香港中文大學
The Chinese University of Hong Kong

3.6 Examples of Applications

- To obtain information identities is WYSIWYG.

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

    because

$$\mu^*(A) \quad \leq \quad \mu^*(A) + \mu^*(B - A)$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

because

$$\mu^*(A) \quad \leq \quad \mu^*(A) + \underline{\mu^*(B - A)}$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \le \mu^*(B)$$

because

$$\mu^*(A) \quad \le \quad \mu^*(A) + \mu^*(B - A)$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

because

$$\mu^*(A) \quad \leq \quad \mu^*(\underline{A}) + \mu^*(\underline{B - A})$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

    because

$$\begin{aligned} \mu^*(A) &\leq \mu^*(A) + \mu^*(B - A) \\ &= \mu^*(\underline{A \cup (B - A)}) \end{aligned}$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \le \mu^*(B)$$

because

$$\begin{aligned} \mu^*(A) &\le \mu^*(A) + \mu^*(B - A) \\ &= \mu^*(A \cup (B - A)) \\ &= \mu^*(B) \end{aligned}$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

because

$$\begin{aligned} \mu^*(A) &\leq \mu^*(A) + \mu^*(B - A) \\ &= \mu^*(A \cup (B - A)) \\ &= \mu^*(B) \end{aligned}$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

    because

$$\begin{aligned} \mu^*(A) &\leq \mu^*(A) + \mu^*(B - A) \\ &= \mu^*(A \cup (B - A)) \\ &= \mu^*(B) \end{aligned}$$

- To obtain information identities is WYSIWYG.
- To obtain information inequalities:
  - If μ^* is nonnegative, then

$$A \subset B \quad \Rightarrow \quad \mu^*(A) \leq \mu^*(B)$$

    because

$$\begin{aligned} \mu^*(A) &\leq \mu^*(A) + \mu^*(B - A) \\ &= \mu^*(A \cup (B - A)) \\ &= \mu^*(B) \end{aligned}$$

  - If μ^* is a signed measure, need to invoke the basic inequalities to compare $\mu^*(A)$ and $\mu^*(B)$.

Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2).$$

Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \qquad (1)$$

1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

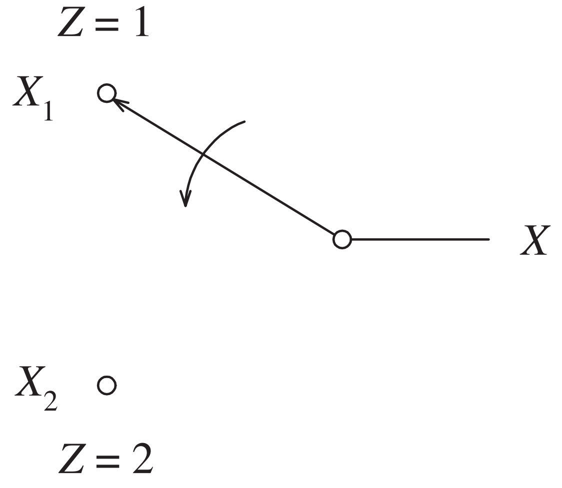
Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

1. Consider the system as shown in which the position of the switch is determined by a random variable Z with



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

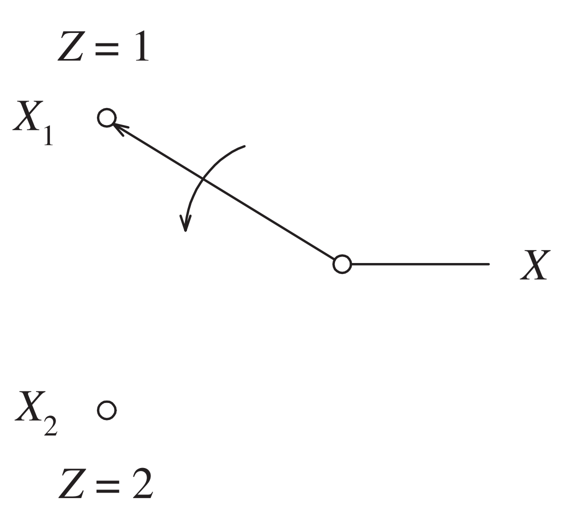
$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

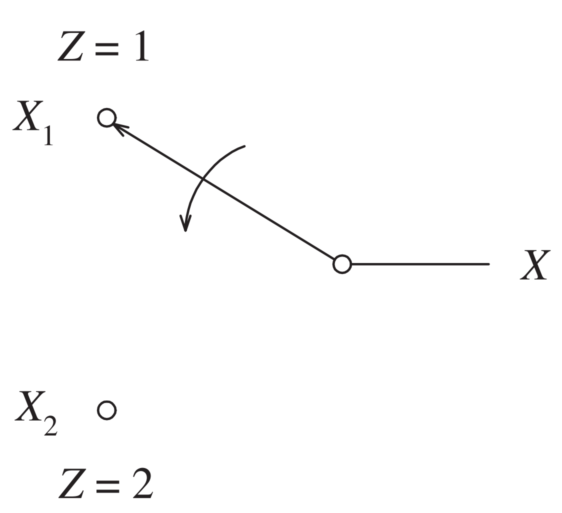
where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

