Information and Coding Theory (CO349)

Transmission of Information

Herbert Wiklicky
herbert@doc.ic.ac.uk

Department of Computing
Imperial College London

Autumn 2018

Channel

What we have

- A code to represent information
- A source which produces information
- A channel which transmits information.

A channel links/is linking a sender and a receiver.

Channel

What we have

- A code to represent information
- A source which produces information
- A channel which transmits information.

A channel links/is linking a sender and a receiver.

Channel

What we have

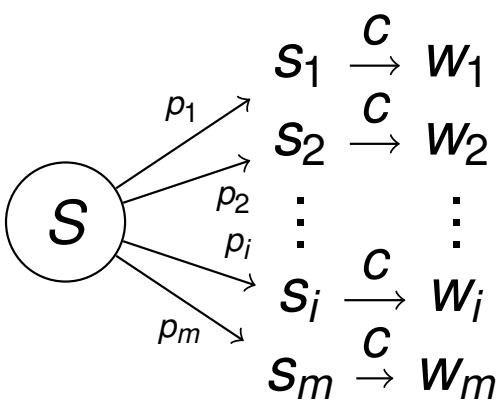
- A code to represent information
- A source which produces information
- A channel which transmits information.

A channel links/is linking a sender and a receiver.

Channel

What we have

- A code to represent information
- A source which produces information
- A channel which transmits information.

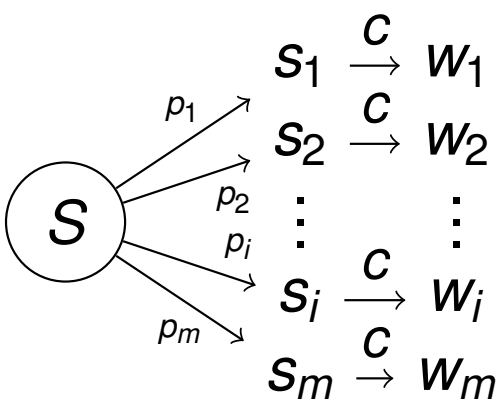


A channel links/is linking a sender and a receiver.

Channel

What we have

- A code to represent information
- A source which produces information
- A channel which transmits information.

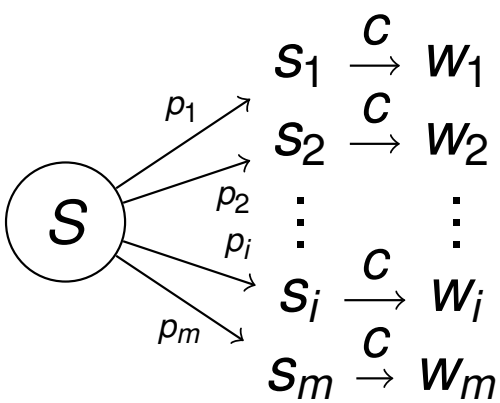


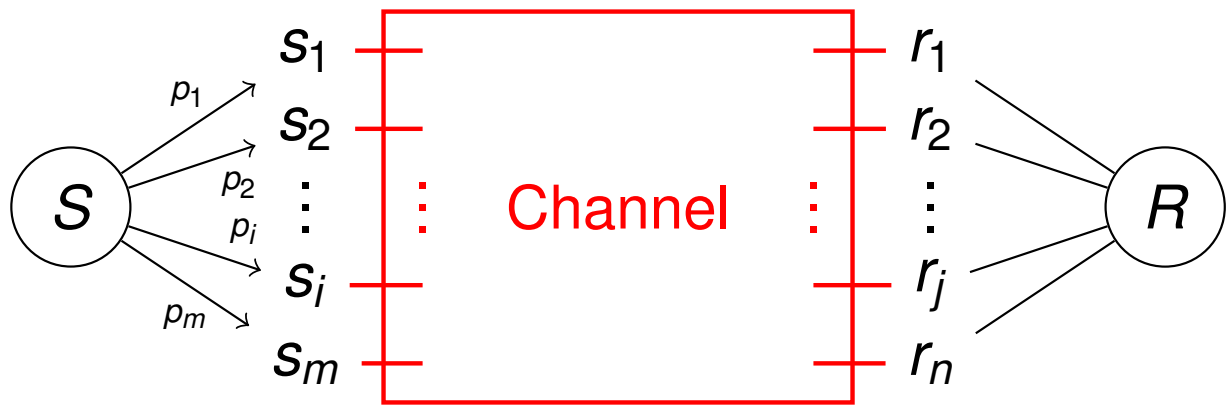
A channel links/is linking a sender and a receiver.

Channel

What we have

- A code to represent information
- A source which produces information
- A channel which transmits information.



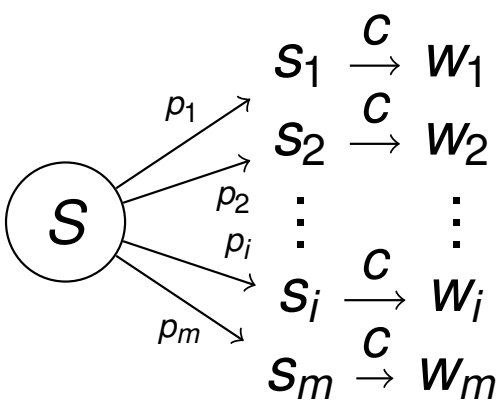


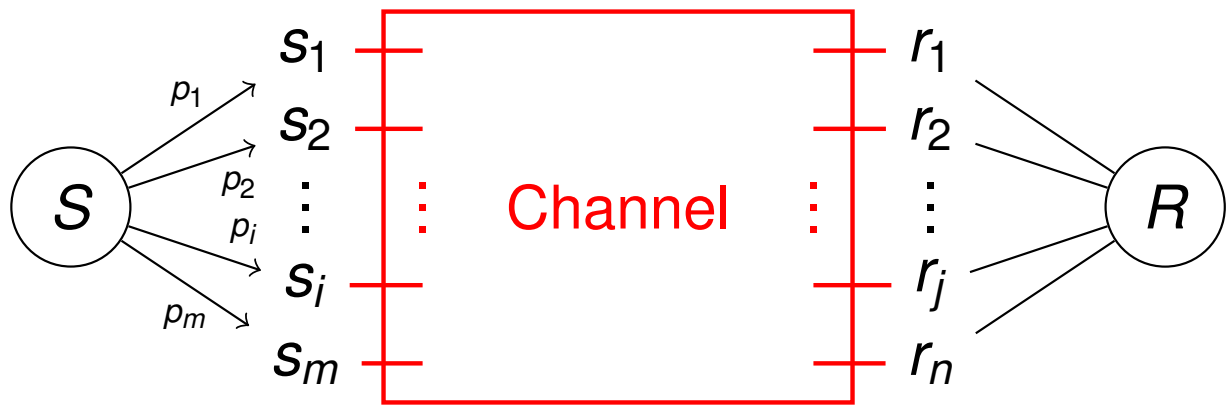
A channel links/is linking a sender and a receiver.

Channel

What we have

- A code to represent information
- A source which produces information
- A channel which transmits information.





A channel links/is linking a sender and a receiver.

Noisy Channel

Definition

A channel (matrix) Γ with input set $I = \{s_1, \ldots, s_m\}$ and output set $J = \{r_1, \ldots, r_n\}$ is a matrix whose entries define the (conditional) probabilities $\Pr(r_j \mid s_i)$ with $i \in I$ and $j \in J$, i.e.

$$\Gamma_{ij} = \Pr(j \mid i) = \Pr(r_j \mid s_i)$$

If at least one $\Gamma_{ij} \neq 0$ with $i \neq j$ then the channel is noisy.

Rows correspond to inputs and columns to outputs.

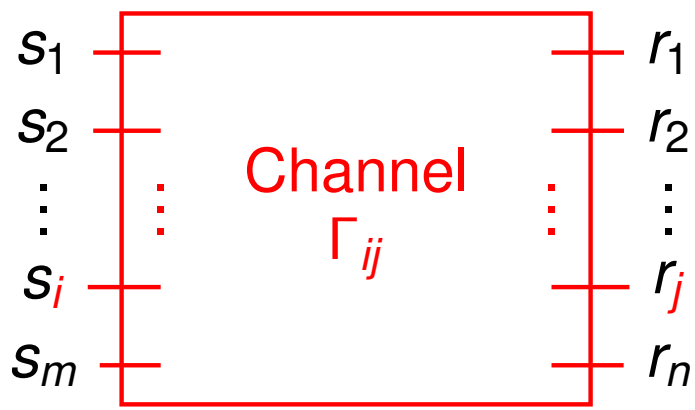
Noisy Channel

Definition

A channel (matrix) Γ with input set $I = \{s_1, \ldots, s_m\}$ and output set $J = \{r_1, \ldots, r_n\}$ is a matrix whose entries define the (conditional) probabilities $\Pr(r_j \mid s_i)$ with $i \in I$ and $j \in J$, i.e.

$$\Gamma_{ij} = \Pr(j \mid i) = \Pr(r_j \mid s_i)$$

If at least one $\Gamma_{ij} \neq 0$ with $i \neq j$ then the channel is noisy.



Rows correspond to inputs and columns to outputs.

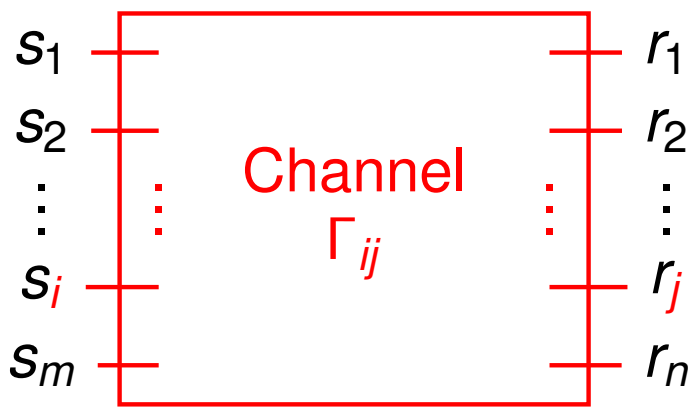
Noisy Channel

Definition

A channel (matrix) Γ with input set $I = \{s_1, \ldots, s_m\}$ and output set $J = \{r_1, \ldots, r_n\}$ is a matrix whose entries define the (conditional) probabilities $\Pr(r_j \mid s_i)$ with $i \in I$ and $j \in J$, i.e.

$$\Gamma_{ij} = \Pr(j \mid i) = \Pr(r_j \mid s_i)$$

If at least one $\Gamma_{ij} \neq 0$ with $i \neq j$ then the channel is noisy.



Rows correspond to inputs and columns to outputs.

Binary Symmetric Channel

Definition

A binary symmetric channel (BSC) corrsponds to the channel matrix of the form:

$$\Gamma = \begin{pmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{pmatrix} = \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix}$$

with error $e > 0$.

