

In questa lezione dimostriamo che il secondo teorema di Shannon non è migliorabile. Ovvero, se vogliamo portare a zero l'errore di decodifica allora il tasso di trasmissione non può superare la capacità del canale. Cominciamo col dimostrare una serie di risultati ausiliari. Per prima cosa, enunciamo una semplice generalizzazione della chain rule per l'entropia.

**Teorema 1 (Chain rule generalizzata per l'entropia)** *Siano  $X_1, \dots, X_n$  delle variabili casuali con distribuzione congiunta  $p(x_1, \dots, x_n)$ . Allora*

$$H(X_1, \dots, X_n) = \sum_{i=1}^n H(X_i | X_1, \dots, X_{i-1}) .$$

Utilizzando questo risultato, possiamo dimostrare il seguente maggiorante sull'informazione mutua.

**Lemma 2** *Dato un canale  $\langle \mathcal{X}, \mathcal{Y}, p(y | x) \rangle$  di capacità  $C$  e detta  $p$  una distribuzione di probabilità su  $\mathcal{X}^n$ , vale che  $I(X^n, Y^n) \leq nC$ .*

DIMOSTRAZIONE.

$$\begin{aligned} I(X^n, Y^n) &= H(Y^n) - H(Y^n | X^n) \\ &= \sum_{i=1}^n \left( H(Y_i | Y_1, \dots, Y_{i-1}) - H(Y_i | Y_1, \dots, Y_{i-1}, X^n) \right) \quad \text{per la chain rule dell'entropia} \\ &\leq \sum_{i=1}^n \left( H(Y_i) - H(Y_i | X_i) \right) \quad \text{dato che il condizionamento non aumenta l'entropia} \\ &= \sum_{i=1}^n I(X_i, Y_i) \\ &\leq nC \quad \text{per definizione di capacità.} \end{aligned}$$

Nella prima disuguaglianza abbiamo usato  $H(Y_i | Y_1, \dots, Y_{i-1}, X^n) = H(Y_i | X_i)$  dato che, per la definizione di canale,  $Y_i$  dipende solo da  $X_i$ .  $\square$

Siamo pronti per dimostrare la converso al secondo teorema di Shannon. Iniziamo con un caso semplice, ovvero assumiamo che ci sia un codice tale che  $p_e^{(n)} = 0$  e deduciamo che  $R \leq C$ .

Nel seguito, adotteremo la seguente notazione per un canale  $\langle \mathcal{X}, \mathcal{Y}, p(y | x) \rangle$  ed un codice canale per  $M$  messaggi.  $W$  denota una variabile casuale con distribuzione uniforme sull'insieme dei messaggi, ovvero  $\mathbb{P}(W = i) = \frac{1}{M}$  per  $i = 1, \dots, M$ .  $X^n(W)$  indica la codifica di  $W$  e  $Y^n(W)$  denota la variabile casuale che rappresenta la sequenza di simboli ottenuta inviando  $X^n(W)$  nel canale.

Detta  $g : \mathcal{Y}^n \rightarrow \{1, \dots, M\}$  la funzione di decodifica, ricordiamo che la probabilità di errore per un codice canale è

$$p_e^{(n)} = \mathbb{P}(g(Y^n(W)) \neq W) .$$

**Teorema 3 (Conversa al secondo teorema di Shannon nel caso  $p_e^{(n)} = 0$ )** Dato un canale  $\langle \mathcal{X}, \mathcal{Y}, p(y | x) \rangle$  di capacità  $C$  e dato un codice canale di tipo  $(\lceil 2^{nR} \rceil, n)$ , se  $p_e^{(n)} = 0$  allora  $R \leq C$ .

DIMOSTRAZIONE. Sia  $g : \mathcal{Y}^n \rightarrow \{1, \dots, M\}$  la funzione di decodifica del codice canale, dove  $M = \lceil 2^{nR} \rceil$  è il numero dei messaggi. Per ipotesi,

$$\mathbb{P}(g(Y^n(W)) \neq W) = p_e^{(n)} = 0$$

quindi  $Y^n(W)$  determina  $W$  univocamente, il che implica  $H(W | Y^n(W)) = 0$ . Notiamo anche che, per definizione di canale,  $W$  e  $Y^n(W)$  sono indipendenti dato  $X^n(W)$ . Quindi, la data processing inequality si applica e fornisce  $I(X^n(W), Y^n(W)) \geq I(W, Y^n(W))$ . Usando questi fatti possiamo scrivere

$$\begin{aligned} nR &\leq \log_2 M \\ &= H(W) \\ &= \underbrace{H(W | Y^n(W))}_{=0} + I(W, Y^n(W)) \\ &\leq I(X^n(W), Y^n(W)) \quad \text{per la data processing inequality (vedi sopra)} \\ &\leq nC \quad \text{per il Lemma 2.} \end{aligned}$$

Quindi concludiamo che  $R \leq C$ . □

Passiamo ora a dimostrare il risultato più forte. Ovvero, dimostriamo che se esiste una sequenza di codici tale che  $p_e^{(n)} \rightarrow 0$  allora, di nuovo, deve valere che  $R \leq C$ .

**Teorema 4 (Conversa al secondo teorema di Shannon nel caso  $p_e^{(n)} \rightarrow 0$ )** Dato un canale  $\langle \mathcal{X}, \mathcal{Y}, p(y | x) \rangle$  di capacità  $C$  e data una sequenza di codici canale di tipo  $(\lceil 2^{nR} \rceil, n)$  per  $n = 1, 2, \dots$ , se  $p_e^{(n)} \rightarrow 0$  per  $n \rightarrow \infty$  allora  $R \leq C$ .

DIMOSTRAZIONE. Dato che  $p_e^{(n)} > 0$  non possiamo dire che  $H(W | Y^n(W)) = 0$  come nel caso precedente. Però possiamo applicare la disuguaglianza di Fano alle variabili  $W$  e  $Y^n(W)$  per dedurre che

$$p_e^{(n)} \geq \frac{H(W | Y^n(W)) - 1}{nR} .$$

Sfruttando questo, possiamo scrivere

$$\begin{aligned} nR &\leq \log_2 M \\ &= H(W) \\ &= H(W | Y^n(W)) + I(W, Y^n(W)) \\ &\leq 1 + p_e^{(n)}nR + I(W, Y^n(W)) \quad \text{per la disuguaglianza di Fano} \\ &\leq 1 + p_e^{(n)}nR + I(X^n(W), Y^n(W)) \quad \text{per la data processing inequality} \\ &\leq 1 + p_e^{(n)}nR + nC \quad \text{per il Lemma 2.} \end{aligned}$$

Dividendo entrambi i membri per  $n$  otteniamo

$$R \leq \frac{1}{n} + p_e^{(n)}R + C .$$

Dato che la disuguaglianza sopra vale per ogni  $n$ , sfruttando l'ipotesi che  $p_e^{(n)} \rightarrow 0$  per  $n \rightarrow \infty$  otteniamo la tesi  $R \leq C$ . □