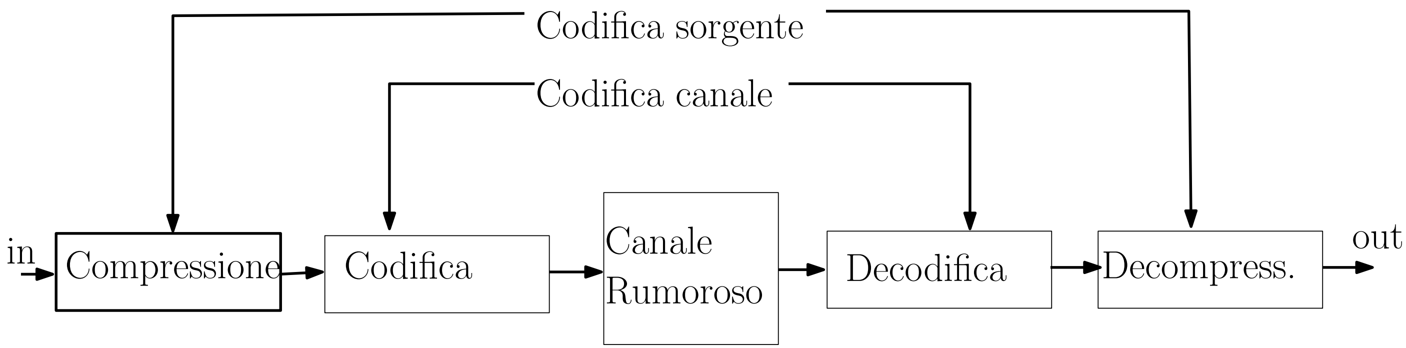
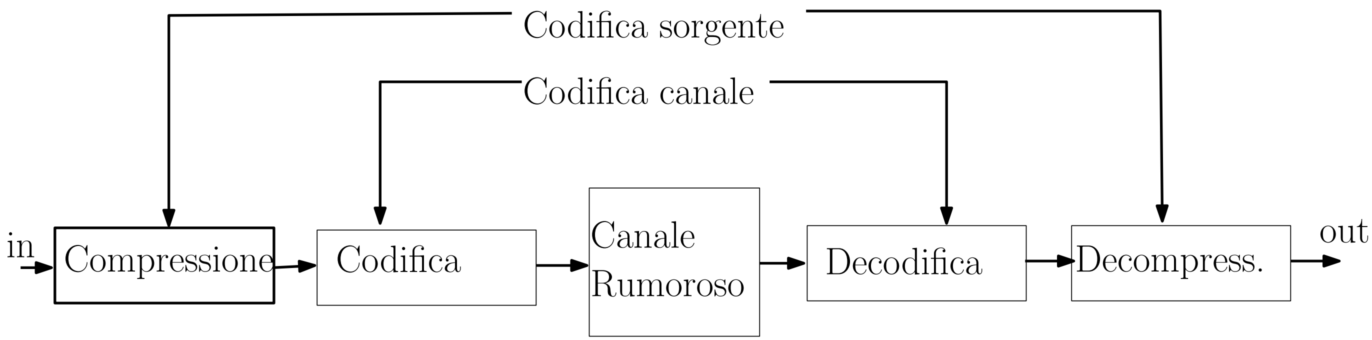
Codifica sorgente e canale



- Codifica sorgente: comprimere i dati per rimuovere ridondanza
- Codifica canale: aggiungere ridondanza per proteggere da errori di trasmissione sul canale
- Comunicazione con successo: out=in

Codifica canale



- Es. effetto rumore.
  Input possibili al canale: sequenze 101 e 111,
  input 101, rumore: secondo bit modificato, output: 111;
  input: 111, rumore: nessuno, output: 111;
  $\Rightarrow$ diverse sequenze in input producono lo stesso output (input confondibili).
- Obiettivo: proteggere l'informazione da eventuali errori di trasmissione legati al rumore

Codifica canale

- Obiettivo: Input NON Confondibili $\equiv$ correzione errori trasmissione

Codifica canale

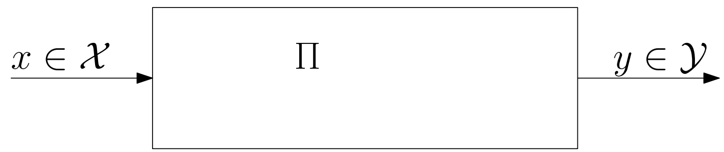
- Obiettivo: Input NON Confondibili $\equiv$ correzione errori trasmissione
- Metodo: Aggiungere ridondanza

Codifica canale

- Obiettivo: Input NON Confondibili $\equiv$ correzione errori trasmissione
- Metodo: Aggiungere ridondanza
- Nota: impossibile eliminare effetto rumore
  vogliamo input non confondibili con alta probabilità

Canali discreti senza memoria

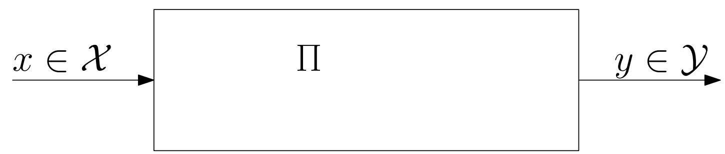
- Canale discreto (alfabeti I/O discreti): $(\mathcal{X}, \Pi, \mathcal{Y})$



$\mathcal{X}$= alfabeto input al canale
$\mathcal{Y}$= alfabeto output al canale
$\Pi = [p(y/x)]$= matrice delle probabilità di transizione

Canali discreti senza memoria

- Canale discreto (alfabeti I/O discreti): $(\mathcal{X}, \Pi, \mathcal{Y})$



  $\mathcal{X}$= alfabeto input al canale
  $\mathcal{Y}$= alfabeto output al canale
  $\Pi = [p(y/x)]$= matrice delle probabilità di transizione

- Canale discreto senza memoria (DMC) (X, Π, Y): probabilità output dipende solo da input corrispondente NON da precedenti input o output.

Capacità per n usi del canale

$$C^{(n)} = \frac{1}{n} \max_{p(x_1 \ldots x_n)} I(X_1 \ldots X_n; Y_1 \ldots Y_n)$$

Capacità per n usi del canale

$$C^{(n)} = \frac{1}{n} \max_{p(x_1 \ldots x_n)} I(X_1 \ldots X_n; Y_1 \ldots Y_n)$$

Per canale discreto senza memoria (DMC)

$$\begin{aligned} I(X_1 \ldots X_n; Y_1 \ldots Y_n) &= H(Y_1 \ldots Y_n) - H(Y_1 \ldots Y_n / X_1 \ldots X_n) \\ &= \sum_{i=1}^{n} H(Y_i / Y_1 \ldots Y_{i-1}) - \sum_{i=1}^{n} H(Y_i / X_1 \ldots X_n, Y_1 \ldots Y_{i-1}) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i / X_i) \qquad \text{regola catena+DMC} \\ &\leq \sum_{i=1}^{n} I(X_i; Y_i) \end{aligned}$$