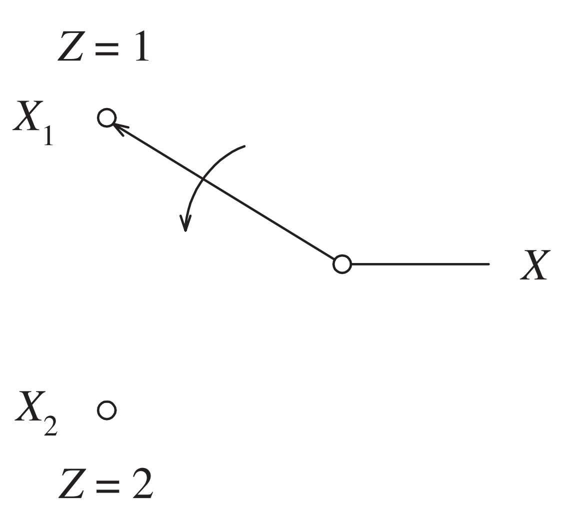
$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$.



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

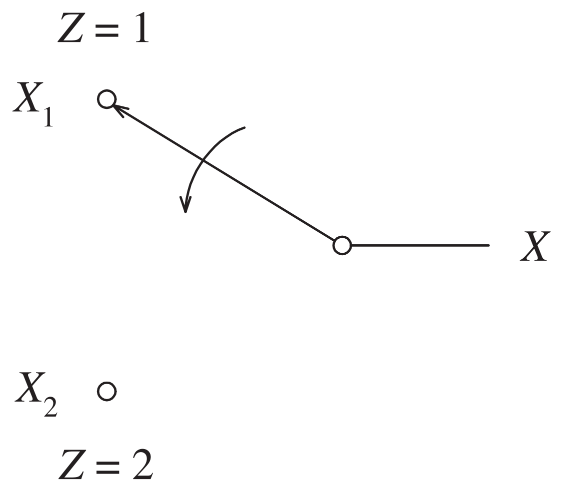
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

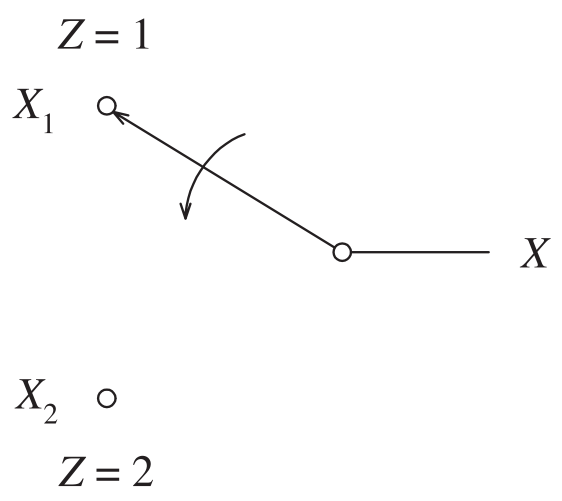
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

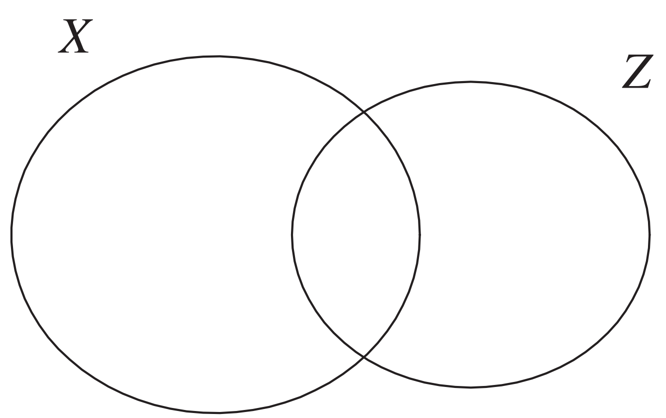
where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

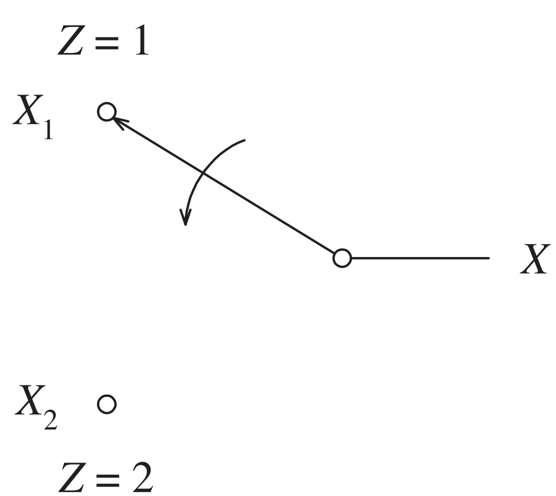
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

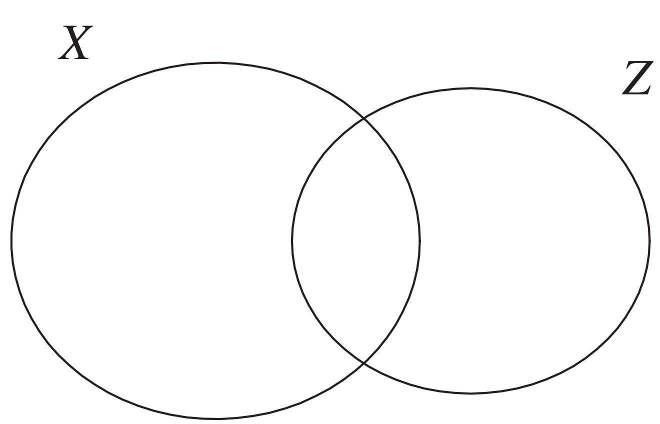
2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

