香港中文大學
The Chinese University of Hong Kong

Chapter 7
Discrete Memoryless Channels

The Chinese University of Hong Kong

In this chapter

In this chapter

- Two equivalent models of a discrete memoryless channel (DMC)

In this chapter

- Two equivalent models of a discrete memoryless channel (DMC)
- How to communicate reliably through a DMC?

In this chapter

- Two equivalent models of a discrete memoryless channel (DMC)
- How to communicate reliably through a DMC?
- The capacity of a DMC

In this chapter

- Two equivalent models of a discrete memoryless channel (DMC)
- How to communicate reliably through a DMC?
- The capacity of a DMC
- Shannon's channel coding theorem: achievability and converse

In this chapter

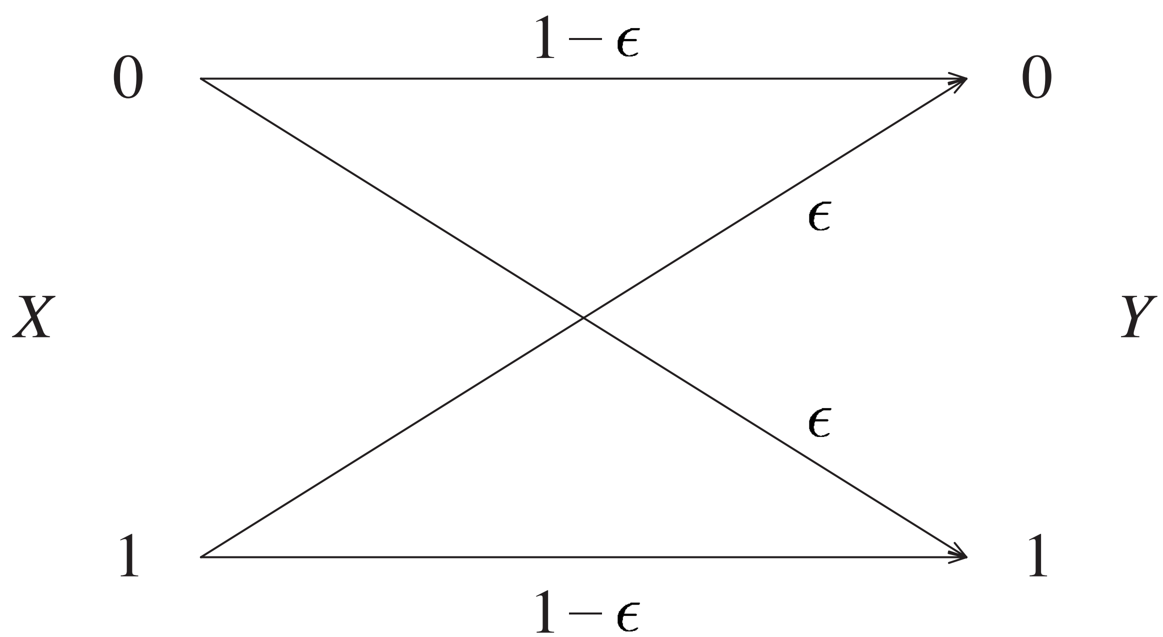
- Two equivalent models of a discrete memoryless channel (DMC)
- How to communicate reliably through a DMC?
- The capacity of a DMC
- Shannon's channel coding theorem: achievability and converse
- The capacity when there is feedback

In this chapter

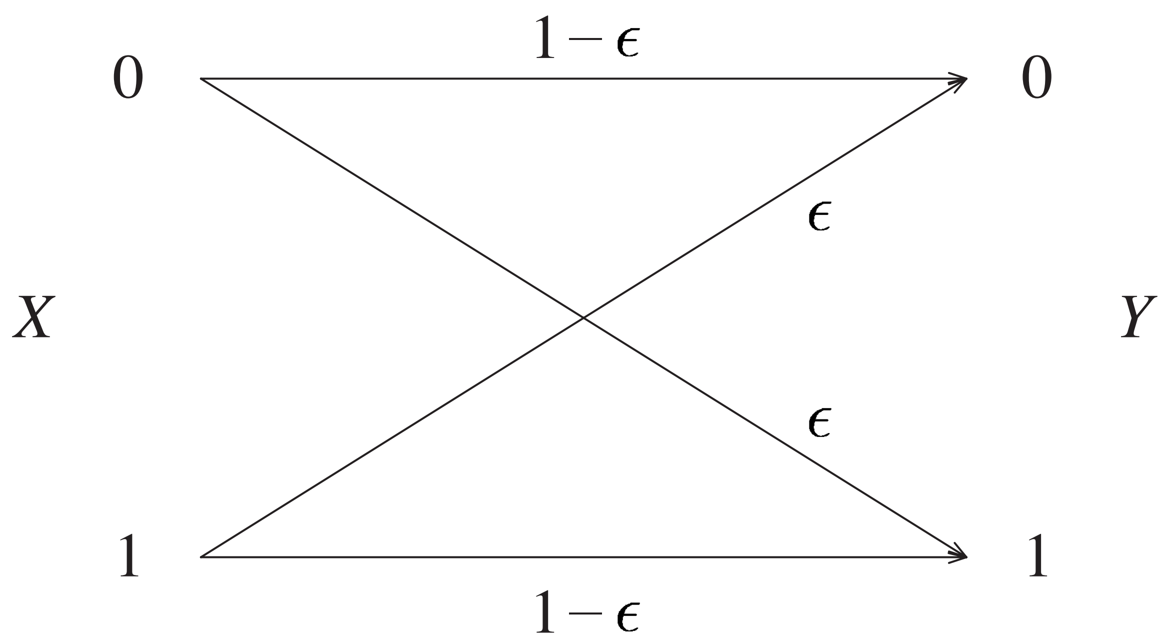
- Two equivalent models of a discrete memoryless channel (DMC)
- How to communicate reliably through a DMC?
- The capacity of a DMC
- Shannon's channel coding theorem: achievability and converse
- The capacity when there is feedback
- Separation theorem for source coding and channel coding

An Informal Discussion

Binary Symmetric Channel (BSC)

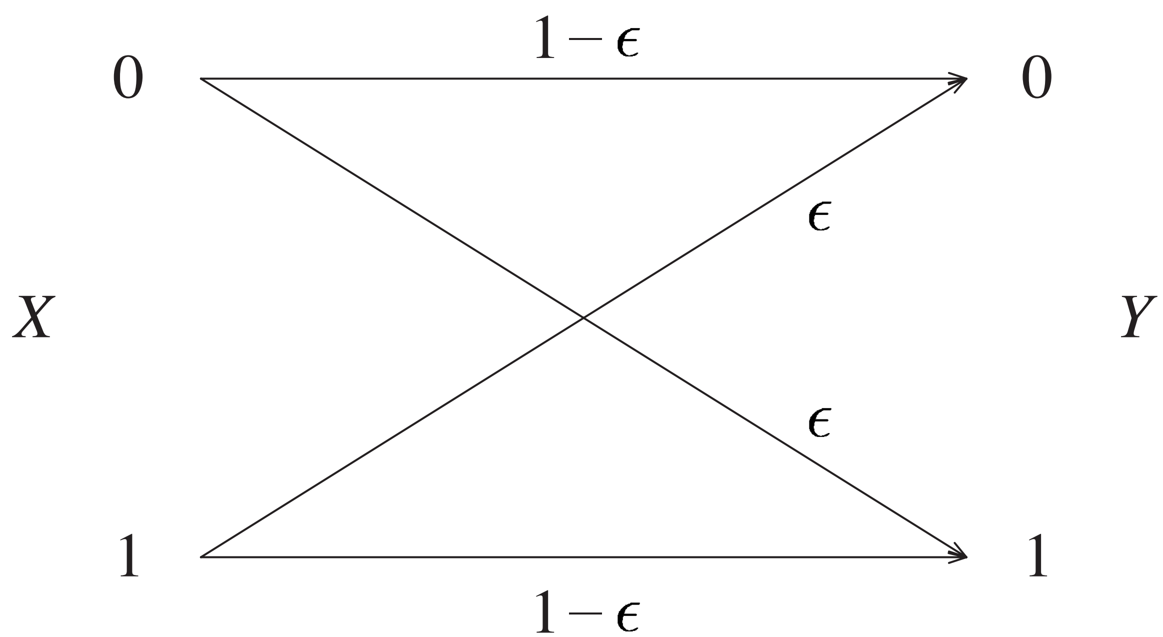


Binary Symmetric Channel (BSC)



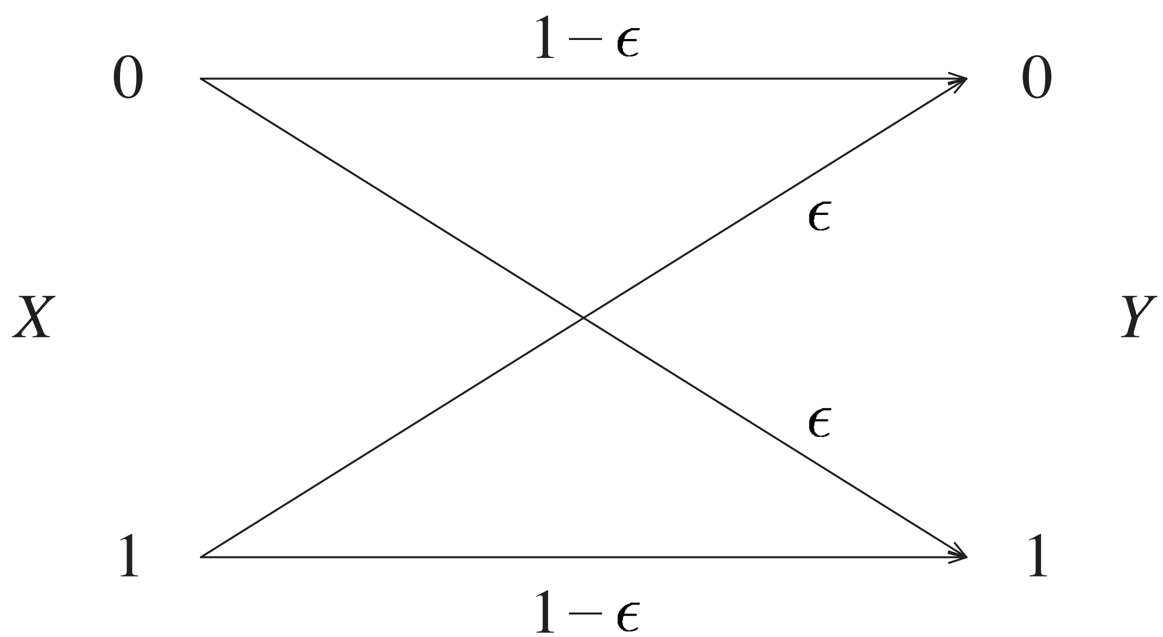
- the simplest channel model

Binary Symmetric Channel (BSC)



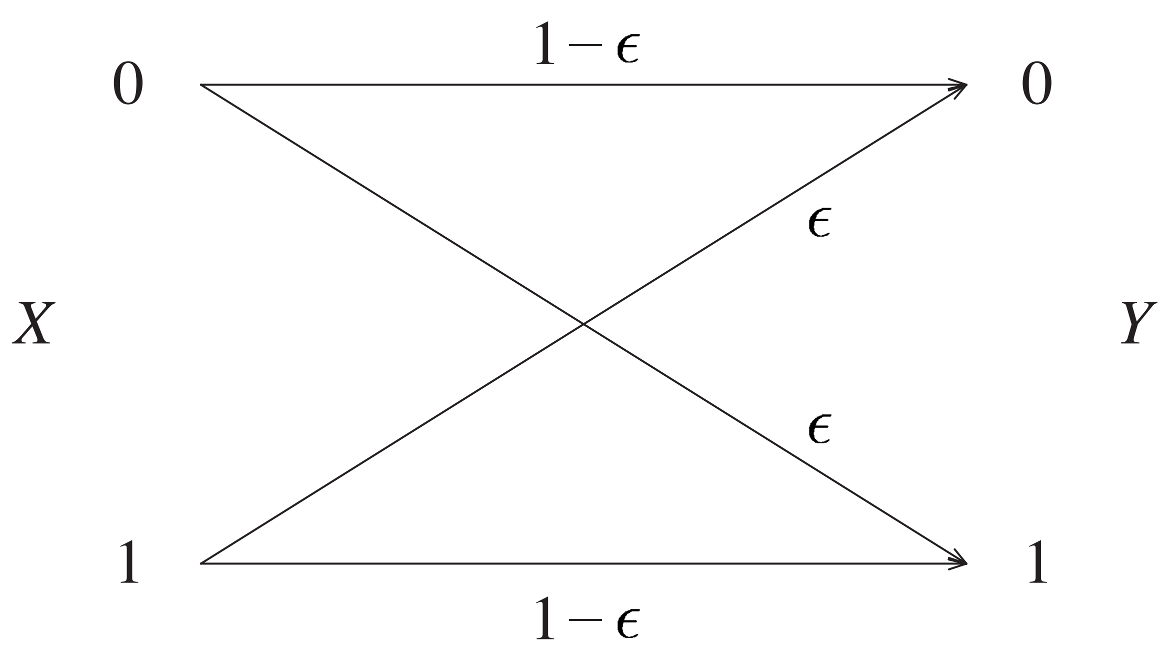
- the simplest channel model
- input alphabet $\mathcal{X} = \{0, 1\}$

Binary Symmetric Channel (BSC)



- the simplest channel model
- input alphabet $\mathcal{X} = \{0, 1\}$
- output alphabet $\mathcal{Y} = \{0, 1\}$

Binary Symmetric Channel (BSC)



- the simplest channel model
- input alphabet $\mathcal{X} = \{0, 1\}$
- output alphabet $\mathcal{Y} = \{0, 1\}$
- Crossover probability $= \epsilon$, $0 \leq \epsilon \leq 1$

A Simple Code

A Simple Code

- Assume $\epsilon < 0.5$.

A Simple Code

- Assume $\epsilon < 0.5$.
- Two possible messages, A and B, are to be sent through the channel.

A Simple Code

- Assume $\epsilon < 0.5$.
- Two possible messages, A and B, are to be sent through the channel.
- **Coding Scheme 1**

$$\text{Encoding} \left\{ \begin{array}{ccc} A & \rightarrow & 0 \\ B & \rightarrow & 1 \end{array} \right. \qquad \text{Decoding} \left\{ \begin{array}{ccc} 0 & \rightarrow & A \\ 1 & \rightarrow & B \end{array} \right.$$

A Simple Code

- Assume $\epsilon < 0.5$.
- Two possible messages, A and B, are to be sent through the channel.
- **Coding Scheme 1**

$$\text{Encoding} \begin{cases} A & \rightarrow & 0 \\ B & \rightarrow & 1 \end{cases} \qquad \text{Decoding} \begin{cases} 0 & \rightarrow & A \\ 1 & \rightarrow & B \end{cases}$$

- A decoding error if and only if a crossover occurs. Therefore, $P_e = \epsilon$.

A Repetition Code

A Repetition Code

- To improve reliability, use the BSC n times for a large n.

A Repetition Code

- To improve reliability, use the BSC n times for a large n.
- Let

$$\begin{aligned} N_0 &= \text{\# 0's received} \\ N_1 &= \text{\# 1's received} \end{aligned}$$

A Repetition Code

- To improve reliability, use the BSC n times for a large n.
- Let

$$\begin{aligned} N_0 &= \#\ 0\text{'s received} \\ N_1 &= \#\ 1\text{'s received} \end{aligned}$$

- **Coding Scheme 2**

$$\text{Encoding} \quad \begin{cases} A & \rightarrow \quad 00\cdots 0 \\ B & \rightarrow \quad 11\cdots 1 \end{cases} \qquad \text{Decoding} \quad \begin{cases} N_0 > N_1 & \rightarrow \quad A \\ N_1 > N_0 & \rightarrow \quad B \end{cases}$$

A Repetition Code

- To improve reliability, use the BSC n times for a large n.
- Let

$$\begin{array}{rcl} N_0 & = & \#\ 0\text{'s received} \\ N_1 & = & \#\ 1\text{'s received} \end{array}$$

- **Coding Scheme 2**

$$\text{Encoding} \left\{ \begin{array}{ccc} A & \to & 00\cdots0 \\ B & \to & 11\cdots1 \end{array} \right. \qquad \text{Decoding} \left\{ \begin{array}{ccc} N_0 > N_1 & \to & A \\ N_1 > N_0 & \to & B \end{array} \right.$$

- If message is A, by WLLN, $N_0 \approx n(1-\epsilon)$ and $N_1 \approx n\epsilon$ w.p. $\to 1$.

A Repetition Code

- To improve reliability, use the BSC n times for a large n.
- Let

$$\begin{aligned} N_0 &= \#\ 0\text{'s received} \\ N_1 &= \#\ 1\text{'s received} \end{aligned}$$

- **Coding Scheme 2**

$$\text{Encoding} \begin{cases} A & \rightarrow \quad 00\cdots 0 \\ B & \rightarrow \quad 11\cdots 1 \end{cases} \qquad \text{Decoding} \begin{cases} N_0 > N_1 & \rightarrow \quad A \\ N_1 > N_0 & \rightarrow \quad B \end{cases}$$

- If message is A, by WLLN, $N_0 \approx n(1-\epsilon)$ and $N_1 \approx n\epsilon$ w.p. $\rightarrow 1$.
- Then $N_0 > N_1$ because $\epsilon < 0.5$.

A Repetition Code

- To improve reliability, use the BSC n times for a large n.
- Let

$$\begin{array}{rcl} N_0 & = & \#\ 0\text{'s received} \\ N_1 & = & \#\ 1\text{'s received} \end{array}$$

- **Coding Scheme 2**

$$\text{Encoding} \left\{ \begin{array}{ccc} A & \rightarrow & 00\cdots0 \\ B & \rightarrow & 11\cdots1 \end{array} \right. \qquad \text{Decoding} \left\{ \begin{array}{ccc} N_0 > N_1 & \rightarrow & A \\ N_1 > N_0 & \rightarrow & B \end{array} \right.$$

- If message is A, by WLLN, $N_0 \approx n(1-\epsilon)$ and $N_1 \approx n\epsilon$ w.p. $\rightarrow 1$.
- Then $N_0 > N_1$ because $\epsilon < 0.5$.
- Therefore decode correctly w.p. $\rightarrow 1$ if message is A (similarly for B).

