realizzazione di 40 simboli dell'ingresso e dell'uscita del canale

$$...ABCABABCABACBCBACAACBB\hat{C}BACABACBABCBCBCAC...$$
$$...ABABBAACCBACCCBABAACBBCBBCACAABCBCBCABAC...$$

Si noti la presenza per circa un terzo di ogni simbolo nell'ingresso e nell'uscita, e come circa il 40% dei simboli sia in errore ($P(e) = 0.4$).

6.1.5 Il canale perfetto

Questo caso particolare di canale quadrato ha matrice pari alla matrice identità

$$\mathbf{P}_c = \begin{bmatrix} 1 & 0 & . & . & 0 \\ 0 & 1 & 0 & . & 0 \\ . & . & . & . & . \\ 0 & . & . & 0 & 1 \end{bmatrix} = \mathbf{I}_n, \tag{6.46}$$

ed è detto *canale perfetto*, o *senza rumore*, per ovvie ragioni, visto che la associazione tra simboli di ingresso e simboli di uscita è univoca e senza errori. La probabilità di errore è ovviamente nulla. Si noti che anche per tutte le matrici di canale che sono permutazioni della matrice identità (matrici di permutazione unitarie) abbiamo una situazione di canale perfetto, con la peculiarità che i simboli vengono associati in maniera diversa, ma non confusi. Per esempio la matrice

$$\mathbf{P}_c = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{6.47}$$

corrisponde ancora ad un canale perfetto secondo la associazione

$$x_1 \leftrightarrow y_2, \quad x_2 \leftrightarrow y_1, \quad x_3 \leftrightarrow y_3. \tag{6.48}$$

6.1.6 Il canale sempre in errore

Questo canale quadrato ha una matrice di canale costituita da zeri sulla diagonale principale. Se si tratta di un canale simmetrico si ha $\frac{1}{n-1}$ per tutti gli altri elementi

$$\mathbf{P}_c = \begin{bmatrix} 0 & \frac{1}{n-1} & . & . & \frac{1}{n-1} \\ \frac{1}{n-1} & 0 & \frac{1}{n-1} & . & \frac{1}{n-1} \\ . & . & . & . & . \\ \frac{1}{n-1} & . & . & \frac{1}{n-1} & 0 \end{bmatrix}. \tag{6.49}$$

In questo ultimo caso i simboli saranno sempre in errore e confusi con uguale probabilità con uno degli altri. La probabilità di errore è ovviamente pari a uno.

6.1.7 Il canale massimamente confuso

In questo esempio di canale (non necessariamente quadrato) i simboli di ingresso si distribuiscono uniformemente sui simboli di uscita. La matrice di canale è

$$\mathbf{P}_c = \begin{bmatrix} 1/m & 1/m & . & . & 1/m \\ 1/m & 1/m & 1/m & . & 1/m \\ . & . & . & . & . \\ 1/m & . & . & 1/m & 1/m \end{bmatrix}. \tag{6.50}$$

In tale canale, la osservazione dei simboli di uscita, non ci da alcuna informazione sui simboli di ingresso se le probabilità a priori sono uguali. Se $m = n$ (canale quadrato) e si pensa alla associazione uno-a-uno dei simboli di ingresso con le uscite, abbiamo che $P(e) = (n-1)/n$.

6.1.8 Canale perfetto con cancellazione

Questo esempio di canale può essere utile a modellare il trasporto di simboli con perdite. La struttura della matrice di dimensioni $n \times (n+1)$ è

$$\mathbf{P}_c = \begin{bmatrix} 1-p_c & 0 & . & . & 0 & p_c \\ 0 & (1-p_c) & 0 & . & 0 & p_c \\ . & . & . & . & . & . \\ 0 & . & . & 0 & (1-p_c) & p_c \end{bmatrix}, \tag{6.51}$$

dove p_c è la probabilità di cancellazione. Tutti i simboli, o sono trasferiti inalterati sull'uscita, o sono persi. La probabilità media di corretta ricezione

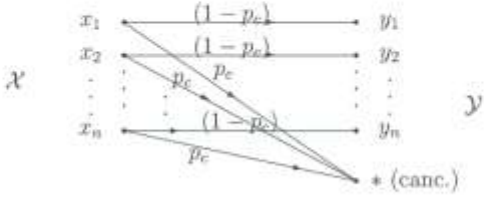


Figura 6.5: Il canale perfetto con cancellazione

è chiaramente $P(c) = 1 - p_c$. La figura 6.5 mostra il grafo relativo al canale.

6.1.9 Il canale senza rumore

Il canale perfetto esaminato precedentemente è un caso particolare del cosiddetto *canale senza rumore*. Un canale senza rumore ha una matrice di canale con un solo elemento diverso da zero in ogni colonna. Ciò significa che ogni simbolo di ingresso si trasforma in un simbolo di uscita, ma nessun altro simbolo di ingresso si trasforma in quel simbolo. In altre parole, dato un simbolo di uscita, è possibile risalire in maniera univoca al simbolo di ingresso che lo ha generato. La figura 6.6 mostra un esempio di canale senza rumore con matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.52}$$

Si noti che per tutti i canali senza rumore le probabilità a posteriori, per tutte le coppie aventi probabilità dirette non nulle sono

$$P(x_i|y_j) = \frac{P(y_j|x_i)P(x_i)}{\sum_{t=1}^{n} P(y_j|x_t)P(x_t)} = \frac{P(y_j|x_i)P(x_i)}{P(y_j|x_i)P(x_i)} = 1. \tag{6.53}$$

Per le altre coppie la probabilità a posteriori è nulla. Pertanto la matrice delle probabilità a posteriori ha la stessa struttura delle matrice di canale ma con elementi unitari. Nell'esempio numerico riportato abbiamo che

$$\mathbf{P}_p = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.54}$$

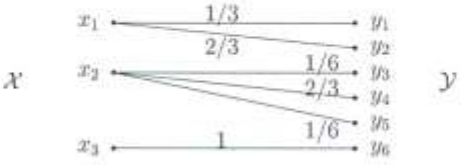


Figura 6.6: Un esempio di canale senza rumore

6.1.10 Il canale deterministico

Questo tipo di canale, contiene anch'esso come caso particolare il canale perfetto e costituisce il caso duale del canale senza rumore. Un canale deterministico ha matrice di canale con un solo elemento diverso da zero in ogni riga (e quindi pari a uno). Ciò significa che ogni simbolo di ingresso si trasforma in un simbolo di uscita senza ambiguità. Lo stesso simbolo di uscita può comunque provenire da più di un simbolo di ingresso. La figura 6.7 mostra un esempio di canale deterministico con matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6.55}$$

Le probabilità a posteriori riflettono la ambiguità nel risalire al simbolo di ingresso, dato un simbolo di uscita, visto che più simboli di ingresso possono averlo generato. Tale ambiguità è chiaramente governata esclusivamente dalle probabilità a priori. Per tutte le coppie ingresso uscita che hanno una probabilità diretta non nulla, abbiamo che

$$P(x_i|y_j) = \frac{P(y_j|x_i)P(x_i)}{\sum_{t=1}^{n} P(y_j|x_t)P(x_t)} = \frac{P(x_i)}{\sum_{t \in \mathcal{X}_j} P(x_t)}, \tag{6.56}$$

dove $\mathcal{X}_j$ indica il sottoinsieme dei simboli di ingresso che possono aver generato y_j. Per le altre coppie la probabilità a posteriori è zero. Nel caso

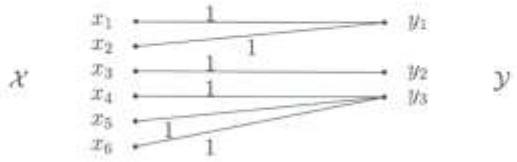


Figura 6.7: Un esempio di canale deterministico

dell'esempio, abbiamo che la matrice delle probabilità a posteriori è

$$\mathbf{P}_p = \begin{bmatrix} \frac{P(x_1)}{P(x_1)+P(x_2)} & 0 & 0 \\ \frac{P(x_2)}{P(x_1)+P(x_2)} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{P(x_4)}{P(x_4)+P(x_5)+P(x_6)} \\ 0 & 0 & \frac{P(x_5)}{P(x_4)+P(x_5)+P(x_6)} \\ 0 & 0 & \frac{P(x_6)}{P(x_4)+P(x_5)+P(x_6)} \end{bmatrix}. \tag{6.57}$$

6.2 Il canale esteso

In analogia alla definizione di sorgente estesa, introdotta a proposito della modellistica delle sorgenti discrete, definiamo qui il *canale esteso di ordine* N. La definizione di canale esteso è di importanza cruciale per i teoremi che seguiranno e consiste nel considerare i simboli che si presentano all'ingresso del canale a gruppi di N. Secondo il modello di canale senza memoria introdotto, le uscite sono anch'esse descritte a gruppi di N simboli e ogni simbolo di uscita dipende solo dal simbolo presente all'ingresso nello stesso istante. La figura 6.8 mostra lo schema di principio del canale esteso. Più formalmente:

