La capacità dei canali asimmetrici ed unidirezionali

Luca G. Tallini*

Sunto: In questo lavoro dopo una presentazione sintetica dei concetti fondamentali della Teoria dell'Informazione di Shannon, si trova un espressione esplicita per la capacità del canale asimmetrico e delle buone limitazioni per la capacità del canale unidirezionale di lunghezza n.

Abstract: In this paper, we first give a sinthetic presentation of Shannon's Information Theory and then we find an explicit expression for the capacity of the asimmetric channel. Further, we find some good bounds for the capacity of the unidirectional channel of length n.

Parole Chiave: Teoria di Shannon, canali di trasmissione, entropia, mutua informazione, capacità.

1. Introduzione

In un canale di trasmissione binario asimmetrico tutti gli errori di trasmissione sono sempre del tipo: $1 \to 0$ (o $0 \to 1$). In un canale unidirezionale sia errori del tipo $1 \to 0$ che errori del tipo $0 \to 1$ possono avvenire, ma per ogni particolare parola di lunghezza n trasmessa tutti gli errori sono dello stesso tipo. Molti

*Dipartimento Di. Tec., Politecnico di Milano, Via Bonardi, 3, 20133 Milano, ITALY. E-mail: luca.tallini@polimi.it

sono i canali fisici che rientrano nel modello dei canali asimmetrici o unidirezionali, tra cui: fibre ottiche, dischi ottici, circuiti e memorie VLSI e Read Only Memories [1]. In questo lavoro, dopo una presentazione sintetica dei concetti fondamentali della Teoria dell'informazione di Shannon, si trovano per la prima volta delle espressioni esplicite per la capacità del canale asimmetrico, C_{Z}, e del canale unidirezionale, $C_{\mathrm{U}(n)}$, di lunghezza n.
In particolare, nella Sezione 2. si presenta sinteticamente la Teoria di Shannon, introducendo i concetti di sorgente di informazione, informazione, entropia, canale di trasmissione, mutua informazione, capacità, eccetera. Sia ϵ la probabilità di errore per bit trasmesso. Nella Sezione 3., si dimostra che

$$C_{\mathrm{Z}}(\epsilon) = \log_2\left(1 + \epsilon^{\epsilon/(1-\epsilon)} - \epsilon^{1/(1-\epsilon)}\right) = \log_2\left(1 + 2^{-h(\epsilon)/(1-\epsilon)}\right),$$

dove $h(x) \stackrel{\mathrm{def}}{=} -x\log_2 x - (1-x)\log_2(1-x)$ è la funzione entropia ddi Shannon. Per finire, nella Sezione 4., si dimostra che

$$C_{\mathrm{Z}}(\epsilon) - \frac{1}{n} \leq C_{\mathrm{U}(n)}(\epsilon) \leq C_{\mathrm{Z}}(\epsilon). \qquad (1)$$

2. Background di Teoria dell'Informazione

L'informatica teorica è la scienza che si occupa del problema di come codificare efficientemente l'informazione, essendo quest'ultima un oggetto definito matematicamente e quindi senza ambiguità logica da Claude Elwood Shannon in un articolo del 1948 [4]. Ovviamente la definizione rigorosa di informazione data da Shannon traduce ciò che comunemente noi tutti intendiamo per informazione.

2.1. La codifica

Sia A un insieme finito e non vuoto, che chiameremo alfabeto. Posto:

$$A^n \stackrel{\mathrm{def}}{=} \{a_1 a_2 \ldots a_n : \quad a_i \in A, \quad \forall i = 1, 2, \ldots, n\}, \quad n \in \mathbb{N}$$

un elemento di A^n si dice parola di lunghezza n sull'alfabeto A, mentre un codice $\mathcal{C}_A$ su A non è altro che un sottoinsieme di $A^+ \stackrel{\text{def}}{=} \bigcup_{n=1}^{\infty} A^n$. Gli elementi di $\mathcal{C}_A$ costituiscono le parole del codice $\mathcal{C}_A$. Un codice è detto a blocchi di lunghezza n se $\mathcal{C}_A \subseteq A^n$, altrimenti è detto a lunghezza variabile.
Dato un insieme S, la cui cardinalità sia minore od uguale a quella del numerabile, una codifica di S è una coppia $(\mathcal{C}_A, \chi)$, dove $\mathcal{C}_A$ è un codice e χ è un'applicazione da S in $\mathcal{C}_A$ iniettiva, detta funzione di codifica.

2.2. La sorgente di informazione

Una sorgente di informazione finita non è altro che un sistema con un numero finito di stati, ognuno codificato dal simbolo di un alfabeto finito A, che assume ogni stato con una certa assegnata probabilità. È quindi possibile schematizzare matematicamente una sorgente mediante una "variabile aleatoria" $X \in A$, essendo A un alfabeto finito, uguale ad esempio a $\mathbf{Z}\!\mathbf{Z}_n = \{0, \ldots, n-1\}$, caratterizzata da una distribuzione di probabilità $\{P(X = x)\}_{x=0,\ldots,n-1}$. D'ora in poi parlando di sorgente di informazione sottintenderemo che è finita.
Un esempio di sorgente di informazione è l'esito del lancio di una moneta non truccata: esso è una variabile aleatoria

$$MONETA \in \{Testa, Croce\},$$

la cui distribuzione di probabilità è:

$$P(MONETA = Testa) = 1/2$$

e

$$P(MONETA = Croce) = 1/2.$$

Un altro esempio di sorgente di informazione è una persona che parla italiano: essa è una variabile aleatoria

$$PERSONA \in \{a, b, c, d, e, \ldots\}$$

caratterizzata da una certa distribuzione di probabilità: quella delle frequenze con cui vengono emesse le varie lettere dell'alfabeto italiano nella lingua parlata.
Data una sorgente di informazione $X \in \mathbb{Z}_n$, caratterizata da una certa distribuzione di probabilità $\{P(X = x)\}_{x=0,\ldots,n-1}$, posto per brevità $P(x) \stackrel{\text{def}}{=} P(X = x)$, si definisce informazione associata al simbolo x la quantità:

$$I(x) \stackrel{\text{def}}{=} \log_2 \frac{1}{P(x)} = -\log_2 P(x).$$

$I(x)$ misura le seguenti cose, tra loro equivalenti:

1. la quantità di informazione fornita dall'occorrenza dell'evento $\{X = x\}$,
2. la nostra incertezza sull'evento $\{X = x\}$,
3. l'aleatorietà dell'evento $\{X = x\}$.

