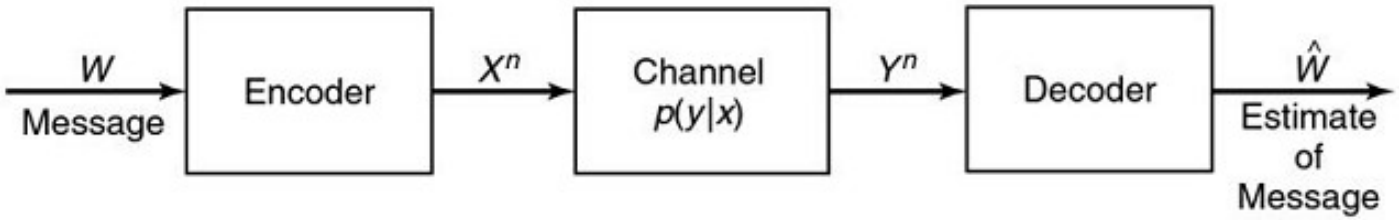
Lecture 9 Communication over a Noisy Channel

Textbook 7.1-7.4

November 26, 2024

Outline

A discrete memoryless channel Q is characterized by an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$, and a set of conditional probability distributions $p(y|x)$, one for each $x \in \mathcal{X}$.



Definition

Let $\mathcal{X}$ and $\mathcal{Y}$ be discrete alphabets, and $p(y|x)$ be a transition matrix from $\mathcal{X}$ to $\mathcal{Y}$. A **discrete channel** $p(y|x)$ is a single inout-single output system with input random variable X taking values in $\mathcal{X}$ and output random variable Y taking values in $\mathcal{Y}$ such that

$$\Pr\{X = x, Y = y\} = \Pr\{X = x\}p(y|x)$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

- We now present an alternative description of a discrete channel. Let $\mathcal{X}$ and $\mathcal{Y}$ be discrete alphabets.
- Let X be a random variable taking values in $\mathcal{X}$ and $p(y|x)$ be any transition matrix from $\mathcal{X}$ to $\mathcal{Y}$.
- Define random variables Z_x with $\mathcal{Z}_x = \mathcal{Y}$ for $x \in \mathcal{X}$ such that

$$\Pr\{\mathcal{Z}_x = y\} = p(y|x)$$

for all $y \in \mathcal{Y}$.

- We assume that Z_x, $x \in \mathcal{X}$ are mutually independent and also independent variable taking values in $\mathcal{Y}$ as

$$Y = Z_x \text{ if } X = x.$$

- Evidently, Y is a function of X and Z.
- Then for $x \in \mathcal{X}$ such that $\Pr\{X = 0\} > 0$, we have

$$\begin{aligned}\Pr\{X = x, Y = y\} &= \Pr\{X = x\}\Pr\{Y = y|X = x\} \\ &= \Pr\{X = x\}\Pr\{Z_x = y|X = x\} \\ &= \Pr\{x = x\}\Pr\{Z_x = y\} \\ &= \Pr\{X = x\}p(y|x).\end{aligned}$$

- For $x \in \mathcal{X}$ such that $\Pr\{X = x\} = 0$, since $\Pr\{X = x\} = 0$ implies $\Pr\{X = x, Y = y\} = 0$.
- Then by regarding X and Y as the input and output random variables, we have obtained an alternative description of the discrete channel $p(y|x)$.

Since Y is a function of X and Z, we can write

$$Y = \alpha(X, Z).$$

Then we have the following equivalent definition of a discrete channel.

Definition

Let $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ be discrete alphabets. Let $\alpha : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$, and Z be a random variable taking values in $\mathcal{Z}$, called the noise variable. A **discrete channel** (α, Z) is a single input-single output system with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. For any input random variable X, the noise variable Z is independent of X, and the output random variable Y is given by

$$Y = \alpha(X, Z).$$

Definition

Two discrete channels $p(y|x)$ and (α, Z) defined on the same input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ are equivalent if

$$\Pr\{\alpha(x, Z) = y\} = p(y|x)$$

for all x and y.

Definition

We define the "information" channel capacity of a discrete memoryless channel as

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

where the maximum is taken over all possible input distributions $p(x)$.

- We shall soon give an operational definition of channel capacity as the highest rate in bits per channel use at which information can be sent with arbitrarily low probability of error.
- Shannon's second theorem establishes that the information channel capacity is equal to the operational channel capacity.
- Thus we drop the word information in most discussions of channel capacity.

- There is a duality between the problems of data compression and data transmission.
- During compression, we remove all the redundancy in the data to form the most compressed version possible, whereas during data transmission, we add redundancy in a controlled fashion to combat errors in the channel.
- Later we show that a general communication system can be broken into two parts and that the problems of data compression and data transmission can be considered separately.