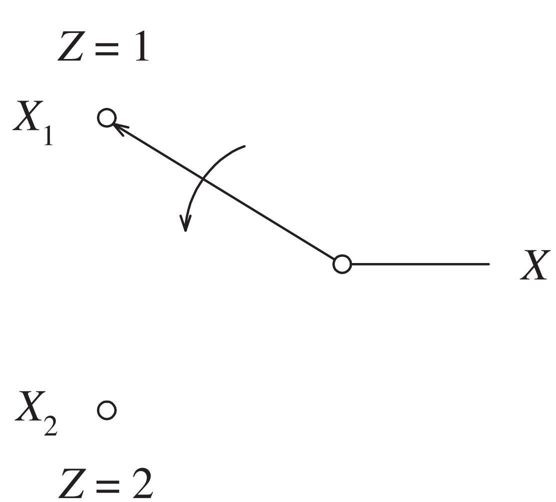
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

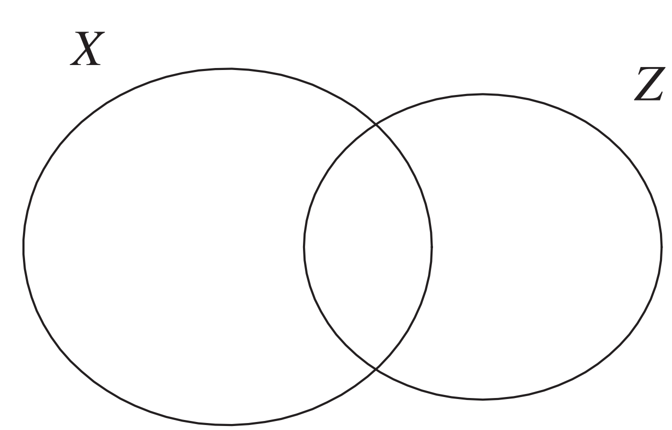
$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

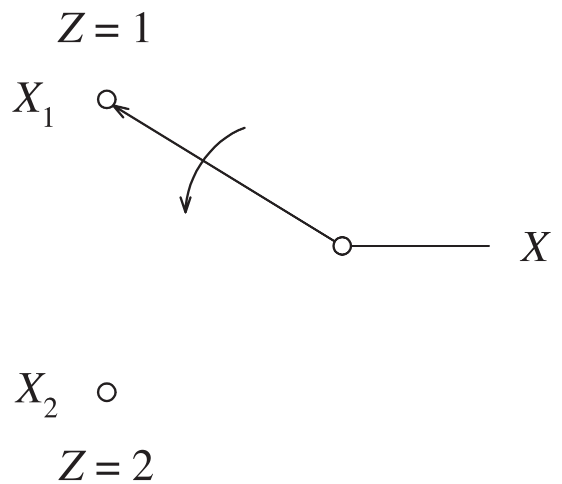
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$

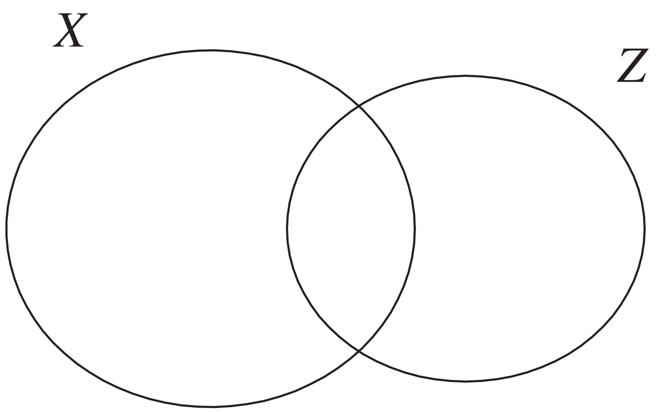


3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \le \lambda \le 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \ge \lambda H(X_1) + \bar{\lambda} H(X_2). \qquad (1)$$

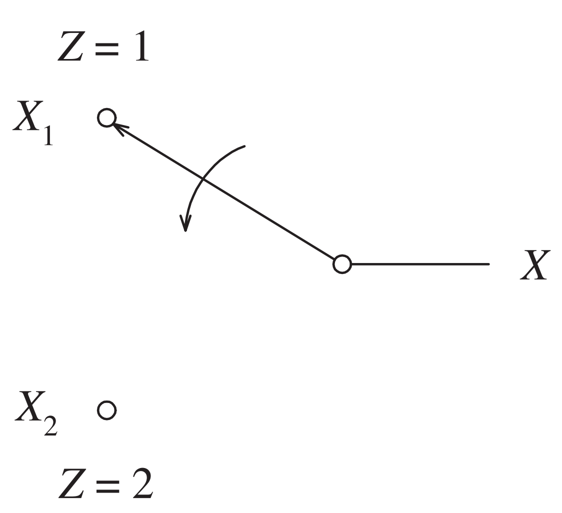
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