Dato un canale discreto senza memoria con alfabeti e matrice di canale:

$$\mathcal{X} = \{x_1, x_2, ..., x_n\}, \quad \mathcal{Y} = \{y_1, y_2, ..., y_m\}, \quad \mathbf{P}_c, \tag{6.58}$$

si definisce canale esteso di ordine N il canale avente come alfabeto di ingresso, l'alfabeto della sorgente estesa di ordine N di $\mathcal{X}$:

$$\mathcal{X}^N = \{x_1^N, x_2^N, ..., x_{M_x}^N\}, \tag{6.59}$$

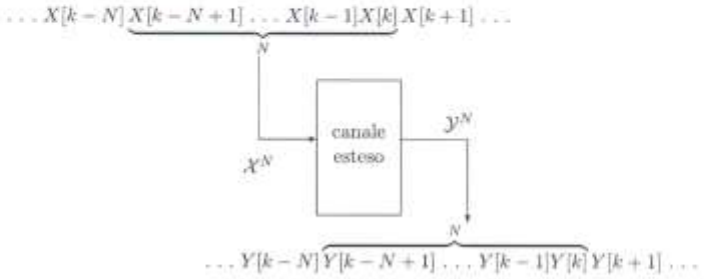


Figura 6.8: Il canale esteso di ordine N

come alfabeto di uscita, l'alfabeto della versione estesa di ordine N di $\mathcal{Y}$:

$$\mathcal{Y}^N = \{y_1^N, y_2^N, ..., y_{M_y}^N\}, \tag{6.60}$$

e come matrice di canale, la matrice:

$$\mathbf{P}_c^N = \left[P\{Y^N = y_j^N | X^N = x_i^N\}\right], i = 1, ..., M_x, \quad j = 1, ..., M_y, \tag{6.61}$$

delle probabilità dirette tra i simboli estesi di $\mathcal{X}^N$ e $\mathcal{Y}^N$. Le cardinalità di $\mathcal{X}^N$ e $\mathcal{Y}^N$ sono rispettivamente: $M_x = n^N$ e $M_y = m^N$.

La definizione di canale esteso non introduce alcun concetto nuovo, ma consiste semplicemente in una riformalizzazione del funzionamento del canale senza memoria su blocchi di lunghezza N della sequenza d'ingresso. Le probabilità condizionate che costituiscono la matrice di canale possono essere calcolate facilmente come prodotti delle probabilità condizionate contenute nella matrice $\mathbf{P}_c$ del canale di partenza. Prima di approfondire il calcolo della matrice di canale, esaminiamo un esempio di canale esteso per chiarirne meglio la definizione.

Esempio 6.5 Il canale binario esteso. Si consideri il canale binario presentato in precedenza. La sua versione estesa con $N = 2$ opera su coppie di simboli binari e ha alfabeti di ingresso e di uscita:

$$\mathcal{X}^2 = \{00, 01, 10, 11\}, \quad \mathcal{Y}^2 = \{00, 01, 10, 11\}. \tag{6.62}$$

La matrice di canale è facilmente ottenibile data la natura senza memoria del canale. Ad esempio:

$$\begin{aligned} P\{Y^2 = 00|X^2 = 00\} &= P\{Y = 0|X = 0\}P\{Y = 0|X = 0\}, \\ P\{Y^2 = 01|X^2 = 00\} &= P\{Y = 0|X = 0\}P\{Y = 1|X = 0\}, \\ P\{Y^2 = 10|X^2 = 01\} &= P\{Y = 1|X = 0\}P\{Y = 0|X = 1\}, \\ \cdots\cdots & \quad \cdots\cdots\cdots \end{aligned} \tag{6.63}$$

poiché l'uscita ad ogni istante è condizionata solo dall'ingresso allo stesso istante. Analogamente si calcolano le probabilità per tutte le altre coppie, per cui la matrice di canale diventa:

$$\begin{aligned} \mathbf{P}_c^2 &= \begin{bmatrix} P(00|00) & P(01|00) & P(10|00) & P(11|00) \\ P(00|01) & P(01|01) & P(10|01) & P(11|01) \\ P(00|10) & P(01|10) & P(10|10) & P(11|10) \\ P(00|11) & P(01|11) & P(10|11) & P(11|11) \end{bmatrix} \\ &= \begin{bmatrix} (1-p_0)^2 & (1-p_0)p_0 & p_0(1-p_0) & p_0^2 \\ (1-p_0)p_1 & (1-p_0)(1-p_1) & p_0p_1 & p_0(1-p_1) \\ p_1(1-p_0) & p_1p_0 & (1-p_1)(1-p_0) & (1-p_1)p_0 \\ p_1^2 & p_1(1-p_1) & (1-p_1)p_1 & (1-p_1)^2 \end{bmatrix}. \end{aligned} \tag{6.64}$$

Analogamente, il canale esteso per $N = 3$ consiste nella associazione di stringhe binarie di lunghezza tre tra l'ingresso e l'uscita. Gli alfabeti sono:

$$\begin{aligned} \mathcal{X}^3 &= \{000, 001, 010, 011, 100, 101, 110, 111\}, \\ \mathcal{Y}^3 &= \{000, 001, 010, 011, 100, 101, 110, 111\}. \end{aligned} \tag{6.65}$$

La matrice di canale si ottiene come nel caso per $N = 2$ e la si lascia al lettore per esercizio.

L'esempio del canale binario suggerisce una formula generale per la matrice di canale del canale esteso. Per semplice ispezione della matrice, è immediato vedere come la struttura generale della matrice $\mathbf{P}_c^2$ sia

$$\begin{aligned} \mathbf{P}_c^2 &= \begin{bmatrix} P(y_1|x_1)\mathbf{P}_c & P(y_2|x_1)\mathbf{P}_c & \cdots & P(y_m|x_1)\mathbf{P}_c \\ P(y_1|x_2)\mathbf{P}_c & P(y_2|x_2)\mathbf{P}_c & \cdots & P(y_m|x_2)\mathbf{P}_c \\ \cdots & \cdots & \cdots & \cdots \\ P(y_1|x_n)\mathbf{P}_c & P(y_2|x_n)\mathbf{P}_c & \cdots & P(y_m|x_n)\mathbf{P}_c \end{bmatrix} \\ &= \begin{bmatrix} P_{11}\mathbf{P}_c & P_{12}\mathbf{P}_c & \cdots & P_{1m}\mathbf{P}_c \\ P_{21}\mathbf{P}_c & P_{22}\mathbf{P}_c & \cdots & P_{2m}\mathbf{P}_c \\ \cdots & \cdots & \cdots & \cdots \\ P_{n1}\mathbf{P}_c & P_{n2}\mathbf{P}_c & \cdots & P_{nm}\mathbf{P}_c \end{bmatrix}. \end{aligned} \tag{6.66}$$

Si noti la struttura a blocchi della matrice di dimensioni $n^2 \times m^2$ risultante. E' riconoscibile, dal calcolo matriciale, il *prodotto di Kronecker* tra la matrice $\mathbf{P}_c$ e se stessa. Pertanto in generale si può scrivere

$$\mathbf{P}_c^N = \underbrace{\mathbf{P}_c \otimes \mathbf{P}_c \otimes ... \otimes \mathbf{P}_c}_{N}, \tag{6.67}$$

dove il simbolo $\otimes$ indica il prodotto di Kronecker. Analogamente è possibile formalizzare in maniera compatta la struttura degli alfabeti di ingresso e uscita. Si consideri dapprima il caso per $N = 2$. Poiché l'alfabeto della sorgente estesa prevede la inclusione di tutte le coppie ottenibili dagli elementi di $\mathcal{X}$, possiamo costruire l'alfabeto di ingresso affiancando ad ogni simbolo di sorgente tutti gli altri simboli in maniera ordinata, ovvero

$$\begin{aligned} \mathcal{X}^2 &= \{x_1(x_1, ..., x_n), x_2(x_1, ..., x_n), ..., x_n(x_1, ..., x_n)\} \\ &= \{x_1x_1, ..., x_1x_n, x_2x_1, ..., x_2x_n, ..., x_nx_1, ..., x_nx_n\}, \end{aligned} \tag{6.68}$$

che è il prodotto cartesiano di $\mathcal{X}$ con se stesso:

$$\mathcal{X}^2 = \mathcal{X} \times \mathcal{X}. \tag{6.69}$$

Generalizzando, per una sorgente estesa di ordine N, abbiamo che

$$\mathcal{X}^N = \underbrace{\mathcal{X} \times \mathcal{X} \times ... \times \mathcal{X}}_{N} \quad , \quad \mathcal{Y}^N = \underbrace{\mathcal{Y} \times \mathcal{Y} \times ... \times \mathcal{Y}}_{N}, \tag{6.70}$$

Quindi usando i prodotti di Kronecker si ha

$$\boldsymbol{\pi}_x^N = \underbrace{\boldsymbol{\pi}_x \otimes \boldsymbol{\pi}_x \otimes ... \otimes \boldsymbol{\pi}_x}_{N}, \tag{6.71}$$

$$\boldsymbol{\pi}_y^N = \underbrace{\boldsymbol{\pi}_y \otimes \boldsymbol{\pi}_y \otimes ... \otimes \boldsymbol{\pi}_y}_{N}, \tag{6.72}$$

dove $\boldsymbol{\pi}_x^N$ e $\boldsymbol{\pi}_y^N$ sono i vettori che contengono le probabilità di Π_X^N e Π_Y^N.