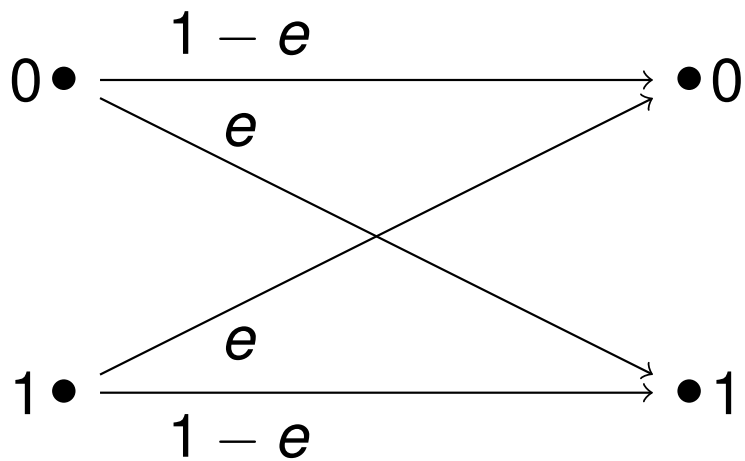
Binary Symmetric Channel

Definition

A binary symmetric channel (BSC) corrsponds to the channel matrix of the form:

$$\Gamma = \begin{pmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{pmatrix} = \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix}$$

with error $e > 0$.



Channel Transmission

Lemma

Each row of a channel matrix Γ *has sum* 1*, i.e. is "stochastic",*

$$\sum_{j \in J} \Gamma_{ij} = 1 \quad \textit{for all } i \in I$$

Theorem

Let Γ *be a channel matrix and* $(I, \mathbf{p})$ *and* $(J, \mathbf{q})$ *be the sources associated to the input and output with distributions:*

$$\mathbf{p} = (p_1, p_2, \ldots, p_m) \quad \textit{and} \quad \mathbf{q} = (q_1, q_2, \ldots, q_n)$$

then

$$\mathbf{q} = \mathbf{p}\Gamma$$

Channel Transmission

Lemma

Each row of a channel matrix Γ *has sum* 1*, i.e. is "stochastic",*

$$\sum_{j \in J} \Gamma_{ij} = 1 \quad \textit{for all } i \in I$$

Theorem

Let Γ *be a channel matrix and* $(I, \mathbf{p})$ *and* $(J, \mathbf{q})$ *be the sources associated to the input and output with distributions:*

$$\mathbf{p} = (p_1, p_2, \ldots, p_m) \quad \textit{and} \quad \mathbf{q} = (q_1, q_2, \ldots, q_n)$$

then

$$\mathbf{q} = \mathbf{p}\Gamma$$

Proof

It is obvious (by definition) that Γ's rows sum up to one.

Proof.

Denote by t_{ij} probability that input is s_i and output is r_j or $t_{ij} = \Pr(\text{input } i \text{ and output } j)$. Therefore, $q_j = \sum_{i \in I} t_{ij}$.

But also, $t_{ij} = \Pr(\text{output } j \mid \text{input } i) \times \Pr(\text{input } i) = \Gamma_{ij} p_i$.

Thus

$$q_j = \sum_{i \in I} t_{ij} = \sum_{i \in I} \Gamma_{ij} p_i \quad \text{or} \quad \mathbf{q} = \mathbf{p}\Gamma.$$

□

Proof

It is obvious (by definition) that Γ's rows sum up to one.

Proof.

Denote by t_{ij} probability that input is s_i and output is r_j or $t_{ij} = \Pr(\text{input } i \text{ and output } j)$. Therefore, $q_j = \sum_{i \in I} t_{ij}$.

But also, $t_{ij} = \Pr(\text{output } j \mid \text{input } i) \times \Pr(\text{input } i) = \Gamma_{ij} p_i$.

Thus

$$q_j = \sum_{i \in I} t_{ij} = \sum_{i \in I} \Gamma_{ij} p_i \quad \text{or} \quad \mathbf{q} = \mathbf{p}\Gamma.$$

□

Example

Example

Consider the BSC (Binary Symmetric Channel) with $e = 0.1$.
Assume an input probability $\mathbf{p} = (0.7, 0.3)$.
In general we have for the output probabilities $\mathbf{q}$ the equation:

$$\begin{pmatrix} q_0 & q_1 \end{pmatrix} = \begin{pmatrix} p_0 & p_1 \end{pmatrix} \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix}$$

or, alternatively:

$$\begin{aligned} q_0 &= p_0(1-e) + p_1 e \\ q_1 &= p_0 e + p_1(1-e) \end{aligned}$$

So for $e = 0.1$ and $\mathbf{p} = (0.7, 0.3)$ we have $\mathbf{q} = (0.66, 0.34)$.

Conditional Entropy

Consider the probability distribution **t** on $I \times J$ as

$$t_{ij} = \Pr(\text{input } i \text{ and output } j) \neq \Pr(\text{output } j \text{ if input } i)$$

We can use this to construct [cf. marginal distributions]:

$$p_i = \sum_j t_{ij} \qquad q_j = \sum_i t_{ij} \qquad \Gamma_{ij} = \frac{t_{ij}}{p_i}$$

Definition

The conditional entropy $\mathbf{H}(\mathbf{p} \mid \mathbf{q})$ (for $\mathbf{q} = \mathbf{p}\Gamma$) is defined as:

$$\mathbf{H}(\mathbf{p} \mid \mathbf{q}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q}) = \mathbf{H}(\Gamma; \mathbf{p})$$

Conditional entropy of $\mathbf{p}$ with respect to transmission through Γ.

Conditional Entropy

Consider the probability distribution **t** on $I \times J$ as

$$t_{ij} = \Pr(\text{input } i \text{ and output } j) \neq \Pr(\text{output } j \text{ if input } i)$$

We can use this to construct [cf. marginal distributions]:

$$p_i = \sum_j t_{ij} \qquad q_j = \sum_i t_{ij} \qquad \Gamma_{ij} = \frac{t_{ij}}{p_i}$$

Definition

The conditional entropy $\mathbf{H}(\mathbf{p} \mid \mathbf{q})$ (for $\mathbf{q} = \mathbf{p}\Gamma$) is defined as:

$$\mathbf{H}(\mathbf{p} \mid \mathbf{q}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q}) = \mathbf{H}(\Gamma; \mathbf{p})$$

Conditional entropy of $\mathbf{p}$ with respect to transmission through Γ.

Conditional Entropy

Consider the probability distribution **t** on $I \times J$ as

$$t_{ij} = \Pr(\text{input } i \text{ and output } j) \neq \Pr(\text{output } j \text{ if input } i)$$

We can use this to construct [cf. marginal distributions]:

$$p_i = \sum_j t_{ij} \qquad q_j = \sum_i t_{ij} \qquad \Gamma_{ij} = \frac{t_{ij}}{p_i}$$

Definition

The conditional entropy $\mathbf{H}(\mathbf{p} \mid \mathbf{q})$ (for $\mathbf{q} = \mathbf{p}\Gamma$) is defined as:

$$\mathbf{H}(\mathbf{p} \mid \mathbf{q}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q}) = \mathbf{H}(\Gamma; \mathbf{p})$$

Conditional entropy of **p** with respect to transmission through Γ.

Example

Example

Consider the BSC (Binary Symmetric Channel) with $e = 0.1$. Assume an input probability $\mathbf{p} = (0.7, 0.3)$.

We can compute $\mathbf{t}$ from Γ and $\mathbf{p}$ as $t_{ij} = p_i \Gamma_{ij}$.

$$t_{00} = 0.63, \quad t_{01} = 0.07, \quad t_{10} = 0.03, \quad t_{11} = 0.27$$

Therefore, $\mathbf{H}(\mathbf{t}) \approx 1.350$. We also know $\mathbf{q} = (0.66, 0.34)$, so $\mathbf{H}(\mathbf{q}) \approx 0.925$. Hence

$$\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q}) \approx 1.350 - 0.925 = 0.425$$

Conditional Entropy for BSC

Theorem

Let Γ *be a BSC with bit-error probability* e*, and* $\mathbf{p}$ *the source distribution* $\mathbf{p} = (p_0, p_1) = (p, 1-p)$ *then*

$$\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q)$$

where $q = p(1-e) + (1-p)e$ *and* $\mathbf{h}$ *is the standard entropy*

$$\mathbf{h}(x) = x \cdot \log_2\left(\frac{1}{x}\right) + (1-x) \cdot \log_2\left(\frac{1}{1-x}\right)$$

Proof

Proof.

We simply can compute $t_{ij} = p_i \Gamma_{ij}$:

$$t_{00} = p(1-e), \quad t_{01} = pe, \quad t_{10} = (1-p)e, \quad t_{11} = (1-p)(1-e)$$

We can also express $\mathbf{t}$ as product of two independent distributions: $\mathbf{t} = \mathbf{p} \otimes \mathbf{e}$ or $t_{ij} = p_i e_j$ with $\mathbf{e} = (1-e, e)$.

From before (product entropy) we know how to compute the entropy of a product of (independent) distributions, so we get:

$$\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q}) = \mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{e}) - \mathbf{H}(\mathbf{q})$$

or in this binary case: $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q)$. □

Proof

Proof.

We simply can compute $t_{ij} = p_i \Gamma_{ij}$:

$$t_{00} = p(1-e), \quad t_{01} = pe, \quad t_{10} = (1-p)e, \quad t_{11} = (1-p)(1-e)$$

We can also express $\mathbf{t}$ as product of two independent distributions: $\mathbf{t} = \mathbf{p} \otimes \mathbf{e}$ or $t_{ij} = p_i e_j$ with $\mathbf{e} = (1-e, e)$.

From before (product entropy) we know how to compute the entropy of a product of (independent) distributions, so we get:

$$\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q}) = \mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{e}) - \mathbf{H}(\mathbf{q})$$

or in this binary case: $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q)$. □

Channel and Source Entropy

Theorem

Let Γ *be a channel and* $\mathbf{p}$ *an input source distribution. Then*

$$\mathbf{H}(\Gamma;\mathbf{p}) \leq \mathbf{H}(\mathbf{p})$$

Equality holds if and only if $\mathbf{p}$ *and* $\mathbf{q} = \mathbf{p}\Gamma$ *are independent.*

Proof.

The marginal distributions of $\mathbf{t}$ are $\mathbf{p}$ and $\mathbf{q}$. Therefore, $\mathbf{H}(\mathbf{t}) \leq \mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{q})$, therefore, as $\mathbf{H}(\Gamma;\mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q})$:

$$\mathbf{H}(\Gamma;\mathbf{p}) + \mathbf{H}(\mathbf{q}) = \mathbf{H}(\mathbf{t}) \leq \mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{q})$$

or $\mathbf{H}(\Gamma;\mathbf{p}) \leq \mathbf{H}(\mathbf{p})$ with equality iff $\mathbf{p}$ and $\mathbf{q}$ are independent. □

Channel and Source Entropy

Theorem

Let Γ *be a channel and* $\mathbf{p}$ *an input source distribution. Then*

$$\mathbf{H}(\Gamma;\mathbf{p}) \leq \mathbf{H}(\mathbf{p})$$

Equality holds if and only if $\mathbf{p}$ *and* $\mathbf{q} = \mathbf{p}\Gamma$ *are independent.*

Proof.

The marginal distributions of $\mathbf{t}$ are $\mathbf{p}$ and $\mathbf{q}$. Therefore, $\mathbf{H}(\mathbf{t}) \leq \mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{q})$, therefore, as $\mathbf{H}(\Gamma;\mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q})$:

$$\mathbf{H}(\Gamma;\mathbf{p}) + \mathbf{H}(\mathbf{q}) = \mathbf{H}(\mathbf{t}) \leq \mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{q})$$

or $\mathbf{H}(\Gamma;\mathbf{p}) \leq \mathbf{H}(\mathbf{p})$ with equality iff $\mathbf{p}$ and $\mathbf{q}$ are independent. □

Capacity of a Channel

Definition

The capacity γ of a channel Γ is the maximum difference $\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma; \mathbf{p})$ over all distributions (on m elements) in $\mathcal{P}$ or

$$\gamma = \gamma(\Gamma) = \max_{\mathbf{p} \in \mathcal{P}} (\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma; \mathbf{p}))$$

It is important to show that the maximum indeed always exists. This can be done using arguments based on the convexity of the entropy function.