Capacità per n usi del canale

$$C^{(n)} = \frac{1}{n} \max_{p(x_1 \ldots x_n)} I(X_1 \ldots X_n; Y_1 \ldots Y_n)$$

Per canale discreto senza memoria (DMC)

$$\begin{aligned}
I(X_1 \ldots X_n; Y_1 \ldots Y_n) &= H(Y_1 \ldots Y_n) - H(Y_1 \ldots Y_n / X_1 \ldots X_n) \\
&= \sum_{i=1}^{n} H(Y_i / Y_1 \ldots Y_{i-1}) - \sum_{i=1}^{n} H(Y_i / X_1 \ldots X_n, Y_1 \ldots Y_{i-1}) \\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i / X_i) \qquad \text{regola catena+DMC} \\
&\leq \sum_{i=1}^{n} I(X_i; Y_i)
\end{aligned}$$

Max $I(X_1 \ldots X_n; Y_1 \ldots Y_n)$ massimizzando ogni $I(X_i; Y_i)$

Ci concentreremo su $\max I(X, Y) \equiv$ singolo uso canale

Capacità Canale

- Capacità del canale senza memoria:

$$C = \max_{p(x)} I(X;Y)$$

- Dimostreremo:
  capacità = massimo numero di bit (di informazione) che possono essere trasmessi per ogni uso del canale.

Canale senza rumore

$$0 \longrightarrow 0$$

$$1 \longrightarrow 1$$

- Trasmette un bit per uso senza errore $\Rightarrow \quad C = 1$

Canale senza rumore

$$0 \longrightarrow 0$$

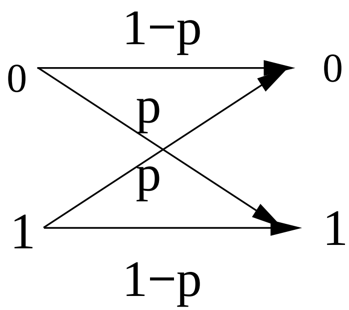
$$1 \longrightarrow 1$$

- Trasmette un bit per uso senza errore $\Rightarrow \quad C = 1$
- Infatti

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(X) - H(X/Y) = \max_{p(x)} H(X) = 1$$

per $p(x) = (1/2, 1/2)$

Canale binario simmetrico

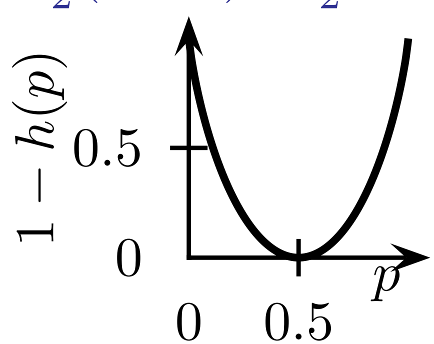


$$\Pi = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y/X) = H(Y) - \sum_x p(x) H(Y/X=x) \\ &= H(Y) - \sum_x p(x) h(p) = H(Y) - h(p) \\ C &= max_{p(X)} I(X;Y) = max_{p(X)} H(Y) - h(p) \leq 1 - h(p) \end{aligned}$$

$$p(X) = (1/2, 1/2) \Rightarrow p(y=1) = \tfrac{1}{2}(1-p) + \tfrac{1}{2}p = \tfrac{1}{2} \Rightarrow H(Y) = 1$$

Quindi $C = 1 - h(p)$bits



Canali Simmetrici

- Canale simmetrico: Ogni riga (risp. colonna) é permutazione di ogni altra riga (risp. colonna)

  Es. $\Pi = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$

Canali Simmetrici

- Canale simmetrico: Ogni riga (risp. colonna) é permutazione di ogni altra riga (risp. colonna)

  Es. $\Pi = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$

- Canale debolmente simmetrico:
  Ogni riga é permutazione di ogni altra riga;
  la somma su ogni colonna é costante

  Es. $\Pi = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix}$

Canali Simmetrici

- Canale simmetrico: Ogni riga (risp. colonna) é permutazione di ogni altra riga (risp. colonna)

  Es. $\Pi = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$

- Canale debolmente simmetrico:
  Ogni riga é permutazione di ogni altra riga;
  la somma su ogni colonna é costante

  Es. $\Pi = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix}$

- Canale simmetrico $\Rightarrow$ Canale debolmente simmetrico

Canali (Debolmente) Simmetrici

- $\mathbf{r}$=riga di Π

$$I(X;Y) = H(Y) - H(Y/X) = H(Y) - H(\mathbf{r}) \leq \log |Y| - H(\mathbf{r})$$

Canali (Debolmente) Simmetrici

- $\mathbf{r}$=riga di Π

$$I(X;Y) = H(Y) - H(Y/X) = H(Y) - H(\mathbf{r}) \leq \log|Y| - H(\mathbf{r})$$

- Ponendo $p(x) = 1/|X|$ per ogni $x \in X$ risulta

$$p(y) = \sum_x p(x)p(y/x) = \sum_x \frac{p(y/x)}{|X|} = \frac{\text{somma colonna}}{|X|} = \frac{1}{|Y|}$$

Canali (Debolmente) Simmetrici

- $\mathbf{r}$=riga di Π

$$I(X;Y) = H(Y) - H(Y/X) = H(Y) - H(\mathbf{r}) \leq \log|Y| - H(\mathbf{r})$$

- Ponendo $p(x) = 1/|X|$ per ogni $x \in X$ risulta

$$p(y) = \sum_x p(x)p(y/x) = \sum_x \frac{p(y/x)}{|X|} = \frac{\text{somma colonna}}{|X|} = \frac{1}{|Y|}$$

- Quindi $C = \log|Y| - H(\mathbf{r})$

Canali (Debolmente) Simmetrici

- $\mathbf{r}$=riga di Π

$$I(X;Y) = H(Y) - H(Y/X) = H(Y) - H(\mathbf{r}) \leq \log|Y| - H(\mathbf{r})$$

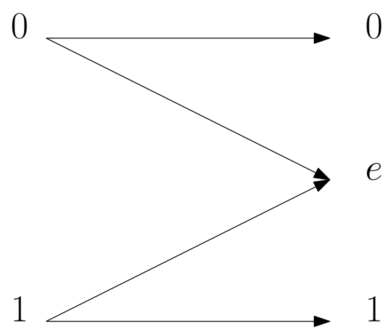
- Ponendo $p(x) = 1/|X|$ per ogni $x \in X$ risulta

$$p(y) = \sum_x p(x)p(y/x) = \sum_x \frac{p(y/x)}{|X|} = \frac{\text{somma colonna}}{|X|} = \frac{1}{|Y|}$$

- Quindi $C = \log|Y| - H(\mathbf{r})$

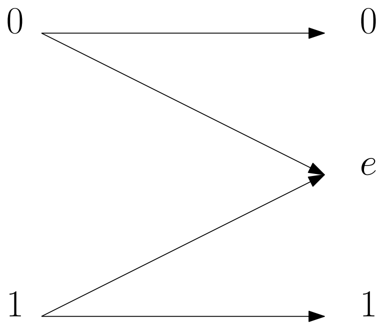
Es. $\Pi = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix}$ $C = \log 3 - H(\frac{1}{3}, \frac{1}{6}, \frac{1}{2}) = \frac{1}{2}\log 3 - \frac{2}{3}$