6.3 L'ambiguità del canale

Seguendo il modello di canale discreto equivalente, abbiamo già visto come la osservazione dell'uscita possa fornirci la possibilità di fare delle supposizioni sul simbolo di ingresso grazie alle probabilità a posteriori. Si possono pertanto mettere a confronto le probabilità a priori di X

$$\{P(x_1), P(x_2), ..., P(x_n)\}, \tag{6.73}$$

polen.22

e le probabilità a posteriori di X una volta che si è osservato un simbolo y_j in uscita

$$\{P(x_1|y_j), P(x_2|y_j), ..., P(x_n|y_j)\}. \tag{6.74}$$

La incertezza, o la informazione, collegata alle due distribuzioni può essere quantificata mediante la funzione entropia applicata ai due casi. Si definisce pertanto *entropia a priori* di X

$$\mathcal{H}(X) = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)}, \tag{6.75}$$

e *entropia a posteriori di* X, *dato* y_j

$$\mathcal{H}(X|Y = y_j) = \sum_{i=1}^{n} P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)}. \tag{6.76}$$

Mediando sui simboli di uscita, abbiamo la *entropia media di* X *dato* Y

$$\mathcal{H}(X|Y) = \sum_{j=1}^{m} P(y_j)\mathcal{H}(X|Y = y_j). \tag{6.77}$$

Più esplicitamente,

$$\begin{aligned} \mathcal{H}(X|Y) &= \sum_{j=1}^{m}\sum_{i=1}^{n} P(y_j)P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i|y_j)} \\ &= -E[\log_2 P(X|Y)]. \end{aligned} \tag{6.78}$$

La quantità $\mathcal{H}(X|Y)$ è anche denominata *ambiguità* (*equivocation*) del canale. Seguendo l'interpretazione tipicamente associata all'entropia, possiamo dire che $\mathcal{H}(X)$ è l'incertezza associata intrinsecamente a X, mentre $\mathcal{H}(X|Y)$ è l'incertezza associata a X una volta che si è osservato Y. Quindi un canale "molto ambiguo" ha un valore di $\mathcal{H}(X|Y)$ vicino alla entropia a priori, ovvero la osservazione della uscita non è molto utile per risalire a X. Viceversa un canale "poco ambiguo" è un canale per cui la osservazione dell'uscita fornisce molta informazione sull'ingresso e ha $\mathcal{H}(X|Y)$ prossima a zero.

6.4 La mutua informazione

Dalla definizione di ambiguità, abbiamo che l'informazione acquisita su X, una volta che si è osservato Y è la differenza

$$I(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y). \tag{6.79}$$

Tale differenza è la *mutua informazione tra* X *e* Y, o *mutua informazione del canale*. La definizione di mutua informazione riveste un ruolo centrale nella teoria dell'informazione in quanto rappresenta l'ammontare di informazione mediamente scambiata tra X e Y. La mutua informazione si misura in bit o nell'unità corrispondente alla base del logaritmo usata. Essa può essere riscritta in vari modi

$$\begin{aligned}
I(X;Y) &= \mathcal{H}(X) - \mathcal{H}(X|Y) \\
&= \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i|y_j)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i)} - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i|y_j)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i|y_j)}{P(x_i)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j|x_i)P(x_i) \log_2 \frac{P(y_j|x_i)}{P(y_j)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j|x_i)P(x_i) \log_2 \frac{P(y_j|x_i)}{\sum_{l=1}^{n} P(y_j|x_l)P(x_l)}.
\end{aligned} \tag{6.80}$$

Per facilità di calcolo, la mutua informatione può anche essere espressa in forma matriciale. Gli elementi all'interno della doppia sommatoria nella espressione della entropia condizionata possono essere organizzate nella matrice

$$\left[P(y_j|x_i) \log_2 \frac{1}{P(x_i|y_j)}\right]_{i,j} = -\mathbf{P}_c \odot (\log_2 \mathbf{P}_p), \tag{6.81}$$

dove $\odot$ è il prodotto termine a termine delle due matrici (detto anche prodotto di Hadamard). Pertanto, moltiplicando per $P(x_i)$ e sommando si ha

$$\mathcal{H}(X|Y) = -\boldsymbol{\pi}_x^T (\mathbf{P}_c \odot (\log_2 \mathbf{P}_p))\, \mathbf{e}_n. \tag{6.82}$$

Una espressione completa della mutua informazione è pertanto

$$I(X;Y) = \boldsymbol{\pi}_x^T \left[(\mathbf{P}_c \odot (\log_2 \mathbf{P}_p))\, \mathbf{e}_n - (\log_2 \boldsymbol{\pi}_x)\right]. \tag{6.83}$$

Prima di approfondire ulteriormente la interpretazione della mutua informazione e la sua applicazione al canale, esaminiamone alcune proprietà.

Proprietà 1: $I(X;Y) = I(Y;X)$. La mutua informazione è simmetrica rispetto ai suoi due argomenti.

Prova: Dalla terz'ultima uguaglianza di equazione (6.80), poichè $P(x_i, y_j) = P(y_j, x_i)$.△

Proprietà 2: $I(X;Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X)$.

Prova: Direttamente da Proprietà 1 e dalla definizione di mutua informazione (6.79). △

Proprietà 3: *Se X e Y sono indipendenti,*

$$\mathcal{H}(X|Y) = \mathcal{H}(X), \quad \mathcal{H}(Y|X) = \mathcal{H}(Y), \quad I(X,Y) = 0. \tag{6.84}$$

Prova: Direttamente dalla definizione di entropia condizionata e dalla definizione di mutua informazione. △

La quantità $\mathcal{H}(Y|X)$ è anche detta "evidenza" del canale, in dualità con la "ambiguità" $\mathcal{H}(X|Y)$.

Proprietà 4: $I(X;Y) \geq 0$.

Prova: Il risultato è immediato in quanto

$$\begin{aligned} -I(X;Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\log_2\frac{P(x_i)P(y_j)}{P(x_i,y_j)} \\ &\leq (\log_2 e)\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\left(\frac{P(x_i)P(y_j)}{P(x_i,y_j)} - 1\right) \\ &= (\log_2 e)\left(\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)P(y_j) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\right) \\ &= 0, \end{aligned} \tag{6.85}$$

dove è stata usata la disuguaglianza $\ln x \leq x - 1$. △

E' utile osservare che, nonostante la mutua informazione sia sempre maggiore o uguale a zero, ciò non vale necessariamente per la quantità

$$\mathcal{H}(X) - \mathcal{H}(X|Y = y_j). \tag{6.86}$$

Alcuni esempi nei problemi chiariranno la differenza.

Proprietà 5(a): *La mutua informazione $I(X;Y)$, fissate le probabilità dirette $\mathbf{P}_c$, è una funzione concava di $\boldsymbol{\pi}_x$.*

Prova: Consideriamo due distribuzioni per X, π_{1x} e π_{2x} aventi probabilità $P_1(x_i)$ e $P_2(x_i)$ per $i = 1, ..., n$. Sia $\lambda \in [0, 1]$ e

$$P_*(x_i) = \lambda P_1(x_i) + (1 - \lambda) P_2(x_i), \quad i = 1, ..., n. \tag{6.87}$$

Valutiamo la quantità $I_* - \lambda I_1 - (1 - \lambda) I_2$, dove I_*, I_1 e I_2 sono le mutue informazioni relative a P_*, P_1 e P_2

$$\begin{aligned}
& \sum_{ij} P(y_j|x_i) P_*(x_i) \log_2 \frac{P(y_j|x_i)}{\sum_l P(y_j|x_l) P_*(x_l)} \\
- \; & \lambda \sum_{ij} P(y_j|x_i) P_1(x_i) \log_2 \frac{P(y_j|x_i)}{\sum_l P(y_j|x_l) P_1(x_l)} \\
- \; & (1 - \lambda) \sum_{ij} P(y_j|x_i) P_2(x_i) \log_2 \frac{P(y_j|x_i)}{\sum_l P(y_j|x_l) P_2(x_l)} \\
= \; & \lambda \sum_{ij} P(y_j|x_i) P_1(x_i) \log_2 \frac{\sum_m P(y_j|x_m) P_1(x_m)}{\sum_l P(y_j|x_l) P_*(x_l)} \\
+ \; & (1 - \lambda) \sum_{ij} P(y_j|x_i) P_2(x_i) \log_2 \frac{\sum_m P(y_j|x_m) P_2(x_m)}{\sum_l P(y_j|x_l) P_*(x_l)} \\
\geq \; & (\log_2 e) \left[\lambda \sum_{ij} P(y_j|x_i) P_1(x_i) \left(1 - \frac{\sum_m P(y_j|x_m) P_*(x_m)}{\sum_l P(y_j|x_l) P_1(x_l)} \right) \right. \\
+ \; & \left. (1 - \lambda) \sum_{ij} P(y_j|x_i) P_2(x_i) \left(1 - \frac{\sum_m P(y_j|x_m) P_*(x_m)}{\sum_l P(y_j|x_l) P_2(x_l)} \right) \right] = 0.
\end{aligned} \tag{6.88}$$

La disuguaglianza, in cui abbiano usato la relazione $\ln x \geq 1 - \frac{1}{x}$, prova la convessità.[5] $\triangle$

Proprietà 5(b): *La mutua informazione $I(X;Y)$, fissate le probabilità Π_x, è una funzione convessa di $\mathbf{P}_c$.*

Prova: Consideriamo due matrici di canale $\mathbf{P}_{1c}$ e $\mathbf{P}_{2c}$ arbitrarie e indichiamo con $P_1(y_j|x_i)$ e $P_2(y_j|x_i)$ le probabilità condizionate in esse contenute. Le probabilità delle uscite sono nei due casi

$$\begin{aligned}
P_1(y_j) &= \sum_{i=1}^{n} P_1(y_j|x_i) P(x_i), \\
P_2(y_j) &= \sum_{i=1}^{n} P_2(y_j|x_i) P(x_i),
\end{aligned} \tag{6.89}$$

[5] La disuguaglianza è facilmente dimostrata usando la relazione $\ln x \leq x - 1$. Sostituendo a x, $\frac{1}{x}$, abbiamo che $\ln \frac{1}{x} \leq \frac{1}{x} - 1$, ovvero $-\ln x \leq \frac{1}{x} - 1$ e quindi $\ln x \geq 1 - \frac{1}{x}$.