A Repetition Code

- To improve reliability, use the BSC n times for a large n.
- Let

$$\begin{aligned} N_0 &= \#\ 0\text{'s received} \\ N_1 &= \#\ 1\text{'s received} \end{aligned}$$

- **Coding Scheme 2**

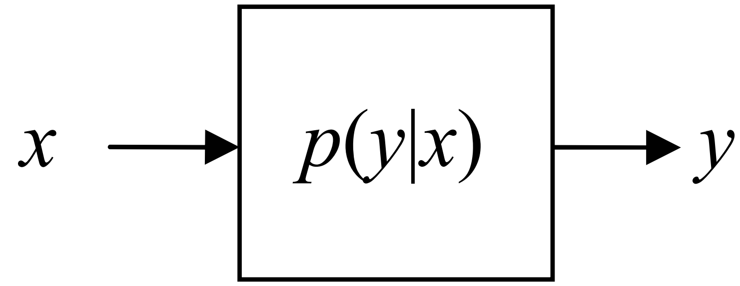
$$\text{Encoding} \begin{cases} A & \rightarrow & 00\cdots0 \\ B & \rightarrow & 11\cdots1 \end{cases} \qquad \text{Decoding} \begin{cases} N_0 > N_1 & \rightarrow & A \\ N_1 > N_0 & \rightarrow & B \end{cases}$$

- If message is A, by WLLN, $N_0 \approx n(1-\epsilon)$ and $N_1 \approx n\epsilon$ w.p. $\rightarrow 1$.
- Then $N_0 > N_1$ because $\epsilon < 0.5$.
- Therefore decode correctly w.p. $\rightarrow 1$ if message is A (similarly for B).
- However, $R = \frac{1}{n}\log 2 \rightarrow 0$ as $n \rightarrow \infty$. :(

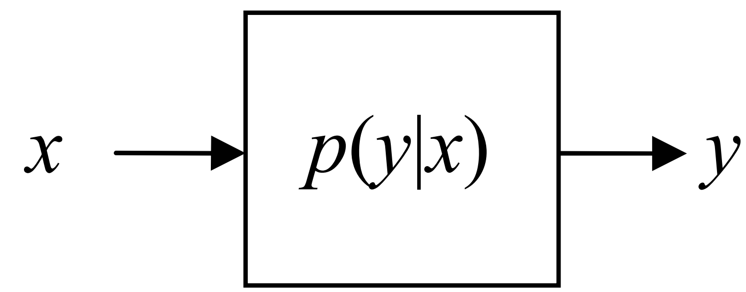
香港中文大學 The Chinese University of Hong Kong

7.1 Definition and Capacity

Definition 7.1 (Discrete Channel I)

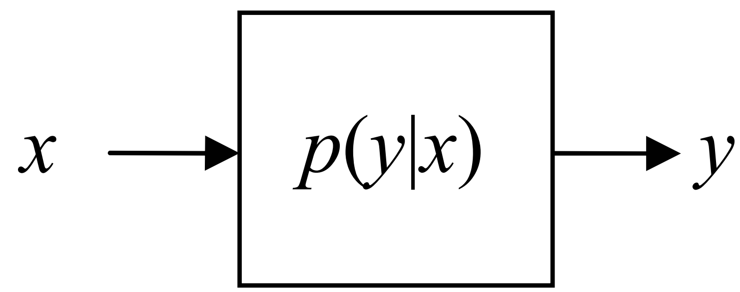


Definition 7.1 (Discrete Channel I)



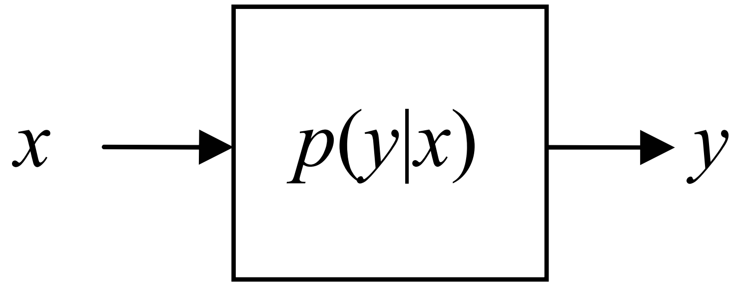
- Input random variable X takes values in discrete alphabet $\mathcal{X}$.

Definition 7.1 (Discrete Channel I)



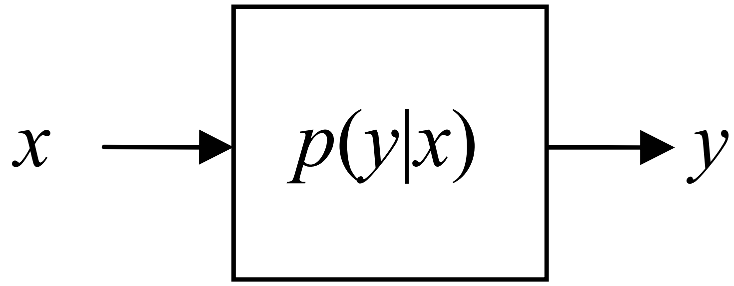
- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete output alphabet $\mathcal{Y}$.

Definition 7.1 (Discrete Channel I)



- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete output alphabet $\mathcal{Y}$.
- The channel is specified by a transition matrix $p(y|x)$ from $\mathcal{X}$ to $\mathcal{Y}$.

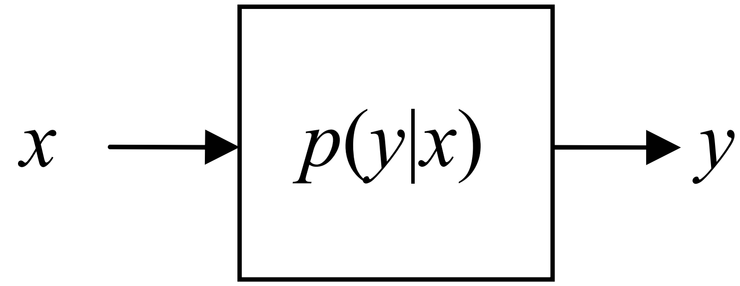
Definition 7.1 (Discrete Channel I)



- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete output alphabet $\mathcal{Y}$.
- The channel is specified by a transition matrix $p(y|x)$ from $\mathcal{X}$ to $\mathcal{Y}$.
- Input-output relation:

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

Definition 7.1 (Discrete Channel I)



- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete output alphabet $\mathcal{Y}$.
- The channel is specified by a transition matrix $p(y|x)$ from $\mathcal{X}$ to $\mathcal{Y}$.
- Input-output relation:

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

- BSC with crossover probability ϵ:

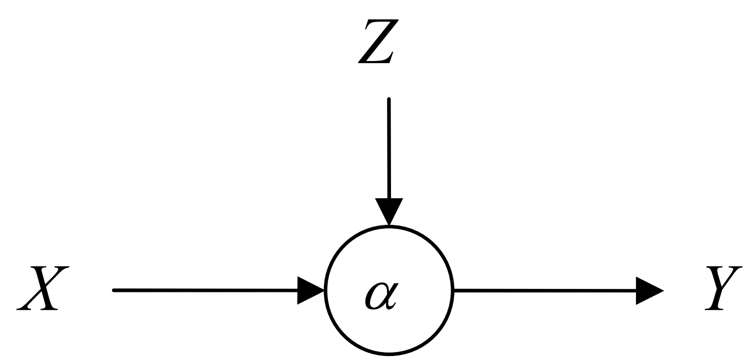
$$[\,p(y|x)] = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix}$$

Definition 7.1 (Discrete Channel I) 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-input 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

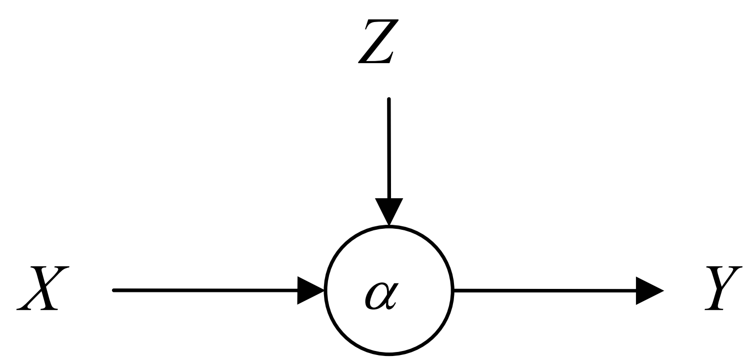
$$\Pr\{X = x, Y = y\} = \Pr\{X = x\}\, p(y|x)$$

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

Definition 7.2 (Discrete Channel II)

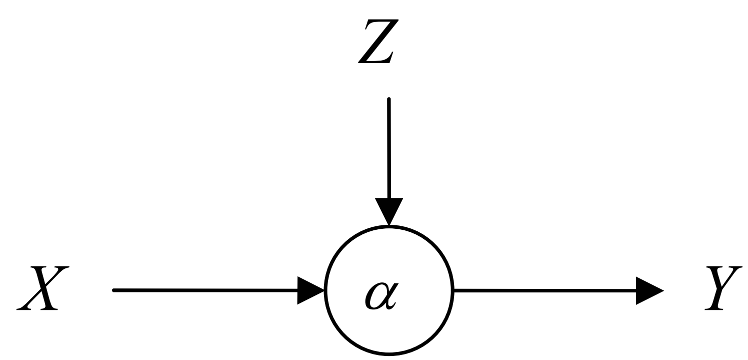


Definition 7.2 (Discrete Channel II)



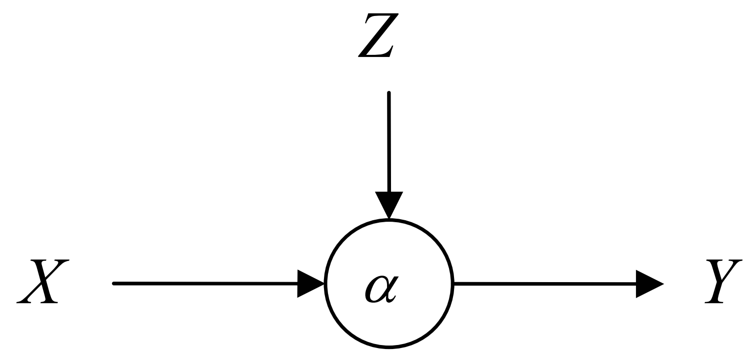
- Input random variable X takes values in discrete alphabet $\mathcal{X}$.

Definition 7.2 (Discrete Channel II)



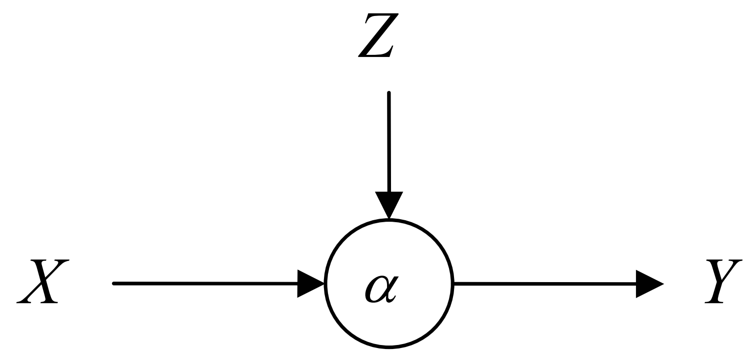
- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete alphabet $\mathcal{Y}$.

Definition 7.2 (Discrete Channel II)



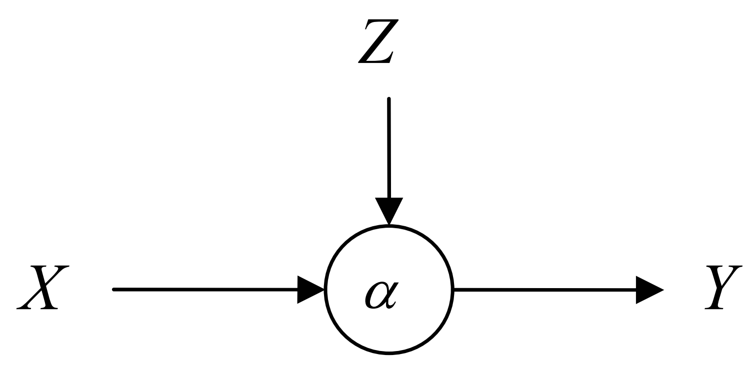
- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete alphabet $\mathcal{Y}$.
- Noise variable Z takes values in discrete alphabet $\mathcal{Z}$.

Definition 7.2 (Discrete Channel II)



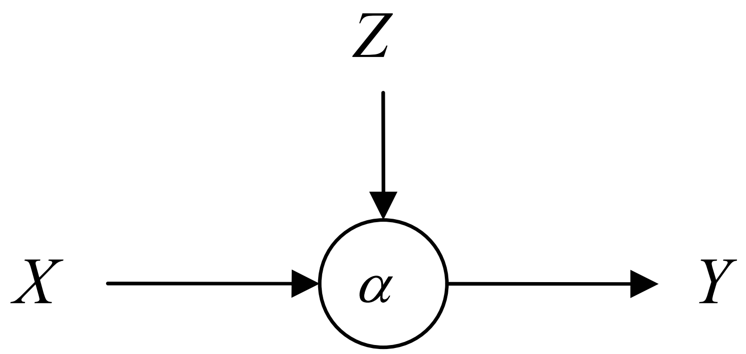
- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete alphabet $\mathcal{Y}$.
- Noise variable Z takes values in discrete alphabet $\mathcal{Z}$.
- Z is independent of X.

Definition 7.2 (Discrete Channel II)



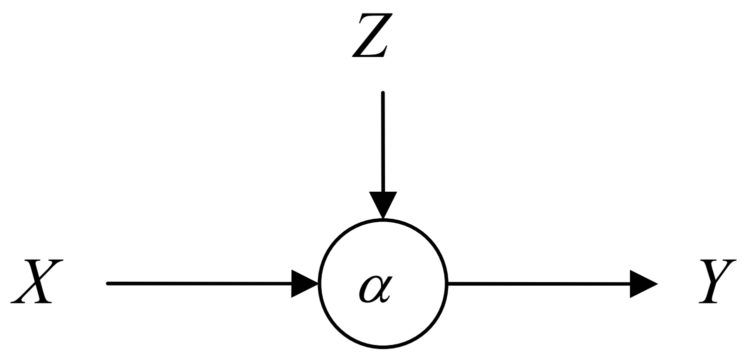
- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete alphabet $\mathcal{Y}$.
- Noise variable Z takes values in discrete alphabet $\mathcal{Z}$.
- Z is independent of X.
- α is a function from $\mathcal{X} \times \mathcal{Z}$ to $\mathcal{Y}$.

Definition 7.2 (Discrete Channel II)



- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete alphabet $\mathcal{Y}$.
- Noise variable Z takes values in discrete alphabet $\mathcal{Z}$.
- Z is independent of X.
- α is a function from $\mathcal{X} \times \mathcal{Z}$ to $\mathcal{Y}$.
- The channel is specified by (α, Z).

Definition 7.2 (Discrete Channel II)



- Input random variable X takes values in discrete alphabet $\mathcal{X}$.
- Output random variable Y takes values in discrete alphabet $\mathcal{Y}$.
- Noise variable Z takes values in discrete alphabet $\mathcal{Z}$.
- Z is independent of X.
- α is a function from $\mathcal{X} \times \mathcal{Z}$ to $\mathcal{Y}$.
- The channel is specified by (α, Z).
- Input-output relation: $Y = \alpha(X, Z)$.

Definition 7.2 (Discrete Channel II) 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

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

Equivalence of
Discrete Channel I & Discrete Channel II

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.
  - Let $Y = Z_x$ if $X = x$, so that $Y = \alpha(X, Z)$.

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.
  - Let $Y = Z_x$ if $X = x$, so that $Y = \alpha(X, Z)$.

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.
  - Let $Y = Z_x$ if $X = x$, so that $Y = \alpha(X, Z)$.
  - Then

Equivalence of
Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.
  - Let $Y = Z_x$ if $X = x$, so that $Y = \alpha(X, Z)$.
  - Then

$$\Pr\{X = x, Y = y\} \quad = \quad \Pr\{X = x\}\Pr\{Y = y|X = x\}$$

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.
  - Let $Y = Z_x$ if $X = x$, so that $Y = \alpha(X, Z)$.
  - Then

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

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.
  - Let $Y = Z_x$ if $X = x$, so that $Y = \alpha(X, Z)$.
  - Then

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

Equivalence of Discrete Channel I & Discrete Channel II

- II $\Rightarrow$ I $(\alpha, Z) \Rightarrow p(y|x)$: Obvious.
- I $\Rightarrow$ II:
  - For $x \in \mathcal{X}$, define Z_x with $\mathcal{Z}_x = \mathcal{Y}$ such that $\Pr\{Z_x = y\} = p(y|x)$.
  - Assume Z_x, $x \in \mathcal{X}$ are mutually independent and also independent of X.
  - Define the noise variable $Z = (Z_x : x \in \mathcal{X})$.
  - Let $Y = Z_x$ if $X = x$, so that $Y = \alpha(X, Z)$.
  - Then

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

Definition 7.3 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.

Discrete Memoryless Channel

Discrete Memoryless Channel

- A discrete channel can be used repeatedly at every time index $i = 1, 2, \cdots$.

Discrete Memoryless Channel

- A discrete channel can be used repeatedly at every time index $i = 1, 2, \cdots$.
- Assume the noise for the transmission over the channel at different time indices are independent of each other.

Discrete Memoryless Channel

- A discrete channel can be used repeatedly at every time index $i = 1, 2, \cdots$.
- Assume the noise for the transmission over the channel at different time indices are independent of each other.
- To properly formulate a DMC, we regard it as a subsystem of a discrete-time stochastic system which will be referred to as "the system".

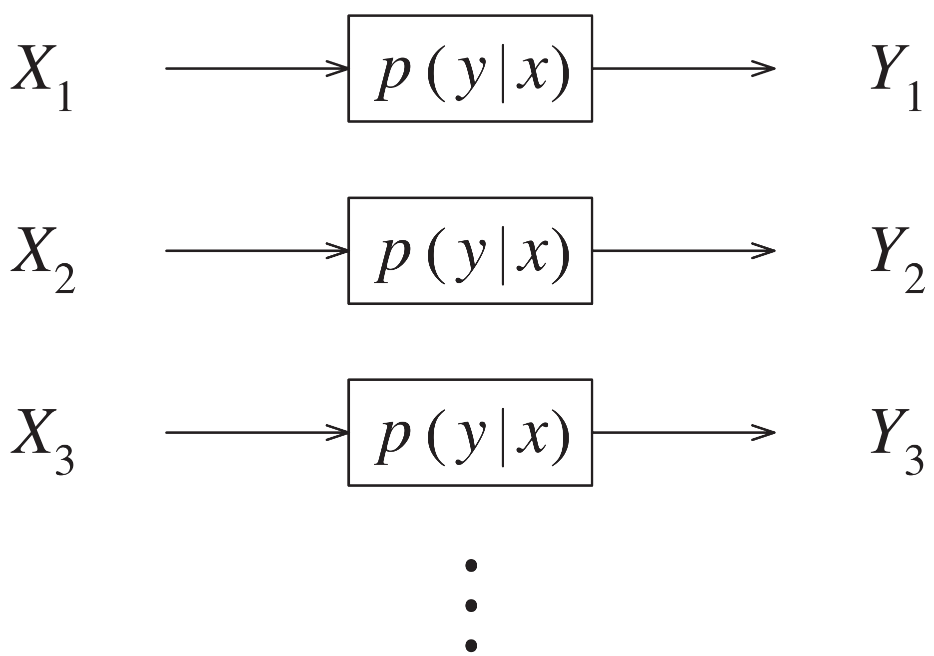
Discrete Memoryless Channel

- A discrete channel can be used repeatedly at every time index $i = 1, 2, \cdots$.
- Assume the noise for the transmission over the channel at different time indices are independent of each other.
- To properly formulate a DMC, we regard it as a subsystem of a discrete-time stochastic system which will be referred to as "the system".
- In such a system, random variables are generated sequentially in discrete-time.