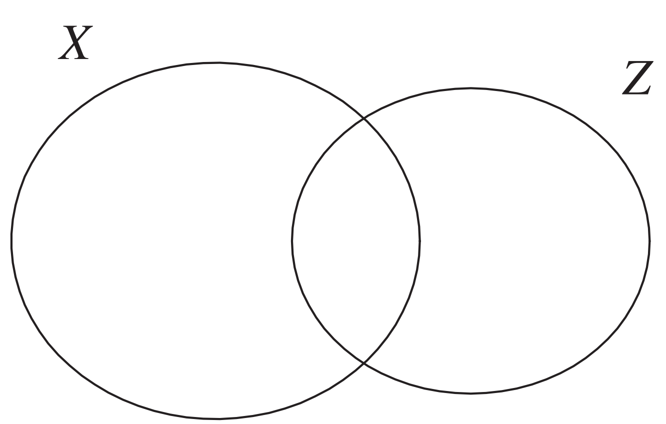
$$\mu^*(\tilde{X}) \ge \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \ge H(X|Z).$$

4. Then

$$H(X) \ge H(X|Z)$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

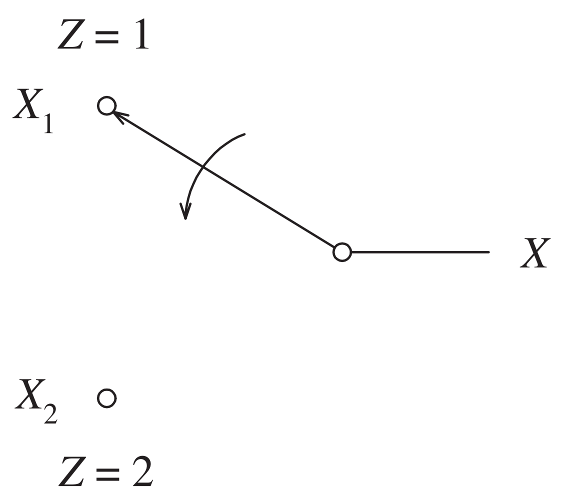
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

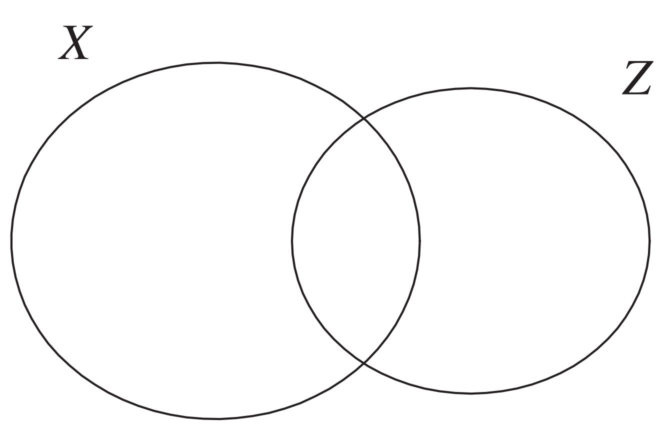
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \geq \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\} H(X|Z = 1) + \Pr\{Z = 2\} H(X|Z = 2) \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

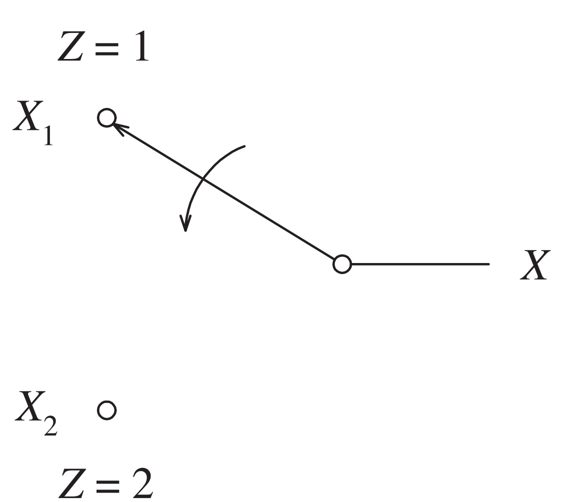
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

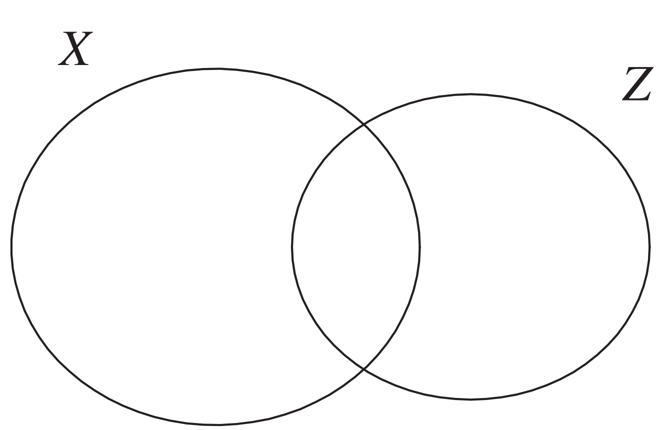
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} &H(X) \\ &\quad \geq \quad H(X|Z) \\ &\quad = \quad \underline{\Pr\{Z = 1\}} H(X|Z = 1) + \Pr\{Z = 2\} H(X|Z = 2) \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