$j = 1, ..., m$, dove le probabilità dei simboli di ingresso $P(x_i)$, $i = 1, ..., n$ sono fissate. Consideriamo un qualunque $\lambda \in [0, 1]$ e la espressione

$$[\lambda P_1(y_j|x_i) + (1-\lambda)P_2(y_j|x_i)] \log_2 \frac{\lambda P_1(y_j|x_i) + (1-\lambda)P_2(y_j|x_i)}{\lambda P_1(y_j) + (1-\lambda)P_2(y_j)}. \tag{6.90}$$

Usando la disuguaglianza (A.7) riportata in Appendice A, con

$$\begin{aligned} \sum a &\leftrightarrow \lambda P_1(y_j|x_i) + (1-\lambda)P_2(y_j|x_i), \\ \sum b &\leftrightarrow \lambda P_1(y_j) + (1-\lambda)P_2(y_j), \end{aligned} \tag{6.91}$$

abbiamo che

$$\begin{aligned} &[\lambda P_1(y_j|x_i) + (1-\lambda)P_2(y_j|x_i)] \log_2 \frac{\lambda P_1(y_j|x_i) + (1-\lambda)P_2(y_j|x_i)}{\lambda P_1(y_j) + (1-\lambda)P_2(y_j)} \\ &\leq \lambda P_1(y_j|x_i) \log_2 \frac{\lambda P_1(y_j|x_i)}{\lambda P_1(y_j)} + (1-\lambda)P_2(y_j|x_i) \log_2 \frac{\lambda P_2(y_j|x_i)}{\lambda P_2(y_j)}. \end{aligned} \tag{6.92}$$

Moltiplicando ambo i membri per $P(x_i)$ e sommando su i e j, abbiamo

$$\begin{aligned} &\sum_{ij} P(x_i) [\lambda P_1(y_j|x_i) + (1-\lambda)P_2(y_j|x_i)] \log_2 \frac{\lambda P_1(y_j|x_i) + (1-\lambda)P_2(y_j|x_i)}{\lambda P_1(y_j) + (1-\lambda)P_2(y_j)} \\ \leq\ & \lambda \sum_{ij} P(x_i)P_1(y_j|x_i) \log_2 \frac{P_1(y_j|x_i)}{P_1(y_j)} \\ &+(1-\lambda) \sum_{ij} P(x_i)P_2(y_j|x_i) \log_2 \frac{P_2(y_j|x_i)}{P_2(y_j)}, \end{aligned} \tag{6.93}$$

che sancisce la convessità cercata. $\triangle$

Va comunque menzionato che le definizioni di mutua informazione e entropia condizionata, qui introdotte con riferimento al canale discreto, hanno validità generale e si applicano a qualunque coppia di variabili aleatorie. [6] Per arricchire il quadro delle definizioni basilari della teoria dell'informazione, è anche possibile definire una *entropia congiunta*

$$\mathcal{H}(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i, y_j)}. \tag{6.94}$$

Si tratta della definizione di entropia già data, ma applicata alla variabile congiunta (X, Y). La entropia congiunta è simmetrica, ovvero $\mathcal{H}(X, Y) =$

[6] Per ulteriori approfondimenti si rimanda il lettore a uno dei libri di testo sulla teoria dell'informazione suggeriti in coda al testo.

polm. 27

$\mathcal{H}(Y,X)$, poichè $P(x_i,y_j) = P(y_j,x_i)$. Le relazioni con $\mathcal{H}(X)$, $\mathcal{H}(Y)$ e $I(X;Y)$ sono facimente ottenibili

$$\begin{aligned}
\mathcal{H}(X,Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\log_2\frac{P(x_i)P(y_j)}{P(x_i,y_j)} + \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\log_2\frac{1}{P(x_i)P(y_j)} \\
&= -I(X;Y) + \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\log_2\frac{1}{P(x_i)} + \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j)\log_2\frac{1}{P(y_j)} \\
&= -I(X;Y) + \sum_{i=1}^{n} P(x_i)\log_2\frac{1}{P(x_i)} + \sum_{j=1}^{m} P(y_j)\log_2\frac{1}{P(y_j)} \\
&= \mathcal{H}(X) + \mathcal{H}(Y) - I(X;Y). \qquad (6.95)
\end{aligned}$$

L'ultima relazione ha una interpretazione interessante: la autoinformazione media (entropia) della coppia (X,Y) è la somma delle autoinformazioni medie marginali delle due variabili, meno la informazione tra esse mediamente scambiata.

Dalla (6.95) e dalla definizione di mutua informazione abbiamo anche che:

$$\mathcal{H}(X,Y) = \mathcal{H}(Y) + \mathcal{H}(X|Y), \qquad (6.96)$$
$$\mathcal{H}(Y,X) = \mathcal{H}(X) + \mathcal{H}(Y|X). \qquad (6.97)$$

La interpretazione di queste relazione è anch'essa interessante: la autoinformazione media (entropia) della coppia (X,Y) è la somma della autoinformazione di X (o di Y), più la autoinformazione media di Y dato X (o di X dato Y). Quindi, se le variabili aleatorie X e Y sono indipendenti, abbiamo che

$$\mathcal{H}(X,Y) = \mathcal{H}(X) + \mathcal{H}(Y). \qquad (6.98)$$

La figura 6.9 mostra un diagramma che descrive sinteticamente le relazioni esistenti tra le quantità informazionali definite: $\mathcal{H}(X)$, $\mathcal{H}(Y)$, $\mathcal{H}(X,Y)$, $\mathcal{H}(X|Y)$, $\mathcal{H}(Y|X)$, $I(X;Y)$.

6.4.1 La mutua informazione per il BSC

Ricordiamo che la matrice di canale per il canale binario simmetrico è

$$\mathbf{P}_c = \begin{bmatrix} P(y_1|x_1) & P(y_2|x_1) \\ P(y_1|x_2) & P(y_2|x_2) \end{bmatrix} = \begin{bmatrix} (1-p) & p \\ p & (1-p) \end{bmatrix}. \qquad (6.99)$$

Le probabilità dell'alfabeto d'ingresso siano

$$P(x_1) = \pi, \quad P(x_2) = 1-\pi. \qquad (6.100)$$

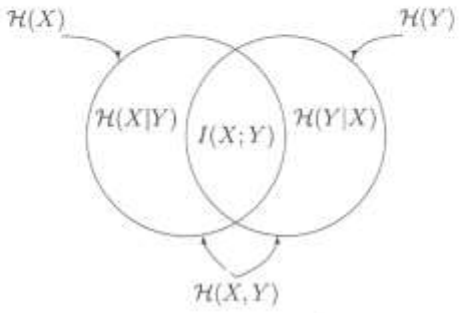


Figura 6.9: Le relazioni tra le quantità informazionali definite.

La mutua informazione può essere calcolata come segue

$$\begin{aligned} I(X;Y) &= \mathcal{H}(Y) - \mathcal{H}(Y|X) \\ &= \mathcal{H}(Y) - \sum_{i=1}^{2}\sum_{j=1}^{2} P(x_i,y_j)\log_2\frac{1}{P(y_j|x_i)} \\ &= \mathcal{H}(Y) - \sum_{i=1}^{2} P(x_i)\sum_{j=1}^{2} P(y_j|x_i)\log_2\frac{1}{P(y_j|x_i)} \\ &= \mathcal{H}(Y) - \sum_{i=1}^{2} P(x_i)\left((1-p)\log_2\frac{1}{1-p} + p\log_2\frac{1}{p}\right) \\ &= \mathcal{H}(Y) - \left((1-p)\log_2\frac{1}{1-p} + p\log_2\frac{1}{p}\right). \end{aligned} \tag{6.101}$$

Poiché le probabilità dei simboli di uscita sono

$$\begin{aligned} P(y_1) &= (1-p)\pi + p(1-\pi), \\ P(y_2) &= p\pi + (1-p)(1-\pi), \end{aligned} \tag{6.102}$$

la mutua informazione diventa

$$\begin{aligned} I(X;Y) &= ((1-p)\pi + p(1-\pi))\log_2\frac{1}{(1-p)\pi + p(1-\pi)} \\ &\quad + (p\pi + (1-p)(1-\pi))\log_2\frac{1}{p\pi + (1-p)(1-\pi)} \end{aligned}$$

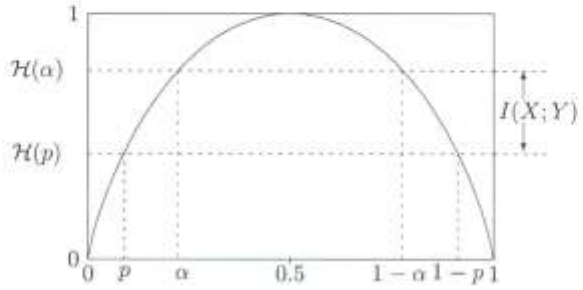


Figura 6.10: La interpretazione grafica della mutua informazione del canale binario simmetrico ($\alpha = (1-p)\pi + p(1-\pi)$).

$$\begin{aligned} & -\left((1-p)\log_2\frac{1}{1-p} + p\log_2\frac{1}{p}\right) \\ = \ & \mathcal{H}(\alpha) - \mathcal{H}(p), \end{aligned} \tag{6.103}$$

dove $\alpha = (1-p)\pi + p(1-\pi)$ e l'ultima uguaglianza esprime la mutua informazione come la differenza della entropia di due sorgenti binarie. La entropia di una sorgente binaria è stata già discussa nel capitolo precedente ed ha la forma riportata in figura 6.10. Poiché al variare di π tra zero e uno:

$$p \leq (1-p)\pi + p(1-\pi) \leq 1-p, \tag{6.104}$$

ovvero $p \leq \alpha \leq 1-p$, sulla figura la mutua informazione può essere valutata graficamente. Si noti anche come la mutua informazione sia sempre maggiore o uguale a zero.

