Codes and Cryptography

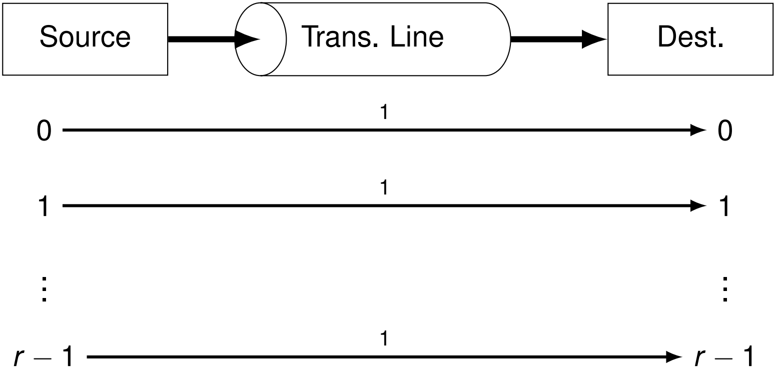
Jorge L. Villar

MAMME, Fall 2015

PART IV

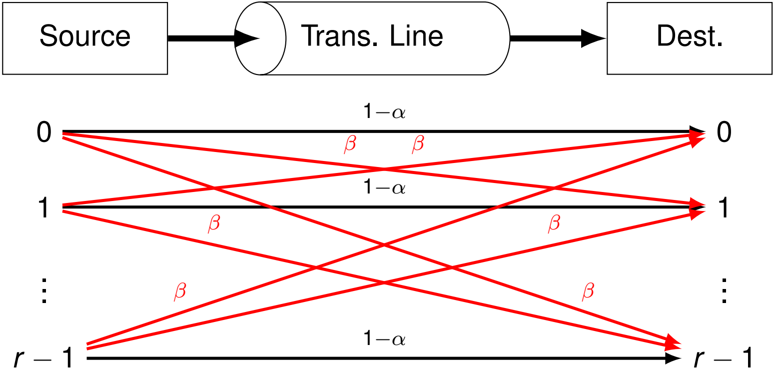
Outline

Transmission Channels



r-ary perfect channel: no transmission errors

Transmission Channels



r-ary symmetric channel $\quad \beta = \frac{\alpha}{r-1}$

Transmission Channels: The Model

Input and output alphabets: $\{0, 1, \ldots, r-1\}$

Transmission probability matrix:

$$T = \begin{pmatrix} t_{0,0} & \cdots & t_{0,r-1} \\ \vdots & \ddots & \vdots \\ t_{r-1,0} & \cdots & t_{r-1,r-1} \end{pmatrix} \text{ where } t_{i,j} = \Pr(\text{receive } j \mid \text{send } i)$$

Transmission Channels: The Model

Input and output alphabets: $\{0, 1, \ldots, r-1\}$

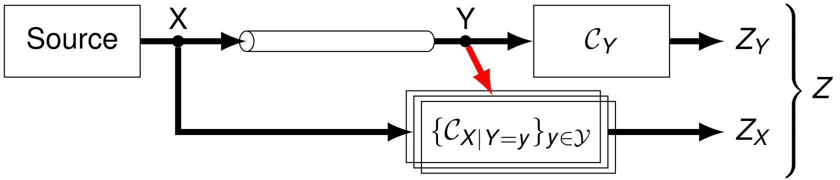
Transmission probability matrix:

$$T = \begin{pmatrix} t_{0,0} & \cdots & t_{0,r-1} \\ \vdots & \ddots & \vdots \\ t_{r-1,0} & \cdots & t_{r-1,r-1} \end{pmatrix} \text{ where } t_{i,j} = \Pr(\text{receive } j \mid \text{send } i)$$

Properties:

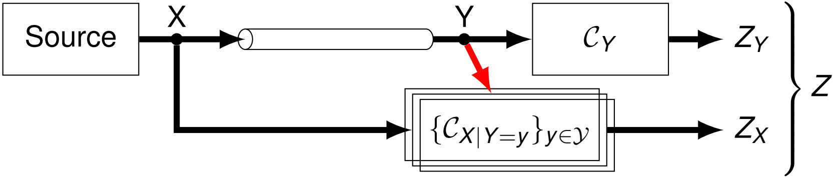
- $T \geq 0$
- $T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$
- If $p_j = \Pr(\text{send } j)$ and $q_i = \Pr(\text{receive } i)$, $(q_0 \ \ldots \ q_{r-1}) = (p_0 \ \ldots \ p_{r-1})T$

Channel Capacity (I)



$$L(\mathcal{C}_{X,Y}, \mathcal{S}_{X,Y}) = L(\mathcal{C}_Y, \mathcal{S}_Y) + \sum_{y\in\mathcal{Y}} \Pr(Y = y) L(\mathcal{C}_{X|Y=y}, \mathcal{S}_{X|Y=y})$$

Channel Capacity (I)



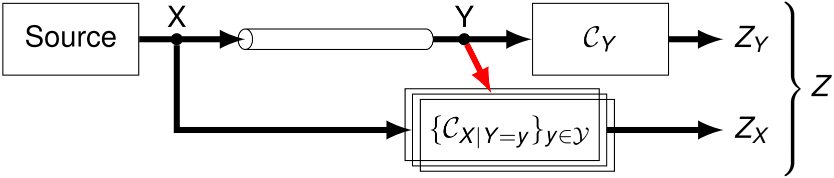
$$L(\mathcal{C}_{X,Y},\mathcal{S}_{X,Y}) = L(\mathcal{C}_Y,\mathcal{S}_Y) + \sum_{y\in\mathcal{Y}} \Pr(Y=y) L(\mathcal{C}_{X|Y=y},\mathcal{S}_{X|Y=y})$$

If all (binary) codes are optimal then

$H(Y) \le L(\mathcal{C}_Y,\mathcal{S}_Y) < H(Y)+1$

$H(X \mid Y=y) \le L(\mathcal{C}_{X|Y=y},\mathcal{S}_{X|Y=y}) < H(X \mid Y=y)+1$

Channel Capacity (I)



$$L(\mathcal{C}_{X,Y},\mathcal{S}_{X,Y}) = L(\mathcal{C}_Y,\mathcal{S}_Y) + \sum_{y\in\mathcal{Y}} \Pr(Y=y) L(\mathcal{C}_{X|Y=y},\mathcal{S}_{X|Y=y})$$

If all (binary) codes are optimal then

$H(Y) \leq L(\mathcal{C}_Y,\mathcal{S}_Y) < H(Y) + 1$

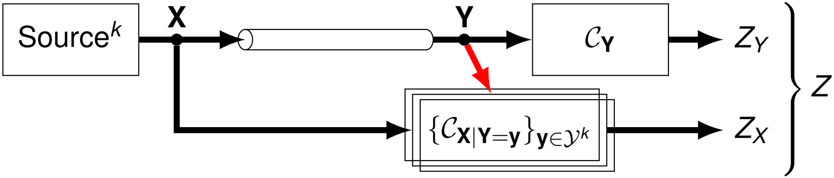
$H(X \mid Y=y) \leq L(\mathcal{C}_{X|Y=y},\mathcal{S}_{X|Y=y}) < H(X \mid Y=y) + 1$