The **capacity** C of a discrete memoryless channel (DMC) can be written in the form

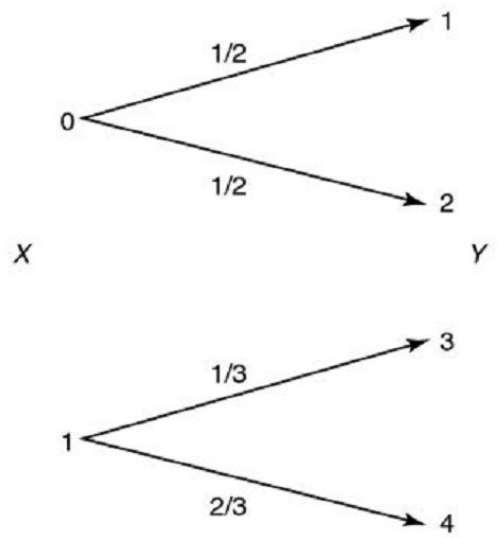
$$C := \max_{Q(0),\cdots,Q(K-1)} \sum_{k,i} Q(k)P(j/k) \log \frac{P(j/k)}{\sum Q(i)P(j/i)}.$$

Outline

Noiseless binary channel

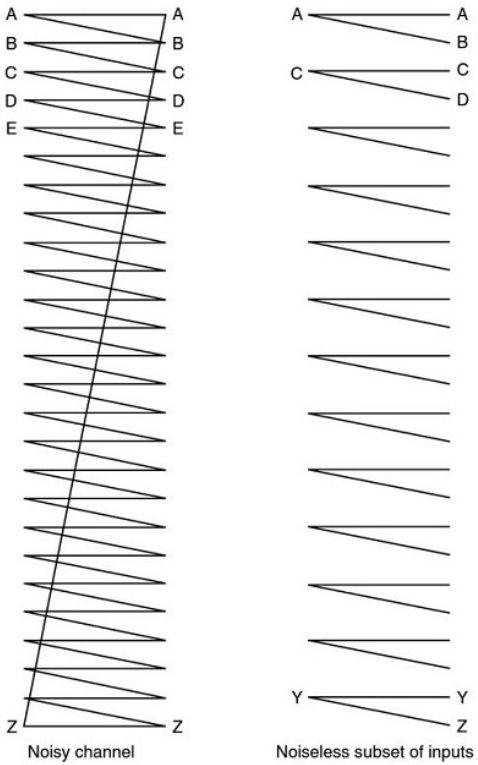
In this case, any transmitted bit is received without error. Hence, one error-free bit can be transmitted per use of the channel, and the capacity is 1 bit. We can also calculate the information capacity $C = \max I(X;Y) = 1$bit, which is achieved by using $p(x) = (\frac{1}{2}, \frac{1}{2})$.

Noisy channel with nonoverlapping outputs



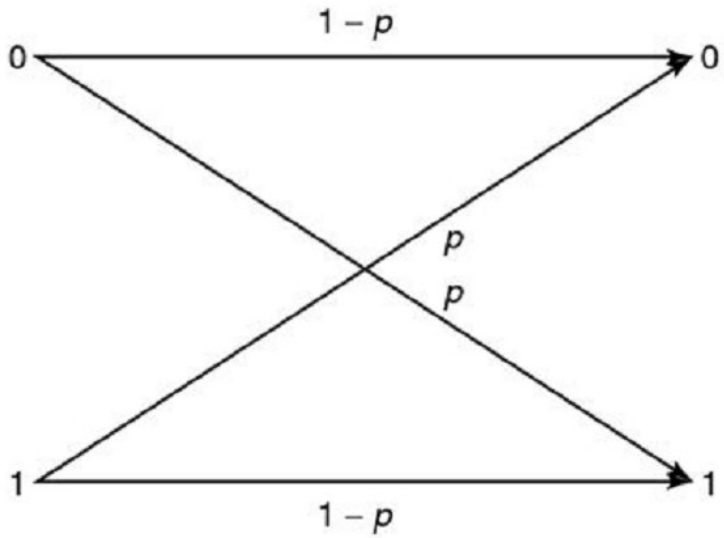
Even though the output of the channel is a random consequence of the input, the input can be determined from the output, and hence every transmitted bit can be recovered without error. The capacity of this channel is also 1 bit per transmission. We can also calculate the information capacity $C = \max I(X;Y) = 1$bits, which is achieved by using $p(x) = (\frac{1}{2}, \frac{1}{2})$.

Noisy typewritter



The input has 26 symbols and we use every alternate input symbol, we can transmit one of 13 symbols without error with each transmission. Hence, the capacity of this channel is $\log 13$ bits per transmission. One can also calculate the information capacity $C = \max I(X;Y) = \max(H(Y) - H(Y|X))$ $= \max H(Y) - 1 = \log 26 - 1 = \log 13$, achieved by using $p(x)$ distributed uniformly over all input.