As $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{H}(\mathbf{p} \mid \mathbf{q}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q})$ we have with $\mathbf{q} = \mathbf{p}\Gamma$:

$$\gamma(\Gamma) = \max_{\mathbf{p}} (\mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{t})).$$

Capacity of a Channel

Definition

The capacity γ of a channel Γ is the maximum difference $\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma; \mathbf{p})$ over all distributions (on m elements) in $\mathcal{P}$ or

$$\gamma = \gamma(\Gamma) = \max_{\mathbf{p} \in \mathcal{P}} (\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma; \mathbf{p}))$$

It is important to show that the maximum indeed always exists. This can be done using arguments based on the convexity of the entropy function.

As $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{H}(\mathbf{p} \mid \mathbf{q}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q})$ we have with $\mathbf{q} = \mathbf{p}\Gamma$:

$$\gamma(\Gamma) = \max_{\mathbf{p}} (\mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{t})).$$

Capacity of a Channel

Definition

The capacity γ of a channel Γ is the maximum difference $\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma; \mathbf{p})$ over all distributions (on m elements) in $\mathcal{P}$ or

$$\gamma = \gamma(\Gamma) = \max_{\mathbf{p} \in \mathcal{P}} (\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma; \mathbf{p}))$$

It is important to show that the maximum indeed always exists. This can be done using arguments based on the convexity of the entropy function.

As $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{H}(\mathbf{p} \mid \mathbf{q}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{q})$ we have with $\mathbf{q} = \mathbf{p}\Gamma$:

$$\gamma(\Gamma) = \max_{\mathbf{p}} (\mathbf{H}(\mathbf{p}) + \mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{t})).$$

BSC Capacity

Theorem

Let Γ *be a BSC with bit-error probability* e *with* $0 \leq e \leq \frac{1}{2}$*, then*

$$\gamma = \gamma(\Gamma) = 1 - \mathbf{h}(e)$$

Proof.

For $\mathbf{p} = (p, 1-p)$ and BSC with e we know the entropy as $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q)$ with $q = p(1-e) + (1-p)e$.

$$\gamma = \max(\mathbf{h}(p) - (\mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q))) = \max(\mathbf{h}(q) - \mathbf{h}(e)).$$

Since e is constant, the maximum is reached for maximum of $\mathbf{h}(q)$ which we get for $q = \frac{1}{2} = p(1-e) + (1-p)e$ or when $(p - \frac{1}{2})(1 - 2e) = 0$. So for $p = \frac{1}{2}$ we have maximum of $\mathbf{h}(q) = 1$, then $\gamma = \max(\mathbf{h}(q) - \mathbf{h}(e)) = 1 - \mathbf{h}(e)$. □

BSC Capacity

Theorem

Let Γ *be a BSC with bit-error probability* e *with* $0 \leq e \leq \frac{1}{2}$*, then*

$$\gamma = \gamma(\Gamma) = 1 - \mathbf{h}(e)$$

Proof.

For $\mathbf{p} = (p, 1-p)$ and BSC with e we know the entropy as $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q)$ with $q = p(1-e) + (1-p)e$.

$$\gamma = \max(\mathbf{h}(p) - (\mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q))) = \max(\mathbf{h}(q) - \mathbf{h}(e)).$$

Since e is constant, the maximum is reached for maximum of $\mathbf{h}(q)$ which we get for $q = \frac{1}{2} = p(1-e) + (1-p)e$ or when $(p - \frac{1}{2})(1-2e) = 0$. So for $p = \frac{1}{2}$ we have maximum of $\mathbf{h}(q) = 1$, then $\gamma = \max(\mathbf{h}(q) - \mathbf{h}(e)) = 1 - \mathbf{h}(e)$. □

BSC Capacity

Theorem

Let Γ *be a BSC with bit-error probability* e *with* $0 \leq e \leq \frac{1}{2}$*, then*

$$\gamma = \gamma(\Gamma) = 1 - \mathbf{h}(e)$$

Proof.

For $\mathbf{p} = (p, 1-p)$ and BSC with e we know the entropy as $\mathbf{H}(\Gamma; \mathbf{p}) = \mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q)$ with $q = p(1-e) + (1-p)e$.

$$\gamma = \max(\mathbf{h}(p) - (\mathbf{h}(p) + \mathbf{h}(e) - \mathbf{h}(q))) = \max(\mathbf{h}(q) - \mathbf{h}(e)).$$

Since e is constant, the maximum is reached for maximum of $\mathbf{h}(q)$ which we get for $q = \frac{1}{2} = p(1-e) + (1-p)e$ or when $(p - \frac{1}{2})(1-2e) = 0$. So for $p = \frac{1}{2}$ we have maximum of $\mathbf{h}(q) = 1$, then $\gamma = \max(\mathbf{h}(q) - \mathbf{h}(e)) = 1 - \mathbf{h}(e)$. □

Conditional Entropy (Redux)

For input $i \in I$ denote $\mathbf{H}(\mathbf{q} \mid i) = \sum_{j \in J} \Gamma_{ij} \log\left(\frac{1}{\Gamma_{ij}}\right)$ with $\mathbf{q} = \mathbf{p}\Gamma$

Theorem

The conditional entropy $\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{p})$ *can also be calculated as:*

$$\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i)$$

Proof $*$.

$$\begin{aligned} \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \mathbf{H}(\mathbf{p}) &= \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \sum_{i \in I} p_i \cdot \log(\frac{1}{p_i}) \\ &= \sum_{i \in I} p_i \cdot \left(\mathbf{H}(\mathbf{q} \mid i) + \log(\frac{1}{p_i})\right) \end{aligned}$$

Conditional Entropy (Redux)

For input $i \in I$ denote $\mathbf{H}(\mathbf{q} \mid i) = \sum_{j \in J} \Gamma_{ij} \log\left(\frac{1}{\Gamma_{ij}}\right)$ with $\mathbf{q} = \mathbf{p}\Gamma$

Theorem

The conditional entropy $\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{p})$ *can also be calculated as:*

$$\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i)$$

Proof $*$.

$$\begin{aligned}
\sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \mathbf{H}(\mathbf{p}) &= \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \sum_{i \in I} p_i \cdot \log(\frac{1}{p_i}) \\
&= \sum_{i \in I} p_i \cdot \left(\mathbf{H}(\mathbf{q} \mid i) + \log(\frac{1}{p_i})\right)
\end{aligned}$$

Conditional Entropy (Redux)

For input $i \in I$ denote $\mathbf{H}(\mathbf{q} \mid i) = \sum_{j \in J} \Gamma_{ij} \log\left(\frac{1}{\Gamma_{ij}}\right)$ with $\mathbf{q} = \mathbf{p}\Gamma$

Theorem

The conditional entropy $\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{p})$ *can also be calculated as:*

$$\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i)$$

Proof $*$.

$$\begin{aligned} \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \mathbf{H}(\mathbf{p}) &= \sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \sum_{i \in I} p_i \cdot \log(\frac{1}{p_i}) \\ &= \sum_{i \in I} p_i \cdot \left(\mathbf{H}(\mathbf{q} \mid i) + \log(\frac{1}{p_i})\right) \end{aligned}$$

Proof (cont.)

Proof (cont).

Applying definition of $\mathbf{H}(\mathbf{q} \mid i)$ and the fact $\sum_j \Gamma_{ij} = 1$ gives:

$$\begin{aligned}\mathbf{H}(\mathbf{q} \mid i) + \log(\frac{1}{p_i}) &= \sum_{j \in J} \Gamma_{ij} \cdot \log(\frac{1}{\Gamma_{ij}}) + \sum_{j \in J} \Gamma_{ij} \cdot \log(\frac{1}{p_i}) \\ &= \sum_{j \in J} \Gamma_{ij} \cdot \log(\frac{1}{p_i \Gamma_{ij}}) = \frac{1}{p_i} \sum_{j \in J} t_{ij} \cdot \log(\frac{1}{t_{ij}})\end{aligned}$$

since $p_i \Gamma_{ij} = t_{ij}$. Therefore, we get

$$\sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \mathbf{H}(\mathbf{p}) = \sum_{i \in I, j \in J} t_{ij} \cdot \log(\frac{1}{t_{ij}}) = \mathbf{H}(\mathbf{t})$$

from which with $\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{p})$ the claim follows. □

Proof (cont.)

Proof (cont).

Applying definition of $\mathbf{H}(\mathbf{q} \mid i)$ and the fact $\sum_j \Gamma_{ij} = 1$ gives:

$$\begin{aligned}\mathbf{H}(\mathbf{q} \mid i) + \log(\frac{1}{p_i}) &= \sum_{j \in J} \Gamma_{ij} \cdot \log(\frac{1}{\Gamma_{ij}}) + \sum_{j \in J} \Gamma_{ij} \cdot \log(\frac{1}{p_i}) \\ &= \sum_{j \in J} \Gamma_{ij} \cdot \log(\frac{1}{p_i \Gamma_{ij}}) = \frac{1}{p_i} \sum_{j \in J} t_{ij} \cdot \log(\frac{1}{t_{ij}})\end{aligned}$$

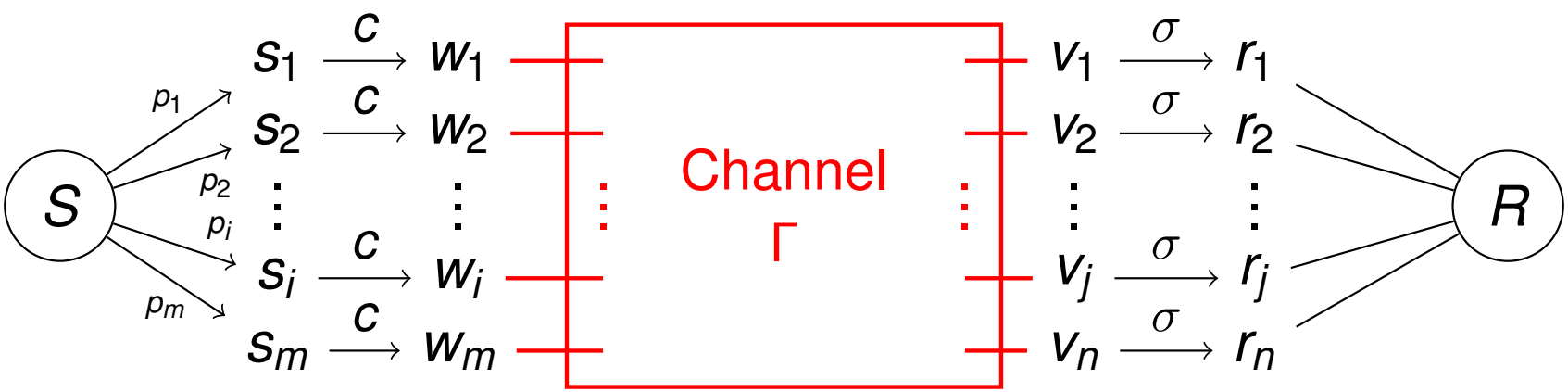
since $p_i \Gamma_{ij} = t_{ij}$. Therefore, we get

$$\sum_{i \in I} p_i \cdot \mathbf{H}(\mathbf{q} \mid i) + \mathbf{H}(\mathbf{p}) = \sum_{i \in I, j \in J} t_{ij} \cdot \log(\frac{1}{t_{ij}}) = \mathbf{H}(\mathbf{t})$$

from which with $\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \mathbf{H}(\mathbf{t}) - \mathbf{H}(\mathbf{p})$ the claim follows. □