Thus,

$H(X,Y) = H(Y) + H(X \mid Y) \leq L(\mathcal{C}_{X,Y},\mathcal{S}_{X,Y}) < H(X,Y) + 2$

Channel Capacity (II)

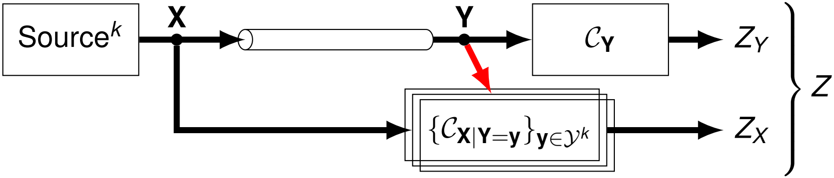
With extended sources,



$$L(\mathcal{C}_{\mathbf{X},\mathbf{Y}}, \mathcal{S}^k_{X,Y}) = L(\mathcal{C}_{\mathbf{Y}}, \mathcal{S}^k_{Y}) + \sum_{\mathbf{y}\in\mathcal{Y}^k} \Pr(\mathbf{Y}=\mathbf{y}) L(\mathcal{C}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}, \mathcal{S}^k_{X|\mathbf{Y}=\mathbf{y}})$$

Channel Capacity (II)

With extended sources,



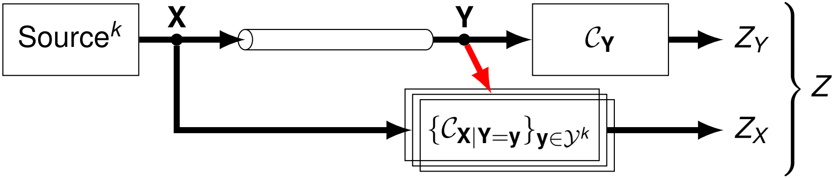
$$L(\mathcal{C}_{\mathbf{X},\mathbf{Y}},\mathcal{S}^k_{X,Y}) = L(\mathcal{C}_{\mathbf{Y}},\mathcal{S}^k_{Y}) + \sum_{\mathbf{y}\in\mathcal{Y}^k} \Pr(\mathbf{Y}=\mathbf{y}) L(\mathcal{C}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}},\mathcal{S}^k_{X|\mathbf{Y}=\mathbf{y}})$$

Now,

$$H(\mathbf{X},\mathbf{Y}) = H(\mathbf{Y}) + H(\mathbf{X}\mid\mathbf{Y}) = kH(Y) + kH(X\mid Y) =$$
$$kH(X,Y) \le L(\mathcal{C}_{\mathbf{X},\mathbf{Y}},\mathcal{S}^k_{X,Y}) < H(\mathbf{X},\mathbf{Y}) + 2 = kH(X,Y) + 2$$

Channel Capacity (II)

With extended sources,



$$L(\mathcal{C}_{\mathbf{X},\mathbf{Y}}, \mathcal{S}^k_{X,Y}) = L(\mathcal{C}_{\mathbf{Y}}, \mathcal{S}^k_{Y}) + \sum_{\mathbf{y}\in\mathcal{Y}^k} \Pr(\mathbf{Y}=\mathbf{y}) L(\mathcal{C}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}, \mathcal{S}^k_{X|\mathbf{Y}=\mathbf{y}})$$

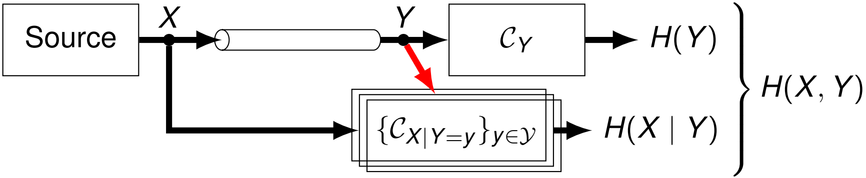
Now,

$H(\mathbf{X},\mathbf{Y}) = H(\mathbf{Y}) + H(\mathbf{X} \mid \mathbf{Y}) = kH(Y) + kH(X \mid Y) =$
$kH(X,Y) \leq L(\mathcal{C}_{\mathbf{X},\mathbf{Y}}, \mathcal{S}^k_{X,Y}) < H(\mathbf{X},\mathbf{Y}) + 2 = kH(X,Y) + 2$

Therefore, asymptotically $H(X,Y) = \dfrac{1}{k} L(\mathcal{C}_{\mathbf{X},\mathbf{Y}}, \mathcal{S}^k_{X,Y})$

Channel Capacity (III)

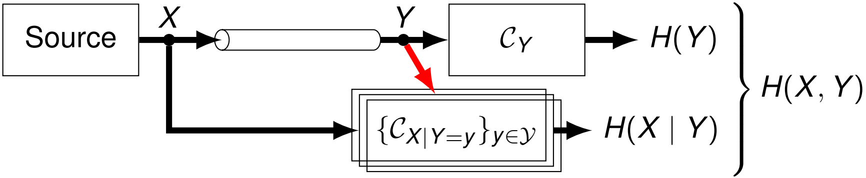
Regarding entropies,



The information about X conveyed by Y is
$H(X) - H(X \mid Y) = H(X) + H(Y) - H(X, Y) = I(X; Y)$

Channel Capacity (III)

Regarding entropies,



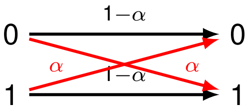
The information about X conveyed by Y is
$H(X) - H(X \mid Y) = H(X) + H(Y) - H(X, Y) = I(X; Y)$

Definition (Channel Capacity)

$$\Lambda = \max_{\mathcal{S}_X}(I(X; Y))$$

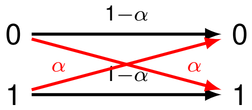
where the maximum is taken over all possible sources $\mathcal{S}_X$

Capacity of the Binary Symmetric Channel



For $\Pr(X = 1) = p$, $H(X) = h_p = -p\log_2 p - (1-p)\log_2(1-p)$

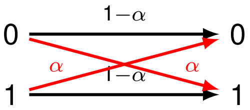
Capacity of the Binary Symmetric Channel



For $\Pr(X=1)=p$, $H(X)=h_p=-p\log_2 p-(1-p)\log_2(1-p)$

$\Pr(Y=1)=p(1-\alpha)+(1-p)\alpha$, $H(Y)=h_{p(1-\alpha)+(1-p)\alpha}$

Capacity of the Binary Symmetric Channel

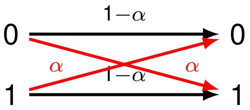


For $\Pr(X = 1) = p$, $H(X) = h_p = -p \log_2 p - (1-p) \log_2(1-p)$

$\Pr(Y = 1) = p(1-\alpha) + (1-p)\alpha$, $H(Y) = h_{p(1-\alpha)+(1-p)\alpha}$

Conditioned to X, $H(Y \mid X = 0) = H(Y \mid X = 1) = h_\alpha$.