Canale binario con cancellazioni



$$\Pi = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix}$$

Canale binario con cancellazioni

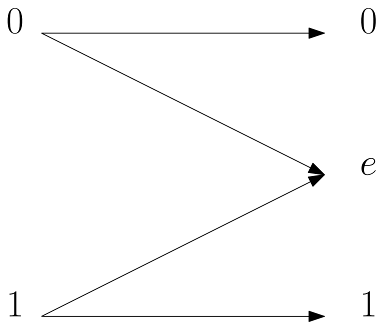


$$\Pi = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix}$$

Sia $p(X=0)=p$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y/X) \\ &= H(Y) - pH(Y/X=0) - (1-p)H(Y/X=1) \\ &= H(Y) - h(\alpha) \end{aligned}$$

Canale binario con cancellazioni



$$\Pi = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix}$$

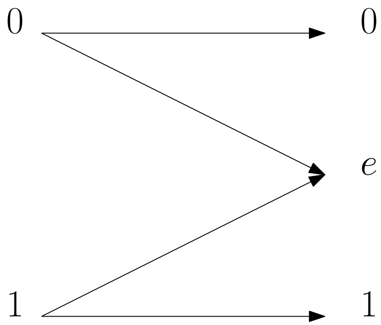
Sia $p(X=0) = p$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y/X) \\ &= H(Y) - pH(Y/X=0) - (1-p)H(Y/X=1) \\ &= H(Y) - h(\alpha) \end{aligned}$$

$p(Y=0) = p(1-\alpha)$, $P(Y=e) = \alpha$

$H(Y) - h(\alpha) = H(p(1-\alpha), (1-p)(1-\alpha), \alpha) - h(\alpha) = (1-\alpha)h(p)$

Canale binario con cancellazioni



$$\Pi = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix}$$

Sia $p(X=0) = p$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y/X) \\ &= H(Y) - pH(Y/X=0) - (1-p)H(Y/X=1) \\ &= H(Y) - h(\alpha) \end{aligned}$$

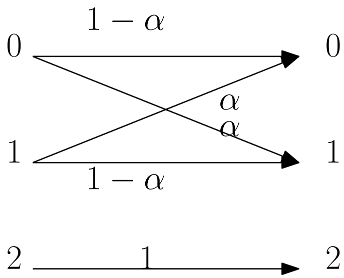
$p(Y=0) = p(1-\alpha)$, $P(Y=e) = \alpha$

$H(Y) - h(\alpha) = H(p(1-\alpha), (1-p)(1-\alpha), \alpha) - h(\alpha) = (1-\alpha)h(p)$

$C = \max_p (1-\alpha)h(p) = 1-\alpha,$ per $p = 1/2$

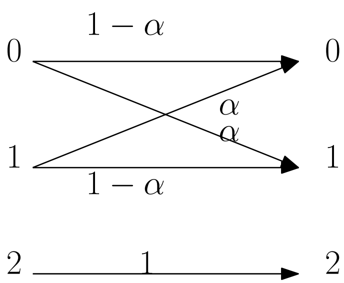
Frazione α bit cancellati $\Rightarrow$ mumero medio bit di info trasmessi é $1-\alpha$.

Canale asimmetrico



$$\Pi = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \\ 0 & 0 & 10 \end{bmatrix}$$

Canale asimmetrico

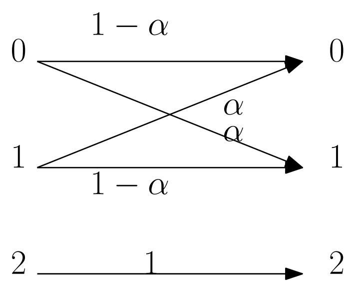


$$\Pi = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \\ 0 & 0 & 10 \end{bmatrix}$$

Sia $p(X) = (p, p, 1-2p) \Rightarrow P(Y) = (p, p, 1-2p)$

$$\begin{aligned} H(Y/X) &= 2ph(\alpha) + (1-2p)h(1) = 2ph(\alpha) \\ H(Y) &= H(p, p, 1-2p) = -2p\log p - (1-2p)\log(1-2p) \end{aligned}$$

Canale asimmetrico



$$\Pi = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \\ 0 & 0 & 10 \end{bmatrix}$$

Sia $p(X) = (p, p, 1-2p) \Rightarrow P(Y) = (p, p, 1-2p)$

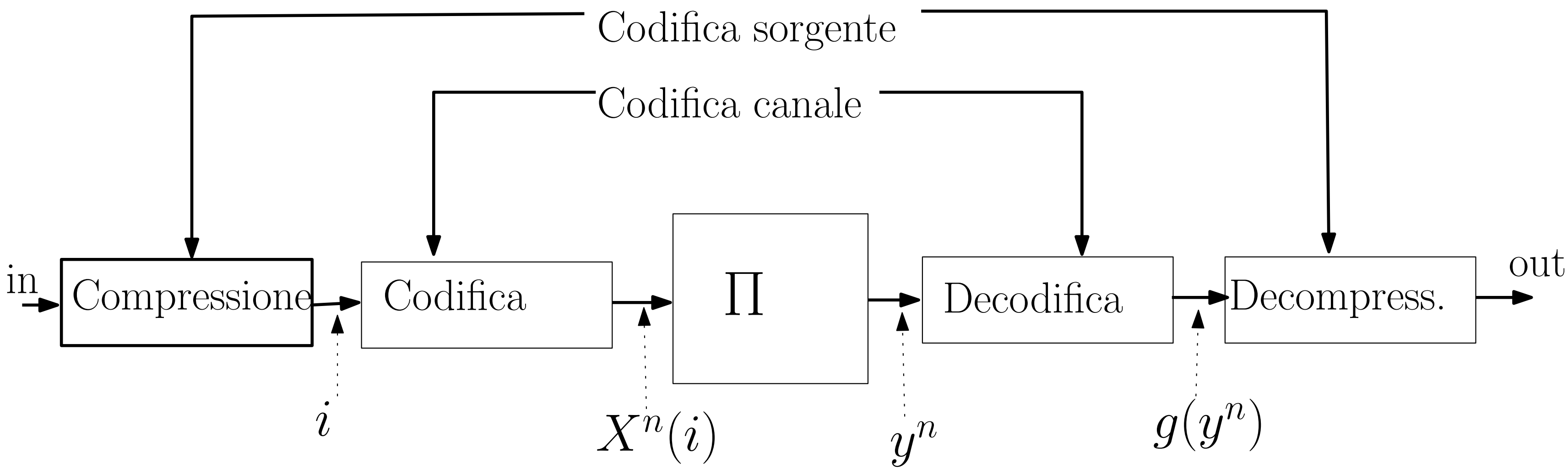
$$\begin{aligned} H(Y/X) &= 2ph(\alpha) + (1-2p)h(1) = 2ph(\alpha) \\ H(Y) &= H(p, p, 1-2p) = -2p\log p - (1-2p)\log(1-2p) \end{aligned}$$