La figura mostra anche l'evoluzione della mutua informazione se si si fissa p (canale) e si fa variare π. Il valore massimo della mutua informazione è attinto per $\alpha = \frac{1}{2}$, ovvero per :

$$\pi = \frac{\frac{1}{2} - p}{1-2p}. \tag{6.105}$$

La massima mutua informazione è in tali condizioni:

$$I(X;Y) = 1 - \mathcal{H}(p). \tag{6.106}$$

Un caso degenere si ha quando $p = 0$ (canale senza rumore) che fornisce $\mathcal{H}(p) = 0$, $\alpha = \pi$ e $I(X;Y) = \mathcal{H}(\pi)$. In tale situazione la massima mutua informazione si ottiene per $\pi = \frac{1}{2}$ ed è pari a uno.

Un altro caso limite si ottiene quando il canale genera la massima confusione, ovvero quando $p = \frac{1}{2}$ per cui $\mathcal{H}(p) = 1$, $\alpha = \frac{1}{2}$, $\mathcal{H}(\alpha) = 1$ e la mutua informazione è nulla indipendentemente dal valore di π.

Esempio 6.6 Consideriamo il *canale senza rumore* definito precedentemente. Si ricordi come tutte le probabilità a posteriori corrispondenti a coppie ingresso-uscita con probabilità diretta non nulla, fossero pari a uno. In tal caso abbiamo che:

$$\begin{aligned} I(X;Y) &= \mathcal{H}(X) - \mathcal{H}(X|Y) \\ &= \mathcal{H}(X) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_i|y_j)} \\ &= \mathcal{H}(X) - \sum_{j=1}^{m} P(y_j) \sum_{i=1}^{n} P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)} \\ &= \mathcal{H}(X) - \sum_{j=1}^{m} P(y_j) \sum_{X_j} 1 \log_2 1 \\ &= \mathcal{H}(X) - 0 = \mathcal{H}(X). \end{aligned} \tag{6.107}$$

Il canale ha ambiguità $\mathcal{H}(X|Y) = 0$ e tutta l'informazione $\mathcal{H}(X)$ prodotta alla sorgente è trasferita all'uscita.

Esempio 6.7 Consideriamo il calcolo della mutua informazione per il *canale deterministico*. In tal caso le probabilità dirette del canale sono o uno o zero. Pertanto

$$\begin{aligned} \mathcal{H}(Y|X) &= \sum_{j=1}^{m}\sum_{i=1}^{n} P(x_i, y_j) \log_2 \frac{1}{P(y_j|x_i)} \\ &= \sum_{i=1}^{n} P(x_i) \sum_{j=1}^{m} P(y_j|x_i) \log_2 \frac{1}{P(y_j|x_i)} \\ &= 0, \end{aligned} \tag{6.108}$$

poiché $1 \log_2 1 = 0$ e $0 \log_2 \frac{1}{0} = 0$. Quindi la mutua informazione per il canale deterministico è semplicemente

$$I(X;Y) = \mathcal{H}(Y). \tag{6.109}$$

Si noti la dualità del risultato con quello ottenuto per il canale senza rumore.

polu. 31

Esempio 6.8 Calcoliamo le mutua informazione per il canale perfetto con cancellazione avente alfabeto binario all'ingresso ($n = 2$). Si assuma che i due simboli di ingresso abbiano probabilità $P(x_1) = \pi$ e $P(x_2) = 1 - \pi$. Dalla matrice di canale abbiamo

$$\begin{aligned}
\mathcal{H}(Y|X) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j|x_i)P(x_i)\log_2\frac{1}{P(y_j|x_i)} \\
&= (1-p_c)\pi\log_2\frac{1}{1-p_c} + p_c\pi\log_2\frac{1}{p_c} \\
&\quad +(1-p_c)(1-\pi)\log_2\frac{1}{1-p_c} + p_c(1-\pi)\log_2\frac{1}{p_c} \\
&= (1-p_c)\log_2\frac{1}{1-p_c} + p_c\log_2\frac{1}{p_c} \\
&= \mathcal{H}(p_c). \qquad (6.110)
\end{aligned}$$

Le probabilità dei simboli di uscita sono

$$\begin{aligned}
P(y_1) &= (1-p_c)\pi, \\
P(y_2) &= (1-p_c)(1-\pi), \\
P(y_3) &= p_c\pi + p_c(1-\pi) = p_c. \qquad (6.111)
\end{aligned}$$

Pertanto l'entropia dell'uscita è

$$\begin{aligned}
\mathcal{H}(Y) &= (1-p_c)\pi\log_2\frac{1}{(1-p_c)\pi} + (1-p_c)(1-\pi)\log_2\frac{1}{(1-p_c)(1-\pi)} \\
&\quad +p_c\log_2\frac{1}{p_c} \\
&= (1-p_c)\pi\log_2\frac{1}{(1-p_c)} + (1-p_c)\pi\log_2\frac{1}{\pi} \\
&\quad +(1-p_c)(1-\pi)\log_2\frac{1}{(1-p_c)} + (1-p_c)(1-\pi)\log_2\frac{1}{(1-\pi)} \\
&\quad +p_c\log_2\frac{1}{p_c} = \ldots \\
&= \mathcal{H}(p_c) + (1-p_c)\mathcal{H}(\pi). \qquad (6.112)
\end{aligned}$$

Pertanto la mutua informazione del canale binario perfetto con cancellazione è

$$I(X;Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X) = (1-p_c)\mathcal{H}(\pi). \qquad (6.113)$$

Seguendo una procedura del tutto simile è possibile dimostrare che la mutua informazione del canale m-ario perfetto con cancellazione è in generale

$$I(X;Y) = (1-p_c)\mathcal{H}(X). \qquad (6.114)$$

6.4.2 La mutua informazione del canale esteso

Il calcolo della mutua informazione del canale esteso è importante per i teoremi sulla codifica di canale che seguiranno. La ipotesi di base sulla natura senza memoria di sorgenti e canali rende il calcolo abbastanza semplice. Ricordiamo come l'entropia di una sorgente senza memoria estesa di ordine N sia

$$\mathcal{H}(X^N) = N\mathcal{H}(X). \tag{6.115}$$

Ricordando la definizione del canale esteso e aiutandosi con la figura 6.8, è immediato osservare che la sequenza di N simboli di sorgente e la sequenza di N simboli di uscita sono mutuamente dipendenti solo ad ogni istante di tempo. Quindi l'ambiguità del canale esteso può essere scritta come

$$\mathcal{H}(X^N|Y^N) = N\mathcal{H}(X|Y). \tag{6.116}$$

Pertanto la mutua informazione del canale esteso per canale e sorgenti senza memoria è semplicemente

$$I(X^N;Y^N) = \mathcal{H}(X^N) - \mathcal{H}(X^N|Y^N) = N\ I(X;Y). \tag{6.117}$$

6.5 Capacità di canale

Il canale discreto, nella sua formalizzazione probabilistica, ha lo scopo di rappresentare il modello equivalente di un sistema di comunicazione che consenta il trasferimento di informazione dall'ingresso all'uscita. Abbiamo visto come la mutua informazione $I(X;Y)$ quantifichi in generale la capacità di fornire informazioni sull'ingresso una volta che sia stata osservata l'uscita (o viceversa). La mutua informazione è quindi l'entità di informazione mediamente scambiata tra ingresso e uscita ed è quindi suscettibile di essere considerata come una misura di *capacità* del canale. Sfortunatamente essa dipende non solo dalla matrice di canale, ma anche dalle probabilità dei simboli d'ingresso. Ne ricordiamo qui di seguito la definizione

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j|x_i)P(x_i)\log_2 \frac{P(y_j|x_i)}{\sum_{i=1}^{n} P(y_j|x_i)P(x_i)}. \tag{6.118}$$

Pertanto la mutua informazione non dipende solo dalle caratteristiche intrinseche del canale, ma anche da come esso viene utilizzato: al variare della distribuzione dei simboli di ingresso, la mutua informazione può variare notevolmente fino ad annullarsi. Per esempio, se la distribuzione di

probabilità sui simboli di ingresso è tale che uno dei simboli ha probabilità pari a uno, abbiamo che $\mathcal{H}(X) = 0$, $\mathcal{H}(X|Y) = 0$ e $I(X;Y) = 0$. Poichè la mutua informazione è sempre maggiore o uguale a zero, il minimo globale è appunto lo zero. Tale situazione di assenza di scambio di informazione tra ingresso e uscita, non è certo la situazione di interesse per l'utilizzo del canale che è tipicamente preposto al trasferimento dell'informazione. La situazione opposta, ovvero quella di massimo scambio di informazione, è quella che riveste maggiore importanza nelle applicazioni. La definizione di capacità di canale riflette proprio questo concetto.

