Appunti di Teoria dell'Informazione

Corso di Telecomunicazioni

(Classe Quinta della specializzazione Elettronica e Telecomunicazioni)

1. La teoria dell'informazione

La teoria dell'informazione descrive il funzionamento dei sistemi di comunicazione sia analogici che digitali ponendo l'attenzione sul contenuto informativo del messaggio.

Essa tratta aspetti basilari quali:

a) La misura dell'informazione associata ad una sorgente.

b) La misura della quantità di informazione trasferibile attraverso un canale di comunicazione.

c) L'uso di codifiche del messaggio per sfruttare al meglio la capacità informativa del canale.

La teoria dell'informazione ha un teorema fondamentale: **il teorema di Shannon.**

Sia data una sorgente S che emetta simboli alla velocità R [bit/sec] ed un canale di capacità C [bit/sec], se R≤C allora esiste un metodo di codifica della sorgente tale che la frequenza degli errori al ricevitore sia arbitrariamente piccola; se, invece, R>C non esiste alcun codice che consenta una equivocazione minore di R-C.

2. Le sorgenti

Le sorgenti sono dispositivi che emettono messaggi di natura DISCRETA o CONTINUA.

In queste pagine saranno considerate solo sorgenti **discrete.** Quindi una sorgente emetterà simboli appartenenti ad un insieme numerabile, detto **alfabeto di sorgente**; allora potremo scrivere per il generico simbolo $\mathbf{x_i}$ che $x_i \in X = \{x_1, x_2, ..., x_m\}$ dove **X** è l'alfabeto di sorgente e **m** la sua dimensione.

La codifica di sorgente, nei modi che si vedranno, ha lo scopo di ridurre la RIDONDANZA e quindi sfruttare al meglio la capacità del canale.

3. La quantità di informazione

Per definire correttamente la quantità di informazione è fondamentale comprendere che essa è collegata da una relazione di proporzionalità inversa con il grado di conoscenza del messaggio. Cioè, l'informazione associata ad un evento noto è **nulla**; un messaggio relativo ad un evento **raro** ha invece **elevato contenuto informativo.**

Quindi trattando il problema della quantità di informazione dal punto di vista probabilistico possiamo procedere nel seguente modo.
Sia **S** una sorgente discreta che emette simboli $\mathbf{x_i}$ appartenenti ad un alfabeto **X** di dimensione m. Ovvero $x_i \in X = \{x_1, x_2, ..., x_m\}$ e sia $\mathbf{P(x_i)}$ la probabilità di emissione del simbolo $\mathbf{x_i}$, ovviamente sarà verificato che $\sum_1^m P(x_i) = 1$.
Abbiamo già detto che, detta $\mathbf{I(x_i)}$ la quantità di informazione associata all'emissione del simbolo $\mathbf{x_i}$, dovrà essere $I(x_i) \propto \frac{1}{P(x_i)}$. Allora si può dare la seguente definizione di quantità di informazione associata al simbolo $\mathbf{x_i}$:

$$\text{def.}\quad I(x_i) = \log_2\left(\frac{1}{P(x_i)}\right) = -\log_2\big(P(x_i)\big)\ \text{[bit]}.$$

La quantità di informazione viene misurata in **bit** e conviene anche ricordare la relazione $\log_2(a) = \frac{\log_{10}(a)}{0{,}301}$.

4. Entropia di una sorgente

La quantità di informazione complessiva emessa da una sorgente è tanto maggiore quanto più elevata è la capacità di scelta della sorgente stessa, ovvero quanto più le probabilità di emissione sono uniformemente distribuite tra tutti i suoi simboli.
Ora l'entropia di sorgente **H(x)** misura il grado di scelta, ovvero di dispersione e di incertezza, di una sorgente. Essa è l'informazione media emessa dalla sorgente ed è definita come:

$$\text{def.}\quad H(x) = \sum_1^m P(x_i) I(x_i) = \sum_1^m P(x_i)\log_2\left(\frac{1}{P(x_i)}\right) = -\sum_1^m P(x_i)\log_2\big(P(x_i)\big)$$

misurata in **[bit/simbolo]**.
È evidente che se la sorgente non sceglie, ad esempio emette un solo simbolo $\mathbf{x_j}$ con $\mathbf{P(x_j)=1}$, allora $\mathbf{H(x)=0}$. Infatti

