### Transmission Channels

**Outline**

1. Transmission Channels
2. Decoding
3. Shannon Theorem

---

**Codes and Cryptography**

Jorge L. Villar

MAMME, Fall 2015

**PART IV**

---

#### Transmission Channels

**r-ary perfect channel:** no transmission errors

\[
\begin{align*}
0 & \quad 1 \\
1 & \quad 1 \\
\vdots & \quad \vdots \\
r-1 & \quad 1 \\
\end{align*}
\]

\[
\begin{align*}
0 & \quad 0 \\
1 & \quad 1 \\
\vdots & \quad \vdots \\
r-1 & \quad r-1 \\
\end{align*}
\]

---

**r-ary symmetric channel**

\[
\beta = \frac{\alpha}{r-1}
\]

---

**Trans. Line**

**Source** – **Dest.**
Transmission Channels: The Model

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

Transmission probability matrix:

\[ T = \begin{pmatrix} t_{0,0} & \cdots & t_{0,r-1} \\ \vdots & \ddots & \vdots \\ t_{r-1,0} & \cdots & t_{r-1,r-1} \end{pmatrix} \]

where \( t_{i,j} = \Pr(\text{receive } j \mid \text{send } i) \)

Properties:

- \( T \geq 0 \)
- \( T \sum_{i} = 1 \)
- If \( p_i = \Pr(\text{send } j) \) and \( q_i = \Pr(\text{receive } i) \),
  \( (q_0 \ldots q_{r-1}) = (p_0 \ldots p_{r-1}) T \)

Channel Capacity (I)

\[ L(C_X, Y, S_X, Y) = L(C_Y, S_Y) + \sum_{y \in Y} \Pr(Y = y) L(C_{X|Y=y}, S_{X|Y=y}) \]

If all (binary) codes are optimal then

\[ H(Y) \leq L(C_Y, S_Y) < H(Y) + 1 \]

\( H(X \mid Y = y) \leq L(C_{X|Y=y}, S_{X|Y=y}) < H(X \mid Y = y) + 1 \)

Thus,

\[ H(X, Y) = H(Y) + H(X \mid Y) \leq L(C_X, Y, S_X, Y) < H(X, Y) + 2 \]

Channel Capacity (II)

With extended sources,

\[ L(C_X, Y, S_X^k, Y) = L(C_Y, S_Y^k) + \sum_{y \in Y^k} \Pr(Y = y) L(C_{X|Y=y}, S_{X|Y=y}^k) \]

Now,

\[ H(X, Y) = H(Y) + H(X \mid Y) = kH(Y) + kH(X \mid Y) = kH(X, Y) \leq L(C_X, Y, S_X^k, Y) < H(X, Y) + 2 = kH(X, Y) + 2 \]

Therefore, asymptotically \( H(X, Y) = \frac{1}{k} L(C_X, Y, S_X^k, Y) \)

Channel Capacity (III)

Regarding entropies,

\[ H(X, Y) = H(Y) + H(X \mid Y) \]

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)

\[ \Lambda = \max_{S_X} \{I(X; Y)\} \]

where the maximum is taken over all possible sources \( S_X \)
Capacity of the Binary Symmetric Channel

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

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

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

1. Transmission Channels
2. Decoding
3. Shannon Theorem

Decision Rules

Decision rule $\Delta$: 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)

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:
\[
Y = 0 \quad \rightarrow \quad \text{if } 1 - \alpha \geq \alpha \text{ then } \hat{X} = 0 \quad (\alpha \leq \frac{1}{2}) \\
Y = 1 \quad \rightarrow \quad \text{if } 1 - \alpha \geq \alpha \text{ then } \hat{X} = 1 \quad (\alpha \leq \frac{1}{2})
\]

\[ p_{\text{corr}} = \Pr(\hat{X} = X) = \max(\alpha, 1 - \alpha) \]

Ideal Observer:
\[
Y = 0 \quad \rightarrow \quad \text{if } (1 - \alpha)(1 - p) \geq \alpha p \text{ then } \hat{X} = 0 \quad (\alpha \leq 1 - p) \\
Y = 1 \quad \rightarrow \quad \text{if } (1 - \alpha)p \geq \alpha(1 - p) \text{ then } \hat{X} = 1 \quad (\alpha \leq p)
\]

\[ p_{\text{corr}} = \max(\alpha, p, 1 - p, 1 - \alpha) \]

Both differ only when the channel is useless!

Improving Channel Reliability (I)

**GOAL:** Make \( p_{\text{corr}} \) 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 \( \Lambda_{\text{BSC}} = 1 - h_\alpha \).

Therefore, we need to expand \( X \) to a length of at least \( \frac{1}{\Lambda_{\text{BSC}}} \) to make \( p_{\text{corr}} \rightarrow 1 \).

BSC Replication (or Repetition Code)

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

Max. Likelihood Decoding Rule: (Assume \( \alpha < \frac{1}{2} \))
\[ Y \quad \rightarrow \quad \text{decide } \hat{X} \text{ by majority vote among } Y_1, \ldots, Y_n \]

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.