Capacity of the Binary Symmetric Channel



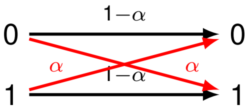
For $\Pr(X = 1) = p$, $H(X) = h_p = -p \log_2 p - (1-p) \log_2 (1-p)$

$\Pr(Y = 1) = p(1-\alpha) + (1-p)\alpha$, $H(Y) = h_{p(1-\alpha)+(1-p)\alpha}$

Conditioned to X, $H(Y \mid X = 0) = H(Y \mid X = 1) = h_\alpha$.

$I(X; Y) = H(Y) - H(Y \mid X) = h_{p(1-\alpha)+(1-p)\alpha} - h_\alpha$

Capacity of the Binary Symmetric Channel



For $\Pr(X=1)=p$, $H(X)=h_p=-p\log_2 p-(1-p)\log_2(1-p)$

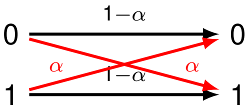
$\Pr(Y=1)=p(1-\alpha)+(1-p)\alpha$, $H(Y)=h_{p(1-\alpha)+(1-p)\alpha}$

Conditioned to X, $H(Y\mid X=0)=H(Y\mid X=1)=h_\alpha$.

$I(X;Y)=H(Y)-H(Y\mid X)=h_{p(1-\alpha)+(1-p)\alpha}-h_\alpha$

The maximum occurs at $p(1-\alpha)+(1-p)\alpha=\frac{1}{2}$, i.e., $p=\frac{1}{2}$ (the uniform source).

Capacity of the Binary Symmetric Channel



For $\Pr(X = 1) = p$, $H(X) = h_p = -p \log_2 p - (1-p)\log_2(1-p)$

$\Pr(Y = 1) = p(1-\alpha) + (1-p)\alpha$, $H(Y) = h_{p(1-\alpha)+(1-p)\alpha}$

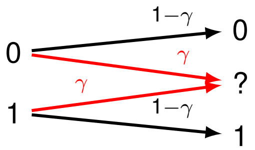
Conditioned to X, $H(Y \mid X = 0) = H(Y \mid X = 1) = h_\alpha$.

$I(X; Y) = H(Y) - H(Y \mid X) = h_{p(1-\alpha)+(1-p)\alpha} - h_\alpha$

The maximum occurs at $p(1-\alpha) + (1-p)\alpha = \frac{1}{2}$, i.e., $p = \frac{1}{2}$ (the uniform source).

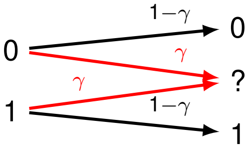
The capacity of the BSC is $\Lambda_{BSC} = 1 - h_\alpha$

Capacity of the Binary Channel With Erasures



For $\Pr(X = 1) = p$, $H(X) = h_p = -p\log_2 p - (1-p)\log_2(1-p)$

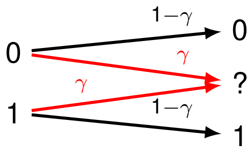
Capacity of the Binary Channel With Erasures



For $\Pr(X=1)=p$, $H(X)=h_p=-p\log_2 p-(1-p)\log_2(1-p)$

$\Pr(Y=0)=p(1-\gamma)$, $\Pr(Y=1)=(1-p)(1-\gamma)$,
$H(Y)=h_\gamma+(1-\gamma)h_p$

Capacity of the Binary Channel With Erasures

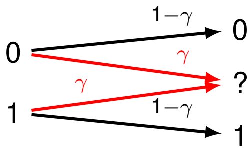


For $\Pr(X = 1) = p$, $H(X) = h_p = -p \log_2 p - (1 - p) \log_2 (1 - p)$

$\Pr(Y = 0) = p(1 - \gamma)$, $\Pr(Y = 1) = (1 - p)(1 - \gamma)$,
$H(Y) = h_\gamma + (1 - \gamma) h_p$

Conditioned to X, $H(Y \mid X = 0) = H(Y \mid X = 1) = h_\gamma$.

Capacity of the Binary Channel With Erasures



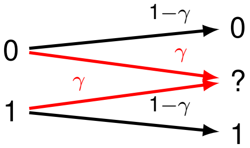
For $\Pr(X = 1) = p$, $H(X) = h_p = -p \log_2 p - (1-p)\log_2(1-p)$

$\Pr(Y = 0) = p(1-\gamma)$, $\Pr(Y = 1) = (1-p)(1-\gamma)$,
$H(Y) = h_\gamma + (1-\gamma)h_p$

Conditioned to X, $H(Y \mid X = 0) = H(Y \mid X = 1) = h_\gamma$.

$I(X; Y) = H(Y) - H(Y \mid X) = (1-\gamma)h_p$

Capacity of the Binary Channel With Erasures



For $\Pr(X = 1) = p$, $H(X) = h_p = -p\log_2 p - (1-p)\log_2(1-p)$

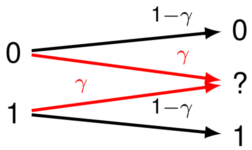
$\Pr(Y = 0) = p(1-\gamma)$, $\Pr(Y = 1) = (1-p)(1-\gamma)$,
$H(Y) = h_\gamma + (1-\gamma)h_p$

Conditioned to X, $H(Y \mid X = 0) = H(Y \mid X = 1) = h_\gamma$.

$I(X;Y) = H(Y) - H(Y \mid X) = (1-\gamma)h_p$

The maximum occurs at $p = \frac{1}{2}$ (the uniform source).

Capacity of the Binary Channel With Erasures



For $\Pr(X=1)=p$, $H(X)=h_p=-p\log_2 p-(1-p)\log_2(1-p)$

$\Pr(Y=0)=p(1-\gamma)$, $\Pr(Y=1)=(1-p)(1-\gamma)$,
$H(Y)=h_\gamma+(1-\gamma)h_p$

Conditioned to X, $H(Y\mid X=0)=H(Y\mid X=1)=h_\gamma$.

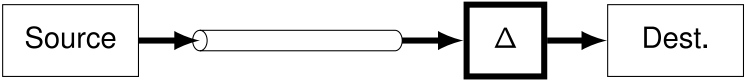
$I(X;Y)=H(Y)-H(Y\mid X)=(1-\gamma)h_p$

The maximum occurs at $p=\frac{1}{2}$ (the uniform source).