La funzione composta

$$I(X) \in \left\{ \log_2 \frac{1}{P(x)} \right\}_{x=0,\ldots,n-1}$$

sarà quindi una variabile aleatoria avente la stessa distribuzione di X.
Si definisce entropia della sorgente X la media della variabile aleatoria $I(X)$ e si indica con $H(X)$. Esplicitamente:

$$H(X) \stackrel{\text{def}}{=} \sum_{x=0}^{n-1} P(x) I(x) = \sum_{x=0}^{n-1} P(x) \log_2 \frac{1}{P(x)}.$$

$H(X)$ può quindi essere pensata come una misura delle seguenti cose, tra loro equivalenti, su X:

1. la quantità di informazione che in media la sorgente X fornisce,
2. la nostra incertezza su X,

3. l'aleatorietà di X.

Si dimostra che $H(X)$ è una funzione convessa $\cap$ di

$(P(0), P(1), \ldots, P(n-1))$

tale che

$$\max_X H(X)$$

$$= \max_{P(x):P(x)\geq 0, \sum_x P(x)=1} H(X)(P(0), P(1), \ldots, P(n-1))$$

$$= H(X)(1/n, 1/n, \ldots, 1/n).$$

Notiamo che se $\{X = x\}$ è un evento eccezionale, per cui cioè $P(x) \simeq 0$, allora $I(x) \simeq +\infty$, mentre se $\{X = x\}$ è un evento quasi certo, per cui cioè $P(x) \simeq 1$, allora $I(x) \simeq 0$.
Nella definizione di I la base del logaritmo è 2 per convenzione; vorrà dire che I sarà misurata in *bit* (binary digits): $1bit$ è, per definizione, la quantità di informazione che si ricava osservando l'esito del lancio di una moneta non truccata.
Come si può vedere I e H traducono matematicamente i concetti intuitivi di informazione e di incertezza. Cosicché, un evento che occorre con probabilità 1 non fornisce informazione, mentre un evento poco probabile e quindi non atteso fornisce una grande quantità di informazione. Ad esempio supponiamo il Sig. Rossi si rivolga ad un *Oracolo* e gli chieda se vivrà piú di 150 anni. Se l'*Oracolo* gli risponderà negativamente il Sig. Rossi riceverà poca informazione, dal momento che una tale longevità è estremamente improbabile. Mentre se l'*Oracolo* gli risponderà affermativamente, l'informazione ricevuta dal Sig. Rossi sarà grandissima.
È pertanto chiaro che $I(x)$ deve essere una funzione di x tramite $P(x)$ tale che se $P(x) \simeq 0$, allora $I(x) \simeq +\infty$, mentre se $P(x) \simeq 1$, allora $I(x) \simeq 0$. Ovviamente $f(t) = \log_2(\frac{1}{t})$ non è l'unica funzione che gode delle proprietà su dette, ma è la piú comoda.
Il I Teorema di Shannon sulla codifica di sorgente afferma che non è possibile rappresentare statisticamente una sorgente $X \in A$

in maniera efficiente codificandola (ovvero codificando $A =$ codominio di X) in modo che la lunghezza media della codifica, $\overline{l}$, sia minore dell'entropia della sorgente, H [4], [2]. Se pero si richiede che $\overline{l}$ sia maggiore di H almeno una codifica efficiente esiste. Quindi, per quanto riguarda la codifica di sorgente, il problema fondamentale della Teoria dell'informazione è un problema di compressione dati; ovvero, è quello di codificare la sequenza di simboli emessa dalla sorgente in modo che $\overline{l}$ sia la più piccola possibile e la più vicina possibile all'entropia H di tale sorgente. Si noti che quando ciò non è possibile si ha che $\overline{l} = H$ (come dopo un ottimo algoritmo di compressione) e quindi si ha che l'entropia è massima. ecco perché la sequenza di simboli che scaturisce da un algoritmo di compressione efficiente è caratteristica di una sorgente ad entropia massima in cui tutti i simboli sono equiprobabili (che nel caso binario coincide con una sorgente che emette 0 con probabilità 1/2 ed 1 con probabilità 1/2).

2.3. Il canale di trasmissione di informazione

Un canale di trasmissione finito e senza memoria è una coppia di sistemi: *input* e *output*, ognuno dei quali assume un numero finito di stati. Ogni stato dell'*input* è codificato da un simbolo di un alfabeto finito A_i e ogni stato dell'*output* è codificato da un simbolo di un alfabeto finito A_o, ed inoltre, se l'*input* è in un certo stato $x \in A_i$ allora l'*output* assume ogni stato di A_o con una certa probabilità dipendente esclusivamente da x. Di solito si pone $A_i = \mathbf{Z}\!\!\mathbf{Z}_n$ e $A_o = \mathbf{Z}\!\!\mathbf{Z}_m$. D'ora in poi quando parleremo di canali sottintenderemo che sono finiti e senza memoria.
È possibile schematizzare matematicamente un canale mediante una matrice stocastica (ovvero una matrice, a valori reali positivi, la cui somma degli elementi di ogni riga è pari ad 1)

$$P_c = (P_c(y|x))_{x \in \mathbf{Z}\!\!\mathbf{Z}_n, y \in \mathbf{Z}\!\!\mathbf{Z}_m} ,$$

detta matrice delle probabilità di transizione, il cui generico elemento $P(y|x) = P_c(y|x)$ rappresenta la probabilità che l'*output* sia il simbolo y dato che l'*input* è stato il simbolo x. Dato un canale,