La capacità di un canale con ingresso X e uscita Y, è il valore massimo della mutua informazione ottenuta al variare della distribuzione di probabilità dei simboli d'ingresso:

$$C = \max_{\Pi_x} I(X;Y). \tag{6.119}$$

La capacità di canale, che si misura in bit (o nell'unità relativa alla base del logaritmo usato), diventa quindi una proprietà intrinseca del canale e costituisce una delle definizioni più importanti della teoria dell'informazione. Implicitamente, si intende che la capacità del canale si riferisce ad *un solo utilizzo del canale*, ovvero alla trasmissione di un solo simbolo della sequenza di ingresso. Quindi se il canale è usato una volta ogni T secondi, la capacità può anche essere espressa come C/T e misurata in bit/sec (o nell'unità relativa alla base del logaritmo usato /sec). Vedremo nel prossimo capitolo come la capacità rappresenti il limite teorico superiore alla velocità di informazione per una trasmissione affidabile attraverso il canale. Il secondo teorema di Shannon, che sarà discusso nel prossimo capitolo, sancirà appunto che la capacità di canale costituisce il limite oltre il quale non è possibile ottenere una trasmissione a probabilità di errore trasurabile. Rimandando al prossimo capitolo l'approfondimento di questo punto, esaminiamo ora alcune proprietà della capacità.

Proprietà 6.1 $C \geq 0$.

Prova: Questa proprietà discende direttamente dalla positività della mutua informazione.

Proprietà 6.2 $C \leq \log_2 n$.

Prova: $C = \max I(X;Y) = \max(\mathcal{H}(X) - \mathcal{H}(X|Y)) \leq \max \mathcal{H}(X) = \log_2 n$.

Proprietà 6.3 $C \leq \log_2 m$.

Prova: $C = \max I(X;Y) = \max(\mathcal{H}(Y) - \mathcal{H}(Y|X)) \leq \max \mathcal{H}(Y) = \log_2 m$.

Da notare che nonostante la capacità sia limitata superiormente da $\log_2 n$ e $\log_2 m$, non è detto che questi limiti siano ottenibili. Negli esempi tale proprietà diventerà maggiormente evidente.

E' opportuno comunque menzionare che il calcolo della capacità di canale non è sempre facile da condurre in maniera analitica. Infatti, in molti casi, probabilmente nella maggioranza dei casi, è necessario fare ricorso a delle procedure numeriche per valutare la capacià di un canale con matrice arbitraria. I casi riportati qui di seguito, si riferiscono ad alcuni canali tipici per cui è possibile ottenere una espressione in forma chiusa. Altri esempi saranno suggeriti nei problemi. Un algoritmo per il calcolo della capacità per una generica matrice di canale è la procedura di Arimoto-Blahut la cui discussione viene omessa per brevità. Si rimanda il lettore alla vasta letteratura sull'argomento (vedi Cover e Thomas, 1991, p.367).

Esempio 6.9 Torniamo al canale binario simmetrico (BSC). Ricordiamo come la mutua informazione sia

$$I(X;Y) = \mathcal{H}((1-p)\pi + p(1-\pi)) - \mathcal{H}(p). \tag{6.120}$$

Abbiamo già visto come fissato p, e facendo variare π (la probabilità del primo simbolo di ingresso), abbiamo che il massimo della mutua informazione è pari a $I(X;Y) = 1 - \mathcal{H}(p)$ e si ottiene per $(1-p)\pi + p(1-\pi) = \frac{1}{2}$. Pertanto la capacità del canale binario simmetrico è

$$C = 1 - \mathcal{H}(p). \tag{6.121}$$

Figura 6.11 mostra l'andamento della capacità al variare di p. Si noti come si tratti di una funzione strettamente convessa e che il minimo, pari a zero, corrisponde al canale che massimamente confonde i simboli di ingresso con $p = \frac{1}{2}$. La massima capacità di canale, pari a uno, la si ottiene quendo $p = 0$, ovvero in condizioni di canale perfetto, o quando $p = 1$ ovvero in condizioni ancora di canale perfetto, ma con i simboli scambiati.

Esempio 6.10 Per il canale senza rumore il calcolo della capacità è molto semplice. Poichè la mutua informazione è semplicemente $I(X;Y) = \mathcal{H}(X)$, il massimo si ottiene quando l'entropia dell'ingresso è massima. Ciò avviene quando i simboli sono equiprobabili. Pertanto la capacità del canale senza rumore è

$$C = \max_{\Pi_x} I(X;Y) = \max_{\Pi_x} \mathcal{H}(X) = \log_2 n. \tag{6.122}$$

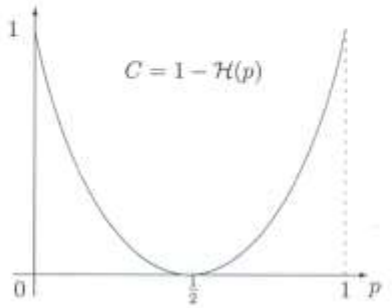


Figura 6.11: La capacità del canale binario simmetrico (BSC).

Il risultato è di facile interpretazione in quanto tale canale è in grado di trasportare tutta l'informazione fornita dalla sorgente.

Esempio 6.11 Per il canale deterministico abbiamo un risultato analogo. Abbiamo già visto come la mutua informazione sia $I(X;Y) = \mathcal{H}(Y)$. Pertanto la massimizzazione della mutua informazione corrisponde alla massimizzazione dell'entropia dell'uscita al variare della distribuzione dei simboli di ingresso. Ciò si ottiene generando una distribuzione dei simboli di ingresso a cui corrisponde una distribuzione uniforme sui simboli di uscita. Per vedere come ciò sia sempre possibile nel canale deterministico si guardi all'esempio numerico di figura 6.7. In tale esempio basta infatti distribuire la probabilità dell'ingresso per un terzo sui primi due simboli, per un terzo sul terzo simbolo e per un terzo sui rimanenti tre simboli. Le proporzioni secondo le quali la probabilità è distribuita all'interno di ogni gruppo di simboli di ingresso è irrilevante. Infatti, se le proporzioni globali sono rispettate i simboli di uscita saranno uniformemente distribuiti. Il ragionamento può essere applicato a qualunque canale deterministico, pertanto in generale la capacità è

$$C = \max_{\Pi_x} I(X;Y) = \max_{\Pi_x} \mathcal{H}(Y) = \log_2 m. \tag{6.123}$$

Esempio 6.12 Come ulteriore esempio, in cui è possibile ottenere una espressione in forma chiusa per la capacità, consideriamo il canale perfetto con cancellazione. Abbiamo già visto come la mutua informazione sia $I(X;Y) = (1 - p_c)\mathcal{H}(X)$. Il valore massimo anche qui è attinto quando l'entropia dell'ingresso è massima, ovvero quando i simboli di ingresso sono equiprobabili. Pertanto la

capacità del canale perfetto con cancellazione è

$$C = \max_{\Pi_x} I(X;Y) = \max_{\Pi_x} (1 - p_c)\mathcal{H}(X) = (1 - p_c)\log_2 n. \tag{6.124}$$

Il risultato è suscettibile di una interpretazione intuitiva: la massima informazione trasferibile dal canale è quella massima emessa della sorgente ($\log_2 n$), meno la frazione persa.

7.1.2 Noisy Channel with Nonoverlapping Outputs

This channel has two possible outputs corresponding to each of the two inputs (Figure 7.3). The channel appears to be noisy, but really is not. Even though the output of the channel is a random consequence of the input, the input can be determined from the output, and hence every transmitted bit can be recovered without error. The capacity of this channel is also 1 bit per transmission. We can also calculate the information capacity $C = \max I(X;Y) = 1$ bit, which is achieved by using $p(x) = (\frac{1}{2}, \frac{1}{2})$.

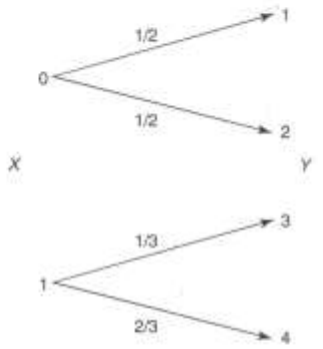


FIGURE 7.3. Noisy channel with nonoverlapping outputs. $C = 1$ bit.