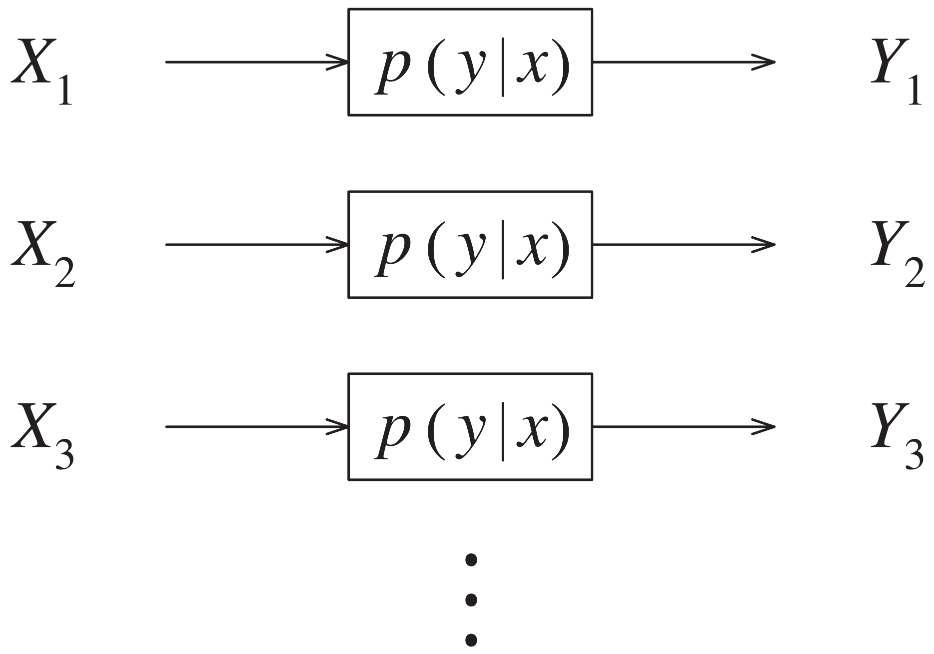
Discrete Memoryless Channel

- A discrete channel can be used repeatedly at every time index $i = 1, 2, \cdots$.
- Assume the noise for the transmission over the channel at different time indices are independent of each other.
- To properly formulate a DMC, we regard it as a subsystem of a discrete-time stochastic system which will be referred to as "the system".
- In such a system, random variables are generated sequentially in discrete-time.
- More than one random variable may be generated instantaneously but sequentially at a particular time index.

Definition 7.4 (DMC I)

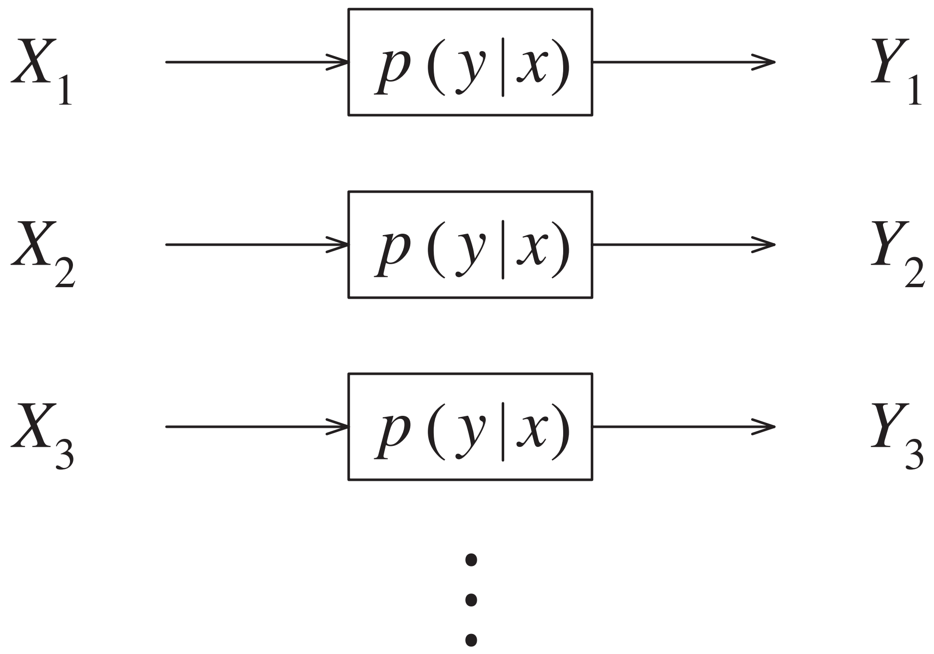


Definition 7.4 (DMC I)



- A DMC specified by $p(y|x)$ is a sequence of replicates of a generic discrete channel $p(y|x)$.

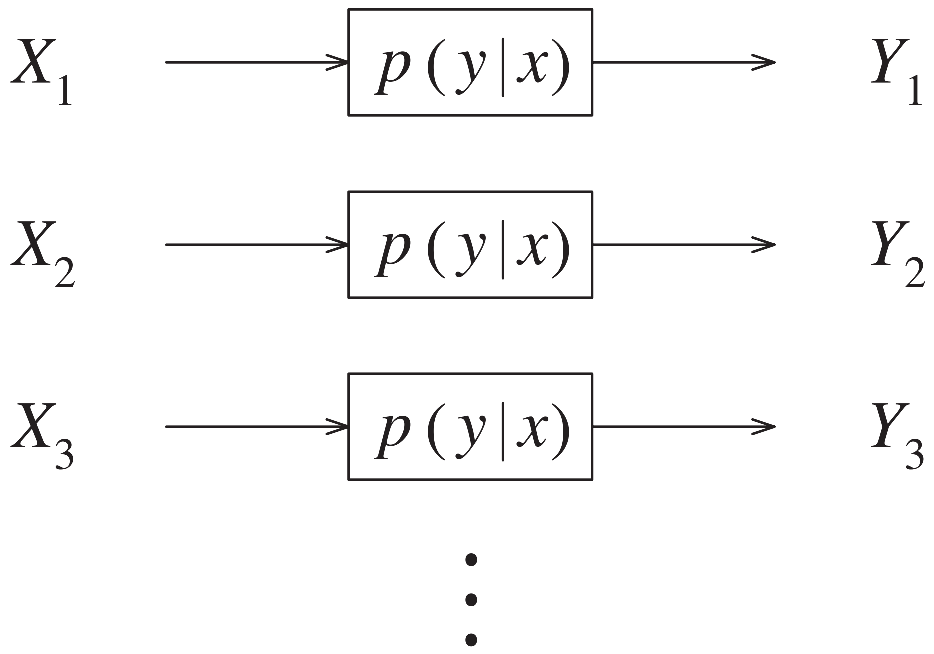
Definition 7.4 (DMC I)



- A DMC specified by $p(y|x)$ is a sequence of replicates of a generic discrete channel $p(y|x)$.

- T_{i-} : all the random variables in the system generated before X_i.

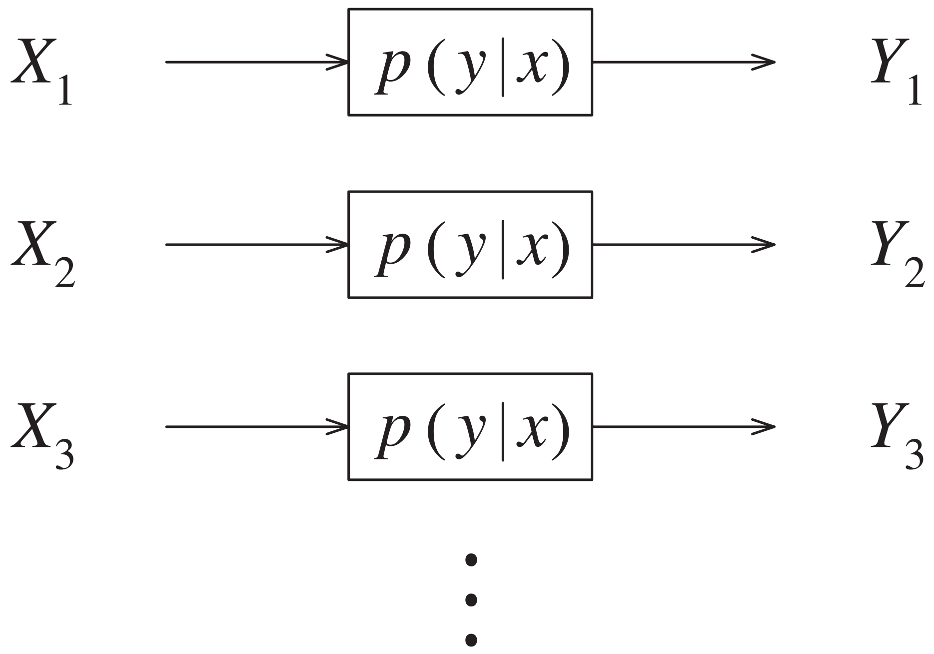
Definition 7.4 (DMC I)



- A DMC specified by $p(y|x)$ is a sequence of replicates of a generic discrete channel $p(y|x)$.
- T_{i-} : all the random variables in the system generated before X_i.
- Memoryless Property (Independent noise):

$$\begin{aligned} &\Pr\{Y_i = y, X_i = x, T_{i-} = t\} \\ &= \Pr\{X_i = x, T_{i-} = t\} p(y|x) \end{aligned}$$

Definition 7.4 (DMC I)

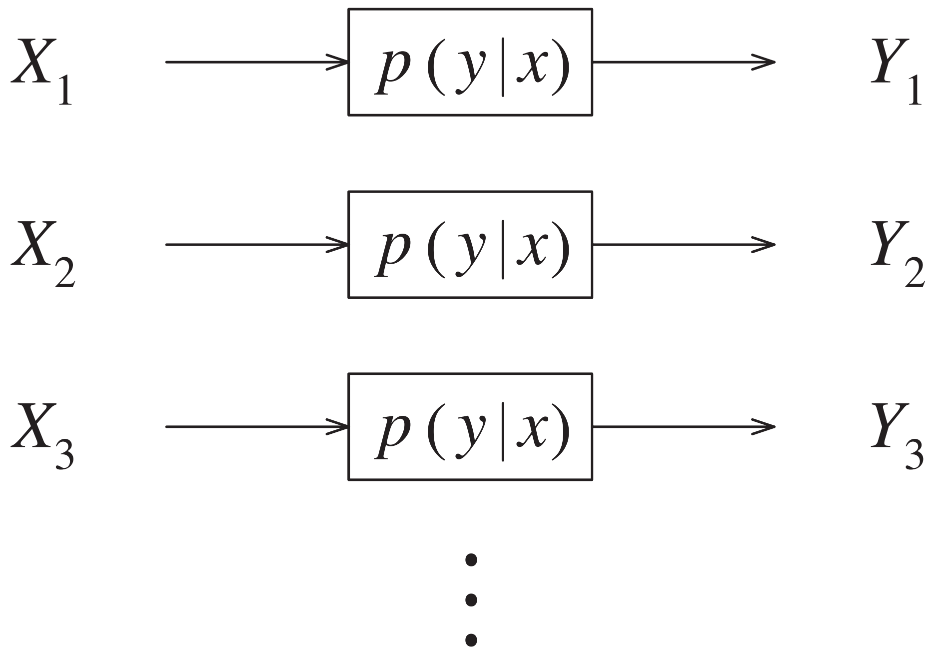


- A DMC specified by $p(y|x)$ is a sequence of replicates of a generic discrete channel $p(y|x)$.
- T_{i-} : all the random variables in the system generated before X_i.
- Memoryless Property (Independent noise):

$$\begin{aligned} &\Pr\{Y_i = y, X_i = x, T_{i-} = t\} \\ = \quad &\Pr\{X_i = x, T_{i-} = t\} p(y|x) \end{aligned}$$

$\Pr\{Y_i = y | X_i = x, T_{i-} = t\}$

Definition 7.4 (DMC I)



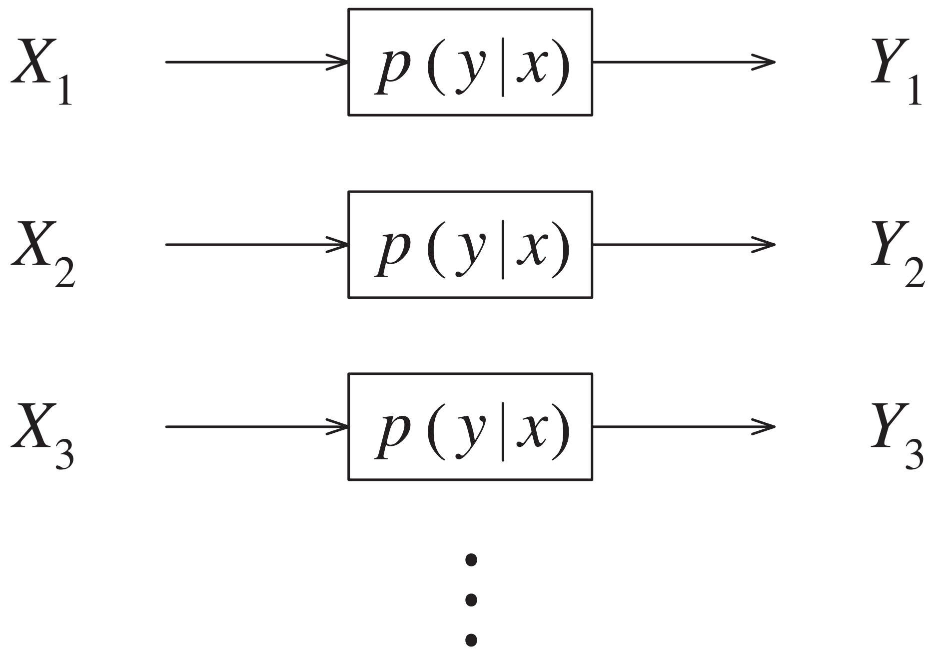
- A DMC specified by $p(y|x)$ is a sequence of replicates of a generic discrete channel $p(y|x)$.
- T_{i-} : all the random variables in the system generated before X_i.
- Memoryless Property (Independent noise):

$$\begin{aligned} & \Pr\{Y_i = y, X_i = x, T_{i-} = t\} \\ = & \Pr\{X_i = x, T_{i-} = t\} p(y|x) \end{aligned}$$

$\Pr\{Y_i = y | X_i = x, T_{i-} = t\}$

- Equivalently, $T_{i-} \to X_i \to Y_i$, or

Definition 7.4 (DMC I)



- A DMC specified by $p(y|x)$ is a sequence of replicates of a generic discrete channel $p(y|x)$.
- T_{i-} : all the random variables in the system generated before X_i.
- Memoryless Property (Independent noise):

$$\Pr\{Y_i = y, X_i = x, T_{i-} = t\} = \Pr\{X_i = x, T_{i-} = t\} p(y|x)$$

$\Pr\{Y_i = y | X_i = x, T_{i-} = t\}$

- Equivalently, $T_{i-} \to X_i \to Y_i$, or

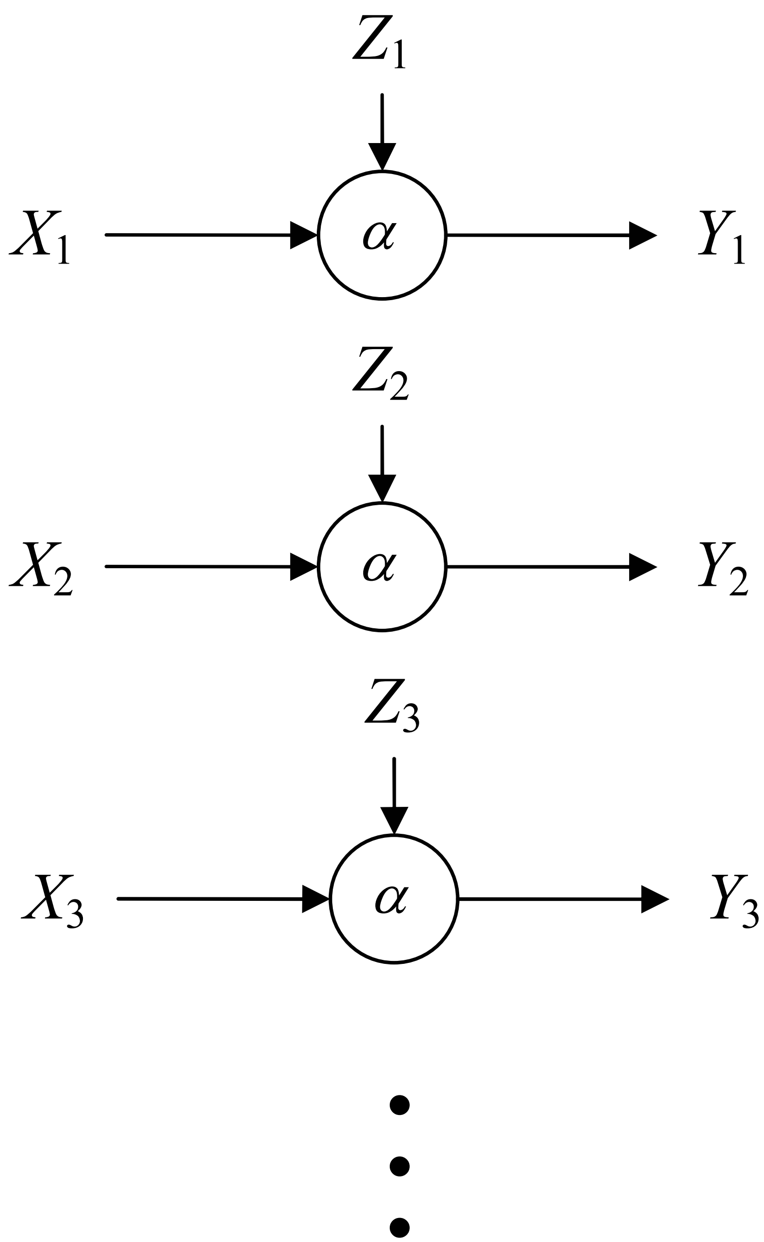
Given X_i, Y_i is independent of everything in the past.

Definition 7.4 (DMC I) A discrete memoryless channel (DMC) $p(y|x)$ is a sequence of replicates of a generic discrete channel $p(y|x)$. These discrete channels are indexed by a discrete-time index i, where $i \geq 1$, with the ith channel being available for transmission at time i. Transmission through a channel is assumed to be instantaneous. Let X_i and Y_i be respectively the input and the output of the DMC at time i, and let T_{i-} denote all the random variables that are generated in the system before X_i. The equality

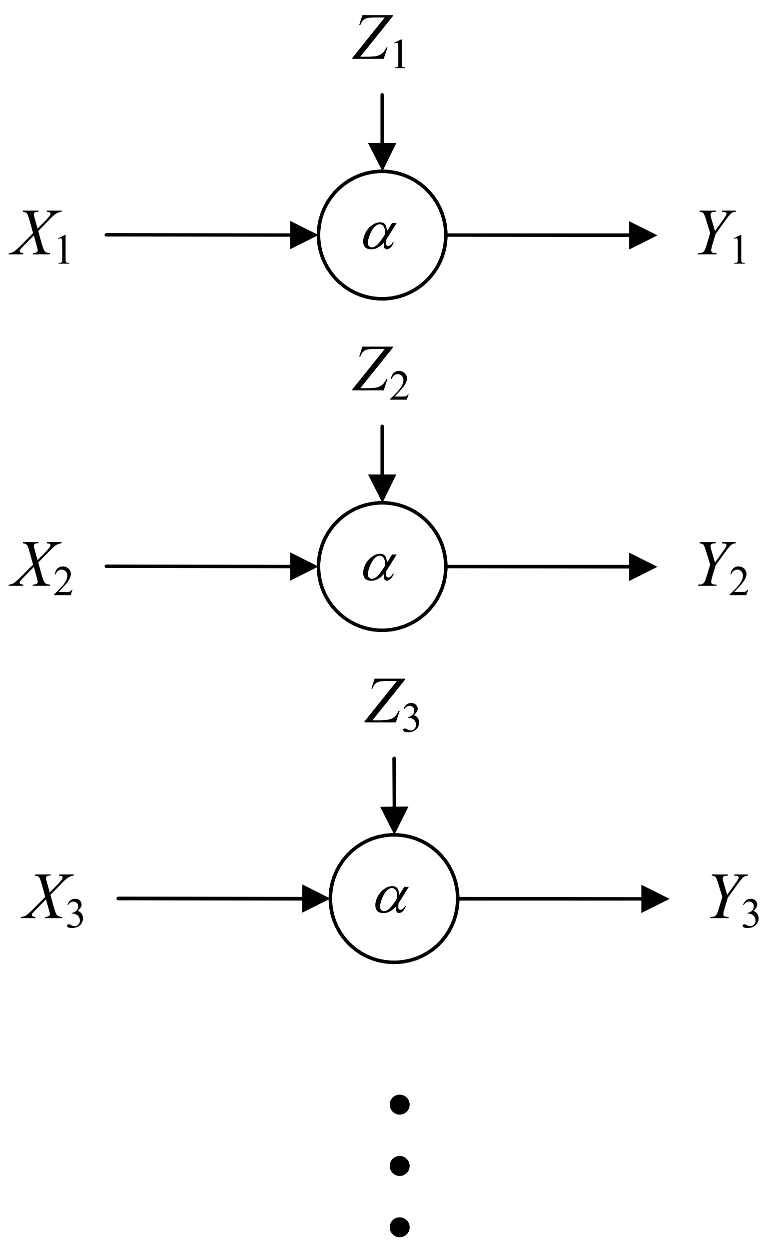
$$\Pr\{Y_i = y, X_i = x, T_{i-} = t\} = \Pr\{X_i = x, T_{i-} = t\} p(y|x)$$

holds for all $(x, y, t) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{T}_{i-}$.