Communication via Noisy Channels

The full problem: transmit information from Sender to Receiver:



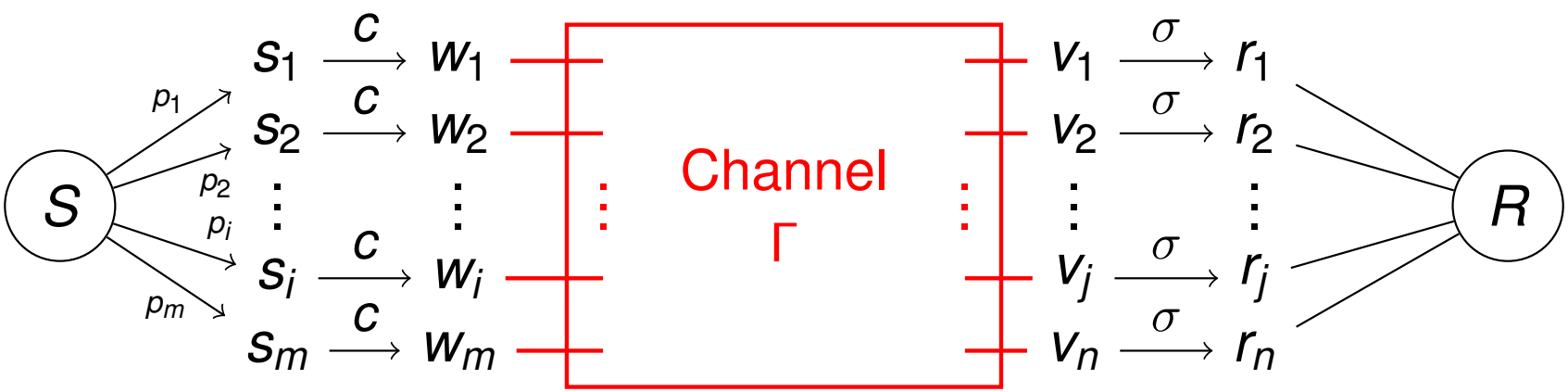
T_1: Original Stream $\to$ Encoded Stream
Coding $c : S \to C \subseteq \mathbb{B}^m$.

T_2: Encoded Stream $\to$ Received Stream
Channel introduces errors $\Gamma : \mathbb{B}^m \to \mathbb{B}^n$.

T_3: Received Stream $\to$ Final Stream
Decision rules $\sigma : \mathbb{B}^n \to C$.

Communication via Noisy Channels

The full problem: transmit information from Sender to Receiver:



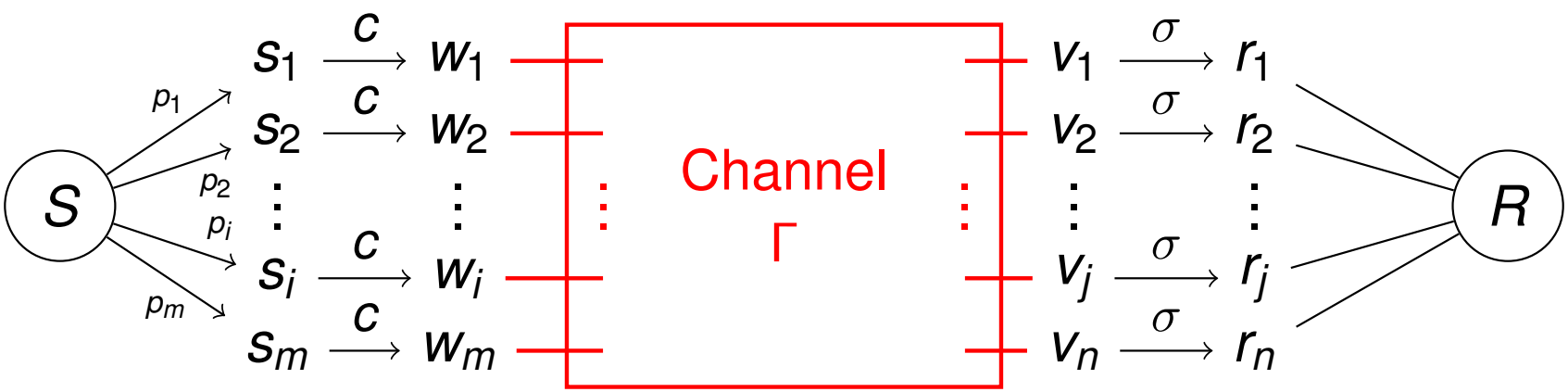
T_1: Original Stream $\to$ Encoded Stream
Coding $c : S \to C \subseteq \mathbb{B}^m$.

T_2: Encoded Stream $\to$ Received Stream
Channel introduces errors $\Gamma : \mathbb{B}^m \to \mathbb{B}^n$.

T_3: Received Stream $\to$ Final Stream
Decision rules $\sigma : \mathbb{B}^n \to C$.

Communication via Noisy Channels

The full problem: transmit information from Sender to Receiver:



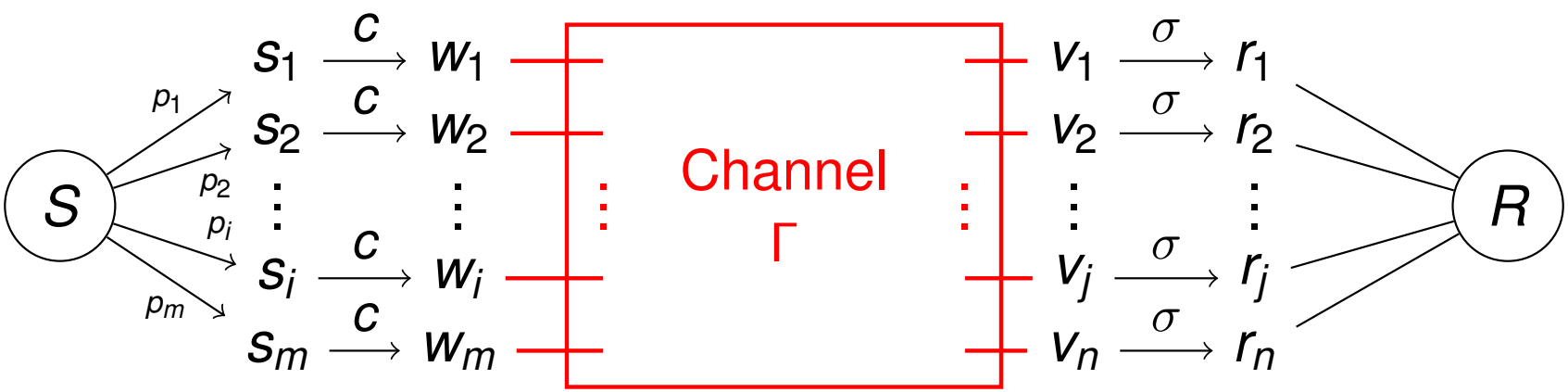
T_1: Original Stream $\to$ Encoded Stream
Coding $c : S \to C \subseteq \mathbb{B}^m$.

T_2: Encoded Stream $\to$ Received Stream
Channel introduces errors $\Gamma : \mathbb{B}^m \to \mathbb{B}^n$.

T_3: Received Stream $\to$ Final Stream
Decision rules $\sigma : \mathbb{B}^n \to C$.

Communication via Noisy Channels

The full problem: transmit information from Sender to Receiver:



T_1: Original Stream $\rightarrow$ Encoded Stream
Coding $c : S \rightarrow C \subseteq \mathbb{B}^m$.

T_2: Encoded Stream $\rightarrow$ Received Stream
Channel introduces errors $\Gamma : \mathbb{B}^m \rightarrow \mathbb{B}^n$.

T_3: Received Stream $\rightarrow$ Final Stream
Decision rules $\sigma : \mathbb{B}^n \rightarrow C$.

Dealing with Errors

Consider the codewords we obtain via a binary code $c : S \to \mathbb{B}^*$, i.e. the set of binary strings $C = c(S)$.

Definition

Given set of binary words $C \subseteq \mathbb{B}^n$. A decision rule for C is a function $\sigma : \mathbb{B}^n \to C$ which assigns to each $z \in \mathbb{B}^n$ a codeword in C.

Definition

We say that a mistake occurs if a codeword in the final stream is different from the codeword in the encoded stream (at corresponding positions).

Dealing with Errors

Consider the codewords we obtain via a binary code $c : S \to \mathbb{B}^*$, i.e. the set of binary strings $C = c(S)$.

Definition

Given set of binary words $C \subseteq \mathbb{B}^n$. A decision rule for C is a function $\sigma : \mathbb{B}^n \to C$ which assigns to each $z \in \mathbb{B}^n$ a codeword in C.

Definition

We say that a mistake occurs if a codeword in the final stream is different from the codeword in the encoded stream (at corresponding positions).

Dealing with Errors

Consider the codewords we obtain via a binary code $c : S \to \mathbb{B}^*$, i.e. the set of binary strings $C = c(S)$.

Definition

Given set of binary words $C \subseteq \mathbb{B}^n$. A decision rule for C is a function $\sigma : \mathbb{B}^n \to C$ which assigns to each $z \in \mathbb{B}^n$ a codeword in C.

Definition

We say that a mistake occurs if a codeword in the final stream is different from the codeword in the encoded stream (at corresponding positions).

Extended Channel

Definition

Given two channels Γ and Γ' with input alphabets I and I' and output alphabets J and J', respectively. The product channel $\Gamma'' = \Gamma \times \Gamma'$ has input alphabet $I \times I'$ and output alphaber $J \times J'$ and channel matrix:

$$\Gamma''_{\langle i,i' \rangle \langle j,j' \rangle} = \Gamma_{ij} \Gamma'_{i'j'} \quad \text{or} \quad \Gamma'' = \Gamma \otimes \Gamma'.$$

Definition

Given a channel Γ with input alphabet I and output alphabet J. The extended channel is defined on words of length n as Γ^n.

Extended Channel

Definition

Given two channels Γ and Γ' with input alphabets I and I' and output alphabets J and J', respectively. The product channel $\Gamma'' = \Gamma \times \Gamma'$ has input alphabet $I \times I'$ and output alphaber $J \times J'$ and channel matrix:

$$\Gamma''_{\langle i,i' \rangle \langle j,j' \rangle} = \Gamma_{ij} \Gamma'_{i'j'} \quad \text{or} \quad \Gamma'' = \Gamma \otimes \Gamma'.$$

Definition

Given a channel Γ with input alphabet I and output alphabet J. The extended channel is defined on words of length n as Γ^n.

Example

Example

Take Γ be the BSC with bit error e, then Γ^2 is represented by the channel matrix;

$$\begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix} \otimes \begin{pmatrix} 1-e & e \\ e & 1-e \end{pmatrix} =$$

$$= \begin{pmatrix} (1-e)^2 & e(1-e) & e(1-e) & e^2 \\ e(1-e) & (1-e)^2 & e^2 & e(1-e) \\ e(1-e) & e^2 & (1-e)^2 & e(1-e) \\ e^2 & e(1-e) & e(1-e) & (1-e)^2 \end{pmatrix}$$

where rows and columns are indexed by $00, 01, 10, 11$.