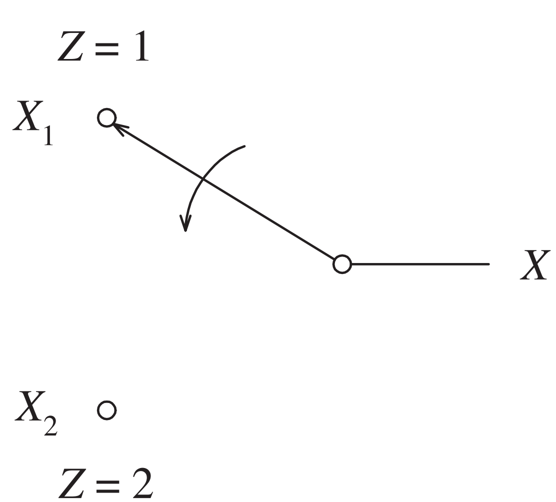
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

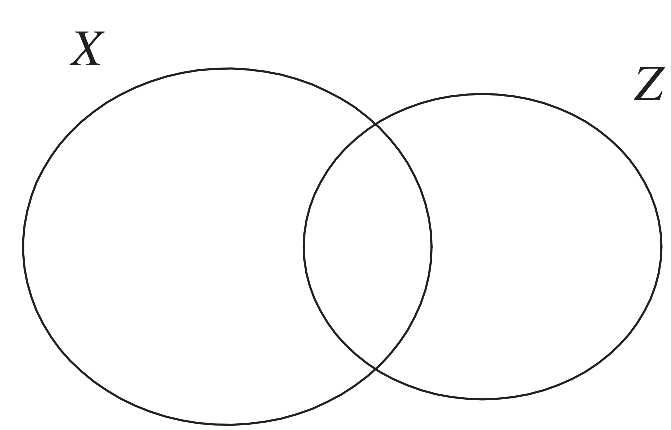
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \geq \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\} H(X|Z = 1) + \Pr\{Z = 2\} H(X|Z = 2) \\ = \quad & \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \le \lambda \le 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \ge \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

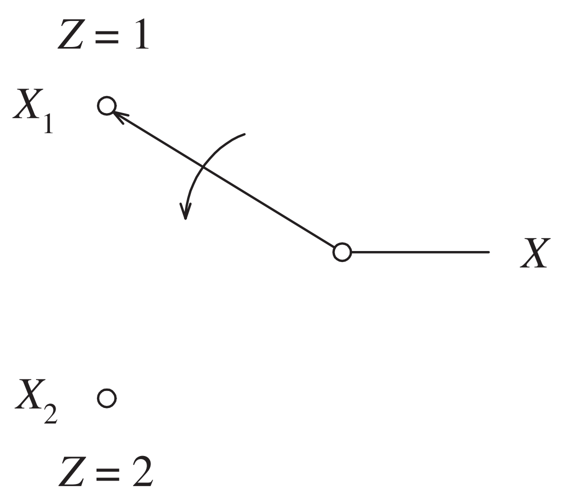
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

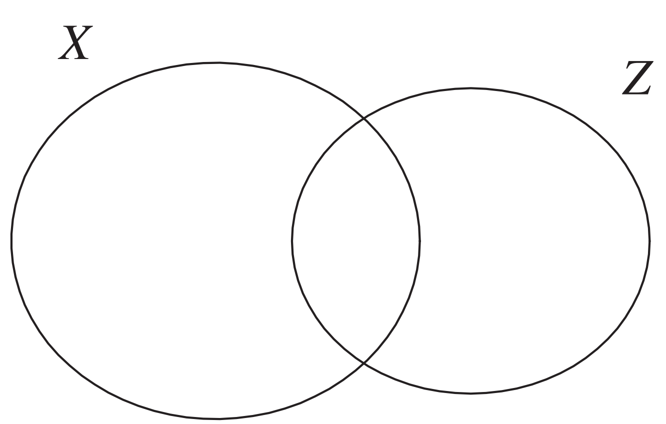
$$\mu^*(\tilde{X}) \ge \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \ge H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \ge \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\} H(X|Z = 1) + \Pr\{Z = 2\} H(X|Z = 2) \\ = \quad & \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

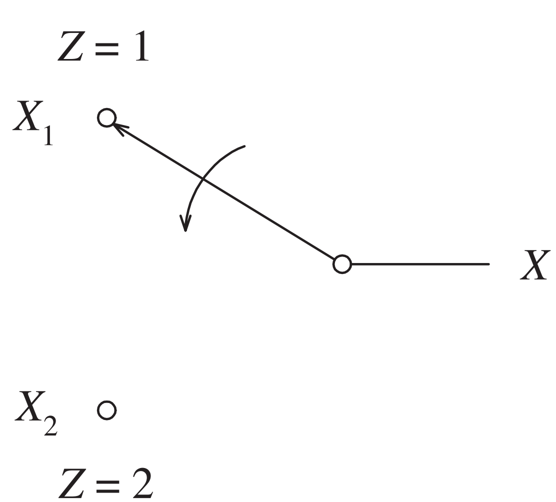
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

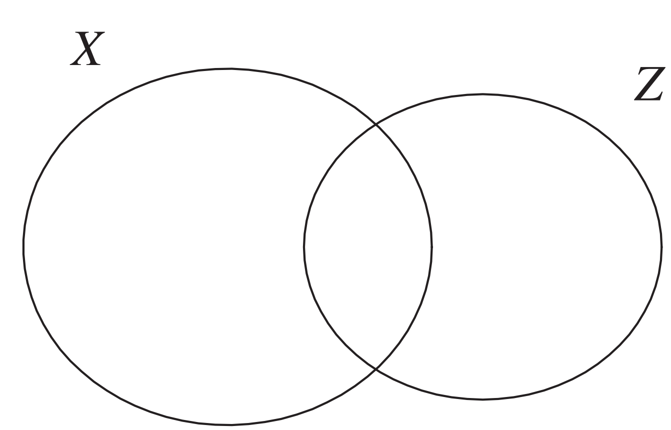
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ &\geq H(X|Z) \\ &= \Pr\{Z = 1\} H(X|Z = 1) + \Pr\{Z = 2\} H(X|Z = 2) \\ &= \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \tag{1}$$