6.5.1 Il canale simmetrico e la sua capacità

Un tipo di canale per cui è possibile ottenere una espressione per la capacità in forma chiusa, è il cosidetto *canale simmetrico*. Un canale simmetrico ha una matrice di canale le cui righe e colonne sono tutte permutazioni l'una dell'altra. Tale canale ha $m = n$ e un esempio per $n = 4$ è dato dalla matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} p_1 & p_2 & p_3 & (1-p_1-p_2-p_3) \\ p_2 & (1-p_1-p_2-p_3) & p_1 & p_3 \\ p_3 & p_1 & (1-p_1-p_2-p_3) & p_2 \\ (1-p_1-p_2-p_3) & p_3 & p_2 & p_1 \end{bmatrix}. \tag{6.125}$$

Si noti che tale matrice è tale che anche la somma delle colonne è pari a uno (le matrici con tale proprietà sono dette matrici doppiamente stocastiche). Definendo $\mathbf{p}$ il vettore delle probabilità contenute nella prima riga della matrice, la entropia condizionata si scrive come

$$\begin{aligned} \mathcal{H}(Y|X) &= \sum_{i=1}^{n} P(x_i) \sum_{j=1}^{n} P(y_j|x_i) \log_2 \frac{1}{P(y_j|x_i)} \\ &= \sum_{i=1}^{n} P(x_i) \mathcal{H}(\mathbf{p}) \\ &= \mathcal{H}(\mathbf{p}). \end{aligned} \tag{6.126}$$

Pertanto la mutua informazione diventa

$$I(X;Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X) = \mathcal{H}(Y) - \mathcal{H}(\mathbf{p}). \tag{6.127}$$

La massimizzazione della mutua informazione richiede che si massimizzi $\mathcal{H}(Y)$ rispetto alle distribuzioni dell'ingresso. Il massimo valore che $\mathcal{H}(Y)$ può assumere è $\log_2 n$ e corrisponde alla distribuzione uniforme. Per usare questo valore come il massimo nel calcolo della capacità è necessario però

accertarsi che esista una distribuzione sui simboli di ingresso che generi una distribuzione uniforme sulle uscite. Ciò in generale non è garantito per qualunque canale, ma nel caso del canale simmetrico basta imporre una distribuzione uniforme sull'ingresso per ottenerne una uniforme sull'uscita

$$P(y_j) = \sum_{i=1}^{n} P(y_j|x_i)P(x_i) = \frac{1}{n}\sum_{i=1}^{n} P(y_j|x_i) = \frac{1}{n}, \quad j = 1, ..., n. \tag{6.128}$$

Pertanto la capacità del canale simmetrico è

$$C = \log_2 n - \mathcal{H}(\mathbf{p}). \tag{6.129}$$

Il canale esaminato è denotato cone nSC (n-ary Symmetric Channel) e si noti come siano casi particolari del canale simmetrico: il canale binario simmetrico (BSC), il canale massimamente confuso, il canale sempre in errore.

Esiste una estensione interessante del risultato sul canale simmetrico in riferimento al cosidetto *canale debolmente simmetrico*. Un tale canale, che può anche non essere quadrato, è tale che tutte le righe della matrice di canale sono una permutazione dell'altra (come nel canale simmetrico), ma con il solo vincolo sulle colonne che esse sommino allo stesso valore. Un esempio numerico è dato dalla matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{bmatrix}. \tag{6.130}$$

Il canale dobolmente simmetrico è definito proprio mediante le due proprietà che ci consentono di ottenere la capacità (6.129). Pertanto la capacità del canale debolmente simmetrico è ancora (6.129).

6.5.2 La capacità del canale uniforme

Un caso particolare del canale simmetrico è il *canale uniforme* avente matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} (1-p_e) & \frac{p_e}{n-1} & \frac{p_e}{n-1} & \cdots & \frac{p_e}{n-1} \\ \frac{p_e}{n-1} & (1-p_e) & \frac{p_e}{n-1} & \cdots & \frac{p_e}{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \frac{p_e}{n-1} & \frac{p_e}{n-1} & \frac{p_e}{n-1} & \cdots & (1-p_e) \end{bmatrix}, \tag{6.131}$$

dove p_e è proprio la probabilità di errore. In tal caso gli errori sono distribuiti in maniera uniforme su tutti i simboli fuori diagonale. Ricordiamo che

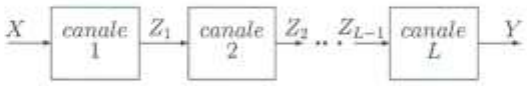


Figura 6.12: La cascata di L canali discreti.

anche il BSC è un caso particolare di canale uniforme. Si noti inoltre che la probabilità di errore all'uscita del canale uniforme non dipende dalla distribuzione dei simboli d'ingresso. Per un tale canale abbiamo

$$\begin{aligned}\mathcal{H}(\mathbf{p}) &= (1-p_e)\log_2\frac{1}{1-p_e} + (n-1)\frac{p_e}{n-1}\log_2\frac{n-1}{p_e} && (6.132)\\ &= (1-p_e)\log_2\frac{1}{1-p_e} + p_e\log_2\frac{n-1}{p_e} && (6.133)\\ &= \mathcal{H}(p_e) + p_e\log_2(n-1). && (6.134)\end{aligned}$$

Pertanto la capacità del canale uniforme è

$$C = \log_2 n - \mathcal{H}(p_e) - p_e\log_2(n-1). \qquad (6.135)$$

6.6 La capacità del canale esteso

La valutazione della capacità C^N del canale esteso è immediata. Infatti, visto che la mutua informazione per un canale esteso senza memoria è $I(X^N;Y^N) = N\ I(X;Y)$, abbiamo che la capacità è semplicemente

$$C^N = \max_{\Pi_x^N} I(X^N;Y^N) = \max_{\Pi_x} N I(X;Y) = N\ C. \qquad (6.136)$$

6.7 Canali in cascata

Esaminiamo ora il comportamento di un canale discreto composto dalla cascata di più canali discreti. La figura 6.12 rappresenta la cascata di L canali discreti indipendenti dove la generica uscita Z_l del canale l-esimo rappresenta l'ingresso del canale $(l+1)$-esimo. Ovviamente le cardinalità degli alfabeti degli elementi della catena devono essere tali che la dimensionalità dell'alfabeto di uscita di ogni canale sia uguale a quella dell'ingresso del canale successivo. Il funzionamento della cascata è tale che ad un simbolo

di sorgente il primo canale associa un simbolo di uscita secondo il suo meccanismo aleatorio. Tale simbolo è poi soggetto ad una nuova trasformazione probabilistica nel canale successivo. Quindi la struttura a cascata implica che

$$\begin{aligned} &Pr\{Y = y_j | X = x_i, Z_1 = z_{1i}, ..., Z_{L-1} = z_{(L-1)i}\} \\ &= Pr\{Y = y_j | Z_{L-1} = z_{(L-1)i}\}. \end{aligned} \tag{6.137}$$

Analogamente per le probabilità a posteriori dell'ingresso

$$\begin{aligned} &Pr\{X = x_i | Z_1 = z_{1i}, ..., Z_{L-1} = z_{(L-1)i}, Y = y_j\} \\ &= Pr\{X = x_i | Z_1 = z_{1i}\}. \end{aligned} \tag{6.138}$$

Utilizzando le relazioni matriciali, abbiamo che le probabilità dei simboli in uscita sono

$$\boldsymbol{\pi}_y = \mathbf{P}_{cL}^T \cdots \mathbf{P}_{c1}^T \boldsymbol{\pi}_x. \tag{6.139}$$

Quindi la matrice di canale complessiva è

$$\mathbf{P}_c = \mathbf{P}_{c1} \cdots \mathbf{P}_{cL}. \tag{6.140}$$

La matrice delle probabilità a posteriori globale, dall'equazione (6.13), diventa

$$\mathbf{P}_p = diag(\boldsymbol{\pi}_x)\mathbf{P}_{c1} \cdots \mathbf{P}_{cL}(diag(\mathbf{P}_{cL}^T \cdots \mathbf{P}_{c1}^T \boldsymbol{\pi}_x))^{-1}. \tag{6.141}$$

Il calcolo della mutua informazione e della capacità di canale si esegue direttamente (generalmente per via numerica) dalla conoscenza della matrice di canale complessiva $\mathbf{P}_c$.

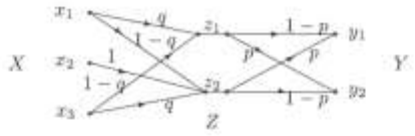


Figura 6.13: Un esempio di cascata di due canali discreti.

Esempio 6.13 La cascata di due canali discreti è mostrata in figura 6.13. I due canali hanno matrici

$$\mathbf{P}_{c1} = \begin{bmatrix} q & 1-q \\ 0 & 1 \\ 1-q & q \end{bmatrix}, \quad \mathbf{P}_{c2} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}. \tag{6.142}$$

Il canale discreto equivalente ha matrice di canale

$$\mathbf{P}_c = \begin{bmatrix} q(1-p)+(1-q)p & pq+(1-p)(1-q) \\ p & 1-p \\ (1-p)(1-q)+pq & p(1-q)+q(1-p) \end{bmatrix}. \tag{6.143}$$

Per canali semplici e cascate di limitata estensione, le probabilità di transizione del canale complessivo possono essere valutate graficamente una per una dal grafo relativo alla cascata: identifica la coppia ingresso-uscita, somma i prodotti relativi a tutti i possibili percorsi nella cascata tra i due terminali.

