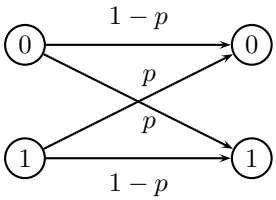
Teoria dell'Informazione/RTI **Anno Accademico 2021–2022**

Lezione 17

Ugo Vaccaro

Il problema che intendiamo studiare in questa lezione è il seguente: Vogliamo spedire messaggi $\mathbf{x} = (x_1, \ldots, x_m) \in \{0,1\}^m$ su di un canale di comunicazione *rumoroso*, ovvero un canale in cui *ciascun* bit trasmesso può essere modificato in un altro bit con probabilità p.

Tipicamente, un siffatto canale viene rappresentato con questo schema:



In altre parole, detto b il bit trasmesso, è come se ricevessimo in output il bit $b' = b \oplus r$, dove r è pari a 1 con probabilità p e 0 con probabilità $(1-p)$, e $\oplus$ denota l'addizione modulo 2.

Se spediamo m bits riceveremo una sequenza di m bits che differisce (in media) in pm bits da quella trasmessa. Possiamo spedire bits con *meno* errori?

Un semplice metodo è il seguente: trasmettiamo 3 copie di ogni bit e decodifichiamo a "maggioranza", ovvero

- se il bit è 0 spediamo 000, se il bit è 1 spediamo 111
- se riceviamo una delle sequenze $000, 001, 010, 100$ diciamo che il bit di partenza è 0
- se riceviamo una delle sequenze $111, 110, 101, 011$ diciamo che il bit di partenza è 1

Ci sarà un errore di decodifica se e solo se il canale commette almeno due *errori* nella trasmissione della tripla. Ad es. per $p = 0.1$ ciò accadrà con probabilità $3p^2(1-p) + p^3 = 3 \times 0.01 \times 0.9 + 0.001 = 0.0028 < 0.1$

Abbiamo quindi ridotto la probabilità di errore, ma abbiamo anche ridotto la velocità di trasmissione (infatti, abbiamo triplicato la lunghezza delle sequenza di bit da trasmettere). Tecnicamente, possiamo anche far andare a 0 la probabilità di errore, spedendo al posto di 0 la sequenza $00\ldots0 \in \{0,1\}^n$, spedendo al posto di 1 la sequenza $11\ldots1 \in \{0,1\}^n$, decodificando a maggioranza e facendo $n \to \infty$ ma in questo modo anche la velocità di trasmissione $\to 0$.

É possibile trasmettere su di un canale rumoroso con probabilità di errore $\to 0$ senza far andare la velocità di trasmissione $\to 0$? Fino al 1948 si pensava di no.

Stabiliamo innanzitutto il modello formale. Denoteremo:

- Messaggi da spedire $\mathbf{s} = (s_1, \ldots, s_m) \in \{0,1\}^m$
- Schema iniettivo di codifica $c : \{0,1\}^m \mapsto C \subseteq \{0,1\}^n$, $n > m$, $|C| = 2^m$
- Schema di decodifica $d : \{0,1\}^n \mapsto C$
- $\frac{m}{n}$ = velocità di trasmissione

- Trasmettiamo $c(\mathbf{s}) \in \{0,1\}^n$ invece che $\mathbf{s} \in \{0,1\}^m$
- Il ricevitore decodifica il mesaggio ricevuto (che in generale sarà diverso da quello trasmesso...)
- $\underbrace{\mathbf{s}}_{m \text{ bits}} \xrightarrow{\text{codifica}} \underbrace{c(\mathbf{s})}_{n \text{ bits}} \xrightarrow{\text{canale}} \underbrace{c(\mathbf{s}) \oplus \mathbf{Z}}_{\text{addiz. mod 2}} \xrightarrow{\text{decodifica}} d(c(\mathbf{s}) \oplus \mathbf{Z})$

  dove $\mathbf{Z} = (Z_1, \ldots, Z_n)$ e $Z_1, \ldots, Z_n$ sono v.c. indipendenti ed identicamente distribuite, con $\Pr\{Z_i = 1\} = p$
- applicando la funzione di decodifica d alla sequenza ricevuta $c(\mathbf{s}) \oplus \mathbf{Z}$ vorremmo che $d(c(\mathbf{s}) \oplus \mathbf{Z}) = c(\mathbf{s})$, da cui poi possiamo ottenere $\mathbf{s} \in \{0,1\}^m$ in quanto la codifica c è iniettiva e quindi invertibile.