Hamming Distance

Definition

Given two binary words $x, y \in \mathbb{B}^n$ as

$$x = x_1 x_2 \cdots x_n \text{ and } y = y_1 y_2 \cdots y_n$$

The Hamming distance $d(x, y)$ is the number of places where x and y differ, i.e. the number of indices i with $x_i \neq y_i$.

Example

Consider the following words in $\mathbb{B}^7$:

$$x = 1010100, \quad y = 0110100, \quad z = 1011110,$$

their Hamming distances are:

$$d(x, y) = 2, \quad d(x, z) = 2, \quad d(y, z) = 4.$$

Hamming Distance

Definition

Given two binary words $x, y \in \mathbb{B}^n$ as

$$x = x_1 x_2 \cdots x_n \text{ and } y = y_1 y_2 \cdots y_n$$

The Hamming distance $d(x, y)$ is the number of places where x and y differ, i.e. the number of indices i with $x_i \neq y_i$.

Example

Consider the following words in $\mathbb{B}^7$:

$$x = 1010100, \quad y = 0110100, \quad z = 1011110,$$

their Hamming distances are:

$$d(x, y) = 2, \quad d(x, z) = 2, \quad d(y, z) = 4.$$

Extended BSC

Theorem

Given two binary words $x, y \in \mathbb{B}^n$. The entry $(\Gamma^n)_{xy}$ in the channel matrix of the extended BSC with bit-error e is given by

$$(\Gamma^n)_{xy} = e^d(1-e)^{n-d} \quad \textit{with} \quad d = d(x,y).$$

Proof.

Take input $x \in \mathbb{B}^n$ and output $y \in \mathbb{B}^n$ with Hamming distance $d(x,y) = d$. That is, they differ on d bits.
Thus d (bit) errors happened. The probability for this is e^d.
The remaining $n-d$ bits are transmitted correctly. This happens with probability $(1-e)^{n-d}$. Thus

$$(\Gamma^n)_{xy} = \Pr(y \mid x) = e^d(1-e)^{n-d}.$$

Extended BSC

Theorem

Given two binary words $x, y \in \mathbb{B}^n$. The entry $(\Gamma^n)_{xy}$ in the channel matrix of the extended BSC with bit-error e is given by

$$(\Gamma^n)_{xy} = e^d(1-e)^{n-d} \quad \textit{with} \quad d = d(x,y).$$

Proof.

Take input $x \in \mathbb{B}^n$ and output $y \in \mathbb{B}^n$ with Hamming distance $d(x,y) = d$. That is, they differ on d bits.
Thus d (bit) errors happened. The probability for this is e^d.
The remaining $n-d$ bits are transmitted correctly. This happens with probability $(1-e)^{n-d}$. Thus

$$(\Gamma^n)_{xy} = \Pr(y \mid x) = e^d(1-e)^{n-d}.$$

Decision Rules

Definition

The ideal observer rule is given by $\sigma(z) = c$ if the probability that z was sent given that c was received is maximal, i.e.

$$\Pr(c \mid z) = \max_{c'} \Pr(c' \mid z)$$

Definition

The maximal likelihood rule is given by $\sigma(z) = c$, i.e. if

$$\Pr(z \mid c) \geq \Pr(z \mid c') \quad \forall c' \in C$$

Definition

The minimum distance (MD) rule is given by $\sigma(z) = c$ such that $d(z, c)$ is minimal,

$$d(z, c) = \min_{c'} d(z, c')$$

Decision Rules

Definition

The ideal observer rule is given by $\sigma(z) = c$ if the probability that z was sent given that c was received is maximal, i.e.

$$\Pr(c \mid z) = \max_{c'} \Pr(c' \mid z)$$

Definition

The maximal likelihood rule is given by $\sigma(z) = c$, i.e. if

$$\Pr(z \mid c) \geq \Pr(z \mid c') \quad \forall c' \in C$$

Definition

The minimum distance (MD) rule is given by $\sigma(z) = c$ such that $d(z, c)$ is minimal,

$$d(z, c) = \min_{c'} d(z, c')$$

Decision Rules

Definition

The ideal observer rule is given by $\sigma(z) = c$ if the probability that z was sent given that c was received is maximal, i.e.

$$\Pr(c \mid z) = \max_{c'} \Pr(c' \mid z)$$

Definition

The maximal likelihood rule is given by $\sigma(z) = c$, i.e. if

$$\Pr(z \mid c) \geq \Pr(z \mid c') \quad \forall c' \in C$$

Definition

The minimum distance (MD) rule is given by $\sigma(z) = c$ such that $d(z, c)$ is minimal,

$$d(z, c) = \min_{c'} d(z, c')$$

Equivalence of Decision Rules

Theorem

For an extended BSC channel Γ^n with bit-error $e < \frac{1}{2}$ the maximal likelihood rule is equivalent to the MD rule.

Proof.

Suppose $c \in C$ with $d(c, z) = i$ then $\Pr(z \mid c) = e^i(1-e)^{n-i}$.
For a second $c' \in C$ s.t $d(c', z) = i' < i$ we have

$$\frac{\Pr(z \mid c)}{\Pr(z \mid c')} = \frac{e^i(1-e)^{n-i}}{e^{i'}(1-e)^{n-i'}} = \left(\frac{1-e}{e}\right)^{i'-i}$$

For $e < \frac{1}{2}$ it follows $\frac{1-e}{e} > 1$. If $i < i'$ then $\left(\frac{1-e}{e}\right)^{i'-i} > 1$ and thus $\Pr(z \mid c) > \Pr(z \mid c')$. Choosing c over c' with $i < i'$ (i.e. closer) corresponds to increased "likelihood" $\Pr(z \mid c) > \Pr(z \mid c')$. □

Equivalence of Decision Rules

Theorem

For an extended BSC channel Γ^n with bit-error $e < \frac{1}{2}$ the maximal likelihood rule is equivalent to the MD rule.

Proof.

Suppose $c \in C$ with $d(c, z) = i$ then $\Pr(z \mid c) = e^i(1-e)^{n-i}$.
For a second $c' \in C$ s.t $d(c', z) = i' < i$ we have

$$\frac{\Pr(z \mid c)}{\Pr(z \mid c')} = \frac{e^i(1-e)^{n-i}}{e^{i'}(1-e)^{n-i'}} = \left(\frac{1-e}{e}\right)^{i'-i}$$

For $e < \frac{1}{2}$ it follows $\frac{1-e}{e} > 1$. If $i < i'$ then $\left(\frac{1-e}{e}\right)^{i'-i} > 1$ and thus $\Pr(z \mid c) > \Pr(z \mid c')$. Choosing c over c' with $i < i'$ (i.e. closer) corresponds to increased "likelihood" $\Pr(z \mid c) > \Pr(z \mid c')$. □

Example

Take the code $C \subseteq \mathbb{B}^8$ with the following seven codewords:

$$c_1 = 00000000, c_2 = 00111000, c_3 = 11000001, c_4 = 00001110,$$
$$c_5 = 10111011, c_6 = 00110110, c_7 = 11001011$$

Applying the MD rule we classify $z = 10101011$ as follows:

c_i	c_1	c_2	c_3	c_4	c_5	c_6	c_7
$d(z, c_i)$	5	4	4	4	1	5	2

so $\sigma_{MD}(z) = c_5 = 10111011$. For $z' = 11001001$ we have

c_i	c_1	c_2	c_3	c_4	c_5	c_6	c_7
$d(z', c_i)$	4	5	1	5	4	8	1

thus we can take either take $\sigma_{MD}(z') = c_3$ or $\sigma_{MD}(z') = c_7$.

Example

Take the code $C \subseteq \mathbb{B}^8$ with the following seven codewords:

$$c_1 = 00000000, c_2 = 00111000, c_3 = 11000001, c_4 = 00001110,$$
$$c_5 = 10111011, c_6 = 00110110, c_7 = 11001011$$

Applying the MD rule we classify $z = 10101011$ as follows:

c_i	c_1	c_2	c_3	c_4	c_5	c_6	c_7
$d(z, c_i)$	5	4	4	4	1	5	2

so $\sigma_{MD}(z) = c_5 = 10111011$. For $z' = 11001001$ we have

c_i	c_1	c_2	c_3	c_4	c_5	c_6	c_7
$d(z', c_i)$	4	5	1	5	4	8	1

thus we can take either take $\sigma_{MD}(z') = c_3$ or $\sigma_{MD}(z') = c_7$.

Example

Take the code $C \subseteq \mathbb{B}^8$ with the following seven codewords:

$$c_1 = 00000000, c_2 = 00111000, c_3 = 11000001, c_4 = 00001110,$$
$$c_5 = 10111011, c_6 = 00110110, c_7 = 11001011$$

Applying the MD rule we classify $z = 10101011$ as follows:

c_i	c_1	c_2	c_3	c_4	c_5	c_6	c_7
$d(z, c_i)$	5	4	4	4	1	5	2

so $\sigma_{MD}(z) = c_5 = 10111011$. For $z' = 11001001$ we have

c_i	c_1	c_2	c_3	c_4	c_5	c_6	c_7
$d(z', c_i)$	4	5	1	5	4	8	1

thus we can take either take $\sigma_{MD}(z') = c_3$ or $\sigma_{MD}(z') = c_7$.

Minimum Distance

Definition

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code). The minimum distance of C is defined as

$$\delta = \min_{c \neq c'} d(c, c').$$

Example

Consider the code $C \subseteq \mathbb{B}^6$ with codewords:

$$c_1 = 000000, \quad c_2 = 111000, \quad c_3 = 001110, \quad c_4 = 110011.$$

We have the hamming distances:

$$d(c_1, c_2) = 3 \quad d(c_1, c_3) = 3 \quad d(c_1, c_4) = 4$$
$$d(c_2, c_3) = 4 \quad d(c_2, c_4) = 3 \quad d(c_3, c_4) = 5$$

Thus the minimum distance is $\delta = 3$.

Minimum Distance

Definition

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code). The minimum distance of C is defined as

$$\delta = \min_{c \neq c'} d(c, c').$$

Example

Consider the code $C \subseteq \mathbb{B}^6$ with codewords:

$$c_1 = 000000, \quad c_2 = 111000, \quad c_3 = 001110, \quad c_4 = 110011.$$

We have the hamming distances:

$$\begin{array}{lll} d(c_1, c_2) = 3 & d(c_1, c_3) = 3 & d(c_1, c_4) = 4 \\ d(c_2, c_3) = 4 & d(c_2, c_4) = 3 & d(c_3, c_4) = 5 \end{array}$$

Thus the minimum distance is $\delta = 3$.

(Potentially) Error Correcting Code

Lemma

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code) with $\delta \geq 2r + 1$. Then for any two (code)words $c, c' \in C$ we have $N_r(c) \cap N_r(c') = \emptyset$ where $N_r(x) = \{y \in \mathbb{B}^n \mid d(x, y) \leq r\}$

Theorem

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code) with $\delta \geq 2r + 1$ used as input and applying the MD rule.
If less than r bit-errors are made during transmission, then there will be no mistakes.

Definition

A code $C \subseteq \mathbb{B}^n$ is r-error-correcting if $\delta \geq 2r + 1$.