Binary symmetric channel



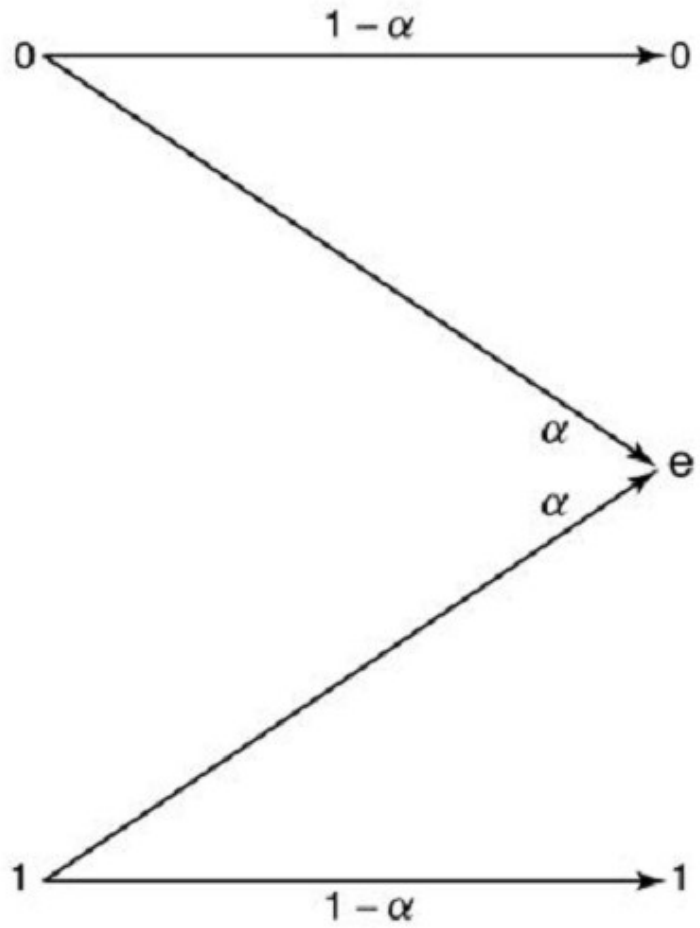
We bound the mutual information by

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum p(x) H(Y|X=x) \\ &= H(Y) - \sum p(x) H(p) \\ &= H(Y) - H(p) \\ &\leq 1 - H(p). \end{aligned}$$

Equality is achieved when the input distribution is uniform. Hence, the information capacity of a binary symmetric channel with parameter p is

$$C = 1 - H(p) \text{ bits.}$$

Binary erasure channel



We calculate the capacity of the binary erasure channel as follows:

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} (H(Y) - H(Y|X)) \\ &= \max_{p(x)} (H(Y) - H(\alpha)). \end{aligned}$$

Let E be the event $\{Y = e\}$, using the expansion

$$H(Y) = H(Y, E) = H(E) + H(Y|E),$$

and letting $\Pr(X = 1) = \pi$, we have

$$H(Y) = H((1-\pi)(1-\alpha), \alpha, \pi(1-\alpha)) = H(\alpha) + (1-\alpha)H(\pi).$$

Hence

$$C = \max_{p(x)} (H(Y) - H(\alpha)) = \max_{\pi} (1-\alpha)H(\pi) = 1 - \alpha.$$

where capacity is achieved by $\pi = \frac{1}{2}$.

- In many practical channels, the sender receives some feedback from the receiver.
- If feedback is available for the binary erasure channel, it is very clear what to do.
- If a bit is lost, retransmit it until it gets through.
- Since the bits get through with probability $1 - \alpha$, the effective rate of transmission is $1 - \alpha$.
- In this way we are easily able to achieve a capacity of $1 - \alpha$ with feedback.
- Later we prove that the rate $1 - \alpha$ is the best that can be achieved both with and without feedback.
- This is one of consequences of the surprising fact that feedback does not increase the capacity of discrete memoryless channels.

Outline

Definition

A channel is said to be symmetric if the rows of the channel transition matrix $p(y|x)$ are permutations of each other and the columns are permutations of each other. A channel is said to be weak symmetric if the every row of the channel transition matrix $p(y|x)$ is a permutation of each other, and all column sums $\sum_x p(y|x)$ are equal.

Example

The channel with transition matrix

$$p(y|x) = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}$$

is weakly symmetric but not symmetric.