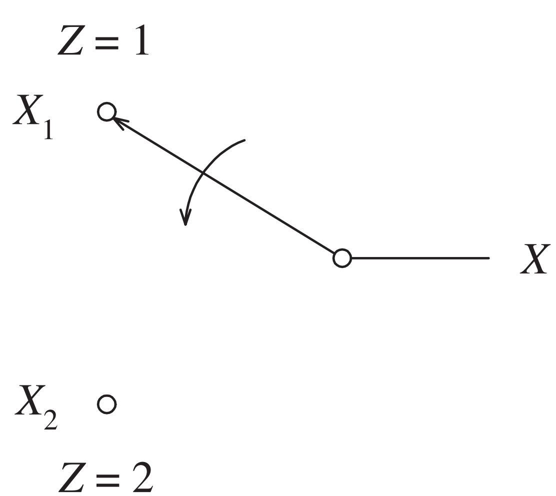
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

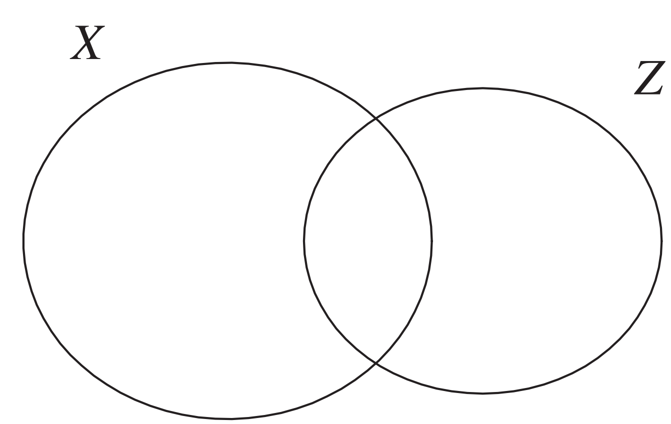
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \geq \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\}\underline{H(X|Z = 1)} + \Pr\{Z = 2\}H(X|Z = 2) \\ = \quad & \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \qquad (1)$$

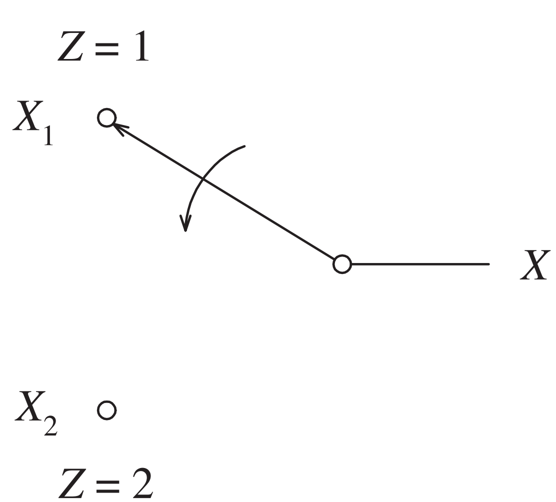
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

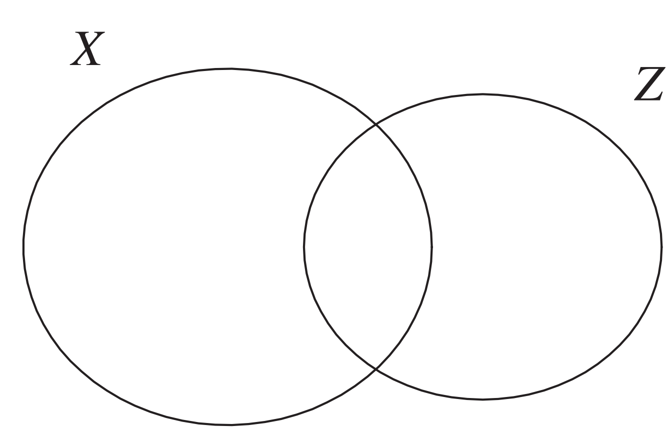
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \geq \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\}\underline{H(X|Z = 1)} + \Pr\{Z = 2\}H(X|Z = 2) \\ = \quad & \lambda \underline{H(X_1)} + \bar{\lambda} H(X_2), \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

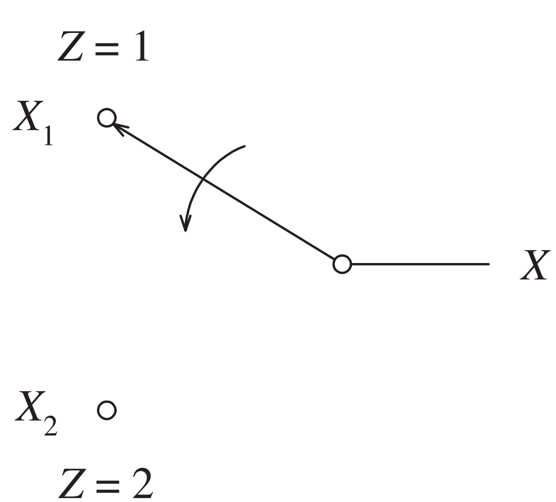
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