(Potentially) Error Correcting Code

Lemma

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code) with $\delta \geq 2r + 1$. Then for any two (code)words $c, c' \in C$ we have $N_r(c) \cap N_r(c') = \emptyset$ where $N_r(x) = \{y \in \mathbb{B}^n \mid d(x, y) \leq r\}$

Theorem

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code) with $\delta \geq 2r + 1$ used as input and applying the MD rule.
If less than r bit-errors are made during transmission, then there will be no mistakes.

Definition

A code $C \subseteq \mathbb{B}^n$ is r-error-correcting if $\delta \geq 2r + 1$.

(Potentially) Error Correcting Code

Lemma

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code) with $\delta \geq 2r + 1$. Then for any two (code)words $c, c' \in C$ we have $N_r(c) \cap N_r(c') = \emptyset$ where $N_r(x) = \{y \in \mathbb{B}^n \mid d(x, y) \leq r\}$

Theorem

Given a set of binary words $C \subseteq \mathbb{B}^n$ (defined by a code) with $\delta \geq 2r + 1$ used as input and applying the MD rule.
If less than r bit-errors are made during transmission, then there will be no mistakes.

Definition

A code $C \subseteq \mathbb{B}^n$ is r-error-correcting if $\delta \geq 2r + 1$.

Packing Bound

Theorem (Packing Bound)

Given a code $C \subseteq \mathbb{B}^n$ with $\delta \geq 2r+1$ (r-error correcting) then

$$|C|\left(1+n+\binom{n}{2}+\cdots+\binom{n}{r}\right) \leq 2^n$$

Proof.

Given $c \in C$, the words z with $d(c,z) = i$ are created by altering any i bits of n possible ones in c. There are $\binom{n}{i}$ such words and thus $|N_r(c)| = 1+n+\binom{n}{2}+\cdots+\binom{n}{r}$.

We have that $\delta \geq 2r+1$ and thus that all $N_r(c)$ are disjoint for all $|C|$ words $c \in C$ which are less than the 2^n words in $\mathbb{B}^n$. $\square$

Packing Bound

Theorem (Packing Bound)

Given a code $C \subseteq \mathbb{B}^n$ with $\delta \geq 2r+1$ (r-error correcting) then

$$|C|\left(1+n+\binom{n}{2}+\cdots+\binom{n}{r}\right) \leq 2^n$$

Proof.

Given $c \in C$, the words z with $d(c,z) = i$ are created by altering any i bits of n possible ones in c. There are $\binom{n}{i}$ such words and thus $|N_r(c)| = 1+n+\binom{n}{2}+\cdots+\binom{n}{r}$.

We have that $\delta \geq 2r+1$ and thus that all $N_r(c)$ are disjoint for all $|C|$ words $c \in C$ which are less than the 2^n words in $\mathbb{B}^n$. □

Information Rate

Definition

Given a code $C \subseteq \mathbb{B}^n$ its information rate is given by

$$\rho = \frac{\log_2 |C|}{n}$$

Example

Given the code $C \subseteq \mathbb{B}^6$

$$C = \{000000, 111000, 001110, 110011\}$$

then $n = 6$ and $|C| = 4 = 2^2$ or $k = \log(4) = 2$ and therefore $\rho = \frac{k}{n} = \frac{2}{6} = \frac{1}{3}$.
So although C uses 6 bit we could transmit information using only 2 bit, i.e. only a third of the potential is utilised.

Information Rate

Definition

Given a code $C \subseteq \mathbb{B}^n$ its information rate is given by

$$\rho = \frac{\log_2 |C|}{n}$$

Example

Given the code $C \subseteq \mathbb{B}^6$

$$C = \{000000, 111000, 001110, 110011\}$$

then $n = 6$ and $|C| = 4 = 2^2$ or $k = \log(4) = 2$ and therefore $\rho = \frac{k}{n} = \frac{2}{6} = \frac{1}{3}$.
So although C uses 6 bit we could transmit information using only 2 bit, i.e. only a third of the potential is utilised.

Probability of a Mistake

Given any decision rule $\sigma : \mathbb{B}^n \to C$. Define for each $c \in C$:

$$F(c) = F_\sigma(c) = \{z \in \mathbb{B}^n \mid \sigma(z) \neq c\}$$

The mistake probability for $c \in C$ is given by:

$$M_c = M_{c,\sigma} = \sum_{z \in F_\sigma(c)} \Pr(z \mid c).$$

Definition

The probability of a mistake when the encoded stream is given by a source $(C, \mathbf{p})$ is given by:

$$M(C, \mathbf{p}) = M_\sigma(C, \mathbf{p}) = \sum_{c \in C} p_c \cdot M_{c,\sigma}$$

Probability of a Mistake

Given any decision rule $\sigma : \mathbb{B}^n \to C$. Define for each $c \in C$:

$$F(c) = F_\sigma(c) = \{z \in \mathbb{B}^n \mid \sigma(z) \neq c\}$$

The mistake probability for $c \in C$ is given by:

$$M_c = M_{c,\sigma} = \sum_{z \in F_\sigma(c)} \Pr(z \mid c).$$

Definition

The probability of a mistake when the encoded stream is given by a source $(C, \mathbf{p})$ is given by:

$$M(C, \mathbf{p}) = M_\sigma(C, \mathbf{p}) = \sum_{c \in C} p_c \cdot M_{c,\sigma}$$

Probability of a Mistake

Given any decision rule $\sigma : \mathbb{B}^n \to C$. Define for each $c \in C$:

$$F(c) = F_\sigma(c) = \{z \in \mathbb{B}^n \mid \sigma(z) \neq c\}$$

The mistake probability for $c \in C$ is given by:

$$M_c = M_{c,\sigma} = \sum_{z \in F_\sigma(c)} \Pr(z \mid c).$$

Definition

The probability of a mistake when the encoded stream is given by a source $(C, \mathbf{p})$ is given by:

$$M(C, \mathbf{p}) = M_\sigma(C, \mathbf{p}) = \sum_{c \in C} p_c \cdot M_{c,\sigma}$$

Probability of a Mistake

Given any decision rule $\sigma : \mathbb{B}^n \to C$. Define for each $c \in C$:

$$F(c) = F_\sigma(c) = \{z \in \mathbb{B}^n \mid \sigma(z) \neq c\}$$

The mistake probability for $c \in C$ is given by:

$$M_c = M_{c,\sigma} = \sum_{z \in F_\sigma(c)} \Pr(z \mid c).$$

Definition

The probability of a mistake when the encoded stream is given by a source $(C, \mathbf{p})$ is given by:

$$M(C, \mathbf{p}) = M_\sigma(C, \mathbf{p}) = \sum_{c \in C} p_c \cdot M_{c,\sigma}$$

Example

Example

Use one of two possible codes defined as follows:

$$C_1 \text{ via } \begin{cases} \textbf{sell} & \mapsto & 0 \\ \textbf{buy} & \mapsto & 1 \end{cases} \quad \text{or} \quad C_3 \text{ via } \begin{cases} \textbf{sell} & \mapsto & 000 \\ \textbf{buy} & \mapsto & 111 \end{cases}$$

Transmit via (extended) BSC with bit-error e and $\mathbf{p} = (p, 1-p)$.

- For C_1 we have $C_1 = \{0, 1\}$, and the MD rule gives

$$\sigma(0) = 0 \quad \text{and} \quad \sigma(1) = 1.$$

The false sets are $F(0) = \{1\}$ and $F(1) = \{0\}$ and mistake probabilities $M_0 = \Pr(1 \mid 0) = e$ and $M_1 = \Pr(0 \mid 1) = e$.

Hence, $M_{MD}(C_1, \mathbf{p}) = pe + (1-p)e = e$.

Example

Example

Use one of two possible codes defined as follows:

$$C_1 \text{ via } \begin{cases} \textbf{sell} & \mapsto & 0 \\ \textbf{buy} & \mapsto & 1 \end{cases} \quad \text{or} \quad C_3 \text{ via } \begin{cases} \textbf{sell} & \mapsto & 000 \\ \textbf{buy} & \mapsto & 111 \end{cases}$$

Transmit via (extended) BSC with bit-error e and $\mathbf{p} = (p, 1-p)$.

- For C_1 we have $C_1 = \{0, 1\}$, and the MD rule gives

$$\sigma(0) = 0 \quad \text{and} \quad \sigma(1) = 1.$$

The false sets are $F(0) = \{1\}$ and $F(1) = \{0\}$ and mistake probabilities $M_0 = \Pr(1 \mid 0) = e$ and $M_1 = \Pr(0 \mid 1) = e$.

Hence, $M_{MD}(C_1, \mathbf{p}) = pe + (1-p)e = e$.

Example (cont.)

- For $C_3 = \{000, 111\}$ the MD rule $\sigma_{MD} : \mathbb{B}^3 \to C_3$ covers

$$\sigma(000) = \sigma(100) = \sigma(010) = \sigma(001) = 000,$$
$$\sigma(111) = \sigma(011) = \sigma(101) = \sigma(110) = 111.$$

The false/fail sets are

$$\begin{aligned} F(000) &= \{011, 101, 110, 111\} \\ F(111) &= \{100, 010, 001, 000\} \end{aligned}$$

The mistake probabilities can be roughly estimated as $M_{000} < 4e^2$ and $M_{111} < 4e^2$, e.g. $\Pr(011|000) = e^2(1-e)$, etc., thus $M_{MD}(C_3, \mathbf{p}) < p(4e^2) + (1-p)(4e^2) = 4e^2$.

Hence, for $e = 0.001$ we have $M_{MD}(C_1, \mathbf{p}) = 0.001$ whereas $M_{MD}(C_3, \mathbf{p}) < 0.000004$, i.e. C_3 makes less mistakes than C_1.

Example (cont.)

- For $C_3 = \{000, 111\}$ the MD rule $\sigma_{MD} : \mathbb{B}^3 \to C_3$ covers

$$\sigma(000) = \sigma(100) = \sigma(010) = \sigma(001) = 000,$$
$$\sigma(111) = \sigma(011) = \sigma(101) = \sigma(110) = 111.$$

The false/fail sets are

$$\begin{aligned} F(000) &= \{011, 101, 110, 111\} \\ F(111) &= \{100, 010, 001, 000\} \end{aligned}$$

The mistake probabilities can be roughly estimated as $M_{000} < 4e^2$ and $M_{111} < 4e^2$, e.g. $\Pr(011|000) = e^2(1-e)$, etc., thus $M_{MD}(C_3, \mathbf{p}) < p(4e^2) + (1-p)(4e^2) = 4e^2$.

Hence, for $e = 0.001$ we have $M_{MD}(C_1, \mathbf{p}) = 0.001$ whereas $M_{MD}(C_3, \mathbf{p}) < 0.000004$, i.e. C_3 makes less mistakes than C_1.

Error Correction and Distance for BSC

Lemma

If $C \subseteq \mathbb{B}^n$ is an r-error-correcting code and $z \in F(c)$ then

$$d(z, c) \geq r + 1.$$

In general, we have for (extended) BSC and code $C \in \mathbb{B}^n$ that $\Pr(z \mid c) = (\Gamma^n)_{cz} = e^{d(c,z)} \cdot (1 - e)^{n - d(c,z)}$ and thus

$$M_c = \sum_{z \in F(c)} e^{d(c,z)} \cdot (1 - e)^{n - d(c,z)}.$$

For an r-error-correcting code each term for M_c is of the form $e^i(1 - e)^{n-i}$ for some $i > r + 1$. For $0 < e < 1$ we have that $e^i(1 - e)^{n-i} < e^i \leq e^{r+1}$.