Definition 7.5 (DMC II)

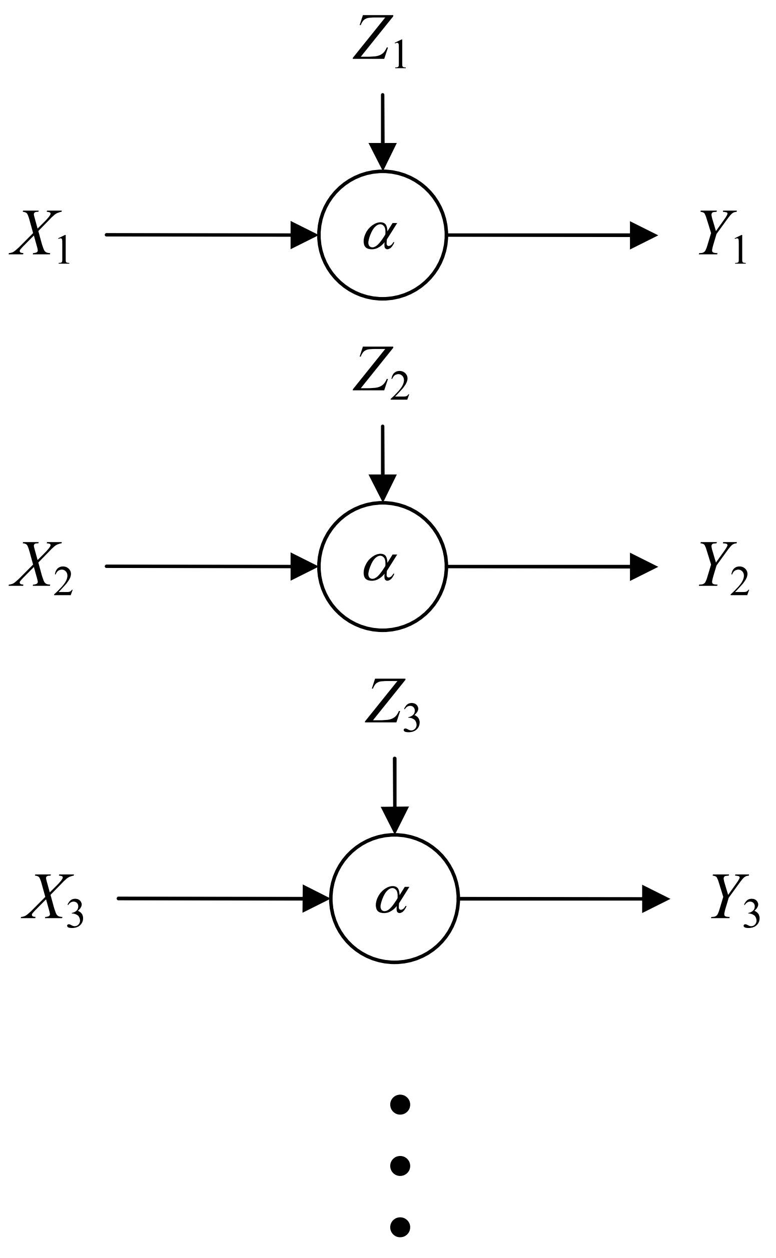


Definition 7.5 (DMC II)



- A DMC specified by (α, Z) is a sequence of replicates of a generic discrete channel (α, Z).

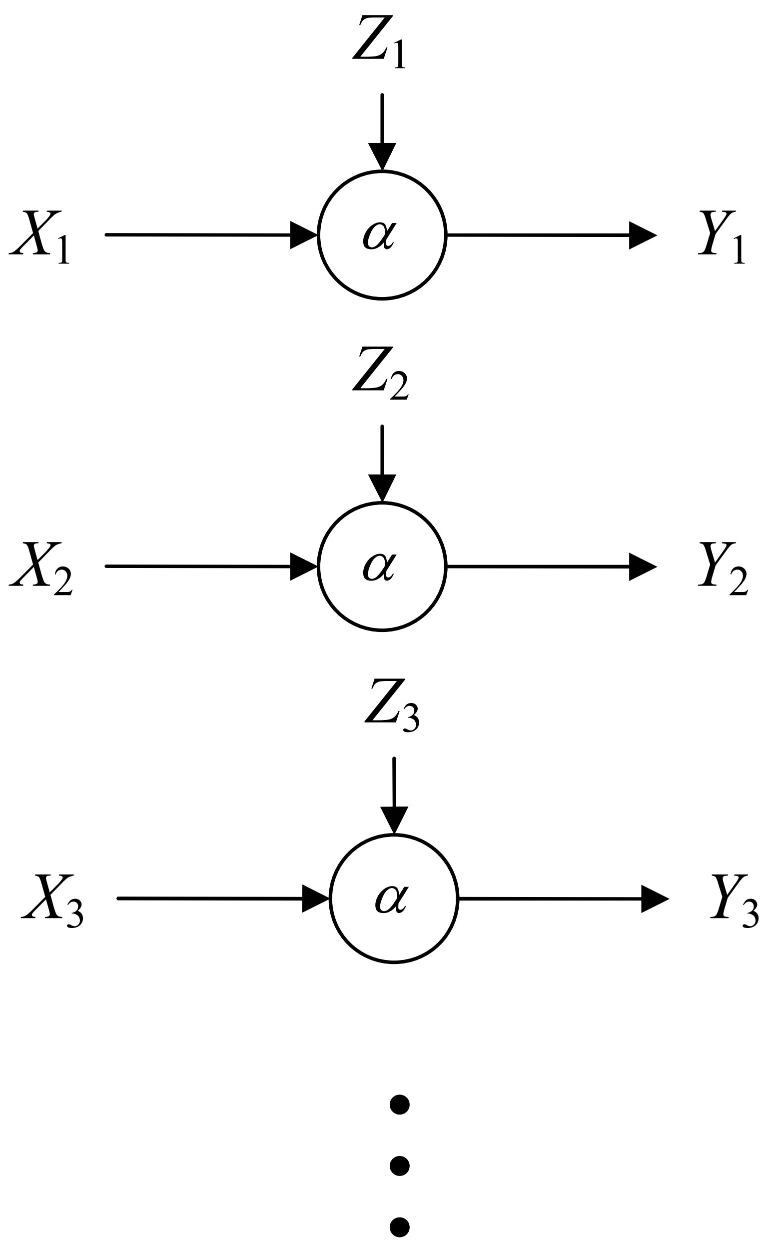
Definition 7.5 (DMC II)



- A DMC specified by (α, Z) is a sequence of replicates of a generic discrete channel (α, Z).

- Z_i is the noise variable for transmission at time i, and has the same distribution as Z.

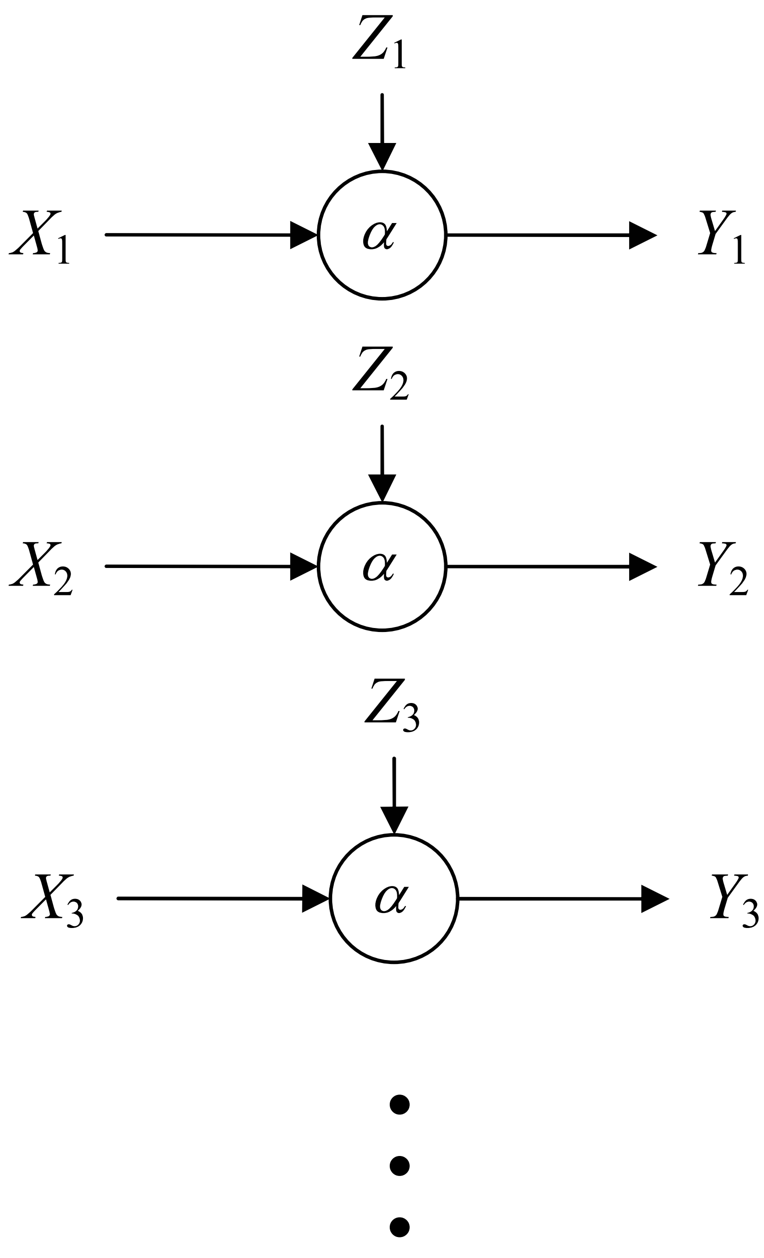
Definition 7.5 (DMC II)



- A DMC specified by (α, Z) is a sequence of replicates of a generic discrete channel (α, Z).
- Z_i is the noise variable for transmission at time i, and has the same distribution as Z.
- Memoryless Property (Independent noise):

$$Z_i \text{ is independent of } (X_i, T_{i-}).$$

Definition 7.5 (DMC II)



- A DMC specified by (α, Z) is a sequence of replicates of a generic discrete channel (α, Z).
- Z_i is the noise variable for transmission at time i, and has the same distribution as Z.
- Memoryless Property (Independent noise):

$$Z_i \text{ is independent of } (X_i, T_{i-}).$$

- That is,

The noise for transmission at time i is independent of both X_i and everything in the past.

Definition 7.5 (DMC II) A discrete memoryless channel (α, Z) is a sequence of replicates of a generic discrete channel (α, Z). These discrete channels are indexed by a discrete-time index i, where $i \geq 1$, with the ith channel being available for transmission at time i. Transmission through a channel is assumed to be instantaneous. Let X_i and Y_i be respectively the input and the output of the DMC at time i, and let T_{i-} denote all the random variables that are generated in the system before X_i. The noise variable Z_i for the transmission at time i is a copy of the generic noise variable Z, and is independent of (X_i, T_{i-}). The output of the DMC at time i is given by

$$Y_i = \alpha(X_i, Z_i).$$

Remark: The equivalence of Definitions 7.4 and 7.5 can be shown. See textbook.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

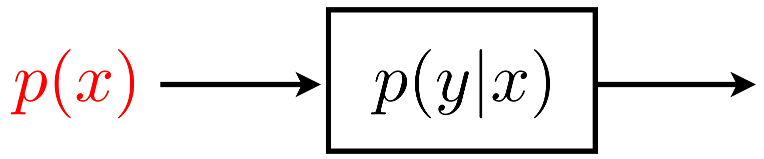
where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

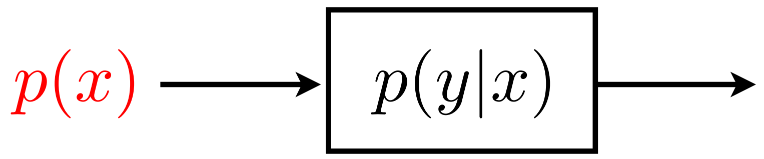


Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.



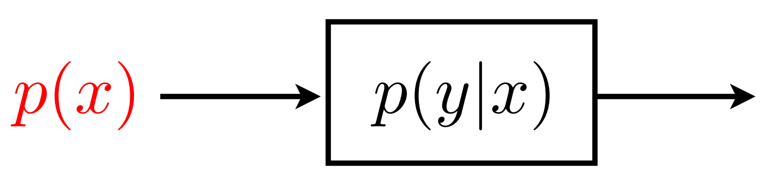
Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

- For each input distribution $p(x)$, we have

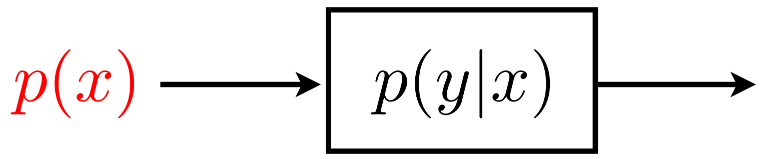


Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.



- For each input distribution $p(x)$, we have

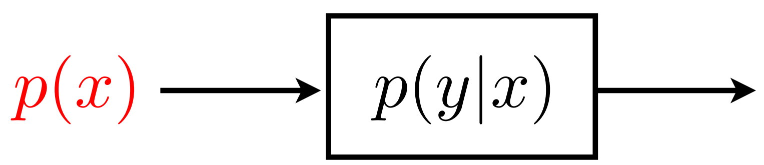
$$p(x, y) = p(x)p(y|x).$$

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.



- For each input distribution $p(x)$, we have

$$p(x, y) = p(x)p(y|x).$$

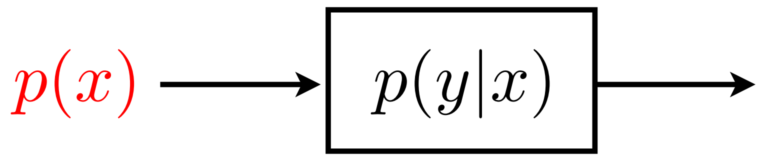
- From $p(x, y)$, we can compute $I(X;Y)$.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.



- For each input distribution $p(x)$, we have

$$p(x,y) = p(x)p(y|x).$$

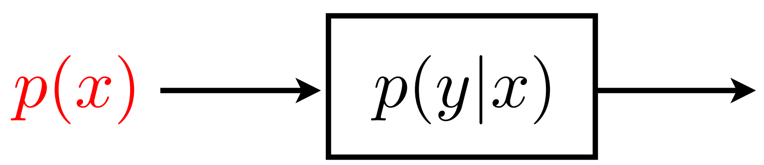
- From $p(x,y)$, we can compute $I(X;Y)$.
- Maximize $I(X;Y)$ over all $p(x)$.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.



- For each input distribution $p(x)$, we have

$$p(x, y) = p(x)p(y|x).$$

- From $p(x, y)$, we can compute $I(X;Y)$.
- Maximize $I(X;Y)$ over all $p(x)$.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Remarks:

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Remarks:

- Since $I(X;Y)$ is a continuous functional of $p(x)$ and the set of all $p(x)$ is a compact set (i.e., closed and bounded) in $\Re^{|\mathcal{X}|}$, the maximum value of $I(X;Y)$ can be attained.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

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

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Remarks:

- Since $I(X;Y)$ is a continuous functional of $p(x)$ and the set of all $p(x)$ is a compact set (i.e., closed and bounded) in $\Re^{|\mathcal{X}|}$, the maximum value of $I(X;Y)$ can be attained.
- Will see that C is in fact the maximum rate at which information can be communicated reliably through a DMC.

Assume both $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Definition 7.6 The capacity of a discrete memoryless channel $p(y|x)$ is defined as

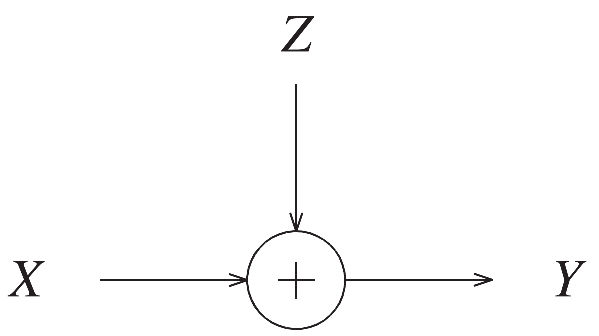
$$C = \max_{p(x)} I(X;Y),$$

where X and Y are respectively the input and the output of the generic discrete channel, and the maximum is taken over all input distributions $p(x)$.

Remarks:

- Since $I(X;Y)$ is a continuous functional of $p(x)$ and the set of all $p(x)$ is a compact set (i.e., closed and bounded) in $\Re^{|\mathcal{X}|}$, the maximum value of $I(X;Y)$ can be attained.
- Will see that C is in fact the maximum rate at which information can be communicated reliably through a DMC.
- Can communicate through a channel at a positive rate while $P_e \to 0$!

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

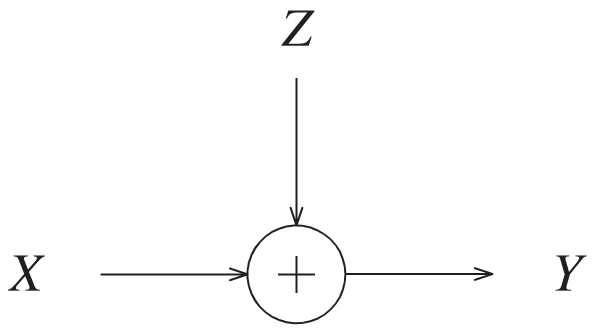
$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

and Z is independent of X.

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

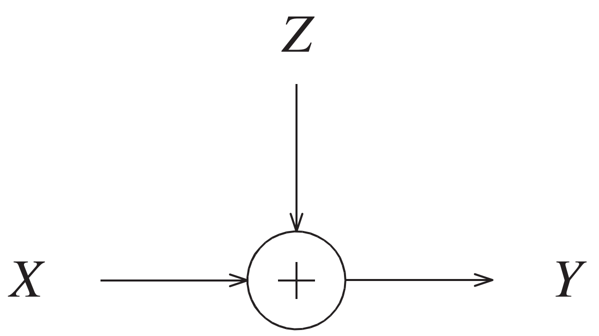
$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

and Z is independent of X.

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \text{ mod } 2$$

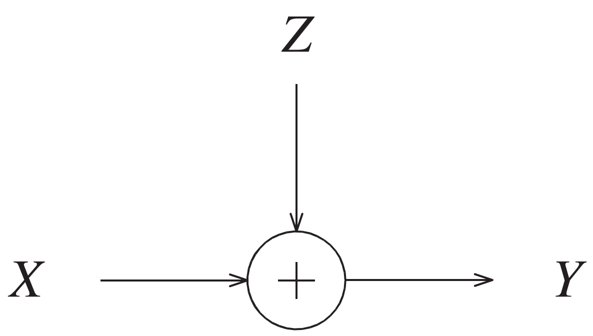
with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

and Z is independent of X.