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

For \( \alpha = 0.1 \), \( \Lambda_{\text{BSC}} = 0.531 \) but for \( p_{\text{corr}} \geq 0.999 \) we need \( n = 9 \)
Transmission Channels

Decoding

Shannon Theorem

\textbf{“Encode then Transmit”}

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

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

\[ Y \xrightarrow{\text{decide}} \hat{X} \text{ by the codeword in } C_k \text{ nearest to } Y \]

\[ p_{\text{corr}} = \Pr(\hat{X} = X) = \sum_{x \in \{0,1\}^k} \Pr(\hat{X} = x \mid X = x) \Pr(X = x) \]

\[ \Pr(\hat{X} = x \mid X = x) = \Pr(Y \in \text{neigh}(E_{C_k}(x)) \mid X = x) \text{ where} \]

\[ \text{neigh}(w) = \{y \in \{0,1\}^n : \forall z \in C_k \setminus \{w\}, d(y, z) > d(y, w)\} \]

\[ \Pr(\hat{X} = x \mid X = x) = \sum_{y \in \text{neigh}(E_{C_k}(x))} \alpha^{d(y, E_{C_k}(x))} (1 - \alpha)^{n - d(y, E_{C_k}(x))} \]

\textbf{Outline}

1. Transmission Channels
2. Decoding
3. Shannon Theorem

\textbf{The Rate of a Constant-Length Code}

\textbf{Definition (Code Rate)}

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

\[ R(C) = \frac{\log_r(|C|)}{n} \]

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

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

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

\textbf{Shannon’s Theorem}

\textbf{Theorem (Shannon)}

\textit{Given an $r$-ary channel (stateless and source independent) with capacity $\Lambda$, 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 $C$ of constant-length $n$, and a decision rule for it such that}

\[ \Lambda - \varepsilon < R(C) \log_2 r < \Lambda \quad \text{and} \quad p_{\text{corr}} > 1 - \delta \]

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 $C$ of constant-length $n$ and rate $R$ (i.e., $|C| = 2^{nR}$). 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 $p_{\text{corr}}$ is

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

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

---

**Sketch of the proof for the BSC (II)**

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

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

and

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

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

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

---

**Sketch of the proof for the BSC (III)**

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

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

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

---

**Proof of the Lemma**

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

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

Thus,$ \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

$$\sum_{i=0}^{d} \binom{n}{i} \leq \left(\alpha^\alpha (1 - \alpha)^{1-\alpha}\right)^n = 2^{n h_\alpha}$$
### Fano Bound (I)

**Theorem (Fano Bound)**

For any channel, r-ary source and decision rule

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

Fano bound is equivalent to

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

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

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

or \(H(X \mid Y) = \sum_{(i,j) \in \Delta} p_{i,t,j} \log_2 \frac{q_j}{p_{i,t,j}} + \sum_{(i,j) \not\in \Delta} p_{i,t,j} \log_2 \frac{q_j}{p_{i,t,j}}\)

---

### Converse of Shannon’s Theorem

**Theorem**

Given an r-ary channel (stateless and source independent) with capacity \(\Lambda\) and a uniform source, for any \(\Lambda_1 > \Lambda_2 > \Lambda\) there exists no sequence of codes \(\{C_n\}_{n \geq 2}\) of constant length \(n\) and rates \(R_n\) such that \(\Lambda_1 > R_n \log_2 r > \Lambda_2\) and \(p_{\text{corr}} \rightarrow 1\) as \(n \rightarrow \infty\).

For \(X\) uniformly distributed in \(C_n\), using Fano bound

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

Thus, \(p_{\text{corr}} \leq \frac{1 + n\Lambda}{nR_n \log_2 r} \leq \frac{1 + n\Lambda}{n\Lambda_2}\)

Finally, for \(n > \frac{2}{\Lambda_2 - \Lambda}\), \(p_{\text{corr}}\) cannot exceed \(1 - \frac{\Lambda_2 - \Lambda}{2\Lambda_2}\).

---

### Fano Bound (II)

\[ p_{\text{corr}} = \sum_{(i,j) \in \Delta} \Lambda_j p_{i,t,j} \quad H(X \mid Y) = \sum_{(i,j) \in \Delta} \Lambda_j p_{i,t,j} \log_2 \Lambda_j + \sum_{(i,j) \not\in \Delta} \Lambda_j p_{i,t,j} \log_2 (1 - p_{\text{corr}}) \log_2 \frac{1 - p_{\text{corr}}}{r - 1} = \]

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

\[ Q \ln 2 \leq p_{\text{corr}} \sum_{(i,j) \in \Delta} q_j - \sum_{(i,j) \not\in \Delta} p_{i,t,j} + \sum_{(i,j) \not\in \Delta} q_j - \sum_{(i,j) \not\in \Delta} p_{i,t,j} \]

But \(\sum_{(i,j) \in \Delta} q_j = 1\) and \(\sum_{(i,j) \not\in \Delta} q_j = r - 1\) both imply \(Q \ln 2 \leq 0\)

---

### Remarks

A transmission channel has a maximum transmission rate, the channel capacity \(\Lambda\).

Large codes with \(p_{\text{corr}} \approx 1\) 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.

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 \(p_{\text{corr}}\) and rate.
END OF PART IV