If we have an estimate for $|F(c)|$ then roughly

$$M_c \leq \sum_{z \in F(c)} e^{r+1} = |F(c)|e^{r+1}.$$

Error Correction and Distance for BSC

Lemma

If $C \subseteq \mathbb{B}^n$ is an r-error-correcting code and $z \in F(c)$ then

$$d(z, c) \geq r + 1.$$

In general, we have for (extended) BSC and code $C \in \mathbb{B}^n$ that $\Pr(z \mid c) = (\Gamma^n)_{cz} = e^{d(c,z)} \cdot (1 - e)^{n-d(c,z)}$ and thus

$$M_c = \sum_{z \in F(c)} e^{d(c,z)} \cdot (1 - e)^{n-d(c,z)}.$$

For an r-error-correcting code each term for M_c is of the form $e^i(1 - e)^{n-i}$ for some $i > r + 1$. For $0 < e < 1$ we have that $e^i(1 - e)^{n-i} < e^i \leq e^{r+1}$.

If we have an estimate for $|F(c)|$ then roughly

$$M_c \leq \sum_{z \in F(c)} e^{r+1} = |F(c)|e^{r+1}.$$

Error Correction and Distance for BSC

Lemma

If $C \subseteq \mathbb{B}^n$ is an r-error-correcting code and $z \in F(c)$ then

$$d(z, c) \geq r + 1.$$

In general, we have for (extended) BSC and code $C \in \mathbb{B}^n$ that $\Pr(z \mid c) = (\Gamma^n)_{cz} = e^{d(c,z)} \cdot (1 - e)^{n-d(c,z)}$ and thus

$$M_c = \sum_{z \in F(c)} e^{d(c,z)} \cdot (1 - e)^{n-d(c,z)}.$$

For an r-error-correcting code each term for M_c is of the form $e^i(1 - e)^{n-i}$ for some $i > r + 1$. For $0 < e < 1$ we have that $e^i(1 - e)^{n-i} < e^i \leq e^{r+1}$.
If we have an estimate for $|F(c)|$ then roughly

$$M_c \leq \sum_{z \in F(c)} e^{r+1} = |F(c)|e^{r+1}.$$

Examples

Example

Suppose we want an information rate $\rho = 0.8 = \frac{4}{5}$ using an 1-error correcting code in $\mathbb{B}^n$. Thus $r = 1$ and $\delta \geq 2r + 1 = 3$.

Denote by $k = \log_2(|C|)$ then $\rho = \frac{k}{n}$. That is $|C| = 2^k$ and by the Packing Bound we have $2^k(1 + n) \leq 2^n$.

We need $\frac{k}{n} \geq \frac{4}{5}$ or $5k \geq 4n$ as well as $n + 1 \leq 2^{n-k}$. The least solution is for $n = 25$ and $k = 20$ (re-coding $\mathbb{B}^k \to \mathbb{B}^n$).

Example

Let C be a code that assigns to each 3-bit block a 6-bit codeword, i.e. $c : \mathbb{B}^3 \to \mathbb{B}^6$ s.t.

$$y_1 y_2 y_3 \mapsto x_1 x_2 x_3 x_4 x_5 x_6$$

Examples

Example

Suppose we want an information rate $\rho = 0.8 = \frac{4}{5}$ using an 1-error correcting code in $\mathbb{B}^n$. Thus $r = 1$ and $\delta \geq 2r + 1 = 3$.

Denote by $k = \log_2(|C|)$ then $\rho = \frac{k}{n}$. That is $|C| = 2^k$ and by the Packing Bound we have $2^k(1 + n) \leq 2^n$.

We need $\frac{k}{n} \geq \frac{4}{5}$ or $5k \geq 4n$ as well as $n + 1 \leq 2^{n-k}$. The least solution is for $n = 25$ and $k = 20$ (re-coding $\mathbb{B}^k \to \mathbb{B}^n$).

Example

Let C be a code that assigns to each 3-bit block a 6-bit codeword, i.e. $c : \mathbb{B}^3 \to \mathbb{B}^6$ s.t.

$$y_1 y_2 y_3 \mapsto x_1 x_2 x_3 x_4 x_5 x_6$$

Examples (cont.)

Example (cont.)

$$\begin{array}{c|l} x_1 = y_1 & x_4 = 0 \text{ if } y_1 = y_2, \quad x_4 = 1 \text{ otherwise} \\ x_2 = y_2 & x_5 = 0 \text{ if } y_2 = y_3, \quad x_5 = 1 \text{ otherwise} \\ x_3 = y_3 & x_6 = 0 \text{ if } y_3 = y_1, \quad x_6 = 1 \text{ otherwise} \end{array}$$

or explicitely:

$$\begin{array}{llll} 000 \mapsto 000000 & 001 \mapsto 001011 & 010 \mapsto 010110 & 011 \mapsto 011101 \\ 100 \mapsto 100101 & 101 \mapsto 101110 & 110 \mapsto 110011 & 111 \mapsto 111000 \end{array}$$

For this code $\delta = 3$ (check!). Therefore, C is a 1-error correcting code. Since $k = 3$ and $n = 6$ its information rate $\rho = \frac{1}{2}$.

Is it possible to chose C such that the probability of a mistake is arbitarily small while information is transmitted at given rate ρ?

Examples (cont.)

Example (cont.)

$$\begin{array}{r|l} x_1 = y_1 & x_4 = 0 \text{ if } y_1 = y_2, \quad x_4 = 1 \text{ otherwise} \\ x_2 = y_2 & x_5 = 0 \text{ if } y_2 = y_3, \quad x_5 = 1 \text{ otherwise} \\ x_3 = y_3 & x_6 = 0 \text{ if } y_3 = y_1, \quad x_6 = 1 \text{ otherwise} \end{array}$$

or explicitely:

$$\begin{array}{llll} 000 \mapsto 000000 & 001 \mapsto 001011 & 010 \mapsto 010110 & 011 \mapsto 011101 \\ 100 \mapsto 100101 & 101 \mapsto 101110 & 110 \mapsto 110011 & 111 \mapsto 111000 \end{array}$$

For this code $\delta = 3$ (check!). Therefore, C is a 1-error correcting code. Since $k = 3$ and $n = 6$ its information rate $\rho = \frac{1}{2}$.

Is it possible to chose C such that the probability of a mistake is arbitarily small while information is transmitted at given rate ρ?

Transmission with Extended BSC

Instead of single symbols use blocks in channel transmission.

$$\begin{aligned}(\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n} &= \Pr(z_1\cdots z_n \mid c_1\cdots c_n)\\ &= \Pr(z_1 \mid c_1)\cdots\Pr(z_n \mid c_n)\\ &= (\Gamma)_{c_1,z_1}\cdots(\Gamma)_{c_n,z_n}.\end{aligned}$$

Lemma

Let Γ be a BSC with bit-error probability e. Then for Γ^n we have

$$\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = n \cdot \mathbf{h}(e).$$

Proof $*$.

We observe that

$$\mathbf{H}(\mathbf{q} \mid c_1\cdots c_n) = \sum_{z_1\ldots z_n} (\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n} \log\left(\frac{1}{(\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n}}\right)$$

Transmission with Extended BSC

Instead of single symbols use blocks in channel transmission.

$$\begin{aligned}(\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n} &= \Pr(z_1\cdots z_n \mid c_1\cdots c_n)\\ &= \Pr(z_1 \mid c_1)\cdots\Pr(z_n \mid c_n)\\ &= (\Gamma)_{c_1,z_1}\cdots(\Gamma)_{c_n,z_n}.\end{aligned}$$

Lemma

Let Γ be a BSC with bit-error probability e. Then for Γ^n we have

$$\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = n \cdot \mathbf{h}(e).$$

Proof $*$.

We observe that

$$\mathbf{H}(\mathbf{q} \mid c_1\cdots c_n) = \sum_{z_1\ldots z_n} (\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n} \log\left(\frac{1}{(\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n}}\right)$$

Transmission with Extended BSC

Instead of single symbols use blocks in channel transmission.

$$\begin{aligned}(\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n} &= \Pr(z_1\cdots z_n \mid c_1\cdots c_n)\\ &= \Pr(z_1 \mid c_1)\cdots\Pr(z_n \mid c_n)\\ &= (\Gamma)_{c_1,z_1}\cdots(\Gamma)_{c_n,z_n}.\end{aligned}$$

Lemma

Let Γ be a BSC with bit-error probability e. Then for Γ^n we have

$$\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = n \cdot \mathbf{h}(e).$$

Proof $*$.

We observe that

$$\mathbf{H}(\mathbf{q} \mid c_1\cdots c_n) = \sum_{z_1\ldots z_n} (\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n} \log\left(\frac{1}{(\Gamma^n)_{c_1\cdots c_n, z_1\cdots z_n}}\right)$$

Proof (cont).

$$
\begin{aligned}
\mathbf{H}(\mathbf{q} \mid c_1 \cdots c_n) = & \\
&= \sum_{z_1 \ldots z_n} (\Gamma^n)_{c_1 \cdots c_n, z_1 \cdots z_n} \log \left(\frac{1}{(\Gamma^n)_{c_1 \cdots c_n, z_1 \cdots z_n}} \right) \\
&= \sum_{z_1 \ldots z_n} (\Gamma)_{c_1, z_1} \cdots (\Gamma)_{c_n, z_n} \log \left(\frac{1}{(\Gamma)_{c_1, z_1} \cdots (\Gamma)_{c_n, z_n}} \right) \\
&= \sum_{z_1} \cdots \sum_{z_n} (\Gamma)_{c_1, z_1} \cdots (\Gamma)_{c_n, z_n} \sum_{i=1}^{n} \log \left(\frac{1}{(\Gamma)_{c_i, z_i}} \right) \\
&= \sum_{i=1}^{n} \sum_{z_i} (\Gamma)_{c_i, z_i} \log \left(\frac{1}{(\Gamma)_{c_i, z_i}} \right)
\end{aligned}
$$

because $\sum_z \Gamma_{c,z} = 1$ for each c.

Proof (cont.)

Proof (cont).

The values z_i are eitrher $z_i = 0$ or $z_i = 1$ so each sum is just:

$$\Gamma_{c,0} \log\left(\frac{1}{\Gamma_{c,0}}\right) + \Gamma_{c,1} \log\left(\frac{1}{\Gamma_{c,1}}\right)$$

As $\Gamma_{c,0} = e$ or $\Gamma_{c,0} = (1 - e)$, and the same for $\Gamma_{c,1}$:

$$\Gamma_{c,0} \log\left(\frac{1}{\Gamma_{c,0}}\right) + \Gamma_{c,1} \log\left(\frac{1}{\Gamma_{c,1}}\right) = \mathbf{h}(e)$$

Thus, for all $c_1, \ldots, c_n$ we have $\mathbf{H}(\mathbf{q} \mid c_1 \cdots c_n) = n \cdot \mathbf{h}(e)$ and so

$$\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = \sum_{c_1 \ldots c_n} \mathbf{p}(c_1 \ldots c_n) \cdot \mathbf{H}(\mathbf{q} \mid c_1 \cdots c_n) = n \cdot \mathbf{h}(e).$$

Capacity of Extended BSC

Theorem

If the capacity of the BSC Γ *is* $\gamma(\Gamma) = 1 - \mathbf{h}(e)$ *then the capacity of the extended BSC* Γ^n *is* $\gamma(\Gamma^n) = n \cdot \gamma(\Gamma) = n \cdot \gamma$.