Determination of C

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

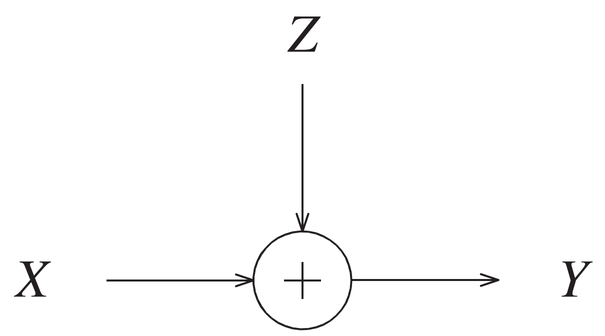
$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

and Z is independent of X.

Determination of C

1. Consider

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

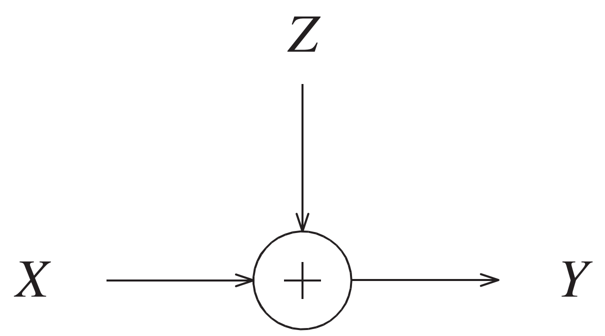
and Z is independent of X.

Determination of C

1. Consider

$$I(X;Y) = H(Y) - H(Y|X)$$

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

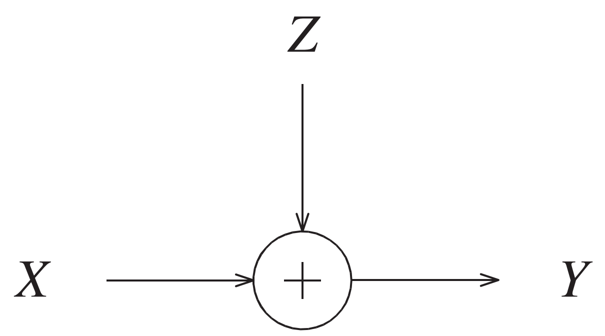
and Z is independent of X.

Determination of C

1. Consider

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

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

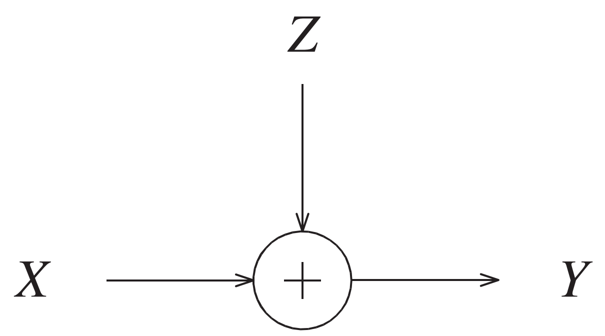
and Z is independent of X.

Determination of C

1. Consider

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_x p(x) H(Y|X = x) \\ &= H(Y) - \sum_x p(x) h_b(\epsilon) \end{aligned}$$

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

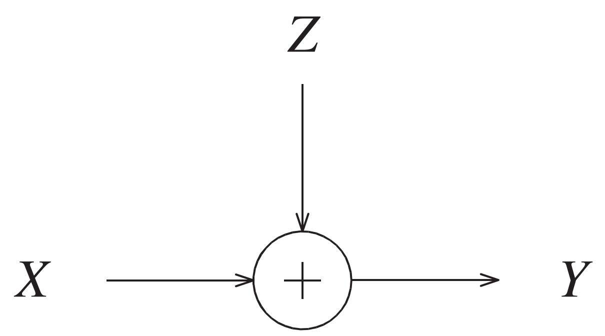
and Z is independent of X.

Determination of C

1. Consider

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_x p(x) H(Y|X = x) \\ &= H(Y) - \sum_x p(x) h_b(\epsilon) \\ &= H(Y) - h_b(\epsilon) \end{aligned}$$

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

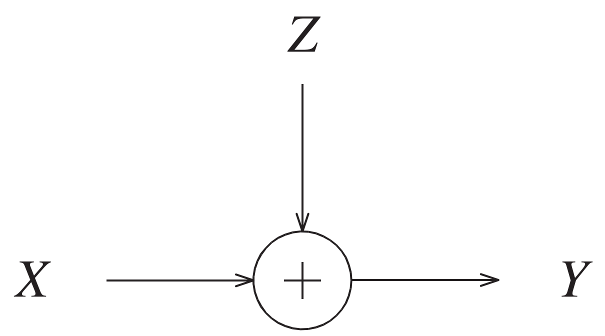
and Z is independent of X.

Determination of C

1. Consider

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

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

and Z is independent of X.

Determination of C

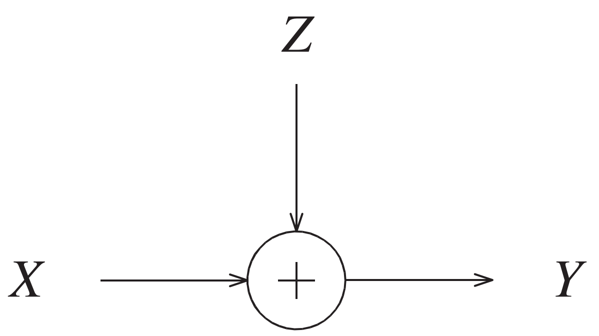
1. Consider

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

2. So,

$$C = \max_{p(x)} I(X;Y) \leq 1 - h_b(\epsilon).$$

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

and Z is independent of X.

Determination of C

1. Consider

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

2. So,

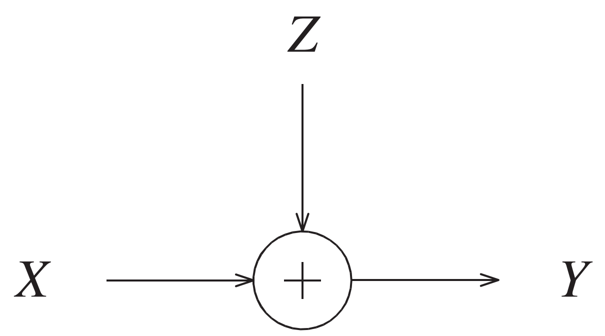
$$C = \max_{p(x)} I(X;Y) \leq 1 - h_b(\epsilon).$$

3. The upper bound on $I(X;Y)$ is tight if

$$H(Y) = 1.$$

This can be achieved by taking the uniform input distribution.

Example 7.7 (BSC)



Alternative representation of a BSC with crossover probability ϵ:

$$Y = X + Z \bmod 2$$

with

$$\Pr\{Z = 0\} = 1 - \epsilon \quad \text{and} \quad \Pr\{Z = 1\} = \epsilon$$

and Z is independent of X.

Determination of C

1. Consider

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

2. So,

$$C = \max_{p(x)} I(X;Y) \leq 1 - h_b(\epsilon).$$

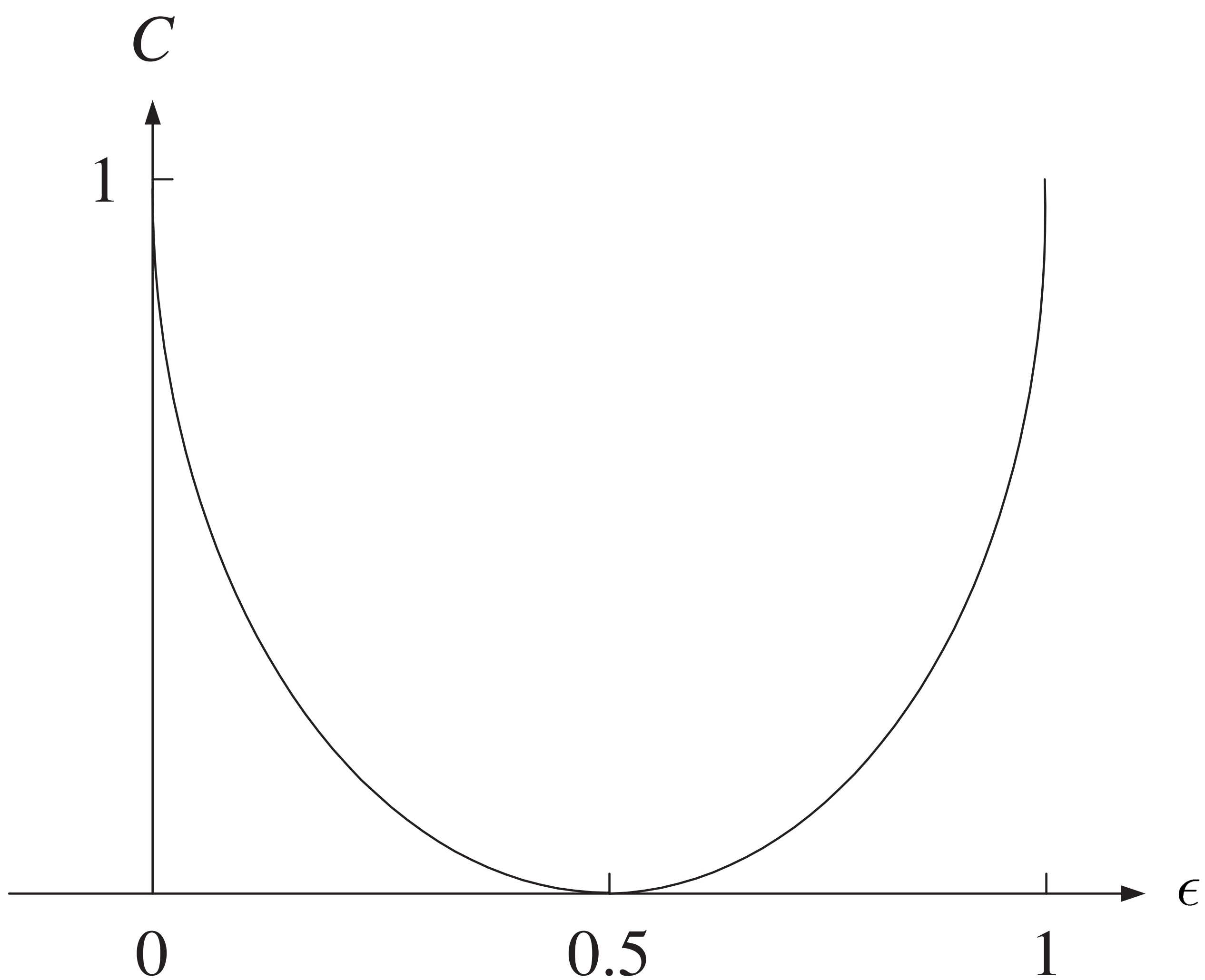
3. The upper bound on $I(X;Y)$ is tight if

$$H(Y) = 1.$$

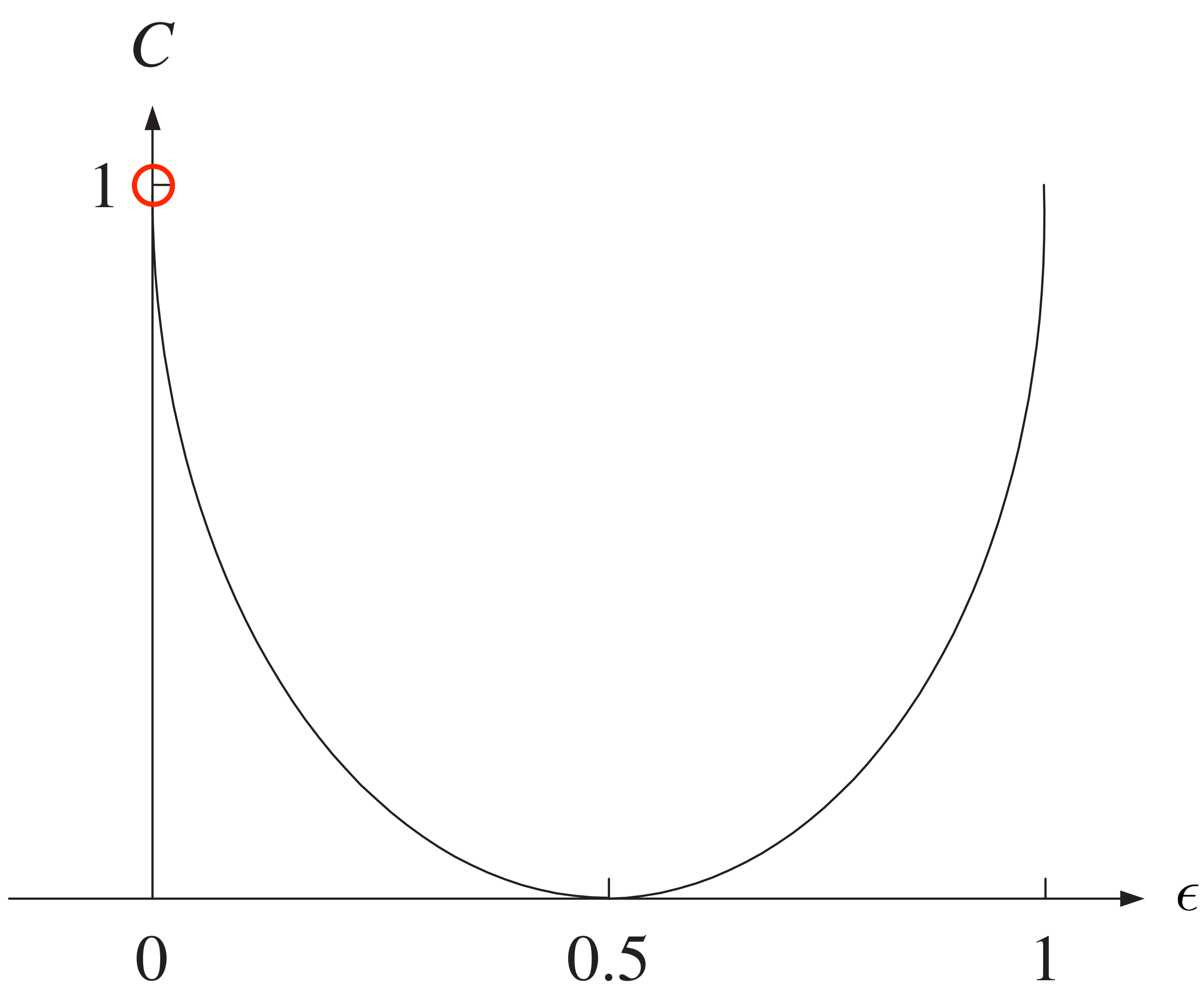
This can be achieved by taking the uniform input distribution.