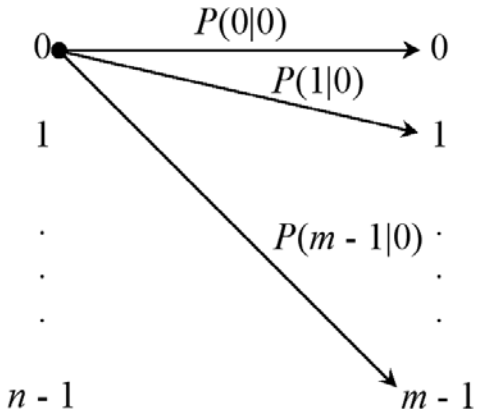

Figura 1: diagramma caratteristico di un canale.

la cui matrice delle probabilità di transizione è P_c, lo si può rappresentare mediante il diagramma in Figura 1. Dati $x \in \mathbb{Z}_n$ e $y \in \mathbb{Z}_m$, se $P(y|x) = 0$ sottintenderemo omessa dal diagramma in questione la freccia che da x va a y. D'ora in poi quando prenderemo in considerazione un canale, lo identificheremo con la sua matrice delle probabilità di transizione P_c. Si noti che poichè il canale è senza memoria, se $\mathbf{X} = x_1, x_2, \ldots, x_l \in \mathbb{Z}_n^l$, è una sequenza di simboli di input al canale e $\mathbf{Y} = y_1, y_2, \ldots, y_l \in \mathbb{Z}_m^l$ è la corrispondente sequenza di simboli di output, allora

$$P_{c^l}(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^{l} P_c(y_i|x_i).$$

Invero, un canale è senza memoria se esiste un'assegnazione di probabilità $P_c(y|x)$ tale che la su scritta equazione è vera per ogni $l \in \mathbb{N}$, per ogni $\mathbf{X}$ e per ogni $\mathbf{Y}$.
Denoteremo con $\mathcal{P}(n \times m)$ l'insieme delle matrici stocastiche con n righe e m colonne. Data una variabile aleatoria $X \in \mathbb{Z}_n$ ed un canale $P_c \in \mathcal{P}(n \times m)$, rimane definita un'altra variabile aleatoria

$Y \in \mathbf{Z}\!\!\mathbf{Z}_m$, usualmente chiamata versione erronea di X tramite il canale P_c, la cui distribuzione delle probabilità è definita da

$$P(y) = \sum_{x \in Z_n} P(y|x)P(x), \qquad y \in \mathbf{Z}\!\!\mathbf{Z}_m.$$

Si definiscono allora le seguenti quantità:

1. Entropia condizionata di X dato y:

   $$H(X|y) \stackrel{\text{def}}{=} \sum_x P(x|y) \log_2 \frac{1}{P(x|y)},$$

   che rappresenta la quantità di incertezza rimasta su X dopo che è stato osservato l'evento $\{Y = y\}$,

2. Entropia condizionata di X dato Y:

   $$H(X|Y) \stackrel{\text{def}}{=} \sum_y P(y)H(X|y) = \sum_{x,y} P(x,y) \log_2 \frac{1}{P(x|y)},$$

   che rappresenta la quantità di incertezza rimasta su X dopo che è stato osservato Y,

3. Mutua informazione tra X e Y:

   $$I(X;Y) \stackrel{\text{def}}{=} H(X) - H(X|Y) = \sum_{x,y} P(x,y) \log_2 \frac{P(x,y)}{P(x)P(y)}$$

   $$= I(Y;X) = H(Y) - H(Y|X) \geq 0,$$

   che rappresenta la quantità di incertezza su X risolta da Y, ovvero la quantità di informazione fornita da Y su X,

Dato un canale $P_c \in \mathcal{P}(n \times m)$ si definisce capacità del canale la seguente quantità:

$$C \stackrel{\text{def}}{=} \max_X I(Y;X) = \max_{P(x):P(x)\geq 0, \sum_x P(x)=1} I(X;Y),$$

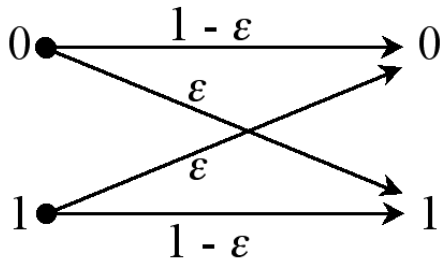

Figura 2: diagramma caratteristico di un canale binario simmetrico. ϵ è la probabilità di errore.

essa rapresenta la quantità massima di informazione media che può passare nel canale per simbolo trasmesso. Invero, ciò è quanto stabilito dal II Teorema di Shannon sulla codifica di canale [4], [2]. È quindi chiaro che, dato un canale di trasmissione, è molto importante conoscere la sua capacità. Essa da un limite teorico sulla quantità massima di informazione che è possibile trasmettere sul canale.
Ad esempio, consideriamo il classico canale binario simmetrico ($n = m = 2$). Esso è caratterizzato dalla seguente matrice delle probabilità di transizione; in cui ϵ è la probabilità di errore.

$$P_{\mathrm{BSC}} = \begin{pmatrix} P(0|0) & P(1|0) \\ P(0|1) & P(1|1) \end{pmatrix} = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}.$$

Tale canale è rappresentato dal diagramma in Figura 2 Si dimostra che la capacità di tale canale è [2]:

$C_{\mathrm{BSC}}(\epsilon) = 1 - h(\epsilon)$,

dove $h : [0,1] \to \mathbf{IR}$ è la funzione entropia di Shannon definita da

$h(x) = -[x \log_2 x + (1-x) \log_2 (1-x)]$.