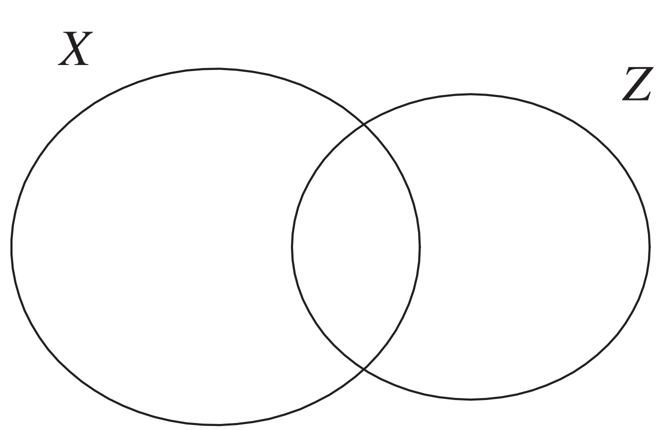
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \geq \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\}H(X|Z = 1) + \Pr\{Z = 2\}H(X|Z = 2) \\ = \quad & \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

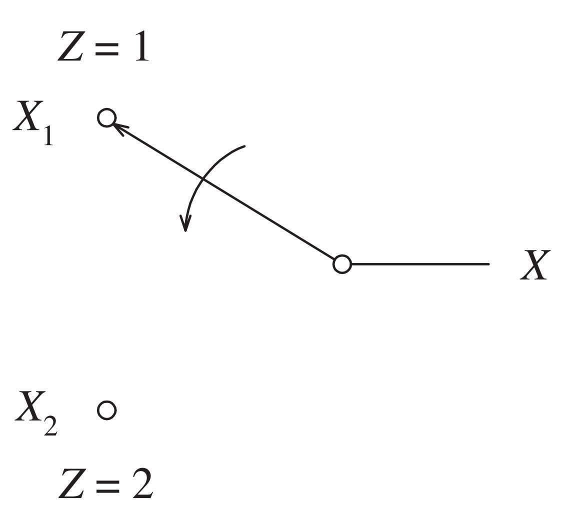
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

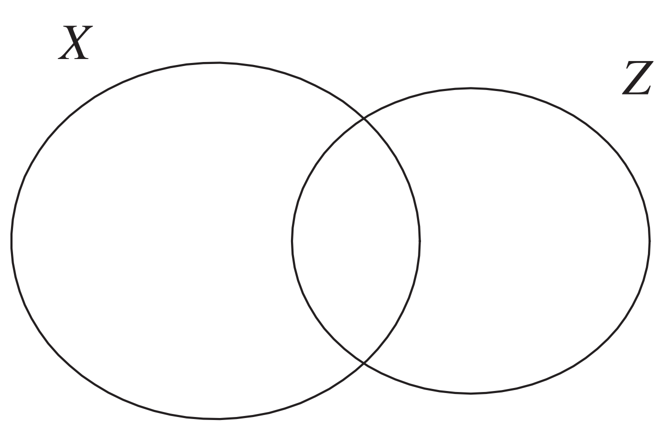
$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \geq \quad & H(X|Z) \\ = \quad & \Pr\{Z=1\}H(X|Z=1) + \Pr\{Z=2\}H(X|Z=2) \\ = \quad & \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \le \lambda \le 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \ge \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

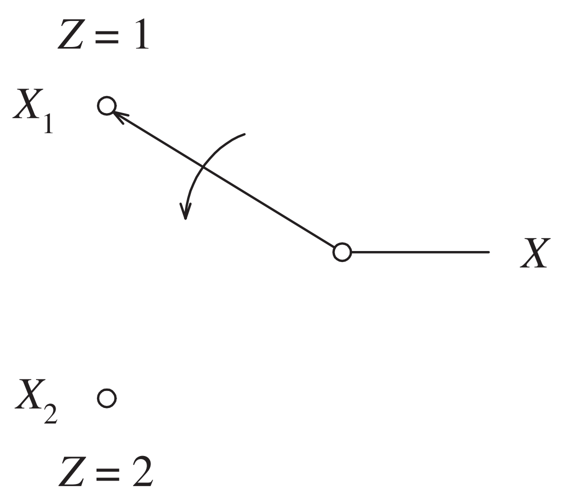
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

$$\mu^*(\tilde{X}) \ge \mu^*(\tilde{X} - \tilde{Z}),$$

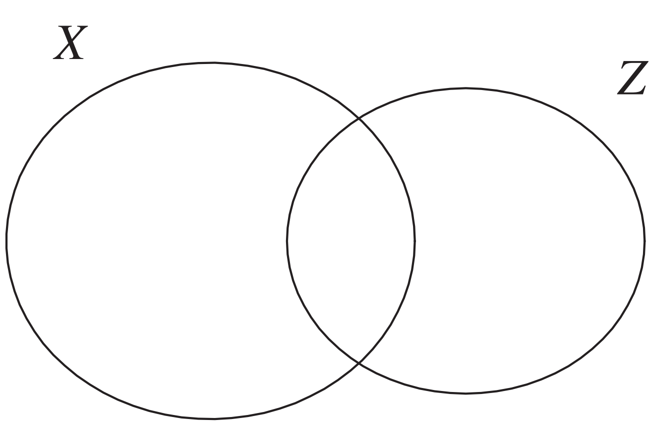
which is equivalent to

$$H(X) \ge H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \ge \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\} H(X|Z = 1) + \Pr\{Z = 2\} H(X|Z = 2) \\ = \quad & \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$

proving (1). This shows that $H(X)$ is a concave functional of $p(x)$.