4. Therefore, $C = 1 - h_b(\epsilon)$ bit per use.

$$C(\epsilon) = 1 - h_b(\epsilon)$$

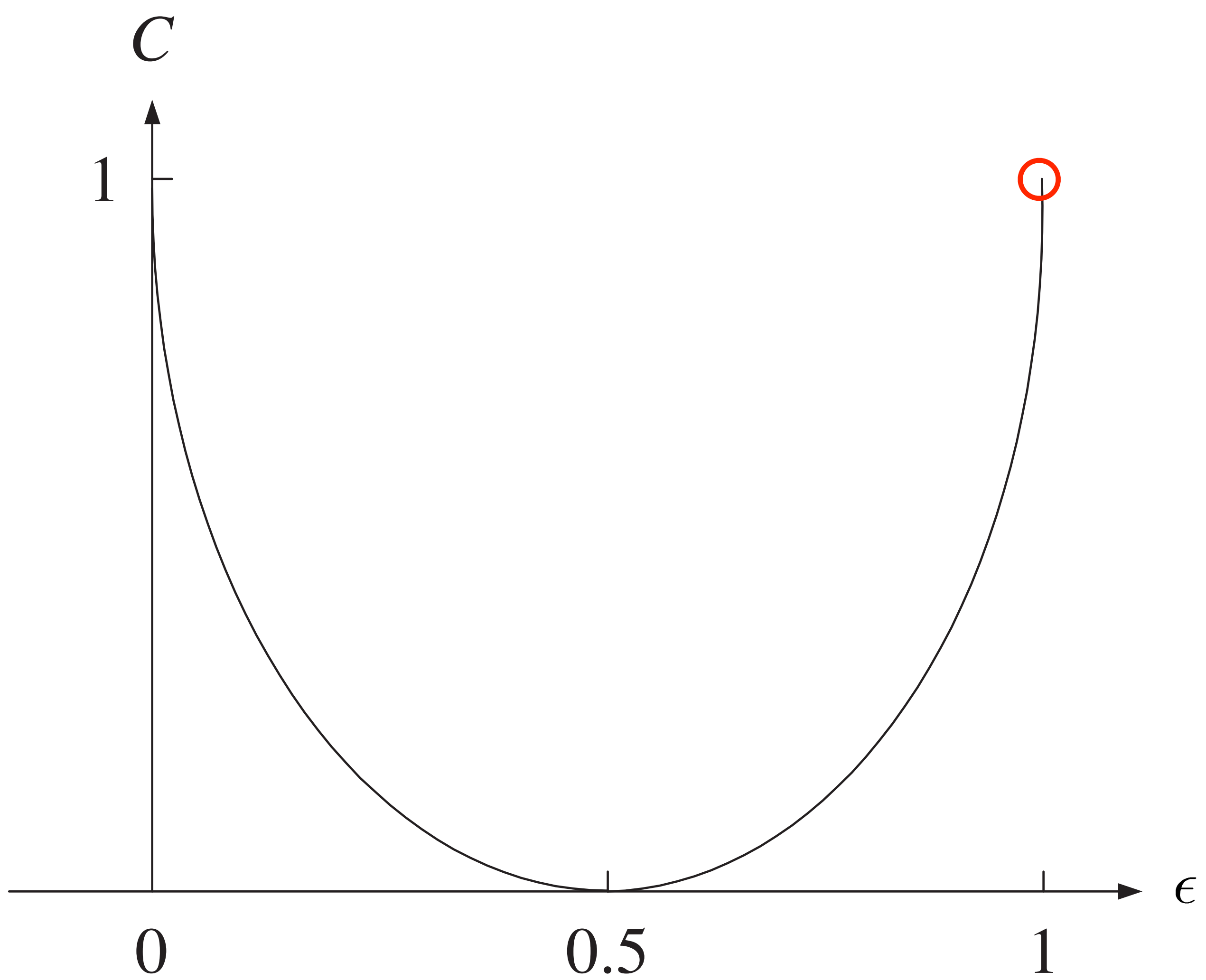


$$C(\epsilon) = 1 - h_b(\epsilon)$$



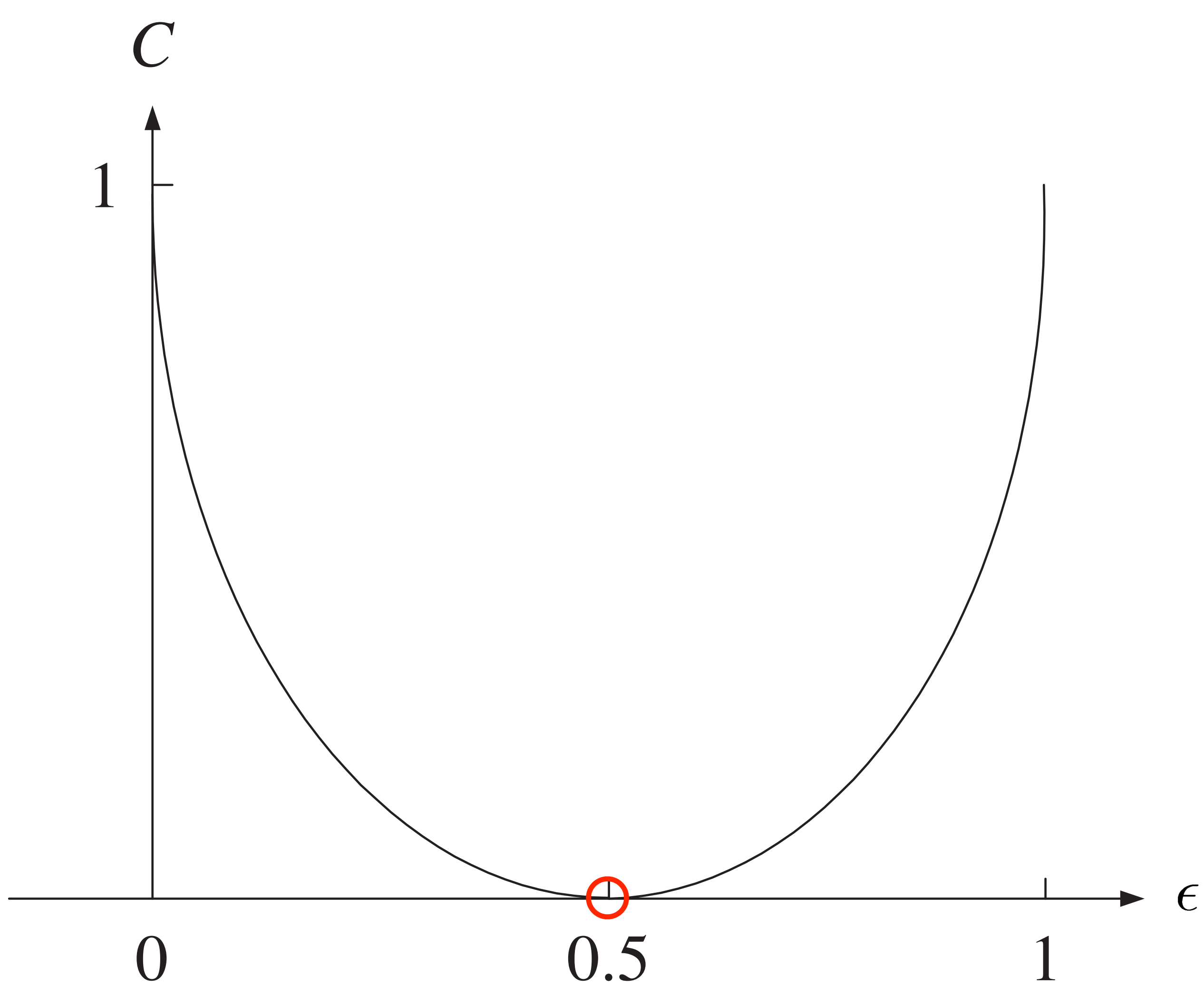
- $C(0) = 1$

$$C(\epsilon) = 1 - h_b(\epsilon)$$



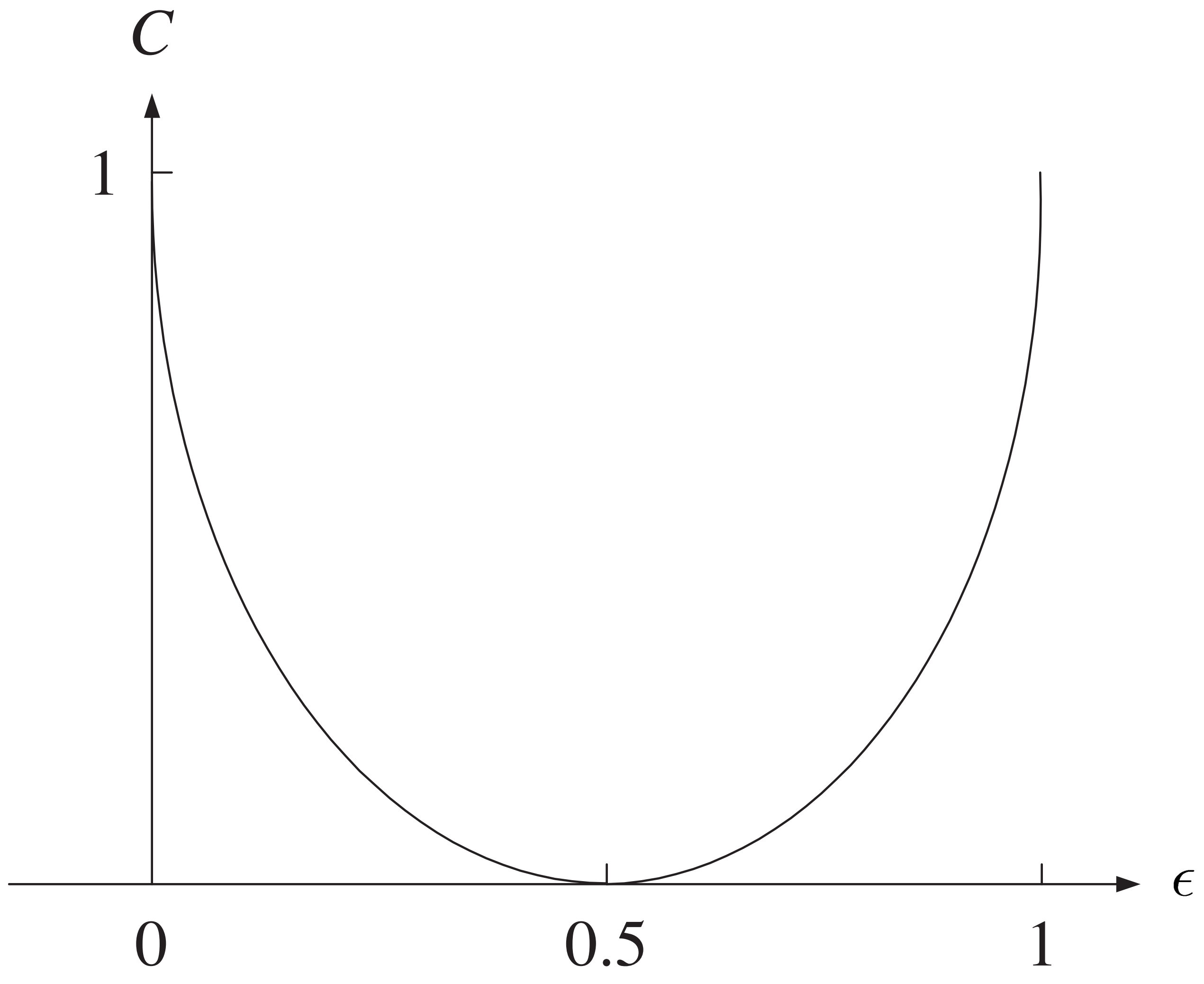
- $C(0) = 1$
- $C(1) = 1$

$$C(\epsilon) = 1 - h_b(\epsilon)$$



- $C(0) = 1$
- $C(1) = 1$
- $C(0.5) = 0$

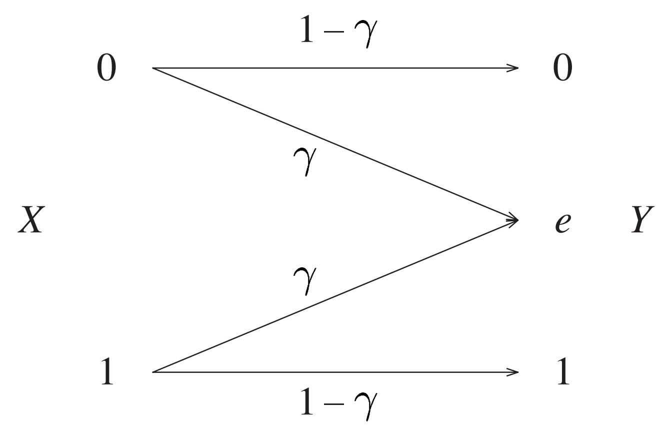
$$C(\epsilon) = 1 - h_b(\epsilon)$$



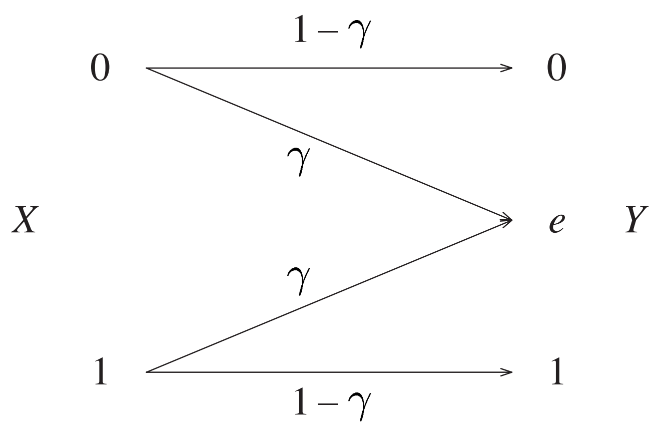
- $C(0) = 1$
- $C(1) = 1$
- $C(0.5) = 0$

Exercise: Show that X and Y are always independent for $\epsilon = 0.5$.

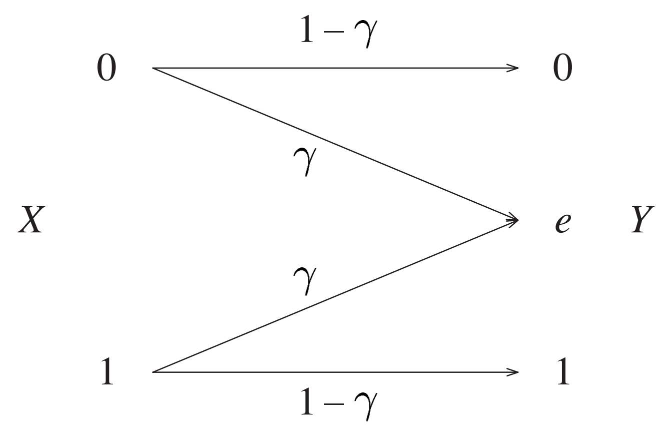
Example 7.8 (Binary Erasure Channel)



Example 7.8 (Binary Erasure Channel)

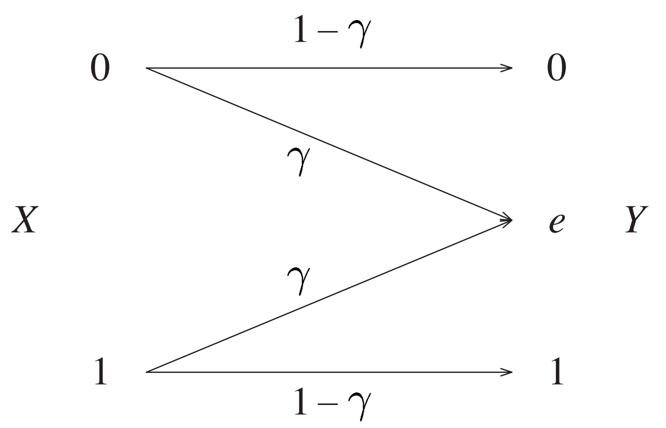


Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

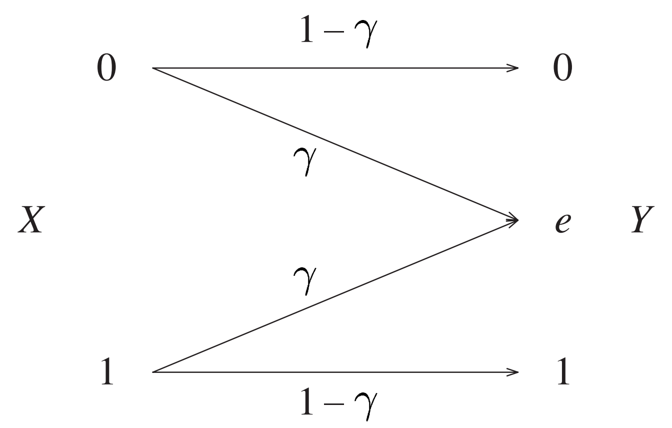
Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes “erasure”.

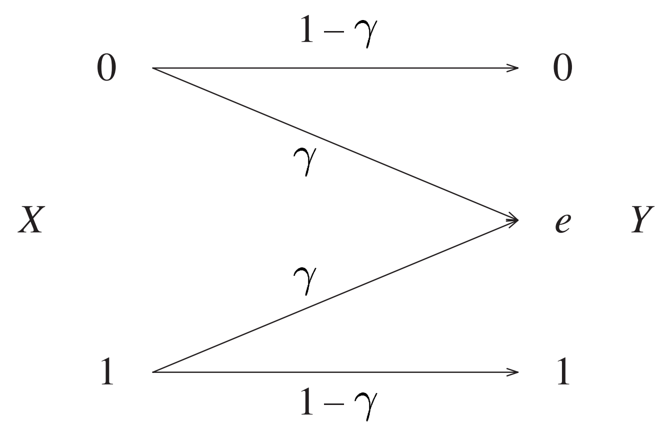
Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$

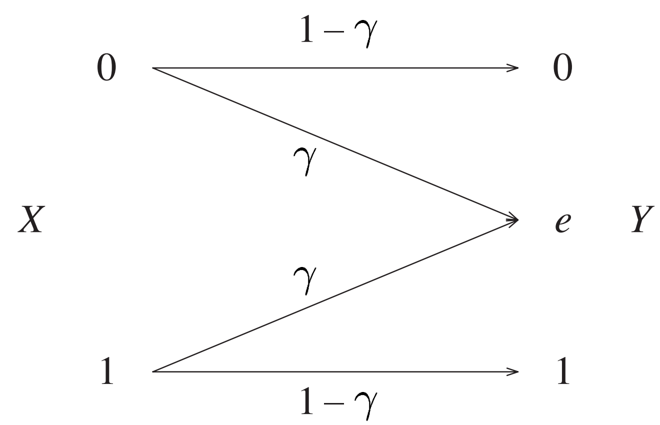
Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).

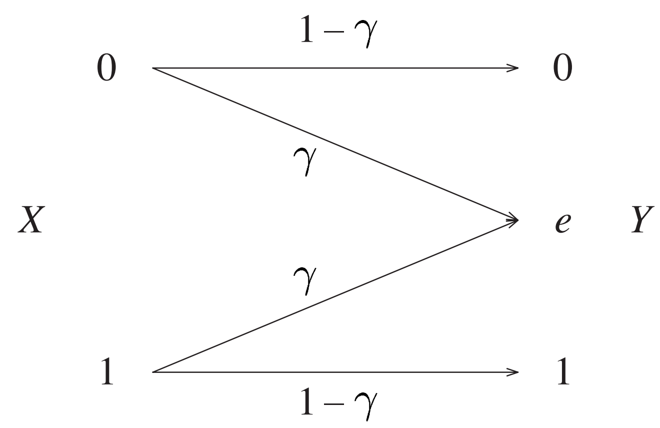
Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).

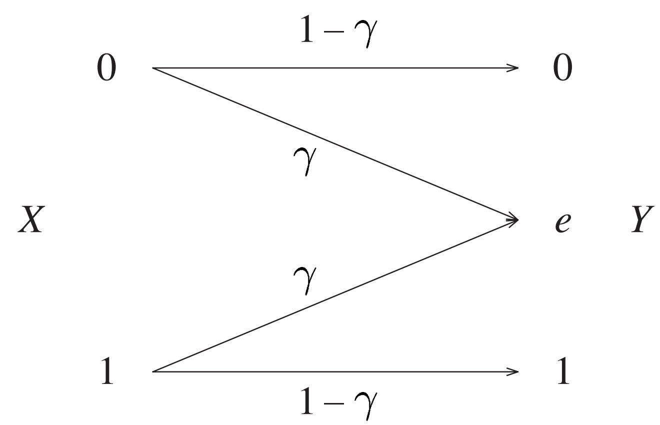
Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes “erasure”.
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

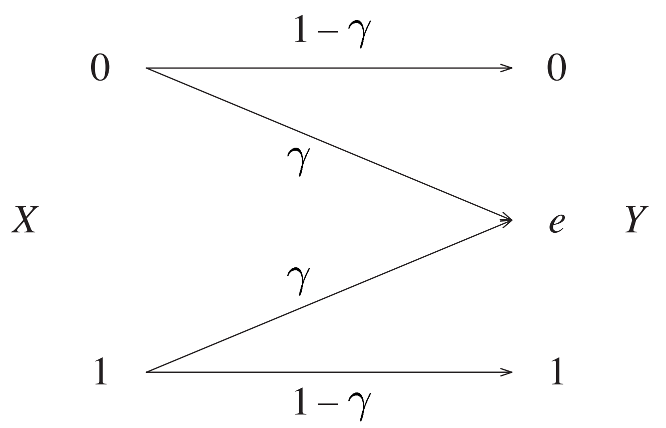
Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Example 7.8 (Binary Erasure Channel)

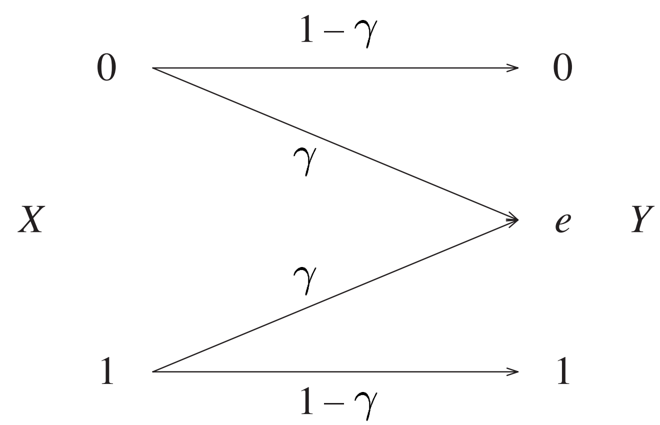


Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

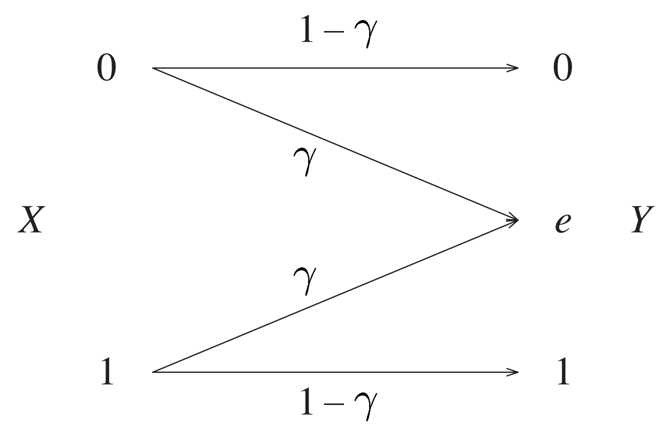
- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

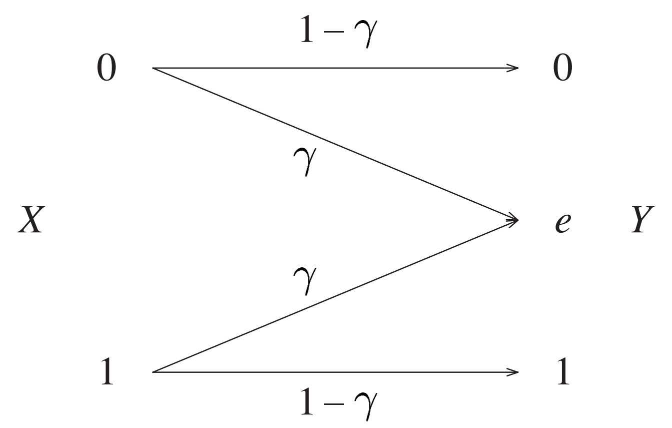
- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