Ad esempio, se $\epsilon = 0,001$ (ogni 1000 bits trasmessi avviene un errore), allora la capacità del canale binario simmetrico è

$C_{\mathrm{BSC}}(0,001) = 0,9885\ldots$

e quindi, per il Teorema di Shannon sulla codifica di canale, è possibile trasmettere in maniera efficiente al più $0,9885\ldots$ bits per uso del canale.
Per quanto riguarda la codifica di canale, il problema fondamentale della Teoria dell'informazione, è quello di codificare la sequenza di simboli fornita da una sorgente di informazione ad entropia massima (ad esempio l'output di un algoritmo di compressione) immettendo della ridondanza controllata, in modo da ottenere la capacità del canale. Ovviamente la ridondanza da immettere deve essere la minima possibile ma sicuramente non può essere minore di quella individuata dal canale. Nell'esempio su riportato, della sequenza di simboli trasmessi, il $0,9885\ldots\%$ deve essere di simboli emessi dalla sorgente (informazione pura) ed almeno il $1 - C_{\mathrm{BSC}}(0,001)\% = 0,0114\ldots\%$ deve essere di simboli aggiunti in maniera opportuna (ridondanza). Con un argomento che va sotto il nome della tecnica del "random coding", il II Teorema di Shannon afferma che in teoria è sempre possibile "raggiungere" la capacità del canale (è essenzialmente un Teorema di esistenza). Come si possa fare ciò in pratica è un problema molto difficile studiato dalla teoria della correzione degli errori.
Riassumendo, "Il problema fondamentale della comunicazione è quello di riprodurre in un punto esattamente o approssimativamente un messaggio scelto in un altro punto" [4]. I due teoremi di Shannon fanno si che la comunicazione può essere effettuata in maniera efficiente se prima di spedire il messaggio 1) lo si comprime (togliendo la ridondanza inutile) e poi 2) ci si aggiunge della ridondanza controllata atta a correggere gli eventuali errori comessi durante la sua trasmissione.

3. La capacità del canale asimmetrico

In un canale di trasmissione binario asimmetrico solo errori del tipo $1 \to 0$ (oppure da $0 \to 1$) possono avvenire, quelli $0 \to 1$ (oppure da $1 \to 0$ rispettivamente) sono impossibili. Ciò implica che se riceviamo 1 (oppure, 0), siamo sicuri che era stato spedito 1 (oppure, 0) e che quindi durante la trasmissione di quell'1 (oppure,

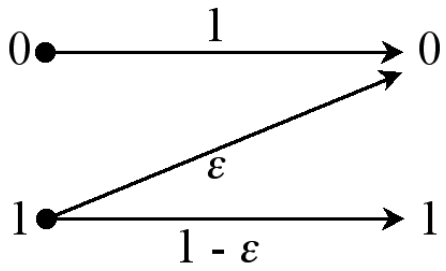

Figura 3: diagramma caratteristico di un canale binario asimmetrico Z. ϵ è la probabilità di errore.

0) non sono avvenuti errori. Il tipo di canale asimmetrico per cui solo errori del tipo $1 \to 0$ sono possibili è caratterizzato dalla seguente matrice delle probabilità di transizione

$$P_{\mathrm{Z}} = \begin{pmatrix} P(0|0) & P(1|0) \\ P(0|1) & P(1|1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \epsilon & 1-\epsilon \end{pmatrix},$$

in cui ϵ è la probabilità di errore, e prende il nome di canale Z in virtù della forma del suo diagramma caratteristico che è riportato in Figura 3. Invece, il canale asimmetrico per cui solo errori del tipo $0 \to 1$ sono possibili è caratterizzato dalla matrice

$$P_{\overline{\mathrm{Z}}} = \begin{pmatrix} P(0|0) & P(1|0) \\ P(0|1) & P(1|1) \end{pmatrix} = \begin{pmatrix} 1-\epsilon & \epsilon \\ 0 & 1 \end{pmatrix},$$

lo indicheremo con $\overline{Z}$ ed è rappresentato in Figura 4.
Ovviamente, a meno di scambiare 1 con 0 il canale Z è uguale al canale $\overline{\mathrm{Z}}$ e quindi tutti e due hanno la stessa capacità. In questa Sezione ci occuperemo del problema di trovare un'espressione analitica del canale Z (e quindi del canale $\overline{\mathrm{Z}}$). Si ha il seguente Teorema.

Teorema 1. La capacità del canale asimmetrco la cui probabilità di errore è ϵ è data da

$$C_{\mathrm{Z}}(\epsilon) = \log_2 \left(1 + \epsilon^{\epsilon/(1-\epsilon)} - \epsilon^{1/(1-\epsilon)}\right)$$

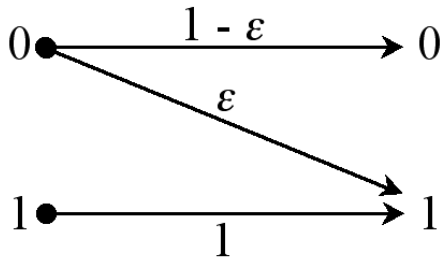

Figura 4: diagramma caratteristico di un canale binario asimmetrico $\overline{\mathrm{Z}}$. ϵ è la probabilità di errore.

$$= \log_2 \left(1 + 2^{-h(\epsilon)/(1-\epsilon)}\right) = C_{\overline{\mathrm{Z}}}(\epsilon). \qquad (2)$$