The capacity of the BSEC is $\Lambda_{BSEC}=1-\gamma$

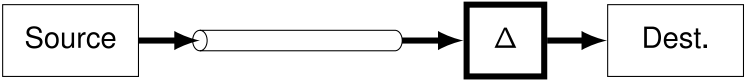
Outline

Decision Rules



Decision rule Δ: Guess the symbol sent from the received symbol

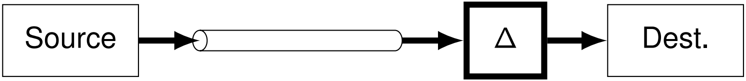
Decision Rules



Decision rule Δ: Guess the symbol sent from the received symbol

- **Ideal Observer:** On receiving i, choose j that maximizes $\Pr(\text{send } j \mid \text{receive } i)$
- **Max. Likelihood:** On receiving i, choose j that maximizes $\Pr(\text{receive } i \mid \text{send } j)$

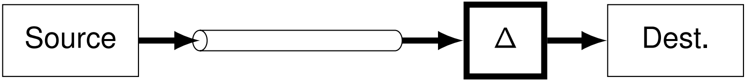
Decision Rules



Decision rule Δ: Guess the symbol sent from the received symbol

- **Ideal Observer:** On receiving i, choose j that maximizes $\Pr(\text{send } j \mid \text{receive } i)$ (Optimal, but depends on the source)
- **Max. Likelihood:** On receiving i, choose j that maximizes $\Pr(\text{receive } i \mid \text{send } j)$ (Not optimal, but independent of the source)

Decision Rules



Decision rule Δ: Guess the symbol sent from the received symbol

- **Ideal Observer:** On receiving *i*, choose *j* that maximizes Pr(send *j* | receive *i*) (Optimal, but depends on the source)
- **Max. Likelihood:** On receiving *i*, choose *j* that maximizes Pr(receive *i* | send *j*) (Not optimal, but independent of the source)

Both are equivalent when the error probabilities are small and the source is balanced enough

The BSC Case

Transmission matrix: $T = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$

Max. Likelihood:

$Y = 0 \quad \longrightarrow \quad \text{if } \ 1-\alpha \geq \alpha \ \text{ then } \ \widehat{X} = 0$

$Y = 1 \quad \longrightarrow \quad \text{if } \ 1-\alpha \geq \alpha \ \text{ then } \ \widehat{X} = 1$

The BSC Case

Transmission matrix: $T = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$

Max. Likelihood:

$$Y = 0 \quad \longrightarrow \quad \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 0$$
$$Y = 1 \quad \longrightarrow \quad \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 1$$

Ideal Observer:

$$Y = 0 \quad \longrightarrow \quad \text{if } (1-\alpha)(1-p) \geq \alpha p \text{ then } \widehat{X} = 0$$
$$Y = 1 \quad \longrightarrow \quad \text{if } (1-\alpha)p \geq \alpha(1-p) \text{ then } \widehat{X} = 1$$

The BSC Case

Transmission matrix: $T = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$

Max. Likelihood:

$Y = 0 \quad \longrightarrow \quad \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 0 \qquad (\alpha \leq \frac{1}{2})$

$Y = 1 \quad \longrightarrow \quad \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 1 \qquad (\alpha \leq \frac{1}{2})$

Ideal Observer:

$Y = 0 \quad \longrightarrow \quad \text{if } (1-\alpha)(1-p) \geq \alpha p \text{ then } \widehat{X} = 0 \qquad (\alpha \leq 1-p)$

$Y = 1 \quad \longrightarrow \quad \text{if } (1-\alpha)p \geq \alpha(1-p) \text{ then } \widehat{X} = 1 \qquad (\alpha \leq p)$

The BSC Case

Transmission matrix: $T = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$

Max. Likelihood:

$$Y = 0 \longrightarrow \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 0 \qquad (\alpha \leq \tfrac{1}{2})$$

$$Y = 1 \longrightarrow \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 1 \qquad (\alpha \leq \tfrac{1}{2})$$

$$p_{corr} = \Pr(\widehat{X} = X) = \max(\alpha, 1-\alpha)$$

Ideal Observer:

$$Y = 0 \longrightarrow \text{if } (1-\alpha)(1-p) \geq \alpha p \text{ then } \widehat{X} = 0 \quad (\alpha \leq 1-p)$$

$$Y = 1 \longrightarrow \text{if } (1-\alpha)p \geq \alpha(1-p) \text{ then } \widehat{X} = 1 \qquad (\alpha \leq p)$$

$$p_{corr} = \max(\alpha, p, 1-p, 1-\alpha)$$

The BSC Case

Transmission matrix: $T = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}$

Max. Likelihood:

$Y = 0 \quad \longrightarrow \quad \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 0 \qquad (\alpha \leq \frac{1}{2})$

$Y = 1 \quad \longrightarrow \quad \text{if } 1-\alpha \geq \alpha \text{ then } \widehat{X} = 1 \qquad (\alpha \leq \frac{1}{2})$

$p_{corr} = \Pr(\widehat{X} = X) = \max(\alpha, 1-\alpha)$

Ideal Observer:

$Y = 0 \quad \longrightarrow \quad \text{if } (1-\alpha)(1-p) \geq \alpha p \text{ then } \widehat{X} = 0 \quad (\alpha \leq 1-p)$

$Y = 1 \quad \longrightarrow \quad \text{if } (1-\alpha)p \geq \alpha(1-p) \text{ then } \widehat{X} = 1 \quad (\alpha \leq p)$

$p_{corr} = \max(\alpha, p, 1-p, 1-\alpha)$

Both differ only when the channel is useless!

Improving Channel Reliability (I)

GOAL: Make p_{corr} as close to 1 as possible maintaining an information transmission rate close to the theoretical channel capacity

Improving Channel Reliability (I)

GOAL: Make p_{corr} as close to 1 as possible maintaining an information transmission rate close to the theoretical channel capacity

Intuitively, one has to add redundancy to the source to be able to correct all errors introduced by the channel.

Improving Channel Reliability (I)

GOAL: Make p_{corr} as close to 1 as possible maintaining an information transmission rate close to the theoretical channel capacity

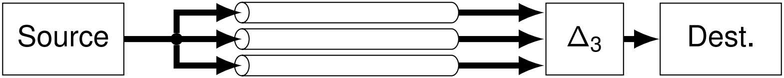
Intuitively, one has to add redundancy to the source to be able to correct all errors introduced by the channel.

E.g., the information rate of the uniform binary source is $H(X) = 1$, but the capacity of the BSC is $\Lambda_{BSC} = 1 - h_\alpha$. Therefore, we need to expand X to a length of at least $\frac{1}{\Lambda_{BSC}}$ to make $p_{corr} \to 1$.

Improving Channel Reliability (II)

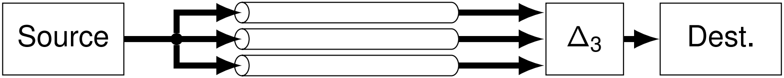
Improving Channel Reliability (II)

Channel replication = Repetition code

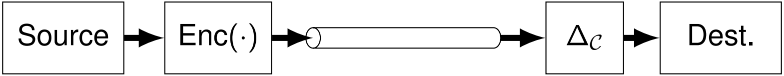


Improving Channel Reliability (II)

Channel replication = Repetition code



"Encode then Transmit"



BSC Replication (or Repetition Code)

Each source symbol is sent $n = 2\ell + 1$ times.