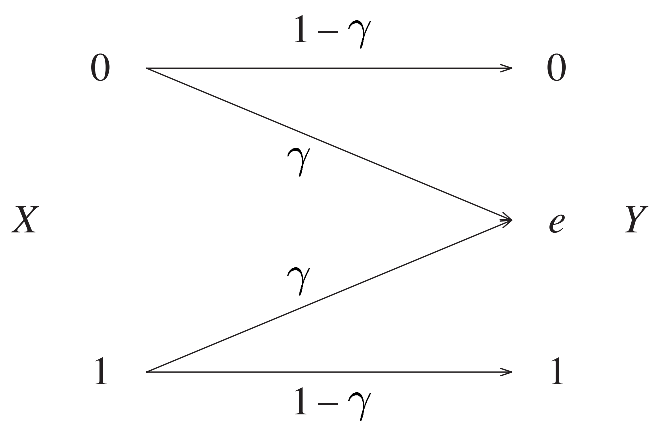
- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

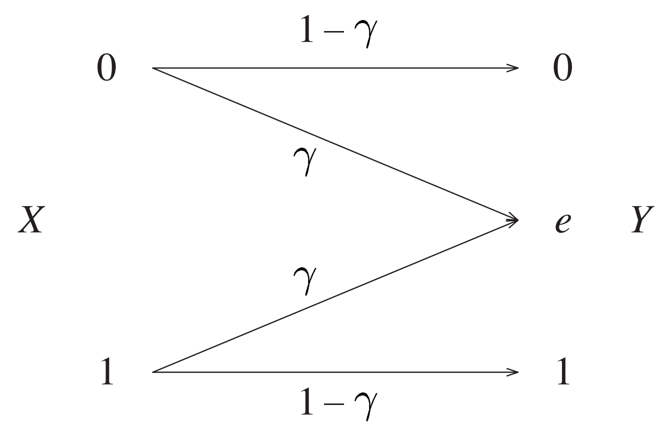
- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

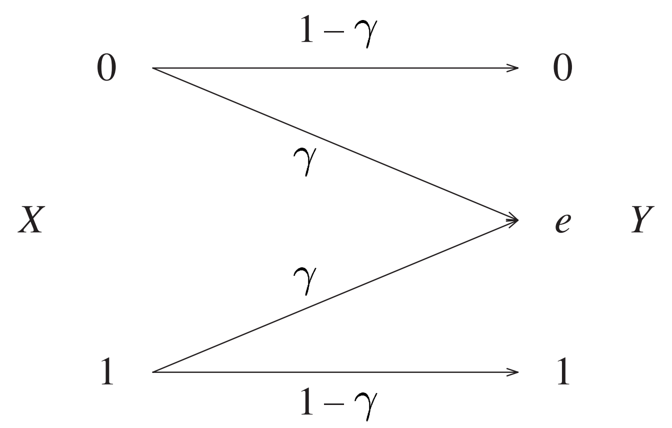
Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

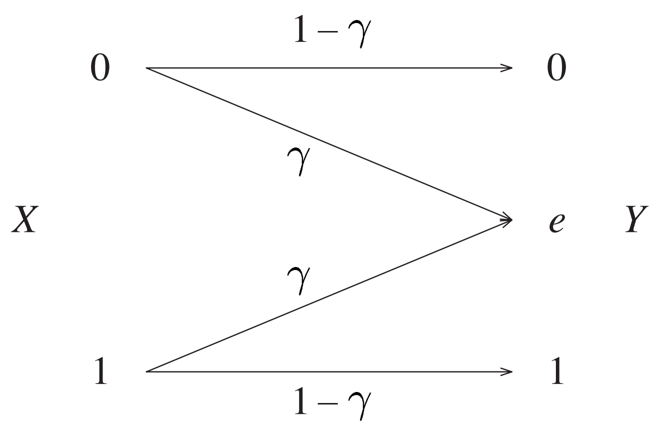
Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

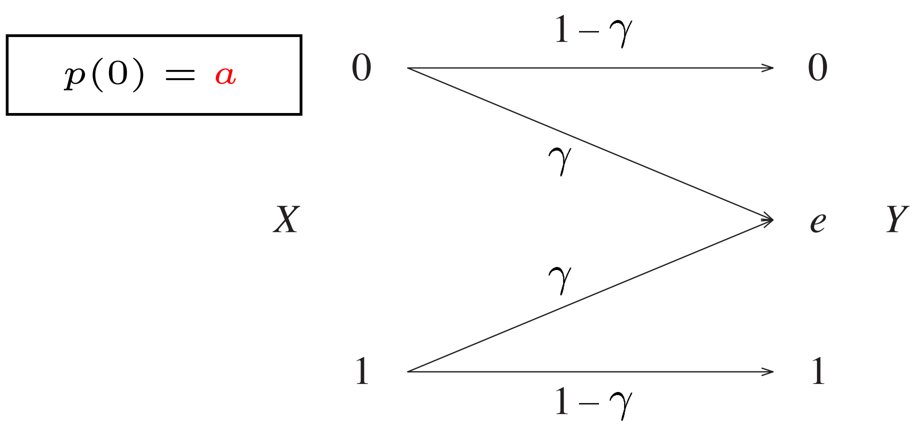
$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} (H(Y) - H(Y|X)) \\
&= \max_{p(x)} H(Y) - h_b(\gamma).
\end{aligned}
$$

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

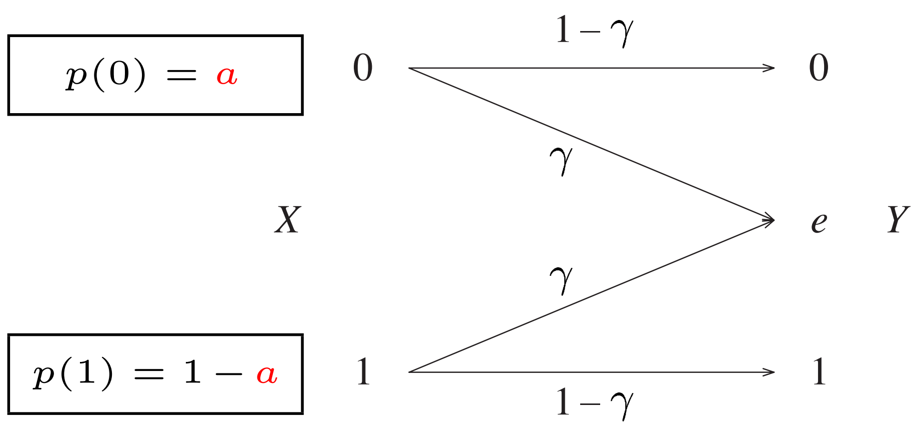
$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} (H(Y) - H(Y|X)) \\
&= \max_{p(x)} H(Y) - h_b(\gamma).
\end{aligned}
$$

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

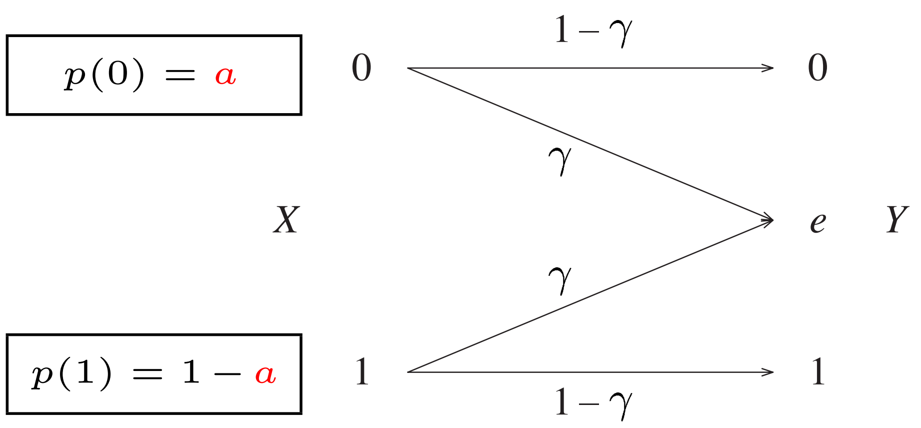
$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} (H(Y) - H(Y|X)) \\ &= \max_{p(x)} H(Y) - h_b(\gamma). \end{aligned}$$

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

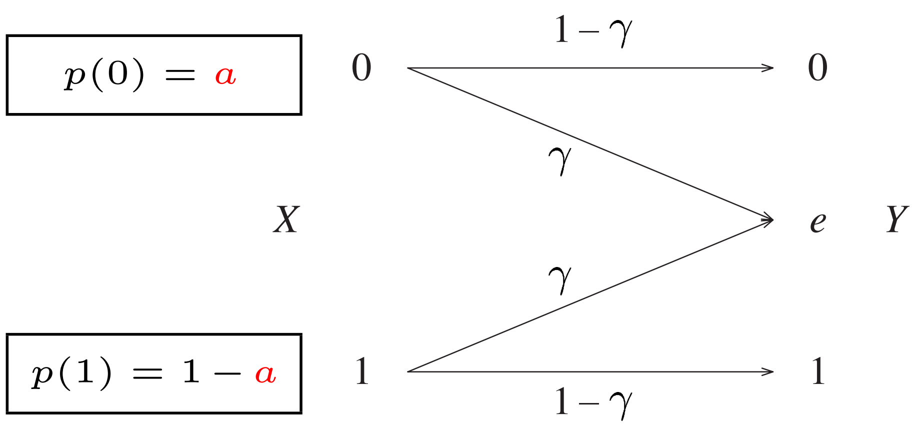
Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

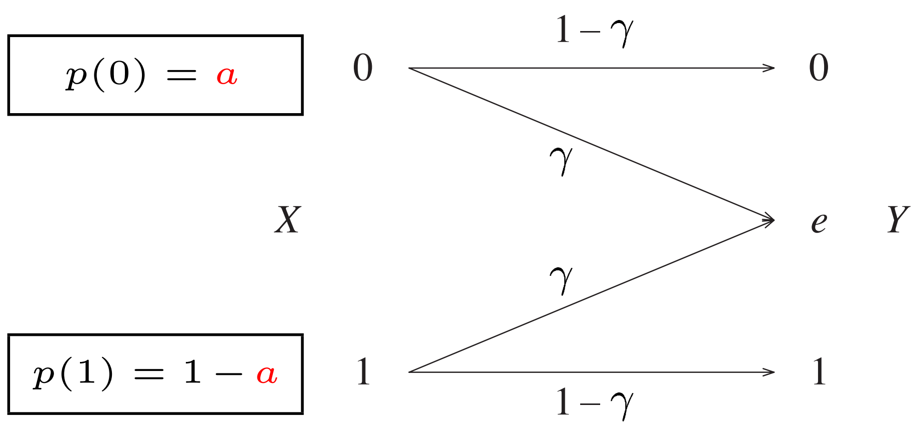
2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

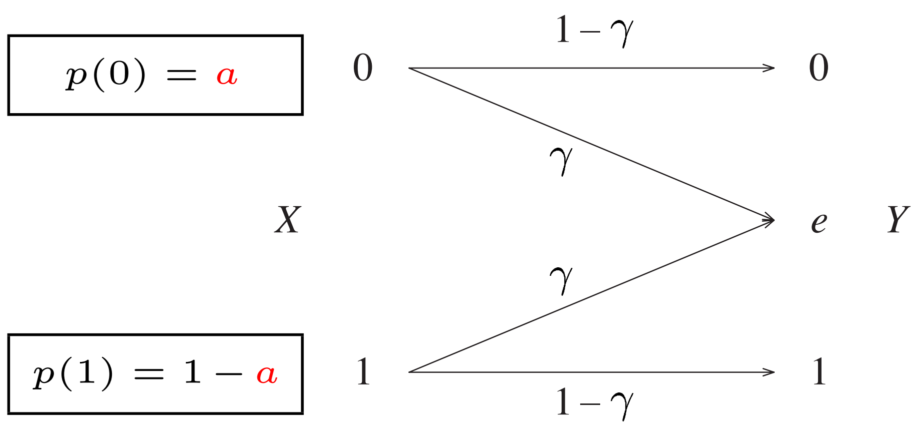
$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

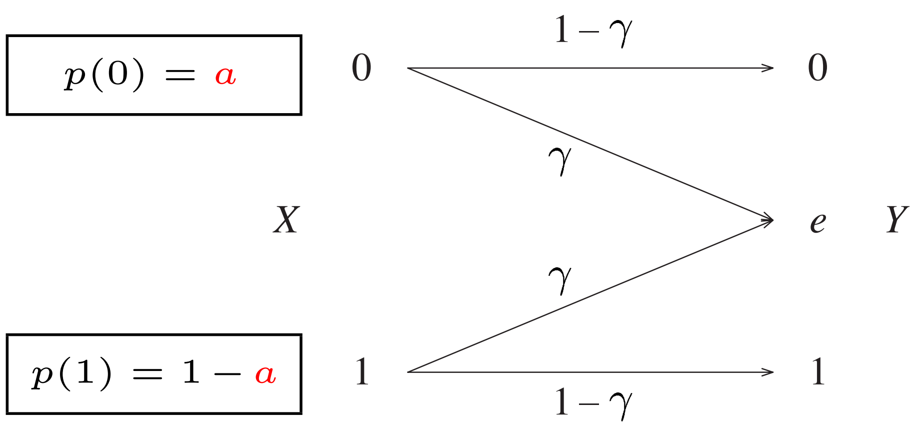
$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

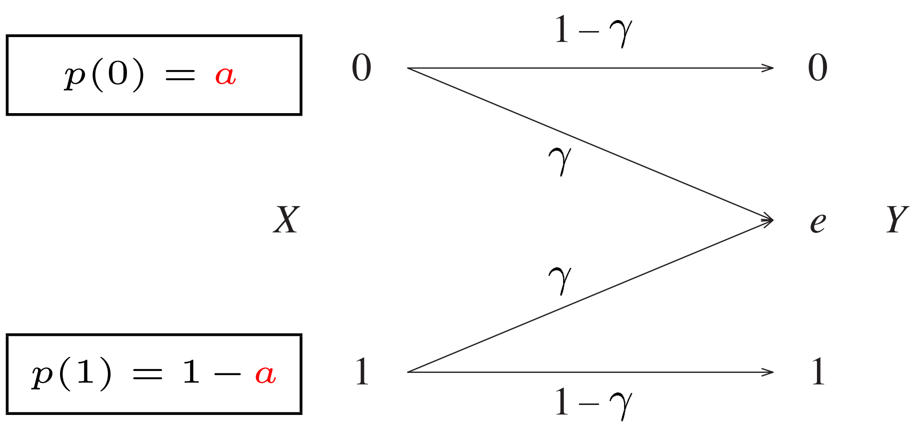
$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

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

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

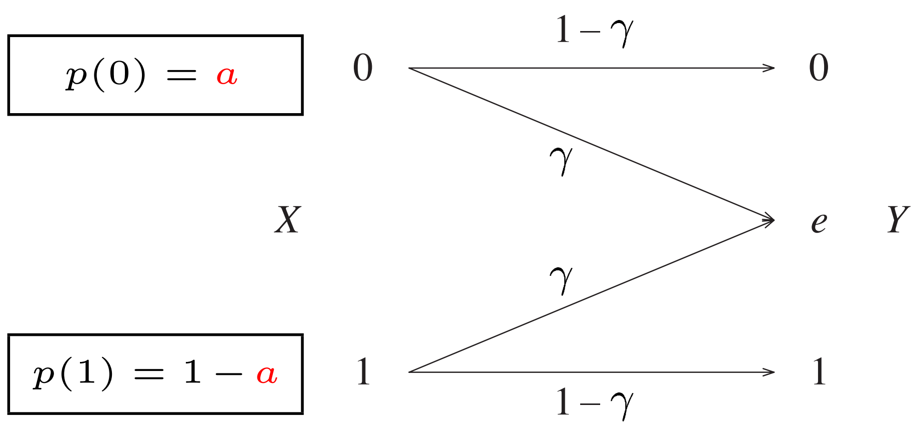
$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

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

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

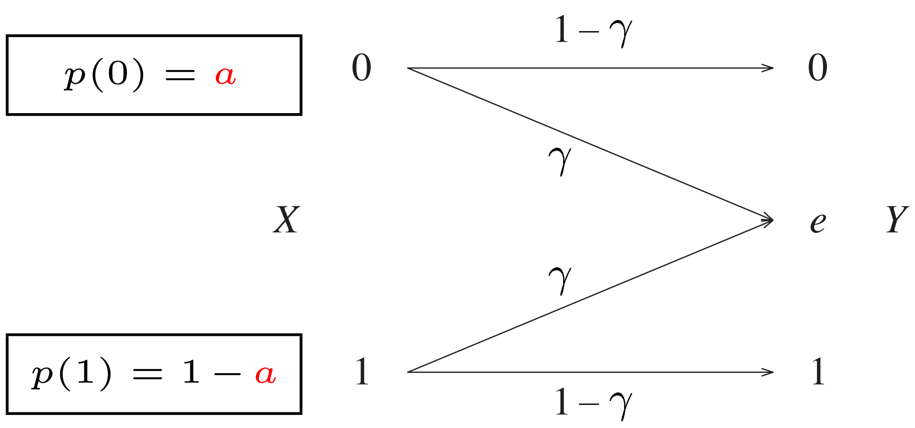
$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

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

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

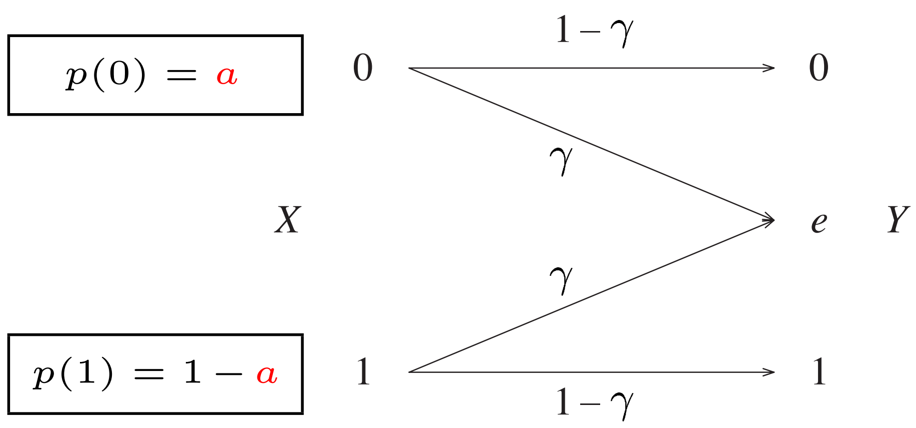
$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

$$\begin{aligned} H(Y) &= H(Y, E) \\ &= \underline{H(E)} + H(Y|E) \\ &= \underline{h_b(\gamma)} + (1-\gamma) h_b(a). \end{aligned}$$

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

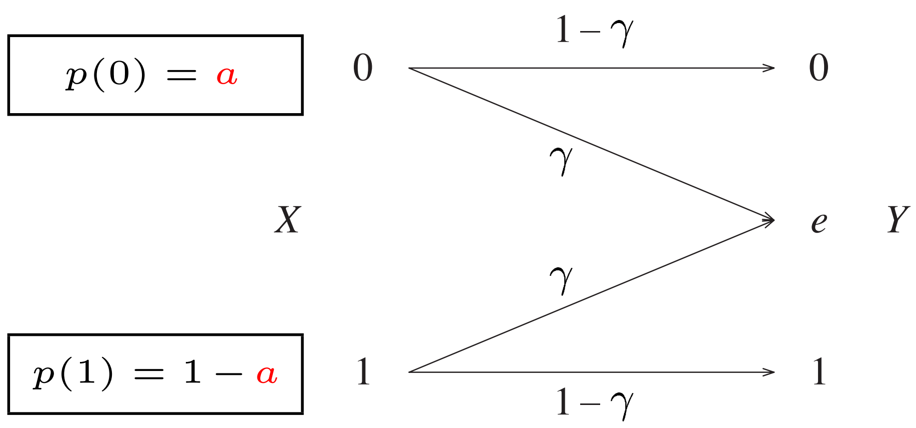
$$E = \begin{cases} 0 & \text{if } Y \ne e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + \underline{H(Y|E)} \\ &= h_b(\gamma) + (1-\gamma) h_b(a). \end{aligned}$$