Dimostrazione: Sia $X \in \{0,1\}$ una sorgente di informazione binaria che assume il valore 1 con probabilità $P(X=1) = q$ ed il valore 0 con probabilità $P(X=0) = 1-q$, e supponiamo che X si data come input ad un canale Z. La variabile aleatoria binaria $Y \in \{0,1\}$ rappresentante l'output del canale assume il valore 1 con probabilità

$$P(Y=1) = q(1-\epsilon)$$

ed il valore 0 con probabilità

$$P(Y=0) = 1 - q(1-\epsilon).$$

Ora, la capacità del canale Z è definita da

$$C_{\mathrm{Z}} = \max_X I(Y;X) = \max_{q\in[0,1]} I(Y;X),$$

dove $I(Y;X)$ è la mutua informazione tra le variabili aleatorie X e Y. Per trovare un espressione analitica per C_{Z} dobbiamo trovare un espressione per $I(Y;X)$ (che è una funzione di ϵ e q) e poi massimizzare l'espressione rispetto a q. Ricordiamo che $h(x) = -[x\log_2 x + (1-x)\log_2(1-x)]$. Si ha

$$I(Y;X) = H(Y) - H(Y|X)$$

dove

$$H(Y) = h(P(Y=1)),$$

è l'entropia della variabile aleatoria Y e

$$H(Y|X) = \sum_x P(x) H(Y|x)$$

$$= P(X=0)H(Y|X=0) + P(X=1)H(Y|X=1)$$

$$= P(X=0)h(P(Y=1|X=0)) + P(X=1)h(P(Y=1|X=1))$$

è l'entropia condizionata di Y dato X. Nel nostro caso, poichè $P(Y=1) = q(1-\epsilon)$, $P(Y=1|X=0) = 0$, $h(0) = 0$, $P(X=1) = q$, $P(Y=1|X=1) = 1-\epsilon$ e $h(x) = h(1-x)$, si ha $H(Y) = h(q(1-\epsilon))$, $H(Y|X) = qh(\epsilon)$ e, quindi,

$$I(Y;X) = h(q(1-\epsilon)) - qh(\epsilon).$$

Consideriamo tale funzione come una funzione di $q \in [0,1]$ e poniamo

$$f_\epsilon(q) = h(q(1-\epsilon)) - qh(\epsilon).$$

Dobbiamo trovare il massimo di $f_\epsilon(q)$ che, per ogni $\epsilon \in [0,1]$, è continua per ogni $q \in [0,1]$ e derivabile per ogni $q \in (0,1)$. Consideriamo prima il caso $\epsilon \in [0,1)$. Tramite le usuali regole del calcolo differenziale, poichè

$$h'(x) = -[\log_2 x - \log_2(1-x)],$$

si ha

$$f'_\epsilon(q) = h'(q(1-\epsilon))(1-\epsilon) - h(\epsilon)$$

$$= (1-\epsilon)\log_2 \frac{1-q(1-\epsilon)}{q(1-\epsilon)} - h(\epsilon) \geq 0$$

$$\iff \quad q \leq \frac{1}{(2^{h(\epsilon)/(1-\epsilon)}+1)(1-\epsilon)} \stackrel{\text{def}}{=} q_{max}(\epsilon).$$

Si può mostrare che, come funzione di $\epsilon \in [0,1)$, $q_{max}(p)$ è una funzione decrescente tale che

$$\lim_{p\to 0} q_{max}(p) = \frac{1}{2} = 0.5$$

e

$$\lim_{p\to 1} q_{max}(p) = \frac{1}{e} = \frac{1}{2.718\ldots} = 0.367\ldots.$$

Quindi, per ogni $\epsilon \in [0,1)$, $q_{max}(\epsilon) \in (0.367, 0.5] \subseteq (0,1)$. Dato ciò, lo studio del segno di $f'_\epsilon(q)$ ci permette di concludere che, per ogni $\epsilon \in [0,1)$ esiste uno ed un solo punto di massimo di $f_\epsilon(q)$ che è assunto per $q = q_{max}(\epsilon)$. Se $\epsilon = 1$ allora $f_1(q) = 0$ e, per ogni $q \in [0,1]$, $f_1(q)$ assume il suo valore massimo; in particolare quando $q = q_{max}(1) = 1/e = 0.367\ldots$. Abbiamo così dimostrato che la capacità del canale Z è data da

$$C_{\mathrm{Z}} = C_{\mathrm{Z}}(\epsilon) = \max_{q\in[0,1]} f_\epsilon(q)$$

$$= h(q_{max}(\epsilon)(1-\epsilon)) - q_{max}(p)h(\epsilon)$$

$$= h\left(\frac{1}{(2^{h(\epsilon)/(1-\epsilon)}+1)}\right) - \frac{h(\epsilon)}{(2^{h(\epsilon)/(1-\epsilon)}+1)\,(1-\epsilon)}.$$

Ora, con un po' di pazienza, portando gli argomenti dei logaritmi dell'espressione appena scritta sotto uno stesso logaritmo , tutto si semplifica e si ottengono le espressioni in (2). ∎

Per concludere questa Sezione, si noti che, poichè

$$\lim_{\epsilon\to 0} \frac{1}{(2^{h(\epsilon)/(1-\epsilon)}+1)} = \frac{1}{2}, \quad \lim_{\epsilon\to 0}(1-\epsilon) = 1 \ \text{ e } \ h(1/2) = 1,$$

quando ϵ è piccolo, $C_{\mathrm{Z}}(\epsilon)$ può essere approssimata dalla seguente semplice formula.

$$C_{\mathrm{Z}}(\epsilon) \approx 1 - \frac{1}{2}h(\epsilon).$$

Si noti che la capacità del canale binario simmetrico con probabilità di errore ϵ è $C_{BSC}(\epsilon) = 1 - h(\epsilon)$.

