Codifica e ridondanza

Confronto tra flusso della sorgente F(X) e capacità del canale K(C) misurata in BPS

- se la sorgente emette informazione a una velocità superiore alla capacità del canale, **K(C)<F(X)**, allora gli effetti del rumore non sono eliminabili
- se invece **K(C)>F(X)**, il rumore presente sul canale può essere "gestito" (corretto).

Si può usare un messaggio ridondante sfruttando la capacità di canale residua

- contiene simboli che in assenza di disturbi non sarebbero necessari al suo corretto riconoscimento (adattamento al canale).
- i simboli aggiuntivi riducono l'incertezza

Esempio: introduzione della ridondanza

$$p(C \rightarrow C) = p(G \rightarrow G) = p(T \rightarrow T) = 1$$

$$p(A \rightarrow A) = 0.75\ (3/4) \qquad p(A \rightarrow C) = 0.25\ (1/4)$$

- Il 25% dei simboli A viene frainteso come C
- Supponendo che la capacità di canale lo permetta, si possono duplicare i simboli in trasmissione :

C --> CC   A --> AA   G --> GG   T --> TT

- Per A il canale avrebbe così il seguente comportamento :

$$p(A \rightarrow AA) = p(A \rightarrow A).p(A \rightarrow A) = \tfrac{3}{4} \times \tfrac{3}{4} = 9/16$$

$$p(A \rightarrow AC) = p(A \rightarrow CA) = p(A \rightarrow C).p(A \rightarrow A) = \tfrac{1}{4} \times \tfrac{3}{4} = 3/16$$

$$p(A \rightarrow CC) = p(A \rightarrow C).p(A \rightarrow C) = \tfrac{1}{4} \times \tfrac{1}{4} = 1/16$$

l'errore (che A raggiunga il destinatario come C) scende da 1/4 a 1/16 !!

Prezzo da pagare: lunghezza doppia del messaggio e quindi tempo doppio di trasmissione

Altri esempi di codifica e ridondanza

- **Bit di parità:** a una sequenza di bit si aggiunge un ulteriore bit in trasmissione per ottenere un numero pari (o dispari) di bit posti a 1

  - **Esempio:** se il segnale generato dalla codifica a lunghezza costante R=4 è **0101.1101** (es. due simboli), il segnale codificato con bit di parità sarà: **01010.11011,** quindi se il segnale diventa **11010.11011** allora l'errore si è verificato sul primo simbolo
  - su simboli di **n** bit ridondanza introdotta di **(n+1)/n** (più bassa del caso precedente **2n**)
  - **poca ridondanza:** permette di rilevare un errore ma non di correggerlo

Altri esempi

- **Linguaggi naturali**
  - I linguaggi naturali sono naturalmente ridondanti a livello sintattico e a livello semantico
  - è generalmente facile comprendere correttamente una parola scritta con grafia errata:

    "trasmixione" viene corretta in "**trasmissione"**
  - Il significato della parola nel contesto è ulteriormente utile a correggere il messaggio a livello semantico

Codici di Hamming (7,4)

Un **codeword** è un codice appartenente all'insieme delle parole lunghe 7 bit (in tutto $128=2^7$) tale che:

- **se** rappresentiamo i suoi 7 bit in tre regioni A, B e C come nella figura di sotto (il primo, terzo, quarto e sesto bit nella regione B; il secondo, terzo, quarto e quinto B nella regione A ed infine il primo, il secondo, il quarto e il settimo bit nella regione C),
- **allora**, la somma dei bit di ogni regione deve essere pari.

Diremo che due codici hanno **distanza 1** se differiscono per 1 bit.
I codeword consentono di correggere tutti gli altri codici che hanno un solo errore (distanza 1).