BSC Replication (or Repetition Code)

Each source symbol is sent $n = 2\ell + 1$ times.

Max. Likelihood Decoding Rule: (Assume $\alpha < \frac{1}{2}$)

$\mathbf{Y} \quad \longrightarrow \quad$ decide $\widehat{X}$ by majority vote among $Y_1, \ldots, Y_n$

BSC Replication (or Repetition Code)

Each source symbol is sent $n = 2\ell + 1$ times.

Max. Likelihood Decoding Rule: (Assume $\alpha < \frac{1}{2}$)

Y $\longrightarrow$ decide $\widehat{X}$ by majority vote among $Y_1, \ldots, Y_n$

Definition (Hamming Distance)

Given two words $x, y \in \{0, \ldots, r-1\}^n$, the Hamming distance $d(x, y)$ is the number positions in which they differ.

BSC Replication (or Repetition Code)

Each source symbol is sent $n = 2\ell + 1$ times.

Max. Likelihood Decoding Rule: (Assume $\alpha < \frac{1}{2}$)

Y $\longrightarrow$ decide $\widehat{X}$ by majority vote among $Y_1, \ldots, Y_n$

Definition (Hamming Distance)

Given two words $x, y \in \{0, \ldots, r-1\}^n$, the Hamming distance $d(x, y)$ is the number positions in which they differ.

$$p_{corr} = \Pr(\widehat{X} = X) = \Pr(d(Y, X^n) \leq \ell) = \sum_{i=0}^{\ell} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}$$

For $\alpha = 0.1$, $\Lambda_{BSC} = 0.531$ but for $p_{corr} \geq 0.999$ we need $n = 9$

"Encode then Transmit"

We use a constant-length binary code $\mathcal{C}_k \subset \{0, 1\}^n$ for the extended source $\mathcal{S}^k$ and send the codewords through a BSC.

“Encode then Transmit”

We use a constant-length binary code $\mathcal{C}_k \subset \{0,1\}^n$ for the extended source $\mathcal{S}^k$ and send the codewords through a BSC.

Max. Likelihood Decoding Rule: (Assume $\alpha < \frac{1}{2}$)

$Y \quad \longrightarrow \quad$ decide $\widehat{X}$ by the codeword in $\mathcal{C}_k$ nearest to Y

"Encode then Transmit"

We use a constant-length binary code $\mathcal{C}_k \subset \{0,1\}^n$ for the extended source $\mathcal{S}^k$ and send the codewords through a BSC.

Max. Likelihood Decoding Rule: (Assume $\alpha < \frac{1}{2}$)

$Y \longrightarrow$ decide $\widehat{X}$ by the codeword in $\mathcal{C}_k$ nearest to Y

$$p_{corr} = \Pr(\widehat{X} = X) = \sum_{\mathbf{x} \in \{0,1\}^k} \Pr(\widehat{X} = \mathbf{x} \mid X = \mathbf{x}) \Pr(X = \mathbf{x})$$

"Encode then Transmit"

We use a constant-length binary code $\mathcal{C}_k \subset \{0,1\}^n$ for the extended source $\mathcal{S}^k$ and send the codewords through a BSC.

Max. Likelihood Decoding Rule: (Assume $\alpha < \frac{1}{2}$)

$Y \quad \longrightarrow \quad$ decide $\widehat{X}$ by the codeword in $\mathcal{C}_k$ nearest to Y

$$p_{corr} = \Pr(\widehat{X} = X) = \sum_{\mathbf{x}\in\{0,1\}^k} \Pr(\widehat{X} = \mathbf{x} \mid X = \mathbf{x}) \Pr(X = \mathbf{x})$$

$$\Pr(\widehat{X} = \mathbf{x} \mid X = \mathbf{x}) = \Pr(Y \in \text{neigh}(E_{\mathcal{C}_k}(\mathbf{x})) \mid X = \mathbf{x}) \ \text{ where}$$

$$\text{neigh}(\mathbf{w}) = \{\mathbf{y} \in \{0,1\}^n \ : \ \forall \mathbf{z} \in \mathcal{C}_k \setminus \{\mathbf{w}\},\ d(\mathbf{y},\mathbf{z}) > d(\mathbf{y},\mathbf{w})\}$$

"Encode then Transmit"

We use a constant-length binary code $\mathcal{C}_k \subset \{0,1\}^n$ for the extended source $\mathcal{S}^k$ and send the codewords through a BSC.

Max. Likelihood Decoding Rule: (Assume $\alpha < \frac{1}{2}$)

$Y \quad \longrightarrow \quad$ decide $\widehat{X}$ by the codeword in $\mathcal{C}_k$ nearest to Y

$$p_{corr} = \Pr(\widehat{X} = X) = \sum_{\mathbf{x} \in \{0,1\}^k} \Pr(\widehat{X} = \mathbf{x} \mid X = \mathbf{x}) \Pr(X = \mathbf{x})$$

$\Pr(\widehat{X} = \mathbf{x} \mid X = \mathbf{x}) = \Pr(Y \in \mathrm{neigh}(E_{\mathcal{C}_k}(\mathbf{x})) \mid X = \mathbf{x})$ where

$$\mathrm{neigh}(\mathbf{w}) = \{\mathbf{y} \in \{0,1\}^n \;:\; \forall \mathbf{z} \in \mathcal{C}_k \setminus \{\mathbf{w}\},\; d(\mathbf{y},\mathbf{z}) > d(\mathbf{y},\mathbf{w})\}$$

$$\Pr(\widehat{X} = \mathbf{x} \mid X = \mathbf{x}) = \sum_{\mathbf{y} \in neigh(E_{\mathcal{C}_k}(\mathbf{x}))} \alpha^{d(\mathbf{y},E_{\mathcal{C}_k}(\mathbf{x}))} (1-\alpha)^{n-d(\mathbf{y},E_{\mathcal{C}_k}(\mathbf{x}))}$$

Outline

The Rate of a Constant-Length Code

Definition (Code Rate)

The rate of a constant-length r-ary code $\mathcal{C} \subset \{0, \ldots, r-1\}^n$ is

$$R(\mathcal{C}) = \frac{\log_r(|\mathcal{C}|)}{n}$$

The Rate of a Constant-Length Code

Definition (Code Rate)

The rate of a constant-length r-ary code $\mathcal{C} \subset \{0, \ldots, r-1\}^n$ is

$$R(\mathcal{C}) = \frac{\log_r(|\mathcal{C}|)}{n}$$

For the r-ary repetition code $\mathcal{C}_{rep} = \{0^n, \ldots, (r-1)^n\}$, $|\mathcal{C}_{rep}| = r$ and $R(\mathcal{C}_{rep}) = \frac{1}{n}$

The Rate of a Constant-Length Code

Definition (Code Rate)

The rate of a constant-length r-ary code $\mathcal{C} \subset \{0, \dots, r-1\}^n$ is

$$R(\mathcal{C}) = \frac{\log_r(|\mathcal{C}|)}{n}$$

For the r-ary repetition code $\mathcal{C}_{rep} = \{0^n, \dots, (r-1)^n\}$, $|\mathcal{C}_{rep}| = r$ and $R(\mathcal{C}_{rep}) = \frac{1}{n}$

For a constant-length r-ary code $\mathcal{C}_k \subset \{0, \dots, r-1\}^n$ for an r-ary extended source $\mathcal{S}^k$, $|\mathcal{C}_k| = r^k$ and $R(\mathcal{C}_k) = \frac{k}{n}$

GOAL(restated)**:** Given an r-ary channel with capacity Λ, find $\mathcal{C}$ with $p_{corr} \approx 1$ and $R(\mathcal{C}) \log_2 r \approx \Lambda$.