4. La capacità del canale unidirezionale

Per un canale di trasmissione unidirezionale la situazione è un po' più complessa, in quanto il canale ha un comportamento aleatorio. Per il canale unidirezionale si assume che la sequenza di simboli emessa dalla sorgente sia suddivisa in parole binarie di lunghezza n le quali in sequenza vengono spedite attraverso il canale. Ora, durante la trasmissione di tutta la sequenza, sia errori del tipo $1 \to 0$ che errori del tipo $0 \to 1$ possono avvenire, ma per ogni particolare parola di lunghezza n trasmessa tutti gli errori sono dello stesso tipo. Per esempio, se $n = 4$ e la sequenza da trasmettere è

$\ldots$ 0010 1001 $\ldots$,

è possibile ricevere erroneamente

$\ldots$ $\underline{1}$010 10$\underline{1}$1 $\ldots$ oppure $\ldots$ 0$\underline{1}$10 100$\underline{0}$ $\ldots$ oppure

$\ldots$ 00$\underline{0}$0 1$\underline{1}$01 $\ldots$ oppure $\ldots$ 00$\underline{0}$0 $\underline{0}$001 $\ldots$,

e non

$\ldots$ 0$\underline{10}$0 $\underline{01}$01 $\ldots$.

Quindi, durante la trasmissione di una parola di lunghezza n il canale unidirezionale si comporta come n canali Z oppure n canali $\overline{\mathrm{Z}}$. Il comportamento è scelto a caso dal sistema prima di trasmettere ogni parola ed indipendentemente da essa: la metà delle volte il canale si comporta come n canali Z e l'altra metà delle volte come n canali $\overline{\mathrm{Z}}$. In maniera più rigorosa diamo la seguente Definizione.

Definizione 1. Siano dati $n \in \mathbb{N}$, $\epsilon \in [0, 1]$ ed una variabile aleatoria $U \in \{\mathrm{Z}, \overline{\mathrm{Z}}\}$, essendo Z e $\overline{\mathrm{Z}}$ i canali asimmetrici della Sezione 3. aventi probabilità di errore ϵ. Sia U tale che

$$P(U = \mathrm{Z}) = P(U = \overline{\mathrm{Z}}) = 1/2.$$

Si definisce canale unidirezionale di lunghezza n e probabilità di errore ϵ (e comportamento aleatorio definito da U), e si indica con $\mathrm{U}(n)$, il canale di trasmissione finito e senza memoria tale che:

1) l'alfabeto di input e l'alfabeto di output sono uguali all'insieme delle n-ple di 0 e 1; ovvero,

   $A_i = A_o = \mathbf{Z}_2^n = \{0,1\}^n$, e

2) $\mathrm{U}(n) = (\underbrace{U, U, \ldots, U}_{n \text{ volte}})$.

Quindi, tale canale è caratterizzato da una matrice delle probabilità di transizione

$$P_{\mathrm{U}(n)} = \left(P_{\mathrm{U}(n)}(\mathbf{Y}|\mathbf{X})\right)_{\mathbf{X},\mathbf{Y}\in\mathbf{Z}_2^n}$$

con 2^n righe e 2^n colonne il cui generico elemento è dato da

$$P_{\mathrm{U}(n)}(\mathbf{Y}|\mathbf{X}) = (1/2)\left[\prod_{i=1}^{n} P_{\mathrm{Z}}(y_i|x_i) + \prod_{i=1}^{n} P_{\overline{\mathrm{Z}}}(y_i|x_i)\right], \quad (3)$$

essendo $\mathbf{X} = x_1x_2\ldots x_n \in \mathbf{Z}_2^n$, $\mathbf{Y} = y_1y_2\ldots y_n \in \mathbf{Z}_2^n$ e, $P_{\mathrm{Z}}(y|x)$ e $P_{\overline{\mathrm{Z}}}(y|x)$ le probabilità di transizione del canale Z e del canale $\overline{\mathrm{Z}}$ rispettivamente. Invero,

1) siccome X ed U sono indipendenti, si ha $P(U|X) = P(U)$;

2) per come è definito il canale $\mathrm{U}(n)$, si ha (con ovvie notazioni)

   $$P(\mathbf{Y}|\mathbf{X}, U = \mathrm{Z}) = P_{\mathrm{Z}^n}(\mathbf{Y}|\mathbf{X})$$

   e

   $$P(\mathbf{Y}|\mathbf{X}, U = \overline{\mathrm{Z}}) = P_{\overline{\mathrm{Z}}^n}(\mathbf{Y}|\mathbf{X});$$

3) e siccome sia il canale Z che il canale $\overline{\mathrm{Z}}$ sono senza memoria, si ha

   $$P_{\mathrm{Z}^n}(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^{n} P_{\mathrm{Z}}(y_i|x_i) \quad \text{e} \quad P_{\overline{\mathrm{Z}}^n}(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^{n} P_{\overline{\mathrm{Z}}}(y_i|x_i).$$