Per trovare C massimizziamo

$$\begin{aligned} f(p) &= H(Y) - H(Y/X) = -2p\log p - (1-2p)\log(1-2p) - 2ph(\alpha) \\ f'(p) &= -2\log e - 2\log p + 2\log e + 2\log(1-2p) - 2h(\alpha) = 0 \\ h(\alpha) &= -\log p + \log(1-2p) \end{aligned}$$

$$C = -2p\log p - (1-2p)\log(1-2p) + (2p\log p - 2p\log(1-2p)) = \log(1-2p)$$

Codifica Canale DMC



Codice canale (M, n) per $(\mathcal{X}, \Pi, \mathcal{Y})$:

- Insieme di indici $\{1, \ldots, M\}$ ($\equiv$ possibili sequenze input)
- Funzione codifica $X^n : \{1, \ldots, M\} \rightarrow \mathcal{X}^n$
- Funzione decodifica $g : \mathcal{Y}^n \rightarrow \{1, \ldots, M\}$

Codifica Canale DMC

- Probabilitá di errore quando si codifica indice i:

$$\lambda_i = Pr\{g(Y^n) \neq i / X^n = i\} = \sum_{y^n} p(y^n / X^n(i)) I(g(y^n) \neq i)$$

Codifica Canale DMC

- Probabilitá di errore quando si codifica indice i:

$$\lambda_i = Pr\{g(Y^n) \neq i/X^n = i\} = \sum_{y^n} p(y^n/X^n(i)) I(g(y^n) \neq i)$$

- Probabilitá massima di errore:

$$\lambda^{(n)} = \max_{1 \leq i \leq M} \lambda_i$$

Codifica Canale DMC

- Probabilitá di errore quando si codifica indice i:

$$\lambda_i = Pr\{g(Y^n) \neq i/X^n = i\} = \sum_{y^n} p(y^n/X^n(i)) I(g(y^n) \neq i)$$

- Probabilitá massima di errore:

$$\lambda^{(n)} = \max_{1 \leq i \leq M} \lambda_i$$

- Probabilitá media di errore:

$$P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$$

Codifica Canali Rumorosi

- Tasso codice (M, n): $R = \frac{\log M}{n} \frac{\text{bit}}{\text{trasm.}}$

Codifica Canali Rumorosi

- Tasso codice (M, n): $R = \frac{\log M}{n} \frac{\text{bit}}{\text{trasm.}}$
- Tasso R é ottenibile se esiste sequenza di codici $(\lceil 2^{nR} \rceil, n)$ con $\lambda^{(n)} \to 0$ per $n \to \infty$

Codifica Canali Rumorosi

- Tasso codice (M, n): $R = \frac{\log M}{n} \frac{\text{bit}}{\text{trasm.}}$
- Tasso R é ottenibile se esiste sequenza di codici $(\lceil 2^{nR} \rceil, n)$ con $\lambda^{(n)} \to 0$ per $n \to \infty$
- Si dimostra che: (Teorema di Codifica Canale)
  (a) ogni tasso $R \leq C$ é ottenibile,
  (b) nessun tasso $R > C$ é ottenibile

Codifica Canali Rumorosi

- Tasso codice (M, n): $R = \frac{\log M}{n} \frac{\text{bit}}{\text{trasm.}}$
- Tasso R é ottenibile se esiste sequenza di codici $(\lceil 2^{nR} \rceil, n)$ con $\lambda^{(n)} \to 0$ per $n \to \infty$
- Si dimostra che: (Teorema di Codifica Canale)
  (a) ogni tasso $R \leq C$ é ottenibile,
  (b) nessun tasso $R > C$ é ottenibile
- Capacitá = limite superiore di tutti i tassi ottenibili

Coppie Tipiche

Definizione L'insieme $A_\epsilon^{(n)}$ delle *coppie tipiche* $\{(x^n, y^n)\}$ rispetto alla distribuzione $p(x,y)$ è definito da:

$$
\begin{aligned}
A_\epsilon^{(n)} \;=\; \Bigg\{ & \;(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \\
& \left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon, \\
& \left| -\frac{1}{n} \log p(y^n) - H(Y) \right| < \epsilon, \\
& \left| -\frac{1}{n} \log p(x^n, y^n) - H(XY) \right| < \epsilon \Bigg\}
\end{aligned}
$$

dove $p(x^n, y^n) = \prod_{i=1}^n p(x_i, x_j)$.

Teorema Siano (X^n, Y^n) seqeuenze di v.c. i.i.d secondo $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, x_j)$. Allora,

1. $Pr((X^n, Y^n) \in A_\epsilon^{(n)}) \to 1$ per $n \to \infty$.

Joint AEP

Teorema Siano (X^n, Y^n) seqeuenze di v.c. i.i.d secondo $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, x_j)$. Allora,

1. $Pr((X^n, Y^n) \in A_\epsilon^{(n)}) \to 1$ per $n \to \infty$.
2. $(1 - e)2^{n(H(X,Y) - \epsilon)} \leq |A_\epsilon^{(n)}| \leq 2^{n(H(X,Y) + \epsilon)}$

Joint AEP

Teorema Siano (X^n, Y^n) seqeuenze di v.c. i.i.d secondo $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, x_j)$. Allora,

1. $Pr((X^n, Y^n) \in A_\epsilon^{(n)}) \to 1$ per $n \to \infty$.
2. $(1-e)2^{n(H(X,Y)-\epsilon)} \leq |A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$
3. Se $\tilde{X^n}$ e $\tilde{Y^n}$ sono indipendenti e scelte secondo $p(x^n)$ e $p(y^n)$, quindi $Pr(\tilde{x^n}, \tilde{y^n}) = p(\tilde{x^n})p(\tilde{y^n})$, allora per $n \to \infty$

$$(1-e)2^{-n(I(X;Y)+3\epsilon)} \leq Pr((\tilde{X^n}, \tilde{X^n}) \in A_\epsilon^{(n)}) \leq 2^{-n(I(X;Y)-3\epsilon)}.$$

Teorema di Codifica Canale

Teorema Per ogni tasso $R < C$, esiste una sequenza di codici $(2^{nR}, n)$ con probabilità massima di errore $\lambda^{(n)} \longrightarrow 0$.

Teorema di Codifica Canale

Teorema Per ogni tasso $R < C$, esiste una sequenza di codici $(2^{nR}, n)$ con probabilità massima di errore $\lambda^{(n)} \to 0$.

Dim. La dimostrazione si basa sulle seguenti idee

- analisi di sequenze lunghe, in modo da sfruttare la legge dei grandi numeri e, specificamente, le proprietà delle coppie tipiche.

Teorema di Codifica Canale

Teorema Per ogni tasso $R < C$, esiste una sequenza di codici $(2^{nR}, n)$ con probabilità massima di errore $\lambda^{(n)} \to 0$.

Dim. La dimostrazione si basa sulle seguenti idee

- analisi di sequenze lunghe, in modo da sfruttare la legge dei grandi numeri e, specificamente, le proprietà delle coppie tipiche.
- Calcolo della probabilità di errore mediata su una scelta random del codice.