Shannon's Theorem

Theorem (Shannon)

Given an r-ary channel (stateless and source independent) with capacity Λ, for any positive $\varepsilon, \delta \in \mathbb{R}^+$ there exists $n_0 \in \mathbb{Z}^+$ such that for all $n > n_0$ there exists an r-ary code $\mathcal{C}$ of constant-length n, and a decision rule for it such that

$$\Lambda - \epsilon < R(\mathcal{C}) \log_2 r < \Lambda \qquad p_{corr} > 1 - \delta$$

Shannon's Theorem

Theorem (Shannon)

Given an r-ary channel (stateless and source independent) with capacity Λ, for any positive $\varepsilon, \delta \in \mathbb{R}^+$ there exists $n_0 \in \mathbb{Z}^+$ such that for all $n > n_0$ there exists an r-ary code $\mathcal{C}$ of constant-length n, and a decision rule for it such that

$$\Lambda - \epsilon < R(\mathcal{C}) \log_2 r < \Lambda \qquad p_{corr} > 1 - \delta$$

The proof builds on the idea that random codes behave well with high probability, but it is not constructive.

Sketch of the proof for the BSC (I)

Consider a random binary code $\mathcal{C}$ of constant-length n and rate R (i.e., $|\mathcal{C}| = 2^{nR}$). Now a random codeword $X = (X_1, \dots, X_n)$ is sent through a BSC with error probability $\alpha < \frac{1}{2}$. The word at reception is $Y = (Y_1, \dots, Y_n)$. We use the *nearest-codeword* decoding rule.

Sketch of the proof for the BSC (I)

Consider a random binary code $\mathcal{C}$ of constant-length n and rate R (i.e., $|\mathcal{C}| = 2^{nR}$). Now a random codeword $X = (X_1, \ldots, X_n)$ is sent through a BSC with error probability $\alpha < \frac{1}{2}$. The word at reception is $Y = (Y_1, \ldots, Y_n)$. We use the *nearest-codeword* decoding rule.

The expected value of p_{corr} is

$$\mathrm{E}(p_{corr}) = \frac{1}{2^n} \sum_{\mathbf{x},\mathbf{y} \in \{0,1\}^n} t_{\mathbf{x},\mathbf{y}} p_{corr}(\mathbf{x}, \mathbf{y})$$

where $t_{\mathbf{x},\mathbf{y}} = \Pr(Y = \mathbf{y} \mid X = \mathbf{x}) = (1 - \alpha)^{n-d(\mathbf{x},\mathbf{y})} \alpha^{d(\mathbf{x},\mathbf{y})}$ and $p_{corr}(\mathbf{x}, \mathbf{y})$ is the probability that $\forall \mathbf{z} \in \mathcal{C} \setminus \{\mathbf{x}\},\ d(\mathbf{z}, \mathbf{y}) > d(\mathbf{x}, \mathbf{y})$.

Sketch of the proof for the BSC (II)

As $p_{corr}(\mathbf{x},\mathbf{y})$ only depends on $d(\mathbf{x},\mathbf{y})$,

$$\mathrm{E}(p_{corr}) = \sum_{d=0}^{n} \binom{n}{d} \alpha^d (1-\alpha)^{n-d} p_{corr}(d)$$

and

$$p_{corr}(d) = \left(1 - \frac{1+n+\ldots+\binom{n}{d}}{2^n}\right)^{|\mathcal{C}|-1}$$

Sketch of the proof for the BSC (II)

As $p_{corr}(\mathbf{x}, \mathbf{y})$ only depends on $d(\mathbf{x}, \mathbf{y})$,

$$E(p_{corr}) = \sum_{d=0}^{n} \binom{n}{d} \alpha^d (1-\alpha)^{n-d} p_{corr}(d)$$

and

$$p_{corr}(d) = \left(1 - \frac{1 + n + \ldots + \binom{n}{d}}{2^n}\right)^{|\mathcal{C}|-1}$$

Recall that $|\mathcal{C}| = 2^{nR}$. Then, for large enough n and $d \leq \frac{n}{2}$

$$p_{corr}(d) \approx e^{-2^{n(R-1)}(1+n+\ldots+\binom{n}{d})}$$

Sketch of the proof for the BSC (III)

Lemma

For any $d \leq \frac{n}{2}$, $\quad 1 + n + \ldots + \binom{n}{d} \leq 2^{nh_{d/n}}$ *where*
$h_\alpha = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$

details...

Sketch of the proof for the BSC (III)

Lemma

For any $d \leq \frac{n}{2}$, $1 + n + \ldots + \binom{n}{d} \leq 2^{n h_{d/n}}$ *where* $h_\alpha = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$

▸ details...

Thus, for $d \leq \frac{n}{2}$, $p_{corr}(d) \geq e^{-2^{n(R-1+h_{d/n})}}$

Sketch of the proof for the BSC (III)

Lemma

For any $d \leq \frac{n}{2}$, $1 + n + \ldots + \binom{n}{d} \leq 2^{nh_{d/n}}$ *where*
$h_\alpha = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$

details...

Thus, for $d \leq \frac{n}{2}$, $p_{corr}(d) \geq e^{-2^{n(R-1+h_{d/n})}}$

Let $\gamma \leq \frac{1}{2}$ be such that $h_\gamma = 1 - R$. If n is large enough, for all $d < \gamma n$, $p_{corr}(d) \approx 1$

Sketch of the proof for the BSC (III)

Lemma

For any $d \leq \frac{n}{2}$, $1 + n + \ldots + \binom{n}{d} \leq 2^{nh_{d/n}}$ *where*
$h_\alpha = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$

details...

Thus, for $d \leq \frac{n}{2}$, $p_{corr}(d) \geq e^{-2^{n(R-1+h_{d/n})}}$

Let $\gamma \leq \frac{1}{2}$ be such that $h_\gamma = 1 - R$. If n is large enough, for all $d < \gamma n$, $p_{corr}(d) \approx 1$ and then

$$E(p_{corr}) \geq \sum_{d=0}^{\lfloor \gamma n \rfloor} \binom{n}{d} \alpha^d (1-\alpha)^{n-d}$$

Sketch of the proof for the BSC (III)

Lemma

For any $d \leq \frac{n}{2}$, $1 + n + \ldots + \binom{n}{d} \leq 2^{nh_{d/n}}$ *where* $h_\alpha = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$

details...