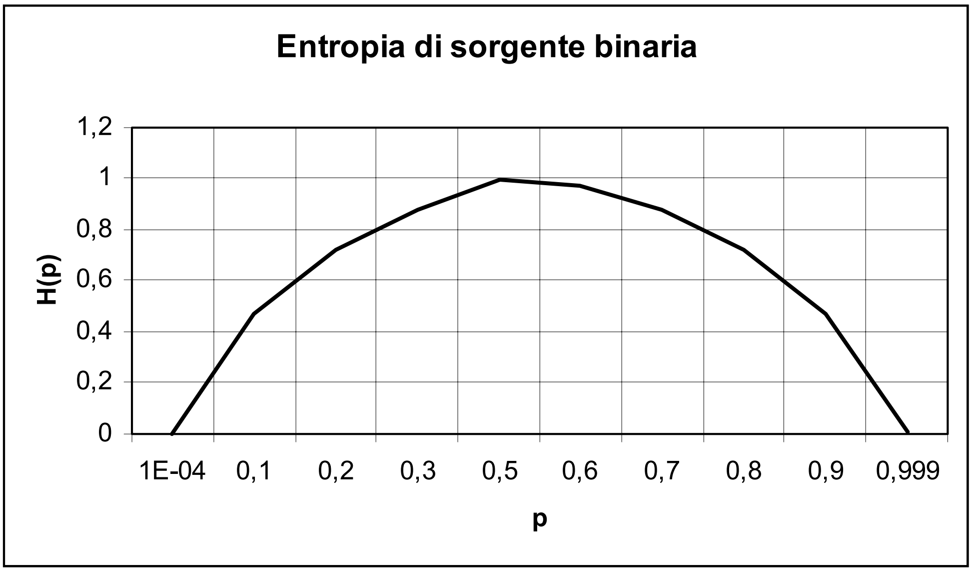
$$H(x) = \sum_1^m P(x_i) I(x_i) = -P(x_j)\log_2\big(P(x_j)\big) = -1 \cdot \log_2(1) = 0\ .$$

Se invece tutti i simboli sono equiprobabili si ha la massima dispersione, ovvero la massima possibilità di scelta, e quindi

$$H(x) = \sum_{1}^{m} P(x_i) I(x_i) = m \cdot \frac{1}{m} \cdot \log_2 \left(\frac{1}{\frac{1}{m}} \right) = \log_2(m) = H_{MAX} \ .$$

Per meglio chiarire quest'ultimo concetto si consideri il caso di una sorgente binaria con $x_i \in \{0,1\}$ e **P(0)=p** e **P(1)=1-p.** In tale caso si ha $H(x) = -p \cdot \log_2(p) - (1-p)\log_2(1-p)$ con valore massimo **H_{MAX}=log$_2$(2)=1 bit/simbolo** quando i due simboli sono equiprobabili.

Facendo variare **p** nell'intervallo **[0,1]** si ottiene il seguente grafico dell'entropia di una sorgente binaria.



5. Ridondanza

La ridondanza di sorgente è quella caratteristica per cui una sorgente di informazione emette messaggi deducibili da informazioni contenute nei messaggi stessi.

La riduzione della ridondanza, mediante opportune codifiche, consente di sfruttare al meglio le caratteristiche di un canale.

Sono i codificatori di sorgente quegli elementi che vengono inseriti dopo la sorgente stessa allo scopo di ridurre la ridondanza del codice emesso; il messaggio è così trasformato in una successione di simboli il più possibile indipendenti tra loro ed equiprobabili.

La definizione matematica di ridondanza è la seguente

$$Rid = 1 - \frac{H(x)}{H_{MAX}} .$$

Si definisce invece Efficienza di Codifica Binaria dei caratteri di sorgente la seguente quantità

$$\eta = \frac{\log_2(m)}{N_{bit}} .$$

Dove per $\mathbf{N_{bit}}$ si intende il numero di bit impiegati per la codifica di sorgente, in genere superiore a quelli strettamente necessari.

6. Capacità informativa di un canale discreto

La capacità informativa di un canale discreto è definita come la massima quantità di informazione trasferibile nel canale, nell'unità di tempo, in modo affidabile.
Si consideri il seguente canale discreto rumoroso

dove **x** sono i simboli entranti e **y** i simboli uscenti.
Sia **H(x)** l'entropia di sorgente, ovvero l'informazione media posseduta in origine; sia **H(x/y)** l'informazione di disturbo dovuta al canale.
Posso quindi definire

$$I(x;y) = H(x) - H(x/y)$$

l'informazione che sarà posseduta alla destinazione.
Si definisce **Capacità di Informazione del Canale** la quantità:

$$C_S = \max\{I(x;y)\}$$

misurata in bit/simbolo.
Conoscendo poi la massima velocità di trasmissione dei simboli consentita nel canale, si può calcolare la **Capacità Informativa del Canale** come **Massima Velocità di Trasferimento su un Canale rumoroso**. E si ha

$$C = s \cdot C_S = s \cdot \max\{I(x;y)\}$$

misurata in bit/sec.
Valore spesso ricavabile dalla nota relazione di Shannon

$$C = 3{,}32B \cdot \log_{10}\left(1 + \frac{S}{N}\right),$$

essendo **B** la banda passante del canale e **S/N** il rapporto segnale-rumore.