Per cui si ha

$$P_{\mathrm{U}(n)}(\mathbf{Y}|\mathbf{X}) = P(\mathbf{Y}, U = \mathrm{Z}|\mathbf{X}) + P(\mathbf{Y}, U = \overline{\mathrm{Z}}|\mathbf{X})$$

$$= P(\mathbf{Y}|\mathbf{X}, U = \mathrm{Z})P(U = \mathrm{Z}|\mathbf{X}) + P(\mathbf{Y}|\mathbf{X}, U = \overline{\mathrm{Z}})P(U = \overline{\mathrm{Z}}|\mathbf{X})$$

$$= P(\mathbf{Y}|\mathbf{X}, U = \mathrm{Z})P(U = \mathrm{Z}) + P(\mathbf{Y}|\mathbf{X}, U = \overline{\mathrm{Z}})P(U = \overline{\mathrm{Z}})$$

$$= (1/2)P_{\mathrm{Z}^n}(\mathbf{Y}|\mathbf{X}) + (1/2)P_{\overline{\mathrm{Z}}^n}(\mathbf{Y}|\mathbf{X})$$

$$= (1/2)\prod_{i=1}^{n} P_{\mathrm{Z}}(y_i|x_i) + (1/2)\prod_{i=1}^{n} P_{\overline{\mathrm{Z}}}(y_i|x_i),$$

e quindi la (3).
Sia $P(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$, per ogni $\mathbf{x} \in \mathbf{Z}\!\!\mathbf{Z}_2^n$. Si definisce capacità del canale unidirezionale $\mathrm{U}(n)$ per bit di dati trasmesso, la quantità

$$\overline{C}_{\mathrm{U}(n)} \stackrel{\mathrm{def}}{=} \frac{C_{\mathrm{U}(n)}}{n} = \frac{1}{n}\max_{\mathbf{X}} I(\mathbf{Y};\mathbf{X})$$

$$= \frac{1}{n}\max_{P(\mathbf{x}):P(\mathbf{x})\geq 0,\sum_{\mathbf{x}} P(\mathbf{x})=1} I(\mathbf{Y};\mathbf{X}).$$

In questa Sezione troveremo delle buone limitazioni per $C_{\mathrm{U}(n)}$ e quindi per $\overline{C}_{\mathrm{U}(n)}$. In particolare dimostreremo il seguente Teorema.

Teorema 2. Per la capacità del canale unidirezionale di lunghezza n e probabilità di errore $\epsilon \in [0, 1]$ si ha:

$$C_{\mathrm{U}(n)}(\epsilon) \leq nC_{\mathrm{Z}}(\epsilon) \tag{4}$$

e

$$C_{\mathrm{U}(n)}(\epsilon) \geq nC_{\mathrm{Z}}(\epsilon) - 1. \tag{5}$$

Dimostrazione: Per trovare delle limitazioni su $C_{\mathrm{U}(n)} = C_{\mathrm{U}(n)}(\epsilon)$ dobbiamo trovare delle limitazioni su $I(\mathbf{Y};\mathbf{X})$.
Diamo prima alcuni risultati di carattere generale. Date tre variabili aleatorie X, Y e Z, si definisce mutua informazione tra X e Y dato Z la seguente quantità

$$I(Y;X|Z) = \sum_z P(z) I(Y;X|Z=z)$$

$$= \sum_z P(z) \sum_{x,y} P(x,y|z) \log_2 \frac{P(x,y|z)}{P(x|z)P(y|z)}$$

$$= \sum_{x,y,z} P(x,y,z) \log_2 \frac{P(x,y|z)}{P(x|z)P(y|z)}.$$

Se X e Z sono indipendenti allora vale la seguente relazione

$$I(Y;X|Z) = I(Y;X) + I(Z;X|Y). \qquad (6)$$

Infatti, dal Teorema 2.5.2 (pag. 22) in [2] per $n = 2$, sostituendo Y ad X_1, Z ad X_2 e X ad Y, si ha

$$I(YZ;X) = I(Y;X) + I(Z;X|Y).$$

Inoltre, da questa relazione, scambiando Y e Z, si ha

$$I(ZY;X) = I(Z;X) + I(Y;X|Z) = I(YZ;X).$$

Siccome X e Z sono indipendenti si ha $I(Z;X) = 0$, e quindi dalle ultime due relazioni segue la (6).
Vale inoltre il seguente Teorema di carattere generale.

Teorema 3 (Teorema 4.2.1 di [3], pag. 75). Dato $n \in \mathbb{N}$, sia $\mathbf{X} = X_1 X_2 \ldots X_n$ una variabile aleatoria rappresentante una sequenza di n inputs ad un canale discreto e senza memoria. Sia $\mathbf{Y} = Y_1 Y_2 \ldots Y_n$ la corrispondente sequenza degli outputs. Allora

$$I(\mathbf{Y};\mathbf{X}) \leq \sum_{i=1}^{n} I(Y_i;X_i)$$

e vale il segno di uguaglianza se, e solo se, le variabili aleatorie X_i sono indipendenti (ovvero, $P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i)$).

Veniamo ora al nostro problema. In virtú del fatto che $\mathbf{X}$ ed U sono indipendenti, dall'equazione (6), applicata alle tre variabili aleatorie $X = \mathbf{X}$, $Y = \mathbf{Y}$ e $Z = U$, si ha

$$I(\mathbf{Y};\mathbf{X}) = I(\mathbf{Y};\mathbf{X}|U) - I(U;\mathbf{X}|\mathbf{Y}). \tag{7}$$

In virtú del fatto che sia il canale Z che il canale $\overline{\mathrm{Z}}$ sono senza memoria dal Teorema 3, si ha

$$I(\mathbf{Y};\mathbf{X}|U) \leq \sum_{i=1}^{n} I(Y_i;X_i|U), \tag{8}$$

valendo il segno di uguaglianza se le variabili aleatorie X_i sono indipendenti.
Dimostriamo la (4). Dalle relazioni (7) e (8), e dal fatto che la mutua informazione (condizionata o meno) è sempre positiva, segue

$$I(\mathbf{Y};\mathbf{X}) \leq I(\mathbf{Y};\mathbf{X}|U) \leq \sum_{i=1}^{n} I(Y_i;X_i|U).$$