Codifica Canale - Parte Diretta

Generiamo 2^{nR} parole codice i.i.d. scegliendone i simboli da $\mathcal{X}$ indipendentemente in accordo ad una fissata d.p. $p(x)$.

- una sequenza x^n è scelta con probabilità $p(x^n) = \prod_{i=1}^{n} p(x_i)$

Codifica Canale - Parte Diretta

Generiamo 2^{nR} parole codice i.i.d. scegliendone i simboli da $\mathcal{X}$ indipendentemente in accordo ad una fissata d.p. $p(x)$.

- una sequenza x^n è scelta con probabilità $p(x^n) = \prod_{i=1}^{n} p(x_i)$
- un codice

$$\mathcal{C} = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{nR}) & x_2(2^{nR}) & \dots & x_n(2^{nR}) \end{bmatrix}.$$

con probabilità $Pr(\mathcal{C}) = \prod_{w=1}^{2^{nR}} \prod_{i=1}^{n} p(x_i(w))$.

Codifica Canale - Il Modello

Consideriamo il seguente modello

- Il codice viene scelto in maniera random (vedi sopra)

Codifica Canale - Il Modello

Consideriamo il seguente modello

- Il codice viene scelto in maniera random (vedi sopra)
- il messaggio W da trasmettere viene scelto uniformemente a caso: $Pr(W = w) = 2^{-nR}$, per ogni $w = 1, 2, \ldots, 2^{nR}$.

Codifica Canale - Il Modello

Consideriamo il seguente modello

- Il codice viene scelto in maniera random (vedi sopra)
- il messaggio W da trasmettere viene scelto uniformemente a caso: $Pr(W = w) = 2^{-nR}$, per ogni $w = 1, 2, \ldots, 2^{nR}$.
- La parola codice $X^n(w)$, corrispondente alla w-esima riga di $\mathcal{C}$ viene spedita sul canale.

Codifica Canale - Il Modello

Consideriamo il seguente modello

- Il codice viene scelto in maniera random (vedi sopra)
- il messaggio W da trasmettere viene scelto uniformemente a caso: $Pr(W = w) = 2^{-nR}$, per ogni $w = 1, 2, \ldots, 2^{nR}$.
- La parola codice $X^n(w)$, corrispondente alla w-esima riga di $\mathcal{C}$ viene spedita sul canale.
- L'output del canale è una sequenza Y^n determinata in accordo alla distribuzione

$$P(y^n|x^n(w)) = \prod_{i=1}^{n} p(y_i|x_i(w)).$$

- La sequenza Y^n viene decodificata come $\tilde{W}$ se
  - $(X^n(\tilde{W}), Y^n)$ formano coppia tipica
  - Non esiste un altro messaggio k t.c. $(X^n(k), Y^n)$ formano coppia tipica.

- La sequenza Y^n viene decodificata come $\tilde{W}$ se
  - $(X^n(\tilde{W}), Y^n)$ formano coppia tipica
  - Non esiste un altro messaggio k t.c. $(X^n(k), Y^n)$ formano coppia tipica.
- se non esiste un tale W o ce ne è più di uno, si emette un segnale di errore.

- La sequenza Y^n viene decodificata come $\tilde{W}$ se
  - $(X^n(\tilde{W}), Y^n)$ formano coppia tipica
  - Non esiste un altro messaggio k t.c. $(X^n(k), Y^n)$ formano coppia tipica.
- se non esiste un tale W o ce ne è più di uno, si emette un segnale di errore.
- Dichiariamo la codifica errata se $\tilde{W} \neq W$, e denotiamo con $\mathcal{E}$ tale evento.

La probabilità di errore. La calcoliamo mediata su tutte le parole del codice, e mediata su tutti i codici possibili:

$$
\begin{aligned}
Pr(\mathcal{E}) &= \sum_{\mathcal{C}} P(\mathcal{C}) P_e^{(n)}(\mathcal{C}) \\
&= \sum_{\mathcal{C}} P(\mathcal{C}) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(\mathcal{C}) \\
&= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_{\mathcal{C}} P(\mathcal{C}) \lambda_w(\mathcal{C})
\end{aligned}
$$

- Poiché mediamo su tutti i codici

$$\sum_{\mathcal{C}} P(\mathcal{C})\lambda_w(\mathcal{C})$$

non dipende da w. Infatti, guardando a tutti codici, la stessa parola appare lo stesso numero di volte con ogni indice.

- Quindi possiamo assumere, senza perdita di generalità che l'indice del messaggio inviato sia $W = 1$, poiché

$$\begin{aligned} P(\mathcal{C}) &= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_{\mathcal{C}} P(\mathcal{C})\lambda_w(\mathcal{C}) = \sum_{\mathcal{C}} P(\mathcal{C})\lambda_1(\mathcal{C}) \\ &= Pr(\mathcal{E}|W = 1). \end{aligned}$$

Sia Y^n la sequenza output quando $X^n(1)$ viene trasmesso (codifichiamo $W = 1$).

- Definiamo $\forall i,$ l'evento "*l' i-esima parola codice e Y^n formano coppia tipica*":

$$E_i = \{(X^n(i), Y^n) \in A_\epsilon^{(n)}\},$$

Sia Y^n la sequenza output quando $X^n(1)$ viene trasmesso (codifichiamo $W = 1$).

- Definiamo $\forall i,$ l'evento "*l' i-esima parola codice e Y^n formano coppia tipica*":

$$E_i = \{(X^n(i), Y^n) \in A_\epsilon^{(n)}\},$$

- Per la decodifica scelta, quando $X^n(1)$ viene trasmessa, si ha errore se una si verifica tra:
  - $(X^n(i), Y^n) \in A_\epsilon^{(n)}$, $i \neq 1$: l' evento E_i;
  - $(X^n(1), Y^n) \notin A_\epsilon^{(n)}$: l'evento $\overline{E_1}$.

- Quindi

$$\begin{aligned} P(\mathcal{E}) &= Pr(\mathcal{E}|W=1) = P(\overline{E_1} \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}) \\ &\leq P(\overline{E_1}) + \sum_{i=2}^{2^{nR}} P(E_i). \end{aligned}$$

- Quindi

$$\begin{aligned} P(\mathcal{E}) &= Pr(\mathcal{E}|W=1) = P(\overline{E_1} \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}) \\ &\leq P(\overline{E_1}) + \sum_{i=2}^{2^{nR}} P(E_i). \end{aligned}$$

- $P(\overline{E_1}) \leq \epsilon$, per $n \to \infty$ (joint AEP 1.);

- Quindi

$$\begin{aligned} P(\mathcal{E}) &= Pr(\mathcal{E}|W=1) = P(\overline{E_1} \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}) \\ &\leq P(\overline{E_1}) + \sum_{i=2}^{2^{nR}} P(E_i). \end{aligned}$$

- $P(\overline{E_1}) \leq \epsilon$, per $n \to \infty$ (joint AEP 1.);
- $X^n(1)$ e $X^n(i)$ indipendenti $\Rightarrow Y^n$ e $X^n(i)$ indipendenti, $\forall i \neq 1$.