Proviamo innanzitutto un'utile risultato che và sotto il nome di Diseguaglinaza di Chernoff-Hoeffding.

Supponiamo di generare casualmente sequenze binarie $\mathbf{z} = z_1 \ldots z_n \in \{0,1\}^n$, ponendo $z_i = 1$ con probabilità pari a p e $z_i = 0$ con probabilità pari ad $1-p$, $0 < p < 1$, per ogni $i = 1, \ldots n$, indipendentemente uno dall'altro. Allora:

a) Per ogni q tale che $1 > q > p$, vale che

$$\Pr\{\mathbf{z} \in \{0,1\}^n : \mathbf{z} \text{ ha } \textit{almeno } qn \text{ bit con valore pari a } 1\} \leq 2^{-nD(q||p)},$$

dove

$$D(q||p) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}$$

è la divergenza informazionale tra la distribuzione di probabilità $(q, 1-q)$ e la distribuzione $(p, 1-p)$.

b) Per ogni q tale che $0 < q < p$, vale che

$$\Pr\{\mathbf{z} \in \{0,1\}^n : \mathbf{z} \text{ ha } \textit{al più } qn \text{ bit con valore pari a } 1\} \leq 2^{-nD(q||p)}.$$

Proviamo la a). La prova della b) è identica.

Poniamo $S = \{\mathbf{z} \in \{0,1\}^n : \mathbf{z} \text{ ha } \textit{almeno } qn \text{ bit con valore pari a } 1\}$.

Sia $\mathbf{z}$ un generico elemento di S, contenente esattamente $k \geq qn$ "1", e calcoliamo il rapporto

$$\frac{q(\mathbf{z})}{p(\mathbf{z})} = \frac{\text{probabilità di aver ottenuto } \mathbf{z} \text{ se i suoi "1" fossero generati ciascheduno con probabilità} = q}{\text{probabilità di aver ottenuto } \mathbf{z} \text{ se i suoi "1" fossero generati ciascheduno con probabilità} = p}.$$

Ovviamente, vale che

$$\frac{q(\mathbf{z})}{p(\mathbf{z})} = \frac{q^k(1-q)^{n-k}}{p^k(1-p)^{n-k}} = \left(\frac{q}{p}\right)^k \cdot \left(\frac{1-q}{1-p}\right)^{n-k}.$$

Osserviamo ora che le ipotesi $q > p$ e $k \geq qn$ ci dicono che $(q/p) > 1$, da cui $(q/p)^k \geq (q/p)^{qn}$ (in quanto, appunto, $k \geq qn$ e la funzione esponenziale a^x è *crescente* in x per $a > 1$). D'altra parte, $q > p$ implica che $(1-q)/(1-p) < 1$, mentre $k \geq qn$ ci dice che $n - k \leq n - qn$. Il tutto ci porta a concludere che $((1-q)/(1-p))^{n-k} \geq ((1-q)/(1-p))^{n-qn}$ (in quanto, appunto, $n - k \leq n - qn$ e la funzione esponenziale a^x è

decrescente in x per $a < 1$). Usando queste due diseguaglianze otteniamo che

$$\begin{aligned}\frac{q(\mathbf{z})}{p(\mathbf{z})} &= \frac{q^k(1-q)^{n-k}}{p^k(1-p)^{n-k}} = \left(\frac{q}{p}\right)^k \cdot \left(\frac{1-q}{1-p}\right)^{n-k} \\ &\geq \left(\frac{q}{p}\right)^{qn} \cdot \left(\frac{1-q}{1-p}\right)^{n(1-q)} \\ &= 2^{\log\left(\frac{q}{p}\right)^{qn}} \cdot 2^{\log\left(\frac{1-q}{1-p}\right)^{n(1-q)}} \\ &= 2^{nq\log\frac{q}{p}} \cdot 2^{n(1-q)\log\frac{1-q}{1-p}} \\ &= 2^{n[q\log\frac{q}{p}+(1-q)\log\frac{1-q}{1-p}]} \\ &= 2^{nD(q||p)}.\end{aligned}$$

Detto in altri termini, per ogni $\mathbf{z} \in S$ vale che $p(\mathbf{z}) \leq q(\mathbf{z})2^{-nD(q||p)}$. Di conseguenza

$$\Pr\{S\} = \sum_{\mathbf{z}\in S} p(\mathbf{z}) \leq 2^{-nD(q||p)} \sum_{\mathbf{z}\in S} q(\mathbf{x}) \leq 2^{-nD(q||p)},$$

in quanto $\sum_{\mathbf{z}\in S} q(\mathbf{z}) \leq 1$, e la prova di a) è completa.