Thus, for $d \leq \frac{n}{2}$, $p_{corr}(d) \geq e^{-2^{n(R-1+h_{d/n})}}$

Let $\gamma \leq \frac{1}{2}$ be such that $h_\gamma = 1 - R$. If n is large enough, for all $d < \gamma n$, $p_{corr}(d) \approx 1$ and then

$$E(p_{corr}) \geq \sum_{d=0}^{\lfloor \gamma n \rfloor} \binom{n}{d} \alpha^d (1-\alpha)^{n-d}$$

that is, the probability that a binomial(n, α) random variable is less than γn, which tends to 1 whenever $\gamma > \alpha$.

Sketch of the proof for the BSC (III)

Lemma

For any $d \leq \frac{n}{2}$, $1 + n + \ldots + \binom{n}{d} \leq 2^{nh_{d/n}}$ *where* $h_\alpha = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$

▸ details...

Thus, for $d \leq \frac{n}{2}$, $p_{corr}(d) \geq e^{-2^{n(R-1+h_{d/n})}}$

Let $\gamma \leq \frac{1}{2}$ be such that $h_\gamma = 1 - R$. If n is large enough, for all $d < \gamma n$, $p_{corr}(d) \approx 1$ and then

$$E(p_{corr}) \geq \sum_{d=0}^{\lfloor \gamma n \rfloor} \binom{n}{d} \alpha^d (1-\alpha)^{n-d}$$

that is, the probability that a binomial(n, α) random variable is less than γn, which tends to 1 whenever $\gamma > \alpha$. Equivalently, $R = 1 - h_\gamma < 1 - h_\alpha = \Lambda_{BSC}$

Fano Bound (I)

Theorem (Fano Bound)

For any channel, r-ary source and decision rule

$$H(X \mid Y) \leq h_{p_{corr}} + (1 - p_{corr}) \log_2(r - 1)$$

Fano Bound (I)

Theorem (Fano Bound)

For any channel, r-ary source and decision rule

$$H(X \mid Y) \leq h_{p_{corr}} + (1 - p_{corr}) \log_2(r - 1)$$

Fano bound is equivalent to

$$H(X \mid Y) \leq -p_{corr} \log_2 p_{corr} - (1 - p_{corr}) \log_2 \frac{1 - p_{corr}}{r - 1}$$

Let's write $(i, j) \in \Delta$ for the decision rule maps $Y = j$ to $X = i$

Fano Bound (I)

Theorem (Fano Bound)

For any channel, r-ary source and decision rule

$$H(X \mid Y) \leq h_{p_{corr}} + (1 - p_{corr}) \log_2 (r-1)$$

Fano bound is equivalent to

$$H(X \mid Y) \leq -p_{corr} \log_2 p_{corr} - (1 - p_{corr}) \log_2 \frac{1 - p_{corr}}{r-1}$$

Let's write $(i,j) \in \Delta$ for the decision rule maps $Y = j$ to $X = i$

Then, $p_{corr} = \sum_{(i,j)\in\Delta} p_i t_{i,j}$ and $H(X \mid Y) = -\sum_{(i,j)} p_i t_{i,j} \log_2 \frac{p_i t_{i,j}}{q_j}$

or $H(X \mid Y) = \sum_{(i,j)\in\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}} + \sum_{(i,j)\notin\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}}$

Fano Bound (II)

$$p_{corr} = \sum_{(i,j)\in\Delta} p_i t_{i,j} \qquad H(X \mid Y) = \sum_{(i,j)\in\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}} + \sum_{(i,j)\notin\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}}$$

$$Q = H(X \mid Y) + p_{corr} \log_2 p_{corr} + (1 - p_{corr}) \log_2 \frac{1 - p_{corr}}{r - 1} =$$

$$\sum_{(i,j)\in\Delta} p_i t_{i,j} \log_2 \frac{q_j p_{corr}}{p_i t_{i,j}} + \sum_{(i,j)\notin\Delta} p_i t_{i,j} \log_2 \frac{q_j(1 - p_{corr})}{p_i t_{i,j}(r - 1)}$$

Fano Bound (II)

$$p_{corr} = \sum_{(i,j)\in\Delta} p_i t_{i,j} \qquad H(X \mid Y) = \sum_{(i,j)\in\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}} + \sum_{(i,j)\notin\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}} \qquad \log_2 Z \le \frac{Z-1}{\ln 2}$$

$$Q = H(X \mid Y) + p_{corr} \log_2 p_{corr} + (1 - p_{corr}) \log_2 \frac{1 - p_{corr}}{r - 1} =$$

$$\sum_{(i,j)\in\Delta} p_i t_{i,j} \log_2 \frac{q_j p_{corr}}{p_i t_{i,j}} + \sum_{(i,j)\notin\Delta} p_i t_{i,j} \log_2 \frac{q_j (1 - p_{corr})}{p_i t_{i,j} (r - 1)}$$

$$Q \ln 2 \le p_{corr} \sum_{(i,j)\in\Delta} q_j - \sum_{(i,j)\in\Delta} p_i t_{i,j} + \frac{(1 - p_{corr})}{(r - 1)} \sum_{(i,j)\notin\Delta} q_j - \sum_{(i,j)\notin\Delta} p_i t_{i,j}$$

Fano Bound (II)

$$p_{corr} = \sum_{(i,j)\in\Delta} p_i t_{i,j} \qquad H(X \mid Y) = \sum_{(i,j)\in\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}} + \sum_{(i,j)\notin\Delta} p_i t_{i,j} \log_2 \frac{q_j}{p_i t_{i,j}} \qquad \log_2 Z \leq \frac{Z-1}{\ln 2}$$

$$Q = H(X \mid Y) + p_{corr} \log_2 p_{corr} + (1-p_{corr}) \log_2 \frac{1-p_{corr}}{r-1} =$$

$$\sum_{(i,j)\in\Delta} p_i t_{i,j} \log_2 \frac{q_j p_{corr}}{p_i t_{i,j}} + \sum_{(i,j)\notin\Delta} p_i t_{i,j} \log_2 \frac{q_j (1-p_{corr})}{p_i t_{i,j}(r-1)}$$

$$Q \ln 2 \leq p_{corr} \sum_{(i,j)\in\Delta} q_j - \sum_{(i,j)\in\Delta} p_i t_{i,j} + \frac{(1-p_{corr})}{(r-1)} \sum_{(i,j)\notin\Delta} q_j - \sum_{(i,j)\notin\Delta} p_i t_{i,j}$$

But $\sum_{(i,j)\in\Delta} q_j = 1$ and $\sum_{(i,j)\notin\Delta} q_j = r-1$ both imply $Q \ln 2 \leq 0$

Converse of Shannon's Theorem

Theorem

Given an r-ary channel (stateless and source independent) with capacity Λ *and a uniform source, for any* $\Lambda_1 > \Lambda_2 > \Lambda$ *there exists no sequence of codes* $\{\mathcal{C}_n\}_{n \geq n_0}$ *of constant length n and rates* R_n *such that* $\Lambda_1 > R_n \log_2 r > \Lambda_2$ *and* $p_{corr} \to 1$ *as* $n \to \infty$.