Per cui,

$$C_{\mathrm{U}(n)}(\epsilon) = \max_{\mathbf{X}} I(\mathbf{Y};\mathbf{X}) \leq \max_{\mathbf{X}} \sum_{i=1}^{n} I(Y_i;X_i|U)$$

$$\leq \sum_{i=1}^{n} \max_{\mathbf{X}} I(Y_i;X_i|U) = \sum_{i=1}^{n} \max_{X_i} I(Y_i;X_i|U)$$

$$= \sum_{i=1}^{n} \max_{X_i} [P(U=\mathrm{Z}) I(Y_i;X_i|U=\mathrm{Z}) + P(U=\overline{\mathrm{Z}}) I(Y_i;X_i|U=\overline{\mathrm{Z}})]$$

$$= \sum_{i=1}^{n} P(U=\mathrm{Z}) \max_{X_i} I(Y_i;X_i|U=\mathrm{Z}) +$$

$$+ P(U=\overline{\mathrm{Z}}) \max_{X_i} I(Y_i;X_i|U=\overline{\mathrm{Z}})$$

$$= \sum_{i=1}^{n} P(U=\mathrm{Z}) C_{\mathrm{Z}}(\epsilon) + P(U=\overline{\mathrm{Z}}) C_{\overline{\mathrm{Z}}}(\epsilon) = n C_{\mathrm{Z}}(\epsilon).$$

L'ultima equazione deriva dal Teorema 1. Così la prima disuguaglianza è dimostrata.
Dimostriamo la (5). Supponiamo le variabili aleatorie X_i siano indipendenti. In questo caso la (8) vale con il segno di uguaglianza; ovvero, si ha

$$I(\mathbf{Y};\mathbf{X}|U) = \sum_{i=1}^{n} I(Y_i; X_i|U). \tag{9}$$

Dalle equazioni (7) e (9) segue che

$$I(\mathbf{Y};\mathbf{X}) = \sum_{i=1}^{n} I(Y_i; X_i|U) - I(U;\mathbf{X}|\mathbf{Y}) \tag{10}$$

Ora, assumendo U solo due valori, si ha

$$-I(U;\mathbf{X}|\mathbf{Y}) = -H(U|\mathbf{Y}) + H(U|\mathbf{X},\mathbf{Y}) \geq -H(U|\mathbf{Y}) \geq -1.$$

Quest'ultima relazione e la (10) danno

$$I(\mathbf{Y};\mathbf{X}) \geq \sum_{i=1}^{n} I(Y_i; X_i|U) - 1; \tag{11}$$

se le variabili aleatorie X_i sono indipendenti. Per cui dalla (11) segue

$$C_{\mathrm{U}(n)}(\epsilon) = \max_{\mathbf{X}} I(\mathbf{Y};\mathbf{X}) \geq \max_{X_1,X_2,\ldots,X_n \text{ indipendenti}} I(\mathbf{Y};\mathbf{X})$$

$$\geq \max_{X_1,X_2,\ldots,X_n \text{ indipendenti}} \sum_{i=1}^{n} I(Y_i; X_i|U) - 1$$

$$= \sum_{i=1}^{n} \max_{X_i} I(Y_i; X_i|U) - 1 = nC_{\mathrm{Z}}(\epsilon) - 1,$$

e quindi anche la seconda disuguaglianza è dimostrata. ∎
Il Teorema appena dimostrato implica la relazione (1) nell'introduzione (in cui si è omesso il simbolo di soprasegnato per uniformità). La Figura 5 riporta i grafici delle capacità del canale binario simmetrico, del canale asimmetrico e della limitazione inferiore per il canale unidirezionale individuato dalla (5).

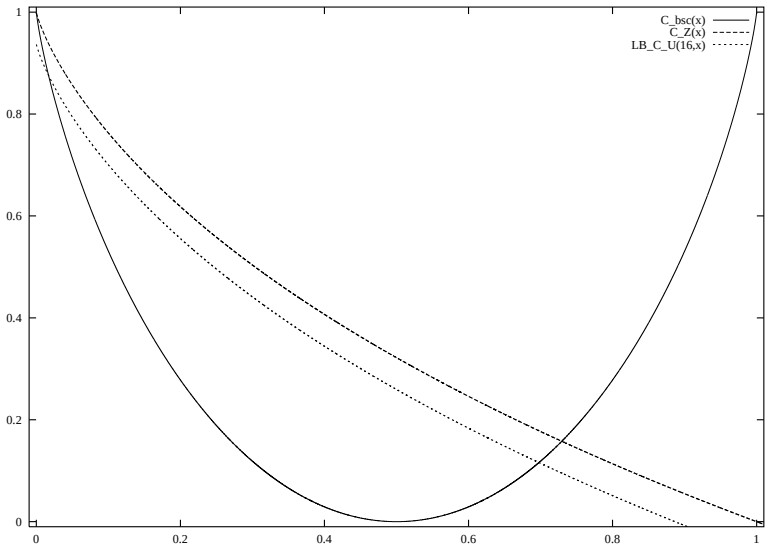

Figura 5: Grafici delle capacità del canale binario simmetrico (denotato con C_bsc(x)), del canale asimmetrico (denotato con C_Z(x)) e della limitazione inferiore per il canale unidirezionale di lunghezza $n = 16$ individuato dalla (5) (denotata con LB_C_U(16,x)). L'asse x rappresenta la probabilità di errore ϵ.

Bibliografia

[1] Blaum M. (1993), *Codes for Detecting and Correcting Unidirectional Errors*, IEEE Computer Society Press.

[2] Covers T. M., Thomas J. A. (1991), *Elements of Information Theory*, John Wiley and Sons Inc., New York.

[3] Gallager R. G. (1968), *Information Theory and Reliable Communication*, John Wiley and Sons Inc., New York.

[4] Shannon C. E. (1948), "A mathematical theory of communication", *Bell System Technical Journal*, 27, 379-423 e 623–656.