Supponiamo ora di trasmettere generiche sequenze binarie $\mathbf{x} \in \{0,1\}^n$ sul "canale" rumoroso che, come già detto, esibisce il seguente comportamento: ogni qualvolta trasmettiamo il bit 0 riceviamo 0 con probabilità pari ad un certo valore $1-p$, e riceviamo 1 con probabilità pari a p (ovvero commettiamo errore). Analogamente, ogni qualvolta trasmettiamo il bit 1 riceviamo 1 con probabilità pari ad un dato valore $1-p$, e riceviamo 0 con probabilità pari a p (ovvero commettiamo errore). Possiamo assumere, senza perdita di generalità, che la probabilità di trasmissione errata p sia tale che $0 < p < 1/2$. Vogliamo provare il seguente importante risultato.

Esistono: un codice $C = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\} \subseteq \{0,1\}^n$, $|C| = 2^m$, uno schema iniettivo di codifica $c : \{0,1\}^m \mapsto C \subseteq \{0,1\}^n$ ed una regola di decodifica $d : \{0,1\}^n \mapsto C \cup \{?\}$ tale che, $\forall \delta > 0$ valgono entrambe le seguenti due proprietà:

i) $\dfrac{m}{n} = \dfrac{\log|C|}{n} \geq 1 - h(p) - \delta = 1 - \left(p\log\dfrac{1}{p} + (1-p)\dfrac{1}{1-p}\right) - \delta;$

ii) Per ogni $\mathbf{x}_i \in C$, la probabilità di decodifica errata $\Pr\{d(\mathbf{x}_i \oplus \mathbf{Z}) \neq \mathbf{x}_i\}$ (dato che si è trasmesso $\mathbf{x}_i$) tende esponenzialmente a 0 al crescere della lunghezza n delle parole codice.

La quantità $C_p = 1 - h(p)$ è detta la capacità del canale.

Il risultato ci dice che per ottenere trasmissione affidabile ci basta trasmettere sequenze binarie lunghe $n \approx \frac{m}{C_p}$, (usando *opportune* regole di codifica/decodifica) e non è necessario che $\frac{m}{n} \xrightarrow[n\to\infty]{} 0$.

Ad es., se $p = 0.1 \Longrightarrow C_p = 0.5310 > \frac{1}{2}$, se $p = 0.25 \Longrightarrow C_p \approx \frac{1}{5}$

Occorre interpretare le sequenze in $\{0,1\}^m$ come l'informazione che vorremmo trasmettere, le $\mathbf{x} \in C$ come le sequenze binarie che effettivamente trasmettiamo (per "combattere" il rumore del canale), e le $\mathbf{x} \oplus \mathbf{z}$ come le sequenze binarie ricevute (ovvero corrotte dal rumore), dove $\mathbf{z}$ è un vettore casuale $\mathbf{z} = z_1 \ldots z_n \in \{0,1\}^n$, per cui $z_i = 1$ con probabilità pari a p e $z_i = 0$ con probabilità pari ad $1-p$, $0 < p < 1$, per ogni $i = 1, \ldots n$, indipendentemente uno dall'altro.

La ii) ci garantisce di poter determinare la generica parola $\mathbf{x}$ trasmessa originalmente sul canale, e poichè la funzione $c : \{0,1\}^m \mapsto C \subseteq \{0,1\}^n$ può assegnare ad ogni distinta sequenza in $\{0,1\}^m$ una distinta parola in C, (in quanto $|C| = 2^m$) dopo aver determinato la parola trasmessa $\mathbf{x}$ possiamo sicuramente determinare la sequenza in $\{0,1\}^m$.

La "costruzione" del codice $C = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\} \subset \{0,1\}^n$ è concettualmente semplice: scegliamo M vettori $\mathbf{x}_1, \ldots, \mathbf{x}_M \in \{0,1\}^n$ in cui ogni componente i-esima (per $i = 1, \ldots, n$) di ciascuna parola $\mathbf{x} \in C$ viene posta a 1 con probabilità pari a $1/2$ (ed analogamente viene posta a 0 con probabilità pari a $1/2$).

La regola di decodifica $d : \mathbf{y} \in \{0,1\}^n \mapsto C \cup \{?\}$ che useremo è la seguente: fissato un $\epsilon > 0$ tale che $p + \epsilon < 1/2$, ogni qualvolta riceviamo uno $\mathbf{y} \in \{0,1\}^n$ lo decodifichiamo con $\mathbf{x} \in C$ *se e solo se* $\mathbf{x}$ è *l'unico* elemento di C che differisce da $\mathbf{y}$ in al più $n(p+\epsilon)$ posizioni, poniamo $d(\mathbf{y}) =?$ altrimenti (oovero, commettiamo deliberatamente errore nella decodifica).