Converse of Shannon's Theorem

Theorem

Given an r-ary channel (stateless and source independent) with capacity Λ and a uniform source, for any $\Lambda_1 > \Lambda_2 > \Lambda$ there exists no sequence of codes $\{\mathcal{C}_n\}_{n \geq n_0}$ of constant length n and rates R_n such that $\Lambda_1 > R_n \log_2 r > \Lambda_2$ and $p_{corr} \to 1$ as $n \to \infty$.

For X uniformly distributed in $\mathcal{C}_n$, using Fano bound

$$nR_n \log_2 r = H(X) = H(X \mid Y) + I(X;Y) \leq$$
$$h_{p_{corr}} + (1 - p_{corr})nR_n \log_2 r + n\Lambda \leq 1 + (1 - p_{corr})nR_n \log_2 r + n\Lambda$$

Converse of Shannon's Theorem

Theorem

Given an r-ary channel (stateless and source independent) with capacity Λ and a uniform source, for any $\Lambda_1 > \Lambda_2 > \Lambda$ there exists no sequence of codes $\{\mathcal{C}_n\}_{n \geq n_0}$ of constant length n and rates R_n such that $\Lambda_1 > R_n \log_2 r > \Lambda_2$ and $p_{corr} \to 1$ as $n \to \infty$.

For X uniformly distributed in $\mathcal{C}_n$, using Fano bound
$nR_n \log_2 r = H(X) = H(X \mid Y) + I(X;Y) \leq$
$h_{p_{corr}} + (1 - p_{corr})nR_n \log_2 r + n\Lambda \leq 1 + (1 - p_{corr})nR_n \log_2 r + n\Lambda$

Thus, $p_{corr} \leq \dfrac{1 + n\Lambda}{nR_n \log_2 r} \leq \dfrac{1 + n\Lambda}{n\Lambda_2}$

Converse of Shannon's Theorem

Theorem

Given an r-ary channel (stateless and source independent) with capacity Λ and a uniform source, for any $\Lambda_1 > \Lambda_2 > \Lambda$ there exists no sequence of codes $\{\mathcal{C}_n\}_{n \geq n_0}$ of constant length n and rates R_n such that $\Lambda_1 > R_n \log_2 r > \Lambda_2$ and $p_{corr} \to 1$ as $n \to \infty$.

For X uniformly distributed in $\mathcal{C}_n$, using Fano bound

$$nR_n \log_2 r = H(X) = H(X \mid Y) + I(X;Y) \leq$$
$$h_{p_{corr}} + (1 - p_{corr})nR_n \log_2 r + n\Lambda \leq 1 + (1 - p_{corr})nR_n \log_2 r + n\Lambda$$

Thus, $p_{corr} \leq \dfrac{1 + n\Lambda}{nR_n \log_2 r} \leq \dfrac{1 + n\Lambda}{n\Lambda_2}$

Finally, for $n > \dfrac{2}{\Lambda_2 - \Lambda}$, p_{corr} cannot exceed $1 - \dfrac{\Lambda_2 - \Lambda}{2\Lambda_2}$.

Remarks

A transmission channel has a maximum transmission rate, the channel capacity Λ.

Large codes with $p_{corr} \approx 1$ and a rate close than $\dfrac{\Lambda}{\log_2 r}$ exist, based on the properties of random codes.

Remarks

A transmission channel has a maximum transmission rate, the channel capacity Λ.

Large codes with $p_{corr} \approx 1$ and a rate close than $\frac{\Lambda}{\log_2 r}$ exist, based on the properties of random codes.

Still, these optimal codes have inefficient decoding.

Remarks

A transmission channel has a maximum transmission rate, the channel capacity Λ.

Large codes with $p_{corr} \approx 1$ and a rate close than $\dfrac{\Lambda}{\log_2 r}$ exist, based on the properties of random codes.

Still, these optimal codes have inefficient decoding.

There are also some combinatorial bounds limiting the existence of codes with practical lengths.

Remarks

A transmission channel has a maximum transmission rate, the channel capacity Λ.

Large codes with $p_{corr} \approx 1$ and a rate close than $\dfrac{\Lambda}{\log_2 r}$ exist, based on the properties of random codes.

Still, these optimal codes have inefficient decoding.

There are also some combinatorial bounds limiting the existence of codes with practical lengths.

The goal is now build not-too-large efficiently decodable codes with good p_{corr} and rate.

Codes and Cryptography

Jorge L. Villar

MAMME, Fall 2015

END OF PART IV

Proof of the Lemma

For any $\alpha \leq \frac{1}{2}$ and $d \leq n$

$$1 = \sum_{i=0}^{n} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} = (1-\alpha)^n \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\alpha}{1-\alpha}\right)^i \geq$$

$$\geq (1-\alpha)^n \sum_{i=0}^{d} \binom{n}{i} \left(\frac{\alpha}{1-\alpha}\right)^d = \alpha^d (1-\alpha)^{n-d} \sum_{i=0}^{d} \binom{n}{i}$$

Proof of the Lemma

For any $\alpha \leq \frac{1}{2}$ and $d \leq n$

$$1 = \sum_{i=0}^{n} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} = (1-\alpha)^n \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\alpha}{1-\alpha}\right)^i \geq$$

$$\geq (1-\alpha)^n \sum_{i=0}^{d} \binom{n}{i} \left(\frac{\alpha}{1-\alpha}\right)^d = \alpha^d (1-\alpha)^{n-d} \sum_{i=0}^{d} \binom{n}{i}$$

Thus, $\displaystyle \sum_{i=0}^{d} \binom{n}{i} \leq \alpha^{-d} (1-\alpha)^{-(n-d)}$

Proof of the Lemma

For any $\alpha \leq \frac{1}{2}$ and $d \leq n$

$$1 = \sum_{i=0}^{n} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} = (1-\alpha)^n \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\alpha}{1-\alpha}\right)^i \geq$$

$$\geq (1-\alpha)^n \sum_{i=0}^{d} \binom{n}{i} \left(\frac{\alpha}{1-\alpha}\right)^d = \alpha^d (1-\alpha)^{n-d} \sum_{i=0}^{d} \binom{n}{i}$$

Thus, $\displaystyle\sum_{i=0}^{d} \binom{n}{i} \leq \alpha^{-d} (1-\alpha)^{-(n-d)}$

Now, if $d \leq \frac{n}{2}$ we can set $\alpha = \frac{d}{n}$, that is $d = n\alpha$, and then

$$\sum_{i=0}^{d} \binom{n}{i} \leq \left(\alpha^\alpha (1-\alpha)^{1-\alpha}\right)^{-n} = 2^{nh_\alpha}$$

go back...