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

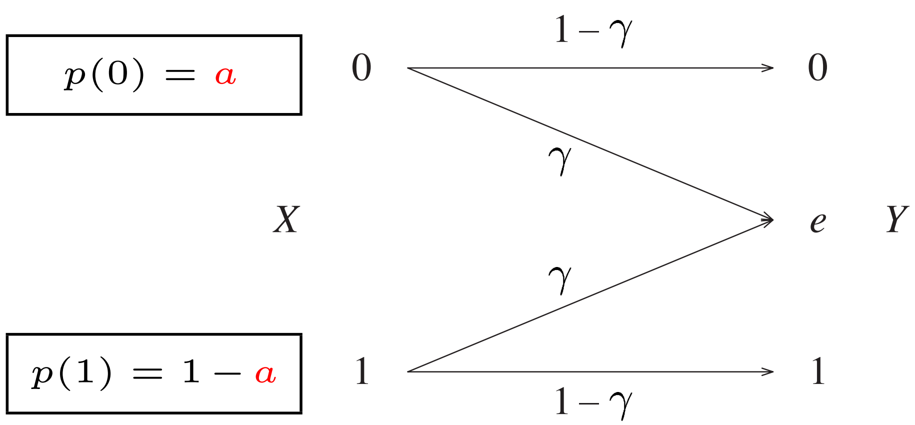
$$
E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}
$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

$$
\begin{aligned}
H(Y) &= H(Y, E) \\
&= H(E) + \underline{H(Y|E)} \\
&= h_b(\gamma) + \underline{(1-\gamma) h_b(a)}.
\end{aligned}
$$

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

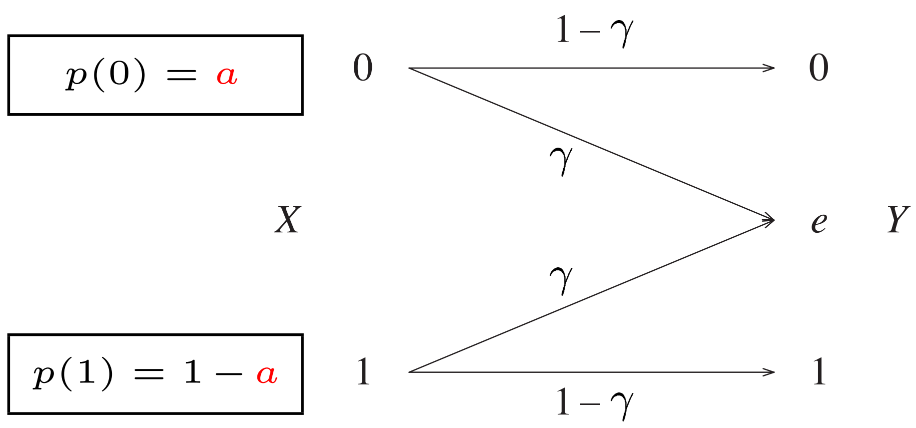
The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + \underline{H(Y|E)} \\ &= h_b(\gamma) + \underline{(1-\gamma)h_b(a)}. \end{aligned}$$

Exercise: Show that $H(Y|E) = (1-\gamma)h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

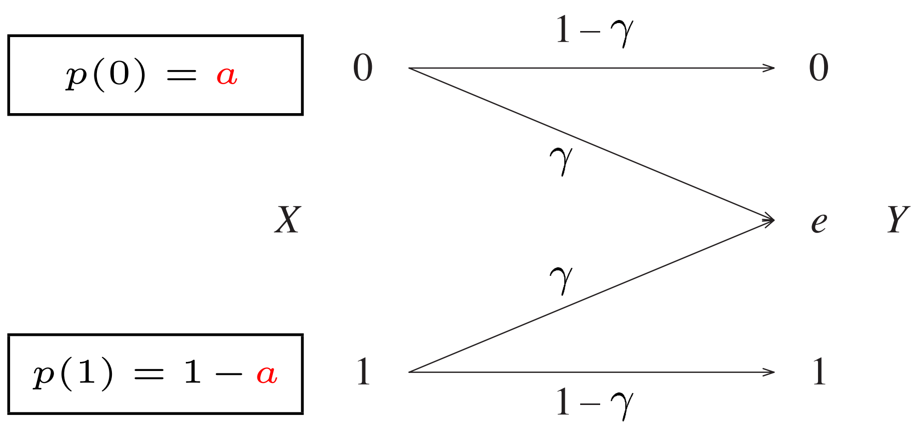
The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1 - \gamma) h_b(a). \end{aligned}$$

Exercise: Show that $H(Y|E) = (1 - \gamma) h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1 - \gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

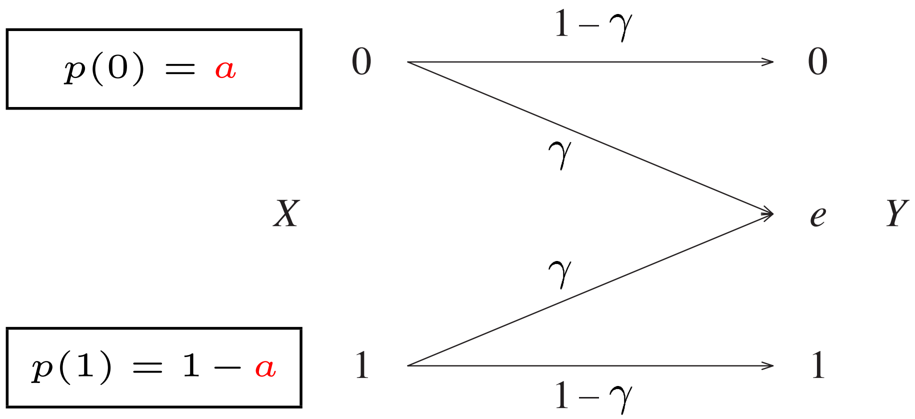
The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1 - \gamma) h_b(a). \end{aligned}$$

Exercise: Show that $H(Y|E) = (1 - \gamma) h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let
$$p(0) = a$$
and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

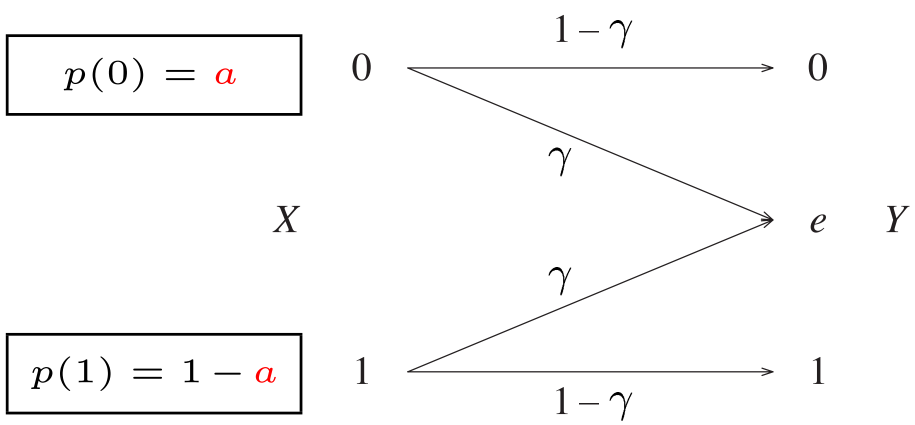
$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1-\gamma)h_b(a). \end{aligned}$$

4. Hence,

$$C = \max_{p(x)} H(Y) - h_b(\gamma)$$

Exercise: Show that $H(Y|E) = (1-\gamma)h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

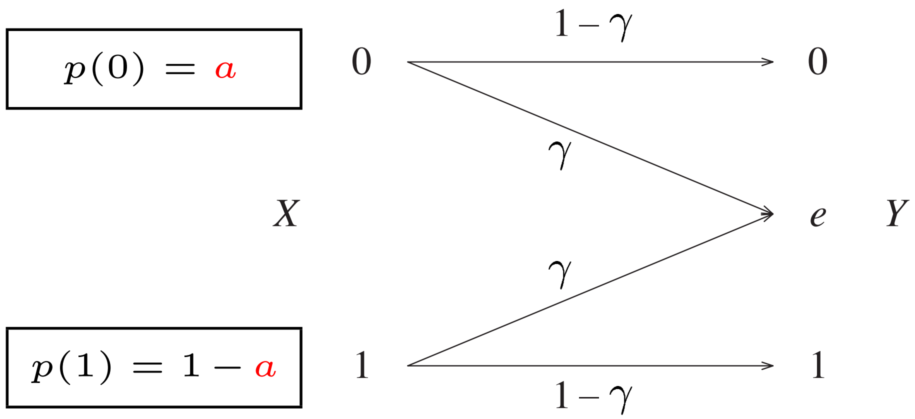
$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1-\gamma) h_b(a). \end{aligned}$$

4. Hence,

$$C = \max_{p(x)} H(Y) - h_b(\gamma)$$

Exercise: Show that $H(Y|E) = (1-\gamma)h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

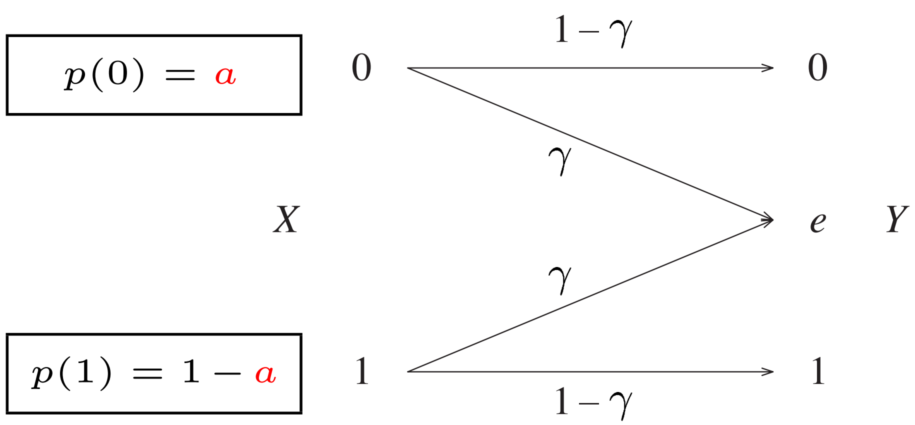
$$
\begin{aligned}
H(Y) &= H(Y, E) \\
&= H(E) + H(Y|E) \\
&= h_b(\gamma) + (1-\gamma) h_b(a).
\end{aligned}
$$

4. Hence,

$$
\begin{aligned}
C &= \max_{p(x)} H(Y) - h_b(\gamma) \\
&= \max_{a} [h_b(\gamma) + (1-\gamma) h_b(a)] - h_b(\gamma)
\end{aligned}
$$

Exercise: Show that $H(Y|E) = (1-\gamma) h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \ne e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

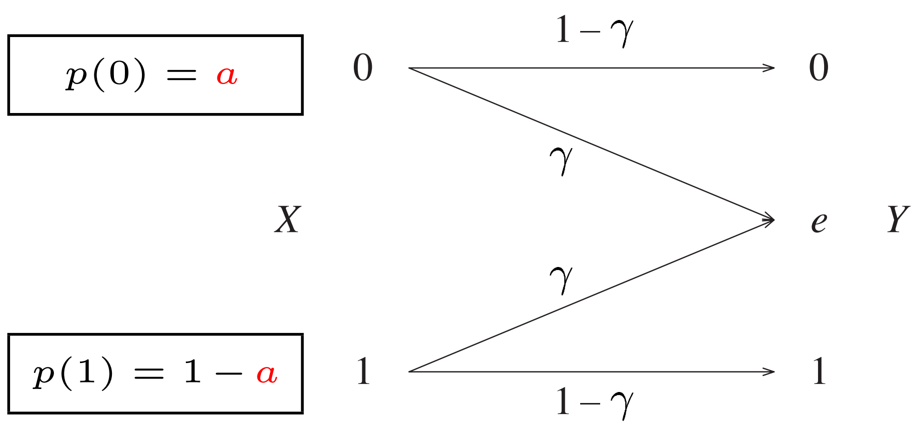
$$\begin{aligned} H(Y) &= H(Y,E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1-\gamma)h_b(a). \end{aligned}$$

4. Hence,

$$\begin{aligned} C &= \max_{p(x)} H(Y) - h_b(\gamma) \\ &= \max_{a}[h_b(\gamma) + (1-\gamma)h_b(a)] - h_b(\gamma) \end{aligned}$$

Exercise: Show that $H(Y|E) = (1-\gamma)h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

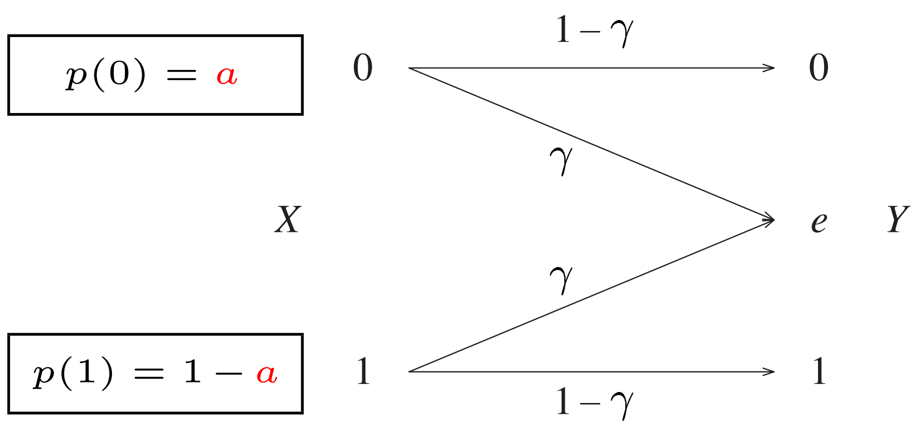
$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1-\gamma)h_b(a). \end{aligned}$$

4. Hence,

$$\begin{aligned} C &= \max_{p(x)} H(Y) - h_b(\gamma) \\ &= \max_{a} [h_b(\gamma) + (1-\gamma)h_b(a)] - h_b(\gamma) \end{aligned}$$

Exercise: Show that $H(Y|E) = (1-\gamma)h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \le \gamma \le 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

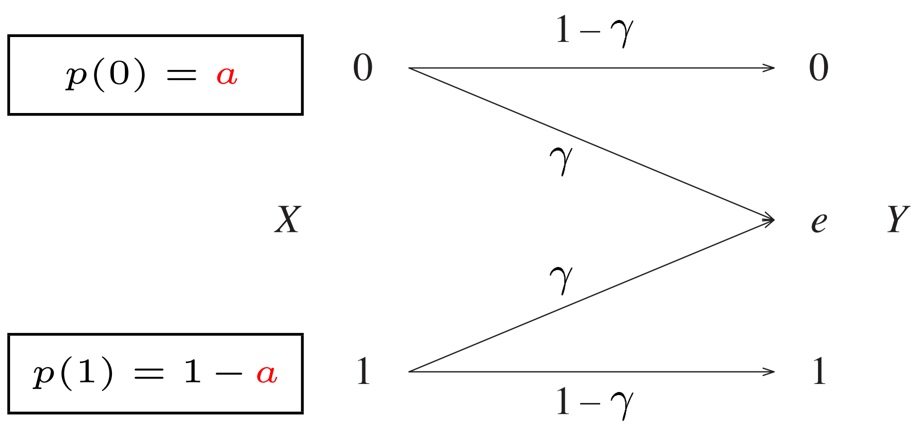
$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1-\gamma) h_b(a). \end{aligned}$$

4. Hence,

$$\begin{aligned} C &= \max_{p(x)} H(Y) - h_b(\gamma) \\ &= \max_{a} [\cancel{h_b(\gamma)} + (1-\gamma) h_b(a)] - \cancel{h_b(\gamma)} \\ &= (1-\gamma) \max_{a} h_b(a) \end{aligned}$$

Exercise: Show that $H(Y|E) = (1-\gamma) h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

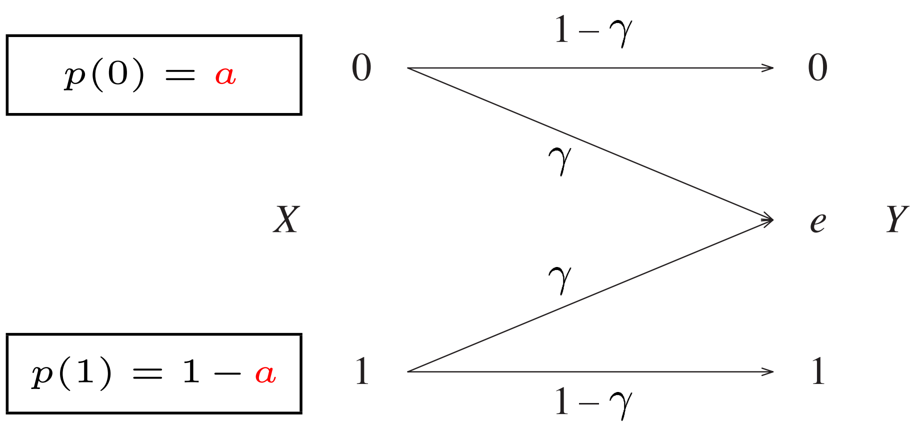
$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1-\gamma) h_b(a). \end{aligned}$$

4. Hence,

$$\begin{aligned} C &= \max_{p(x)} H(Y) - h_b(\gamma) \\ &= \max_{a} [h_b(\gamma) + (1-\gamma) h_b(a)] - h_b(\gamma) \\ &= (1-\gamma) \max_{a} h_b(a) \\ &= (1-\gamma) \text{ bit per use,} \end{aligned}$$

Exercise: Show that $H(Y|E) = (1-\gamma)h_b(a)$.

Example 7.8 (Binary Erasure Channel)



Binary Erasure Channel:

- The output symbol e denotes "erasure".
- γ : Erasure probability, $0 \leq \gamma \leq 1$
- With probability $1-\gamma$, $Y = X$ (regardless of whether $X = 0$ or $X = 1$).
- With probability γ, $Y = e$ (erasure).
- If $X = 0$, Y cannot be 1. If $X = 1$, Y cannot be 0.

Determination of C

1. Consider

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

Thus we only have to maximize $H(Y)$ over all $p(x)$.

2. Let

$$p(0) = a$$

and define a binary random variable E by

$$E = \begin{cases} 0 & \text{if } Y \neq e \\ 1 & \text{if } Y = e. \end{cases}$$

The random variable E indicates whether an erasure has occurred, and it is a function of Y.

3. Then

$$\begin{aligned} H(Y) &= H(Y, E) \\ &= H(E) + H(Y|E) \\ &= h_b(\gamma) + (1-\gamma)h_b(a). \end{aligned}$$

4. Hence,

$$\begin{aligned} C &= \max_{p(x)} H(Y) - h_b(\gamma) \\ &= \max_{a} [\cancel{h_b(\gamma)} + (1-\gamma)h_b(a)] - \cancel{h_b(\gamma)} \\ &= (1-\gamma) \max_{a} h_b(a) \\ &= (1-\gamma) \text{ bit per use,} \end{aligned}$$

where the capacity is achieved by letting $a = 0.5$, i.e., the input distribution is uniform.

Exercise: Show that $H(Y|E) = (1-\gamma)h_b(a)$.