- Quindi

$$
\begin{aligned}
P(\mathcal{E}) &= Pr(\mathcal{E}|W=1) = P(\overline{E_1} \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}) \\
&\leq P(\overline{E_1}) + \sum_{i=2}^{2^{nR}} P(E_i).
\end{aligned}
$$

- $P(\overline{E_1}) \leq \epsilon$, per $n \to \infty$ (joint AEP 1.);
- $X^n(1)$ e $X^n(i)$ indipendenti $\Rightarrow Y^n$ e $X^n(i)$ indipendenti, $\forall i \neq 1$.
- $\Rightarrow P(E_i) \leq 2^{-n(I(X;Y)-3\epsilon)}$ (joint AEP 3.).

Otteniamo

$$
\begin{aligned}
P(\mathcal{E}) &\leq P(\overline{E_1}) + \sum_{i=2}^{2^{nR}} P(E_i) \\
&\leq \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)} \\
&= \epsilon + (2^{nR}-1)2^{-n(I(X;Y)-3\epsilon)} \\
&\leq \epsilon + 2^{3n\epsilon}2^{-n(I(X;Y)-R)} \\
&\leq 2\epsilon,
\end{aligned}
$$

se scegliamo n sufficientemente grande e $R < I(X;Y) - 3\epsilon$.

- Se $R < I(X;Y)$, possiamo scegliere ϵ e n in modo da rendere la media (su tutti i codici) di $P_e^{(n)} < 2\epsilon$.
- Che possiamo dire della probabilità massima di errore?

- scegliamo $p(x) = p^*(x) = \max_{p(x)} I(X;Y)$ cioè quella che ottiene la capacità

- scegliamo $p(x) = p^*(x) = \max_{p(x)} I(X;Y)$ cioè quella che ottiene la capacità
- quindi possiamo sostituire $R < C$ ad $R < I(X;Y)$.

- scegliamo $p(x) = p^*(x) = \max_{p(x)} I(X;Y)$ cioè quella che ottiene la capacità
- quindi possiamo sostituire $R < C$ ad $R < I(X;Y)$.
- Se la media (su tutti i codici) di $P_e^{(n)}(\mathcal{C})$ è $\leq 2\epsilon$, allora esiste un codice $\mathcal{C}^*$ tale che $P_e^{(n)}(\mathcal{C}) \leq 2\epsilon$.

- scegliamo $p(x) = p^*(x) = \max_{p(x)} I(X;Y)$ cioè quella che ottiene la capacità
- quindi possiamo sostituire $R < C$ ad $R < I(X;Y)$.
- Se la media (su tutti i codici) di $P_e^{(n)}(\mathcal{C})$ è $\leq 2\epsilon$, allora esiste un codice $\mathcal{C}^*$ tale che $P_e^{(n)}(\mathcal{C}) \leq 2\epsilon$.
- eliminiamo da $\mathcal{C}^*$ ogni parola i con $\lambda_i > 4\epsilon$ (sono meno della metá, altr. $P_e^{(n)}(\mathcal{C}) > \frac{1}{2^{nR}} \frac{2^{nR}}{2} 4\epsilon = 2\epsilon$)

- scegliamo $p(x) = p^*(x) = \max_{p(x)} I(X;Y)$ cioè quella che ottiene la capacità
- quindi possiamo sostituire $R < C$ ad $R < I(X;Y)$.
- Se la media (su tutti i codici) di $P_e^{(n)}(\mathcal{C})$ è $\leq 2\epsilon$, allora esiste un codice $\mathcal{C}^*$ tale che $P_e^{(n)}(\mathcal{C}) \leq 2\epsilon$.
- eliminiamo da $\mathcal{C}^*$ ogni parola i con $\lambda_i > 4\epsilon$ (sono meno della metá, altr. $P_e^{(n)}(\mathcal{C}) > \frac{1}{2^{nR}} \frac{2^{nR}}{2} 4\epsilon = 2\epsilon$)
- allora

$$2\epsilon \geq \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \lambda_i(\mathcal{C}^*) \Rightarrow \exists (i_1, \ldots, i_{2^{nR-1}}) \ s.t. \ \lambda_{i_j}(\mathcal{C}^*) \leq 4\epsilon.$$

Creiamo un nuovo codice che contiene solo tali parole di $\mathcal{C}^*$ con prob. di errore piccola

- $\tilde{\mathcal{C}}^* = \{X^n(i_j) \in \mathcal{C}^* \mid j = 1, 2, \ldots, 2^{nR-1}\}$.

Creiamo un nuovo codice che contiene solo tali parole di $\mathcal{C}^*$ con prob. di errore piccola

- $\tilde{\mathcal{C}}^* = \{X^n(i_j) \in \mathcal{C}^* \mid j = 1, 2, \ldots, 2^{nR-1}\}$.
- Tale codice contiene ovviamente 2^{nR-1} parole, quindi il suo tasso è $R - \frac{1}{n}$ che per n grande non differisce significativamente da R.

Creiamo un nuovo codice che contiene solo tali parole di $\mathcal{C}^*$ con prob. di errore piccola

- $\tilde{\mathcal{C}}^* = \{X^n(i_j) \in \mathcal{C}^* \mid j = 1, 2, \ldots, 2^{nR-1}\}$.
- Tale codice contiene ovviamente 2^{nR-1} parole, quindi il suo tasso è $R - \frac{1}{n}$ che per n grande non differisce significativamente da R.
- Concludendo: per ogni $R < C$, possiamo scegliere un codice di tasso $R' = R - \frac{1}{n}$, con probabilità massima di errore $\lambda^{(n)} \leq 4\epsilon$. $\square$

Codifica Canale - Osservazioni

- La scelta random del codice serve per la prova non per la codifica

Codifica Canale - Osservazioni

- La scelta random del codice serve per la prova non per la codifica
- Mediando proviamo che esiste almeno un codice con le proprietà desiderate

Codifica Canale - Osservazioni

- La scelta random del codice serve per la prova non per la codifica
- Mediando proviamo che esiste almeno un codice con le proprietà desiderate
- Tale codice può essere trovato (ricerca esaustiva !?!) ed il processo di codifica e decodifica rimane completamente deterministico.

Codifica Canale - Osservazioni

- La scelta random del codice serve per la prova non per la codifica
- Mediando proviamo che esiste almeno un codice con le proprietà desiderate
- Tale codice può essere trovato (ricerca esaustiva !?!) ed il processo di codifica e decodifica rimane completamente deterministico.
- La ricerca di tale codice è esponenziale

Codifica Canale - Osservazioni

- La scelta random del codice serve per la prova non per la codifica
- Mediando proviamo che esiste almeno un codice con le proprietà desiderate
- Tale codice può essere trovato (ricerca esaustiva !?!) ed il processo di codifica e decodifica rimane completamente deterministico.
- La ricerca di tale codice è esponenziale
- Possiamo sceglierlo random e avere buone *chance* di trovarne uno con le caratteristiche richieste. Però la decodifica risulta altamente inefficiente.