Letting **r** be a row of the transition matrix, we have

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(\mathbf{r}) \\ &\leq \log|\mathcal{Y}| - H(\mathbf{r}) \end{aligned}$$

with equality if the output distribution is uniform. But $p(x) = \frac{1}{|\mathcal{X}|}$ achieves a uniform distribution on Y, as seen from

$$p(y) = \sum_{x\in\mathcal{X}} p(y|x)p(x) = \frac{1}{\mathcal{X}} \sum_{x\in\mathcal{X}} p(y|x) = c\frac{1}{|\mathcal{X}|} = \frac{1}{|\mathcal{Y}|},$$

where c is the sum of the entries in one column of the probability transition matrix.

Theorem

For a weakly symmetric channel,

$$C = \log |\mathcal{Y}| - H(\textit{row of transition matrix})$$

Outline

1. $C \geq 0$ since $I(X;Y) \geq 0$.
2. $C \leq \log |\mathcal{X}|$ since $C = \max I(X;Y) \leq \max h(X) = \log |\mathcal{X}|$.
3. $C \leq \log |\mathcal{Y}|$ for the same reason.
4. $I(X;Y)$ is a continuous function of $p(x)$.
5. $I(X;Y)$ is a concave function of $p(x)$. Since $I(X;Y)$ is a concave function over a closed convex set, a local maximum is global maximum. From Properties 2 and 3, the maximum is finite, and we are justified in using the term maximum rather than supremum in the definition of capacity.

Definition

A discrete channel, denoted by $(\mathcal{X}, p(y|x), \mathcal{Y})$, consists of two finite sets $\mathcal{X}$ and $\mathcal{Y}$ and a collection of probability mass functions $p(y|x)$, one for each $x \in \mathcal{X}$, such that for every x and y, $p(y|x) \geq 0$, and for every x, $\sum_x p(y|x) = 1$, with the interpretation that X is the input and Y is the output of the channel.

Definition

The nth extension of the discrete memoryless channel (DMC) is the channel $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$, where

$$p(y_k|x^k, y^{k-1}) = p(y_k|x_k), \ k = 1, 2, \ldots, n.$$

Remark

If the channel is used without feedback [i.e., if the input symbols do not depend on the past output symbols, namely, $p(x_k|x^{k-1}, y^{k-1}) = p(x_k|x^{k-1})$*], the channel transition function for the* n*th extension of the discrete memoryless channel reduces to*

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

To see this, note that

$$p(y^n|x^n) = p(y_n|y^{n-1}, x^n)p(y^{n-1}, x^n) = p(y_n|x_n)p(y^{n-1}|x^n)$$

by the fact that the channel is memoryless.

Remark (Continued)

To compute $p(y^{n-1}|x^n)$*, we shall use the following:*

$$p(y^{n-1}|x^n)p(x_n|x^{n-1}) = p(y^{n-1}, x_n|x^{n-1}) = p(x_n|y^{n-1}, x^{n-1})p(y^{n-1}|x^n$$

Note that the channel is used without feedback,
$p(x_n|x^{n-1}) = p(x_n|y^{n-1}, x^{n-1})$ *and so*
$p(y^{n-1}|x^n) = p(y^{n-1}|x^{n-1})$ *and the claim follows by induction.*

When we refer to the discrete memoryless channel, we mean the discrete memoryless channel without feedback unless we state explicitly otherwise.

Definition

An (M, n) code for the channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ consists of the following:

1. An index set $\{1, 2, \ldots, M\}$.
2. An encoding function $X^n : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, yielding codewords $x^n(1), x^n(2), \ldots, x^n(M)$. The set of codewords is called the cordbook.
3. A decoding function

$$g : \mathcal{Y}^n \to \{1, 2, \ldots, M\},$$

which is a deterministic rule that assigns a guess to each possible received vector.

Conditional probability of error

Let

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

be the conditional probability of error given that index i was sent, where $I(\cdot)$ is the indicator function.

Definition

The maximal probability of error $\lambda^{(n)}$ for an (M, n) code is defined as

$$\lambda^{(n)} = \max_{i \in \{1,2,\ldots,M\}} \lambda_i.$$

Definition

The (arithmetic) average probability of error $P_e^{(n)}$ for an (M, n) code is defined as

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

Note that if the index W is chosen according to a uniform distribution over the set $\{1, 2, \ldots, M\}$, and $X^n = x^n(W)$, then

$$P_e^{(n)} = \mathsf{Pr}(W \neq g(Y^n)).$$

Also, obviously,

$$P_e^{(n)} \leq \lambda^{(n)}.$$

Definition

The *rate* of an (M, n) code is

$$R = \frac{\log M}{n} \text{ bits per transmission.}$$

Definition

A rate R is said to be *achievable* if there exists a sequence of $(\lceil 2^{nR} \rceil,\ n)$ codes such that $\lambda^{(n)}$ tends to 0 as $n \to \infty$.

Definition

The *capacity* of a channel is the supremum of all achievable rates.

Thus, rates less than capacity yield arbitrarily small probability of error for sufficiently large block lengths.