Sotto queste condizioni e sotto l'ipotesi di aver trasmesso la parola $\mathbf{x}_i \in C$, la decodifica sarà errata se e solo se accade *uno qualsiasi* dei seguenti eventi (o anche entrambi...):

(A) riceviamo $\mathbf{y}$ che differisce da $\mathbf{x}_i$ in almeno $n(p+\epsilon)$ posizioni;

(B) esiste almeno una parola codice $\mathbf{x}_j \in C$, $j \neq i$, che differisce dal vettore ricevuto $\mathbf{y}$ in al più $n(p+\epsilon)$ posizioni.

Quindi, siamo interessati a stimare la probabilità $\Pr\{A \cup B\}$ della loro unione, che sappiamo essere al più pari alla somma $\Pr\{A\} + \Pr\{B\}$ delle singole probabilità di ciascun evento.

Iniziamo con il valutare $\Pr\{A\}$ che corrisponde a valutare la probabilità di $\mathbf{y}$ che differisce da $\mathbf{x}_i$ in almeno $n(p+\epsilon)$ posizioni. Ricordando come opera il canale (che commette un errore con probabilità p), la $\Pr\{A\}$ è equivalente a stimare la probabilità che siano accaduti almeno $n(p+\epsilon) = nq$ errori, dove abbiamo posto $q = p + \epsilon$. Detto in altri termini, la $\Pr\{A\}$ è la probabilità che la variabile casuale $\mathbf{Z}$ assuma come valore un qualsiasi $\mathbf{z} \in \{0,1\}^n$ che contiene almeno nq "1".

Dalla a) della Diseguaglianza di Chernoff-Hoeffding otteniamo che

$$\Pr\{A\} \leq 2^{-nD(q||p)}.$$

Essendo $D(q||p) > 0$ (in quanto $q = p+\epsilon \neq p$) otteniamo che al crescere di n la $\Pr\{A\}$ va a zero (esponenzialmente in n).

Stimiamo ora la quantità $\Pr\{B\}$. Essendo l'evento B esso stesso unione di $M-1$ eventi (ognuno relativo ad una parola $\mathbf{x}_j \neq \mathbf{x}_i$) ed usando il fatto che la probabilità dell'unione di più eventi è pari al più alla somma delle probabilità dei songoli eventi, possiamo innanzitutto dire che

$$\Pr\{B\} < M \cdot \Pr\{\mathbf{x}_j \text{ differisce dal vettore } \mathbf{y} \text{ ricevuto in al più } n(p+\epsilon) \text{ posizioni}\}. \quad (1)$$

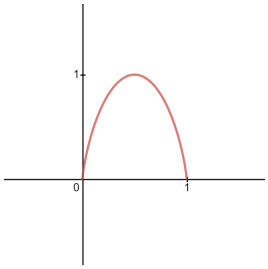
Ricordiamo che ogni bit dei vettori $\mathbf{x}_j \in C$ venivano generati con probabilità pari a $1/2$, che è anche la probabilità con cui il generico i-esimo bit di $\mathbf{x}_j$ differisca dal corrispondente bit i-esimo del vettore ricevuto $\mathbf{y}$. Pertanto, ponendo $q = (p+\epsilon)$ ed osservando che $q+\epsilon < 1/2$, possiamo usare la b) della Diseguaglianza di Chernoff-Hoeffding e la diseguaglianza (1) per limitare ulteriormente la $\Pr\{B\}$ nel modo seguente:

$$\Pr\{B\} < M \cdot 2^{-nD(q||1/2)}. \quad (2)$$

Scegliamo adesso il numero M delle parole codice pari a $M = 2^{n(1-h(p+2\epsilon))}$, dove la funzione $h(\cdot)$ è stata ricordata nella i). Sostituendo questo valore nella (2) otteniamo:

$$\begin{aligned}
\Pr\{B\} &< M \cdot 2^{-nD(q||1/2)} \\
&= 2^{n(1-h(p+2\epsilon))} \times 2^{-nD(q||1/2)} \\
&= 2^{n(1-h(p+2\epsilon))} \times 2^{-n[(p+\epsilon)\log\frac{p+\epsilon}{1/2}+(1-(p+\epsilon))\log\frac{1-(p+\epsilon)}{1/2}]} \\
&= 2^{n(1-h(p+2\epsilon))} \times 2^{-n[1-(p+\epsilon)\log\frac{1}{p+\epsilon}-(1-(p+\epsilon))\log\frac{1}{1-(p+\epsilon)}]} \\
&= 2^{n(1-h(p+2\epsilon))} \times 2^{-n[1-h(p+\epsilon)]} \\
&= 2^{n} \times 2^{(1-h(p+2\epsilon))} \times 2^{-n} \times 2^{h(p+\epsilon)} \\
&= 2^{-n[h(p+2\epsilon)-h(p+\epsilon)]}.
\end{aligned}$$