Codifica Canale - Osservazioni

- La scelta random del codice serve per la prova non per la codifica
- Mediando proviamo che esiste almeno un codice con le proprietà desiderate
- Tale codice può essere trovato (ricerca esaustiva !?!) ed il processo di codifica e decodifica rimane completamente deterministico.
- La ricerca di tale codice è esponenziale
- Possiamo sceglierlo random e avere buone *chance* di trovarne uno con le caratteristiche richieste. Però la decodifica risulta altamente inefficiente.
- Un problema fondamentale: trovare codici con tasso prossimo a C e con una struttura che mantenga la decodifica efficiente

Codifica Canale - Parte Inversa

- Ci rimane da dimostrare che per ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \rightarrow n$ deve valere $R < C$.
- Cominceremo con il dimostrare due lemmi che ci serviranno per la dimostrazione.
- $I(X^n, Y^n) \leq \sum_{i=1}^{n} I(X_i, Y_i)$
- $H(X^n|Y^n) \leq 1 + P_e^{(n)} nR$. (Disuguaglianza di Fano)

Lemma (Disuguaglianza di Fano) Consideriamo un DMC. Sia il messaggio in input W scelto in accordo alla distribuzione uniforme tra 2^{nR} messaggi. Sia $\mathcal{C}$ il codice, Y^n la parola ricevuta in output al canale, $g(\cdot)$ la funzione di decodifica e $P_e^{(n)} = Pr(W \neq g(Y^n))$. Allora

$$H(X^n|Y^n) \leq 1 + P_e^{(n)} nR.$$

Disuguaglianza di Fano

Dim. Definiamo $E = \begin{cases} 1, & se\ g(Y^n) \neq W, \\ 0, & se\ g(Y^n) = W. \end{cases}$

- Espandiamo $H(E, W|Y^n)$ in due modi diversi

$$\begin{aligned} H(E, W|Y^n) &= H(W|Y^n) + H(E|W, Y^n) \\ &= H(E|Y^n) + H(W|E, Y^n) \end{aligned}$$

Disuguaglianza di Fano

Dim. Definiamo $E = \begin{cases} 1, & se\ g(Y^n) \neq W, \\ 0, & se\ g(Y^n) = W. \end{cases}$

- Espandiamo $H(E, W|Y^n)$ in due modi diversi

$$\begin{aligned} H(E, W|Y^n) &= H(W|Y^n) + H(E|W, Y^n) \\ &= H(E|Y^n) + H(W|E, Y^n) \end{aligned}$$

- Notiamo che $H(E|W, Y^n) = 0$, $H(E|Y^n) \leq H(E) \leq 1$, e

$$\begin{aligned} H(W|E, Y^n) &= \sum_{i=0}^{1} P(E = i) H(W|Y^n, E = i) \\ &= (1 - P_e^{(n)})0 + P_e^{(n)} log(2^{nR} - 1) \leq P_e^{(n)} nR. \end{aligned}$$

Dim. (cont.)

- Ne consegue che

$$H(W|Y^n) \leq 1 + P_e^{(n)} nR$$

$\square$

Dim. (cont.)

- Ne consegue che

$$H(W|Y^n) \leq 1 + P_e^{(n)} nR$$

- da cui segue la tesi, in quanto $H(X^n|Y^n) \leq H(W|Y^n)$, poiché X^n è funzione di W.

$\square$

Lemma Sia Y^n l'output di un DMC per input X^n. Allora, per ogni distribuzione $p(x^n)$, vale $I(X^n;Y^n) \leq \sum_{i=1}^n I(X_i;Y_i)$.

Dim.

$$
\begin{aligned}
I(X^n, Y^n) &= H(Y^n) - H(Y^n|X^n) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, \ldots, Y_{i-1}, X^n) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \ \ (no\ memoria) \\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) = \sum_{i=1}^{n} I(X_i;Y_i).
\end{aligned}
$$

$\square$

Lemma Sia Y^n l'output di un DMC per input X^n. Allora, per ogni distribuzione $p(x^n)$, vale $I(X^n;Y^n) \leq \sum_{i=1}^n I(X_i;Y_i)$.

Corollario Sia Y^n l'output di un DMC per input X^n. Allora, per ogni distribuzione $p(x^n)$, vale $I(X^n;Y^n) \leq nR$.

Codifica Canale - Parte Inversa

Teorema Ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$, deve avere $R < C$.

Dim.

- $\lambda^{(n)} \to 0 \Rightarrow P_e^{(n)} \to 0.$

Codifica Canale - Parte Inversa

Teorema Ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$, deve avere $R < C$.

Dim.

- $\lambda^{(n)} \to 0 \Rightarrow P_e^{(n)} \to 0.$
- Consideriamo il messaggio W scelto uniformemente in $\{1, 2, \ldots, 2^{nR}\}$, (quindi $H(W) = nR$)

Codifica Canale - Parte Inversa

Teorema Ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \longrightarrow 0$, deve avere $R < C$.

Dim.

- $\lambda^{(n)} \longrightarrow 0 \Rightarrow P_e^{(n)} \longrightarrow 0$.
- Consideriamo il messaggio W scelto uniformemente in $\{1, 2, \ldots, 2^{nR}\}$, (quindi $H(W) = nR$)
- $P_e^{(n)} = Pr(g(Y^n) \neq W)$.

Codifica Canale - Parte Inversa

Teorema Ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$, deve avere $R < C$.

Dim.

- $\lambda^{(n)} \to 0 \Rightarrow P_e^{(n)} \to 0$.
- Consideriamo il messaggio W scelto uniformemente in $\{1, 2, \ldots, 2^{nR}\}$, (quindi $H(W) = nR$)
- $P_e^{(n)} = Pr(g(Y^n) \neq W)$.
- Allora,

$$
\begin{aligned}
nR &= H(W) = H(W|Y^n) + I(W;Y^n) \quad [W \to X^n(W) \to Y^n \\
&\leq H(W|Y^n) + I(X^n(W);Y^n) \\
&\leq 1 + P_e^{(n)} nR + nC.
\end{aligned}
$$

Teorema Ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$, deve avere $R < C$.
Dim. (cont.)

- $nR \leq 1 + P_e^{(n)} nR + nC.$

Teorema Ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$, deve avere $R < C$.

Dim. (cont.)

- $nR \leq 1 + P_e^{(n)} nR + nC.$
- Dividendo per n otteniamo

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

Teorema Ogni sequenza di codici $(2^{nR}, n)$ con $\lambda^{(n)} \to 0$, deve avere $R < C$.

Dim. (cont.)

- $nR \leq 1 + P_e^{(n)} nR + nC.$
- Dividendo per n otteniamo

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

- e per $n \to \infty$ abbiamo la tesi, usando

$$P_e^{(n)} \to 0 \;\; e \;\; 1/n \to 0.$$

- Riscriviamo la disuguaglianza $R \leq \frac{1}{n} + P_e^{(n)} R + C$ come

$$P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}$$

- Riscriviamo la disuguaglianza $R \leq \frac{1}{n} + P_e^{(n)} R + C$ come

$$P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}$$

- Questo mostra che per $R > C$ la probabilità di errore si mantiene > 0 per $n \to \infty$.