Proof.

Capacity is the maximum value for

$$\mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{q} \mid \mathbf{p})$$

From before we have $\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = n \cdot \mathbf{h}(e)$ so we need the maximum of $\mathbf{H}(\mathbf{q})$. We know that $\mathbf{H}(\mathbf{q}) \leq \log(2^n) = n$ and that the maximum is attained iff we have a uniform distribution.

Hence (for the maximum, i.e. uniform distribution)

$$\max \mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{q} \mid \mathbf{p}) = n - n \cdot \mathbf{h}(e) = n \cdot \gamma.$$

Capacity of Extended BSC

Theorem

If the capacity of the BSC Γ *is* $\gamma(\Gamma) = 1 - \mathbf{h}(e)$ *then the capacity of the extended BSC* Γ^n *is* $\gamma(\Gamma^n) = n \cdot \gamma(\Gamma) = n \cdot \gamma$.

Proof.

Capacity is the maximum value for

$$\mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{q} \mid \mathbf{p})$$

From before we have $\mathbf{H}(\mathbf{q} \mid \mathbf{p}) = n \cdot \mathbf{h}(e)$ so we need the maximum of $\mathbf{H}(\mathbf{q})$. We know that $\mathbf{H}(\mathbf{q}) \leq \log(2^n) = n$ and that the maximum is attained iff we have a uniform distribution.

Hence (for the maximum, i.e. uniform distribution)

$$\max \mathbf{H}(\mathbf{q}) - \mathbf{H}(\mathbf{q} \mid \mathbf{p}) = n - n \cdot \mathbf{h}(e) = n \cdot \gamma.$$

Rate vs Capacity

Theorem

Given a code $C \subseteq \mathbb{B}^n$ with information rate ρ and let $\mathbf{p}^*$ be the uniform probability distribution on C. Consider source $(C, \mathbf{p}^*)$ trough an extended BSC Γ^n with capacity $\gamma(\Gamma) = \gamma$. Then

$$\mathbf{H}(\Gamma^n; \mathbf{p}^*) \geq n \cdot (\rho - \gamma).$$

Proof.

The capacity of Γ^n is the maximum for $\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma^n; \mathbf{p})$.
With $\mathbf{p} = \mathbf{p}^*$ we have $\mathbf{H}(\mathbf{p}^*) - \mathbf{H}(\Gamma^n; \mathbf{p}^*) \leq n \cdot \gamma$.

As we have $\mathbf{H}(\mathbf{p}^*) = \log(|C|)$ we get $\mathbf{H}(\Gamma^n; \mathbf{p}^*) \geq \log(|C|) - n \cdot \gamma$.

Furthermore the rate is $\rho = \frac{\log(|C|)}{n}$. □

Rate vs Capacity

Theorem

Given a code $C \subseteq \mathbb{B}^n$ with information rate ρ and let $\mathbf{p}^*$ be the uniform probability distribution on C. Consider source $(C, \mathbf{p}^*)$ trough an extended BSC Γ^n with capacity $\gamma(\Gamma) = \gamma$. Then

$$\mathbf{H}(\Gamma^n; \mathbf{p}^*) \geq n \cdot (\rho - \gamma).$$

Proof.

The capacity of Γ^n is the maximum for $\mathbf{H}(\mathbf{p}) - \mathbf{H}(\Gamma^n; \mathbf{p})$.
With $\mathbf{p} = \mathbf{p}^*$ we have $\mathbf{H}(\mathbf{p}^*) - \mathbf{H}(\Gamma^n; \mathbf{p}^*) \leq n \cdot \gamma$.

As we have $\mathbf{H}(\mathbf{p}^*) = \log(|C|)$ we get $\mathbf{H}(\Gamma^n; \mathbf{p}^*) \geq \log(|C|) - n \cdot \gamma$.

Furthermore the rate is $\rho = \frac{\log(|C|)}{n}$. □

Fano's Inequality

Theorem (Fano's Inequality)

Given a code $C \subseteq \mathbb{B}^n$ and $M = M(C, \mathbf{p})$ be the probability of a mistake for the source $(C, \mathbf{p})$ being transmitted through the extended BSC Γ^n using the MD rule. Then

$$\mathbf{H}(\Gamma^n; \mathbf{p}) \leq \mathbf{h}(M) + M \cdot \log(|C| - 1)$$

For a proof see Biggs, Section 7.6.

Fano's Inequality

Theorem (Fano's Inequality)

Given a code $C \subseteq \mathbb{B}^n$ and $M = M(C, \mathbf{p})$ be the probability of a mistake for the source $(C, \mathbf{p})$ being transmitted through the extended BSC Γ^n using the MD rule. Then

$$\mathbf{H}(\Gamma^n; \mathbf{p}) \leq \mathbf{h}(M) + M \cdot \log(|C| - 1)$$

For a proof see Biggs, Section 7.6.

Rate Exceeding Capacity

Theorem

Suppose, for an infinite sequence $n = n_1, n_2 \ldots$, we have construct codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$. Let $\mathbf{p}^*$ be the uniform distribution on C_n. If $\rho > \gamma$ then

$$\lim_{n\to\infty} M(C_n, \mathbf{p}^*) \neq 0.$$

Proof $*$.

Let $|C_n| = 2^{k_n}$ with $k_n \geq \rho n$ and $M_n = M(C_n, \mathbf{p}^*)$. Fano's Inequality states

$$\mathbf{H}(\Gamma^n; \mathbf{p}) \leq \mathbf{h}(M_n) + M_n \log(|C_n| + 1).$$

Since $\mathbf{h}(M_n) \leq 1$ and $\log(|C_n| - 1) < \log(|C_n|) = k_n$ we have for any $\mathbf{p}$ (also $\mathbf{p}^*$) that $\mathbf{H}(\Gamma^n; \mathbf{p}) < 1 + M_n k_n$.

Rate Exceeding Capacity

Theorem

Suppose, for an infinite sequence $n = n_1, n_2 \ldots$, we have construct codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$. Let $\mathbf{p}^*$ be the uniform distribution on C_n. If $\rho > \gamma$ then

$$\lim_{n \to \infty} M(C_n, \mathbf{p}^*) \neq 0.$$

Proof $*$.

Let $|C_n| = 2^{k_n}$ with $k_n \geq \rho n$ and $M_n = M(C_n, \mathbf{p}^*)$. Fano's Inequality states

$$\mathbf{H}(\Gamma^n; \mathbf{p}) \leq \mathbf{h}(M_n) + M_n \log(|C_n| + 1).$$

Since $\mathbf{h}(M_n) \leq 1$ and $\log(|C_n| - 1) < \log(|C_n|) = k_n$ we have for any $\mathbf{p}$ (also $\mathbf{p}^*$) that $\mathbf{H}(\Gamma^n; \mathbf{p}) < 1 + M_n k_n$.

Proof (cont.)

Proof (cont).

On the other hand $\mathbf{H}(\Gamma^n; \mathbf{p}) \geq \log(|C|) - n \cdot \gamma = k_n - n \cdot \gamma$.

Combining these inequalities leads us to the constraint:

$$1 + M_n k_n > k_n - n \cdot \gamma$$

and so

$$M_n > 1 - \frac{n \cdot \gamma + 1}{k_n} \geq 1 - \frac{n \cdot \gamma + 1}{n \cdot \rho}$$

For $n \to \infty$ we have $\lim(1 - \frac{n \cdot \gamma + 1}{n \cdot \rho}) = 1 - \frac{\gamma}{\rho}$ which for $\rho > \gamma$ is strictly positive. Thus $\lim M_n > 0$ (if it exists). □

Shannon's Theorem

Theorem (Shannon, 1948)

For $\rho < \gamma$ it is possible to construct a sequence of codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$ and

$$\lim_{n\to\infty} M(C_n, \mathbf{p}) = 0.$$

- Consider γ of given channel and choose $\rho < \gamma$.
- Choose $C_n \subseteq \mathbb{B}^n$ with $|C_n| = 2^k$ and $M(C_n, \mathbf{p})$ small.
- Encode k blocks of original stream using C_n.
- Transmit and apply MD rule.

Problem: Shannon does not provide a practical method for constructing the required code(s) C_n.

Shannon's Theorem

Theorem (Shannon, 1948)

For $\rho < \gamma$ it is possible to construct a sequence of codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$ and

$$\lim_{n\to\infty} M(C_n, \mathbf{p}) = 0.$$

- Consider γ of given channel and choose $\rho < \gamma$.
- Choose $C_n \subseteq \mathbb{B}^n$ with $|C_n| = 2^k$ and $M(C_n, \mathbf{p})$ small.
- Encode k blocks of original stream using C_n.
- Transmit and apply MD rule.

Problem: Shannon does not provide a practical method for constructing the required code(s) C_n.

Shannon's Theorem

Theorem (Shannon, 1948)

For $\rho < \gamma$ it is possible to construct a sequence of codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$ and

$$\lim_{n\to\infty} M(C_n, \mathbf{p}) = 0.$$

- Consider γ of given channel and choose $\rho < \gamma$.
- Choose $C_n \subseteq \mathbb{B}^n$ with $|C_n| = 2^k$ and $M(C_n, \mathbf{p})$ small.
- Encode k blocks of original stream using C_n.
- Transmit and apply MD rule.

Problem: Shannon does not provide a practical method for constructing the required code(s) C_n.

Shannon's Theorem

Theorem (Shannon, 1948)

For $\rho < \gamma$ it is possible to construct a sequence of codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$ and

$$\lim_{n\to\infty} M(C_n, \mathbf{p}) = 0.$$

- Consider γ of given channel and choose $\rho < \gamma$.
- Choose $C_n \subseteq \mathbb{B}^n$ with $|C_n| = 2^k$ and $M(C_n, \mathbf{p})$ small.
- Encode k blocks of original stream using C_n.
- Transmit and apply MD rule.

Problem: Shannon does not provide a practical method for constructing the required code(s) C_n.

Shannon's Theorem

Theorem (Shannon, 1948)

For $\rho < \gamma$ it is possible to construct a sequence of codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$ and

$$\lim_{n\to\infty} M(C_n, \mathbf{p}) = 0.$$

- Consider γ of given channel and choose $\rho < \gamma$.
- Choose $C_n \subseteq \mathbb{B}^n$ with $|C_n| = 2^k$ and $M(C_n, \mathbf{p})$ small.
- Encode k blocks of original stream using C_n.
- Transmit and apply MD rule.

Problem: Shannon does not provide a practical method for constructing the required code(s) C_n.

Shannon's Theorem

Theorem (Shannon, 1948)

For $\rho < \gamma$ it is possible to construct a sequence of codes $C_n \subseteq \mathbb{B}^n$ such that $|C_n| \geq 2^{\rho n}$ and

$$\lim_{n\to\infty} M(C_n, \mathbf{p}) = 0.$$

- Consider γ of given channel and choose $\rho < \gamma$.
- Choose $C_n \subseteq \mathbb{B}^n$ with $|C_n| = 2^k$ and $M(C_n, \mathbf{p})$ small.
- Encode k blocks of original stream using C_n.
- Transmit and apply MD rule.

Problem: Shannon does not provide a practical method for constructing the required code(s) C_n.