Ricordiamo ora il grafico della funzione $h(x) = -x\log x - (1-x)\log(1-x)$, di sotto riportato



da cui si evince che $h(x)$ è strettamente crescente nell'intervallo $[0, 1/2]$, il che ci permette di concludere che $h(p+2\epsilon) - h(p+\epsilon) > 0$. Pertanto, la quantità $2^{-n[h(p+2\epsilon)-h(p+\epsilon)]}$ tende esponenzialmente a 0, al crescere di n, e lo stesso accade quindi anche a $\Pr\{B\}$.

Riassumendo, abbiamo quindi visto che entrambe le quantità $\Pr\{A\}$ e $\Pr\{B\}$ tendono esponenzialmente a zero, al crescere di n, e quindi anche la probabilità di decodifica errata (che era limitata superiormente da $\Pr\{A\}+\Pr\{B\}$). Quindi la ii) è provata. D'altra parte, per come abbiamo scelto il valore di $M = |C| = 2^m$, vale che

$$\frac{m}{n} = \frac{\log|C|}{n} = 1 - h(p+2\epsilon)$$

e prendendo ϵ piccolo abbastanza proviamo anche la i) (sempre perchè la funzione h è crescente nell'intervallo $[0, 1]$).

Proviamo che meglio di come abbiamo fatto **non** si può fare. Sia $\mathbf{X}$ una vc che assume valori in $\{0,1\}^m$ con probabilità uniforme (quindi $H(\mathbf{X}) = m$) ($\mathbf{X}$ è la sorgente di informazione), $\mathbf{Z} = (Z_1, \ldots, Z_n)$, con $Z_1, \ldots, Z_n$ v.c. indipendenti ed identicamente distribuite, con $\Pr\{Z_i = 1\} = p$, che descrivono il rumore del canale, $f : \{0,1\}^m \mapsto \{0,1\}^n$ e $g : \{0,1\}^n \mapsto \{0,1\}^m$ *arbitrarie* funzioni di codifica e decodifica, e sia $\mathbf{Y} = f(\mathbf{X}) \oplus \mathbf{Z}$.

Intendiamo provare che nello schema di trasmissione sopra descritto *necessariamente* vale che:

$$\underline{Se}\ \Pr\{g(\mathbf{Y}) \neq \mathbf{X}\} \leq \epsilon\ \underline{allora}\ n \geq \frac{m(1-\epsilon)-1}{(1-h(p))} \tag{3}$$

Ricordiamo che abbiamo già provato che si può ottenere trasmissione affidabile con $n \approx \frac{m}{(1-h(p))}$.

Proviamo la (3). Lo schema della trasmissione è:

$$\underbrace{\mathbf{X}}_{m \text{ bits}} \longrightarrow \underbrace{f(\mathbf{X})}_{n \text{ bits}} \longrightarrow \underbrace{f(\mathbf{X}) \oplus \mathbf{Z}}_{\mathbf{Y}} \longrightarrow \underbrace{g(f(\mathbf{X}) \oplus \mathbf{Z})}_{g(\mathbf{Y})}$$

Dalla Diseguglianza di Fano otteniamo $H(\mathbf{X}|\mathbf{Y}) \leq h(\epsilon) + \epsilon \log(2^m - 1) \leq 1 + \epsilon \log 2^m = 1 + \epsilon m$.

Vale che $I(\mathbf{X}; \mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) \geq m - \epsilon m - 1 = m(1 - \epsilon) - 1$. Inoltre $H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{Z}) = nh(p)$ in quanto nota la $\mathbf{X}$, l'unica incertezza che ci rimane sulla $\mathbf{Y} = f(\mathbf{X}) \oplus \mathbf{Z}$ è sul valore della $\mathbf{Z}$. Quindi

$$m(1 - \epsilon) - 1 \leq I(\mathbf{X}; \mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \leq n - nh(p) = n(1 - h(p))$$

in quanto $H(\mathbf{Y}) \leq \log 2^n = n$. Abbiamo quindi provato la (3)