- Riscriviamo la disuguaglianza $R \leq \frac{1}{n} + P_e^{(n)} R + C$ come

$$P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}$$

- Questo mostra che per $R > C$ la probabilità di errore si mantiene > 0 per $n \to \infty$.
- ma deve valere per ogni n. Infatti, se avessimo codici con $P_e^{(n)} = 0$ per n piccoli potremmo estenderli a codici di lunghezza maggiore per concatenazione.

- Riscriviamo la disuguaglianza $R \leq \frac{1}{n} + P_e^{(n)} R + C$ come

$$P_e^{(n)} \geq 1 - \frac{C}{R} - \frac{1}{nR}$$

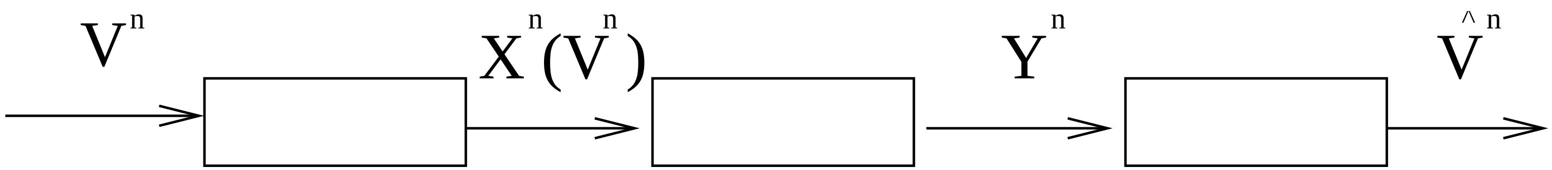
- Questo mostra che per $R > C$ la probabilità di errore si mantiene > 0 per $n \to \infty$.
- ma deve valere per ogni n. Infatti, se avessimo codici con $P_e^{(n)} = 0$ per n piccoli potremmo estenderli a codici di lunghezza maggiore per concatenazione.
- In conclusione, non si può ridurre arbitrariamente la probabilità d'errore a tassi superiori alla capacità.

Codifica Sorgente–Canale

Teorema Se $V_1, \ldots, V_n$ soddisfano la PEA allora esiste una codice sorgente-canale con $P_e^{(n)} \to 0$ se $H(V) < C$.
Se $H(V) > C$ la probabilitá di errore non puó essere resa arbitrariamente piccola.

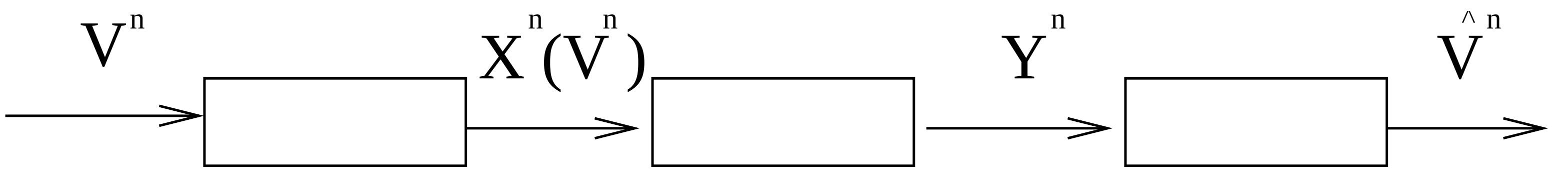
Codifica Sorgente–Canale

Teorema Se $V_1, \ldots, V_n$ soddisfano la PEA allora esiste una codice sorgente-canale con $P_e^{(n)} \to 0$ se $H(V) < C$.
Se $H(V) > C$ la probabilitá di errore non puó essere resa arbitrariamente piccola.
Dim.



Codifica Sorgente–Canale

Teorema Se $V_1, \ldots, V_n$ soddisfano la PEA allora esiste una codice sorgente-canale con $P_e^{(n)} \to 0$ se $H(V) < C$.
Se $H(V) > C$ la probabilitá di errore non puó essere resa arbitrariamente piccola.
Dim.



PEA $\Rightarrow |A_\epsilon^{(n)}| \leq 2^{n(H(V)+\epsilon)}$ e $P(A_\epsilon^{(n)}) > 1 - \epsilon$
Codifichiamo solo sequenze tipiche $\Rightarrow 2^{n(H(V)+\epsilon)}$ indici
$\Rightarrow$ se $R = H(V) + \epsilon < C$ possiamo trasmettere sul canale con prob. err $< \epsilon$
Quindi

$$P_e^{(n)} = P(V^n \neq \hat{V}^n)$$

$$\leq P(V^n \notin A_\epsilon^{(n)}) + P(g(Y^n) \neq V^n | V^n \in A_\epsilon^{(n)}) \leq \epsilon + \epsilon = 2\epsilon$$

Parte inversa. Dalla disuguaglianza di Fano

$$H(V^n|\hat{V}^n) \leq 1 + P_e^{(n)} \log|\mathcal{V}^n| = 1 + P_e^{(n)} n \log|\mathcal{V}|$$

Parte inversa. Dalla disuguaglianza di Fano

$H(V^n|\hat{V}^n) \leq 1 + P_e^{(n)} \log|\mathcal{V}^n| = 1 + P_e^{(n)} n \log|\mathcal{V}|$

Quindi

$$
\begin{aligned}
H(V) &= \frac{H(V_1 \ldots V_n)}{n} = \frac{H(V^n)}{n} = \frac{1}{n} H(V^n|\hat{V}^n) + \frac{1}{n} I(V^n;\hat{V}^n) \\
&\leq \frac{1}{n}(1 + P_e^{(n)} n \log|\mathcal{V}|) + \frac{1}{n} I(V^n;\hat{V}^n) \\
&\leq \frac{1}{n}(1 + P_e^{(n)} n \log|\mathcal{V}|) + \frac{1}{n} I(X^n;Y^n) \\
&\leq \frac{1}{n} + P_e^{(n)} \log|\mathcal{V}| + C
\end{aligned}
$$

Parte inversa. Dalla disuguaglianza di Fano

$H(V^n|\hat{V}^n) \leq 1 + P_e^{(n)} \log|\mathcal{V}^n| = 1 + P_e^{(n)} n \log|\mathcal{V}|$

Quindi

$$
\begin{aligned}
H(V) &= \frac{H(V_1 \ldots V_n)}{n} = \frac{H(V^n)}{n} = \frac{1}{n} H(V^n|\hat{V}^n) + \frac{1}{n} I(V^n;\hat{V}^n) \\
&\leq \frac{1}{n}(1 + P_e^{(n)} n \log|\mathcal{V}|) + \frac{1}{n} I(V^n;\hat{V}^n) \\
&\leq \frac{1}{n}(1 + P_e^{(n)} n \log|\mathcal{V}|) + \frac{1}{n} I(X^n;Y^n) \\
&\leq \frac{1}{n} + P_e^{(n)} \log|\mathcal{V}| + C
\end{aligned}
$$

Per $n \to \infty$, se $P_e^{(n)} \to 0$ si ha $H(V) \leq C$.