7. Velocità di emissione dell'informazione

Si consideri una sorgente che emetta simboli $x_i \in X$, dove **X** è un alfabeto di dimensione **m.** Siano note le probabilità **P(x$_i$)** di emissione dei simboli.
Sia **H(x)** l'entropia della sorgente, ovvero l'informazione media per simbolo.
Si supponga, ora, che la sorgente emetta **n** simboli con velocità **v [simboli/sec].** In tale caso si potrà affermare che **l'informazione totale trasferita** è

$$I(x) = n \cdot H(x)$$

espressa in **bit.**
Posso, quindi, definire la **velocità media di emissione dell'informazione** come

$$R = \frac{\text{informazione trasmessa}}{\text{tempo necessario}}.$$

Allora, nel generico caso in esame, si avrà

$$R = \frac{I(x)}{t} = \frac{n \cdot H(x)}{\frac{n}{v}} = v \cdot H(x)$$

misurata in **bit/sec.**
Chiaramente se l'informazione è emessa in un canale dovrà essere sempre **R≤C.**

- ***Esercizio***
  Una sorgente emette sequenze di n=4 simboli con velocità v=2000 simboli/sec. I simboli sono tutti equiprobabili ed appartengono ad un alfabeto di dimensione m=8. Calcolare la minima capacità di canale necessaria.

  Calcolo l'entropia
  $$H(x) = \log_2(8) = 3\ \textit{bit/simbolo}$$
  Calcolo l'informazione emessa nella sequenza
  $$I(x) = n \cdot H(x) = 4 \cdot 3 = 12\ \textit{bit}$$
  Calcolo la capacità minima
  $$C_{\min} = R = v \cdot H(x) = 2000 \cdot 3 = 6000\ \textit{bit/sec}.$$

- ***Esercizio***
  Una sorgente emette simboli indipendenti, appartenenti ad un alfabeto di dimensione 5, con velocità media di emissione della informazione

R=7500bit/sec. Noto che: $P_1 = \frac{1}{2}P_2$; $P_3 = \frac{1}{2}P_2$; $P_4 = \frac{1}{2}P_3$; $P_4 = P_5$ *calcolare l'entropia di sorgente e la velocità di emissione dei simboli. Quindi calcolare R nell'ipotesi di simboli equiprobabili.*

Noto che $\sum_1^5 P_i = 1$, *sfruttando le condizioni indicate nella traccia, si calcolano le singole probabilità che risultano essere:* $P_1=P_3=0{,}2$ $P_2=0{,}4$ $P_4=P_5=0{,}1$.
Quindi si calcola l'entropia della sorgente come

$$H(x) = -\sum_1^5 P_i \log_2(P_i) = -\frac{\sum_1^5 P_i \log_{10}(P_i)}{0{,}301} = 2{,}12\ bit/simbolo.$$

Di conseguenza la velocità di emissione di sorgente sarà

$$v = \frac{R}{H(x)} = \frac{7500}{2{,}12} = 3538\ simboli/sec.$$

Nell'ipotesi di simboli equiprobabili si ha $P_i = \frac{1}{m} = \frac{1}{5} = 0{,}2$ *e, quindi,*

$$H_{MAX} = \log_2(5) = \frac{\log_{10}(5)}{0{,}301} = 2{,}32\ bit/simbolo.$$

In questo caso la velocità di emissione dell'informazione sarà $R = v \cdot H_{MAX} = 3538 \cdot 2{,}32 = 8208\ bit/sec.$

8. Un esempio di codifica di sorgente: **la codifica di Huffman**

L'operazione di codifica di sorgente viene effettuata per ridurre la ridondanza ed alzare l'efficienza di codifica.
Ovviamente nel momento in cui si procede alla codifica di simboli non equiprobabili si genereranno delle sequenze codificate di lunghezze diverse, questo allo scopo di ottenere un codice ottimale senza equivocazione.
Iniziamo con il dare le seguenti definizioni:

- $\mathbf{N_i}$ = lunghezza della parola di codice associata al simbolo $\mathbf{x_i}$
- $\mathbf{P_i}$ = probabilità di emissione del simbolo $\mathbf{x_i}$

- **m** = numero dei simboli da codificare
- $\overline{N}$ = lunghezza media della parola di codice, calcolata come

$$\overline{N}=\sum_{1}^{m}P_i\cdot N_i\ .$$

Si osservi che un codice risulterà tanto più efficiente quanto più $\overline{N}$ risulterà piccola.
Un codice sarà detto OTTIMALE se rispetterà la seguente condizione

$$H(x)\le\overline{N}\le H(x)+1\ .$$

Si definirà poi EFFICIENZA DI CODIFICA la quantità

$$\eta=\frac{H(x)}{\overline{N}}\le 1\ ,$$

e perché si abbia l'inequivocabilità della codifica dovrà essere soddisfatta la seguente condizione (di Kraft):

$$K=\sum_{1}^{m}2^{-N_i}\le 1\ .$$

In sintesi, un codice potrà dirsi ottimale se saranno verificate le seguenti quattro condizioni:

$\overline{N}=H(x)$
$\eta=1$
K=1
$N_i=I(x_i)=-\log_2(P_i)$

Non rimane che esaminare il caso della codifica di Huffman e conviene affrontarlo proprio con un esempio.
Si consideri una sorgente che emette simboli appartenenti ad un alfabeto di dimensione **m=6** con probabilità di emissione indicate nella seguente tabella

simboli	**probabilità**
X_1	$P(x_1)=0,3$
X_2	$P(x_2)=0,25$
X_3	$P(x_3)=0,2$
X_4	$P(x_4)=0,1$
X_5	$P(x_5)=0,1$
X_6	$P(x_6)=0,05$

Per operare la codifica di Huffman deve essere costruita una particolare tabella, illustrata di seguito, seguendo le regole di seguito indicate.

Step 1	Step 2	Step 3	Step 4	Step 5
X_1=0,3	X_1=0,3	X_1=0,3	X_{3456}=0,45	X_{12}=0,55
X_2=0,25	X_2=0,25	X_2=0,25	X_1=0,3	X_{3456}=0,45
X_3=0,2	X_3=0,2	X_{456}=0,25	X_2=0,25	
X_4=0,1	X_{56}=0,15	X_3=0,2		
X_5=0,1	X_4=0,1			
X_6=0,05				
$\mathbf{X_{56}}=X_5+X_6$ =0,15	$\mathbf{X_{456}}=X_4+X_{56}$ =0,25	$\mathbf{X_{3456}}=X_3+X_{456}$ =0,45	$\mathbf{X_{12}}=X_1+X_2$ =0,55	

1) Costruisci la prima colonna (step 1) ordinando i simboli per probabilità decrescenti.
2) Poi crea un nuovo simbolo accoppiando gli ultimi due, tale simbolo avrà una probabilità pari allo somma delle due che lo costituiscono.
3) Usando il nuovo simbolo crea una seconda colonna (step 2), ordinando i simboli sempre per probabilità decrescenti.
4) Ripeti le operazioni dei passi 2) e 3), creando di volta in volta nuove colonne, fino a quando non rimangono solo due simboli raggruppati e ordinati per probabilità decrescenti. Nel nostro caso il tutto termina allo step 5.

Adesso la tabella va riletta da destra verso sinistra, come una struttura ad albero, e ad ogni biforcazione si assegna uno 0 ed un 1 fino a tornare ai simboli originari.

Simb.	*Valore*	*Simb.*	*Valore*	*Simb.*	*Valore*	*Simb.*	*Valore*
X_{12}	**0**	$\mathbf{X_1}$	**0**				
		$\mathbf{X_2}$	**1**				
X_{3456}	**1**	$\mathbf{X_3}$	**0**				
		X_{456}	**1**	$\mathbf{X_4}$	**0**		
				X_{56}	**1**	$\mathbf{X_5}$	**0**
						$\mathbf{X_6}$	**1**

Adesso, per ogni simbolo originario, non si deve fare altro che leggere il codice leggendo, da sinistra a destra, la sequenza di bit che parte dall'inizio del raggruppamento fino al simbolo in esame.

Nel nostro esempio si verrà a generare il seguente codice:

simbolo	codice	N_i
X_1	**00**	2
X_2	**01**	2
X_3	**10**	2
X_4	**110**	3
X_5	**1110**	4
X_6	**1111**	4

Si osservi come il codice generato è assolutamente privo di equivocazione. Inoltra nella tabella sono indicate anche le lunghezze delle singole parole di codice.
Calcolando l'entropia della sorgente in esame si ottiene H(x)=2,366 bit/simbolo. Ora conoscendo il valore di $\overline{N} = 2 \cdot 0{,}3 + 2 \cdot 0{,}25 + 2 \cdot 0{,}2 + 3 \cdot 0{,}1 + 4 \cdot 0{,}1 + 4 \cdot 0{,}05 = 2{,}4$ bit/simbolo si vede che è verificata la condizione $H(x) \le \overline{N} \le H(x) + 1$.
Inoltre risulta $\eta = \frac{H(x)}{\overline{N}} \approx 99\%$ ed è soddisfatta la condizione di Kraft (sull'inequivocabilità), ovvero risulta $K = 2^{-2} + 2^{-2} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-4} = \frac{3}{4} + \frac{1}{8} + \frac{2}{16} = 1$. Quindi il codice è ottimale.