Esempio 6.14 Consideriamo ora la cascata di tre BSC. Ognuno ha probabilità di errore condizionata pari a p. La cascata è mostrata in figura 6.14. Le cascate di

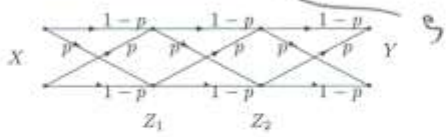


Figura 6.14: Un esempio di cascata di tre BSC

due e di tre BSC sono entrambe ancora BSC. La probabilità di errore condizionata per un singolo BSC è p. Per due BSC, semplicemente guardando il grafo e sommando i contributi dei due percorsi è $2(1-p)p$. Analogamente per tre BSC, guardando al grafo relativo alla cascata del BSC equivalente a due BSC, e un altro BSC, abbiamo

$$2p(1-p)^2 + p(1-2p(1-p)) = 4p^3 - 6p^2 + 3p. \tag{6.144}$$

Le mutue informazioni sono pertanto

$$I(X;Z_1) = 1-\mathcal{H}(p), \;\; I(X;Z_2) = 1-\mathcal{H}(2(1-p)p), \;\; I(X;Y) = 1-\mathcal{H}(4p^3-6p^2+3p). \tag{6.145}$$

Si noti che grazie alla simmetria del canale risultante, le mutue informazioni non dipendono dalla distribuzione dei simboli d'ingresso. Quindi l'espressione della mutua informazione della cascata è anche una espressione per la capacità del canale risultante.

polin.42

6.7.1 Il teorema del trattamento dati

Nello studiare la cascata di più canali discreti emerge l'intuizione che in qualche modo il livello di contaminazione dell'informazione proveniente dalla sorgente sia proporzionale al numero di stadi della catena. Infatti andando ad analizzare la mutua informazione e la capacità del canale composto si può dimostrare che queste non possone che diminuire all'aumentare del numero di elementi della cascata. Per fissare le idee, consideriamo la cascata di due canali di informazione c1 e c2, rappresentati schematicamente come

$$X \xrightarrow{c1} Z \xrightarrow{c2} Y. \tag{6.146}$$

Le tre variabili e gli alfabeti relativi a ingresso, uscita del primo stadio, e uscita del secondo stadio sono

$$X \in \mathcal{X} = \{x_1, ..., x_n\}, \quad Z \in \mathcal{Z} = \{z_1, ..., z_r\} \quad Y \in \mathcal{Y} = \{y_1, ..., y_m\}. \tag{6.147}$$

Vogliamo ora confrontare la ambiguità del canale composto, con quella del primo stadio della catena.

Teorema 6.1 (Teorema del trattamento dati). *Data la cascata di due canali discreti, e siano X,Z e Y rispettivamente l'ingresso del primo canale, l'ingresso al secondo canale e l'uscita della cascata,*

$$I(X;Z) \geq I(X;Y). \tag{6.148}$$

Il teorema esprime il fatto che la informazione mediamente scambiata ai capi della catena è inferiore a quella scambiata ai capi di un solo blocco. In altre parole, l'informazione nel passare nei vari stadi della catena tende a perdersi. Da qui il nome del teorema (*data processing theorem*).

Prova: Consideriamo la differenza tra le due ambiguità

$$\begin{aligned}
\mathcal{H}(X|Y) - \mathcal{H}(X|Z) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(x_j|y_j)} \\
&\quad - \sum_{i=1}^{n}\sum_{l=1}^{r} P(x_i, z_l) \log_2 \frac{1}{P(x_i|z_l)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{r} P(x_i, y_j, z_l) \log_2 \frac{1}{P(x_i|y_j)} \\
&\quad - \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{r} P(x_i, y_j, z_l) \log_2 \frac{1}{P(x_i|z_l)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{r} P(x_i, y_j, z_l) \log_2 \frac{P(x_i|z_l)}{P(x_i|y_j)}.
\end{aligned} \tag{6.149}$$

Usando la regola a catena e l'indipendenza condizionata pelm. 63

$$p(x_i, y_j, z_\ell) = p(y_j | z_\ell x_i)\, p(z_\ell | x_i)\, p(x_i)$$
$$= p(y_j | z_\ell)\, p(z_\ell | x_i)\, p(x_i)$$

Ma $p(z_\ell | x_i) = \dfrac{p(x_i | z_\ell)\, p(z_\ell)}{p(x_i)}$

Quindi

$$p(x_i, y_j, z_\ell) = p(y_j | z_\ell) \frac{p(x_i | z_\ell)\, p(z_\ell)}{\cancel{p(x_i)}} \cancel{p(x_i)}$$

che sostituito in (6.149) fornisce

$$H(X|Y) - H(X|Z) = \sum_\ell \sum_j p(y_j | z_\ell)\, p(z_\ell) \sum_i p(x_i | z_\ell) \log \frac{p(x_i | z_\ell)}{p(x_i | y_j)}$$

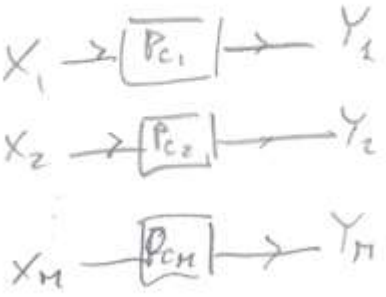
divergenza tra $p(x_i|z_\ell)$ e $p(x_i|y_j)$ vale sempre

$$\sum_p p \log \frac{p}{q} = D(p||q) \geq 0$$

≥ 0

□

CANALI IN PARALLELO

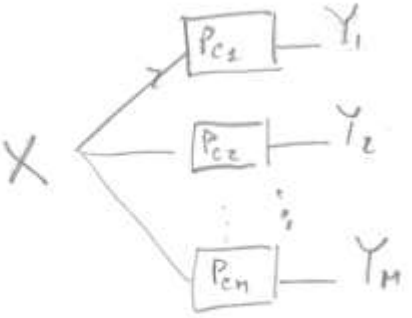


$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_M$$

$$\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \dots \times \mathcal{Y}_M$$

$$P_C = P_{C_1} \otimes P_{C_2} \otimes \dots \otimes P_{C_M}$$

CANALE BROADCAST



$$\mathcal{Y} = \underbrace{\mathcal{Y}_1}_{M_1} \times \underbrace{\mathcal{Y}_2}_{M_2} \times \dots \times \underbrace{\mathcal{Y}_M}_{M_M}$$

$$\mathcal{X}$$

$|\mathcal{X}| = 2$ ...

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} (P_{C_1} \otimes P_{C_2} \otimes \dots \otimes P_{C_M})$$

....

6.8 Problemi

Problema 6.1 Un canale binario, che introduce solo cancellazioni con probabilità $p_c = 0.01$, ha probabilità di cancellazione condizionate rispetto ai due simboli di ingresso uguali. I simboli di ingresso hanno distribuzione $\{0.3, 0.7\}$.

(a) Valutare esplicitamente la matrice di canale;

(b) Calcolare la mutua informazione ingresso-uscita;

(c) Valutare la capacità di canale.

Problema 6.2 Un canale (moltiplicativo) ha all'ingresso due sorgenti binarie indipendenti $X \in \{0,1\}$ e $Y \in \{0,1\}$. La distribuzione dei simboli per le due sorgenti sono uguali e pari a $\{p, 1-p\} = \{0.3, 0.7\}$. L'uscita è il simbolo binario $Z = X \cdot Y$ (il prodotto va inteso come prodotto aritmetico).

(a) Valutare la matrice del canale che ha all'ingresso la sorgente composta $C = (X, Y)$ costituita dall'insieme delle due sorgenti, e all'uscita Z;

(b) Valutare la matrice delle probabilità a posteriori e suggerire uno schema di decodifica;

(c) Valutare la matrice del canale equivalente che ha all'ingresso X e all'uscita Z.

Problema 6.3 Un modulatore numerico ha al suo ingresso la variabile aleatoria binaria $X \in \{0,1\}$. Dopo la modulazione e la trasmissione sul canale fisico, nel ricevitore all'ingresso del rivelatore a soglia si presenta una variabile R tale che:

$$R = \begin{cases} A + N & \text{se X} = 1 \\ -A + N & \text{se X} = 0, \end{cases} \tag{6.155}$$

dove A è una costante positiva e N è una variabile aleatoria gaussiana a media nulla e varianza σ^2. Il rivelatore a soglia "decide" per il simbolo Y ricevuto come segue:

$$Y = \begin{cases} 1 & \text{se } R > 0 \\ 0 & \text{se } R < 0 \end{cases} \tag{6.156}$$

Valutare la matrice del canale $X \to Y$ equivalente.

Problema 6.4 Ripetere il problema 6.3 per un rivelatore a soglia (multipla) che decide come segue:

$$Y = \begin{cases} 1 & \text{se } R > \frac{A}{5} \\ * \text{ (cancellazione)} & \text{se } -\frac{A}{5} < R < \frac{A}{5} \\ 0 & \text{se } R < -\frac{A}{5} \end{cases}. \tag{6.157}$$

Problema 6.5 Un canale ha all'ingresso la variabile binaria $X \in \{0,1\}$ e all'uscita la variabile $Y = X + N$, dove N è un'altra variabile binaria con $N \in \{0,1\}$. La somma é da intendersi come somma aritmetica e le due variabili X e N sono indipendenti e a distribuzione uniforme. Valutare la matrice del canale equivalente $X \to Y$.

Problema 6.6 Valutare la matrice di un canale costituito da due BSC in cascata aventi probabilità di errore rispettivamente p_1 e p_2.

PROBLEMA
VALUTARE LA MATRICE DI CANALE EQUIVALENTE
DI UNA CASCATA DI N BSC E MOSTRARE COSA
SUCCEDE QUANDO $N \to \infty$