Example 3.12 (Concavity of Entropy) Let $X_1 \sim p_1(x)$ and $X_2 \sim p_2(x)$. Let

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x),$$

where $0 \leq \lambda \leq 1$ and $\bar{\lambda} = 1 - \lambda$. Show that

$$H(X) \geq \lambda H(X_1) + \bar{\lambda} H(X_2). \quad (1)$$

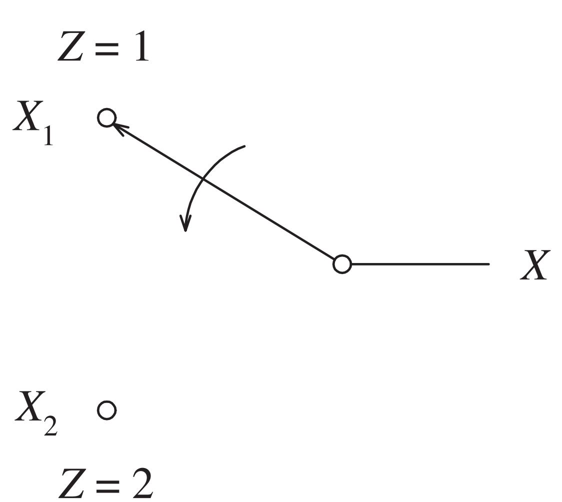
1. Consider the system as shown in which the position of the switch is determined by a random variable Z with

$$\Pr\{Z = 1\} = \lambda \quad \text{and} \quad \Pr\{Z = 2\} = \bar{\lambda},$$

where Z is independent of X_1 and X_2.

2. The switch takes position i if $Z = i$, $i = 1, 2$. The random variable Z is called a mixing random variable for the probability distributions $p_1(x)$ and $p_2(x)$. Then

$$X \sim p(x) = \lambda p_1(x) + \bar{\lambda} p_2(x).$$



3. From the information diagram for X and Z, we see that $\tilde{X} - \tilde{Z}$ is a subset of $\tilde{X}$. Since μ^* is nonnegative for two random variables, we can conclude that

$$\mu^*(\tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}),$$

which is equivalent to

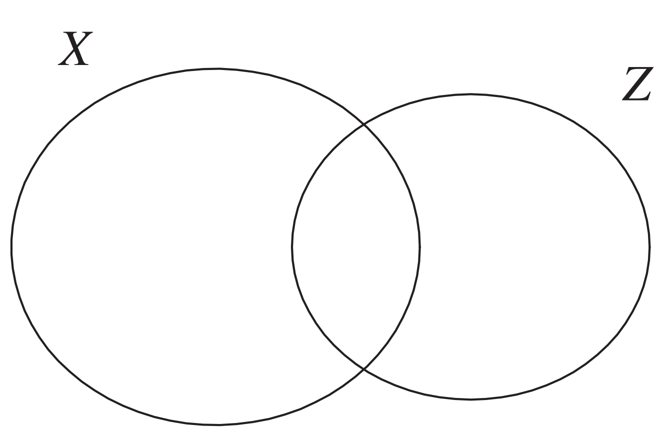
$$H(X) \geq H(X|Z).$$

4. Then

$$\begin{aligned} H(X) & \\ \geq \quad & H(X|Z) \\ = \quad & \Pr\{Z = 1\}H(X|Z = 1) + \Pr\{Z = 2\}H(X|Z = 2) \\ = \quad & \lambda H(X_1) + \bar{\lambda} H(X_2), \end{aligned}$$

proving (1). This shows that $H(X)$ is a concave functional of $p(x)$.

Interpretation The entropy of a mixture of distributions is at least equal to the mixture of the corresponding entropies.



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

Example 3.13 (Convexity of Mutual Information)
Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

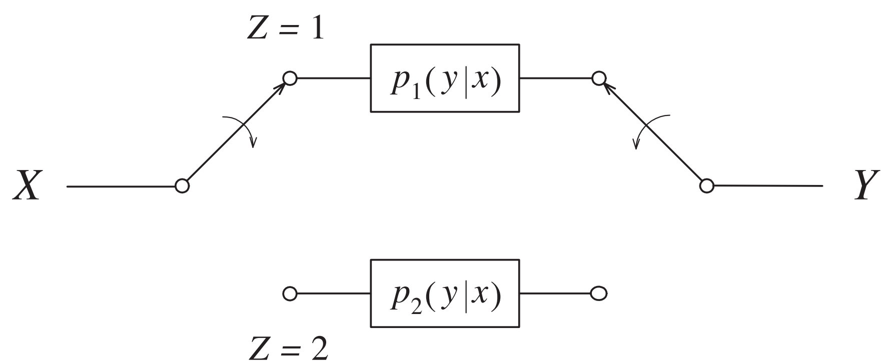
Example 3.13 (Convexity of Mutual Information)
Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,



Example 3.13 (Convexity of Mutual Information)
Let

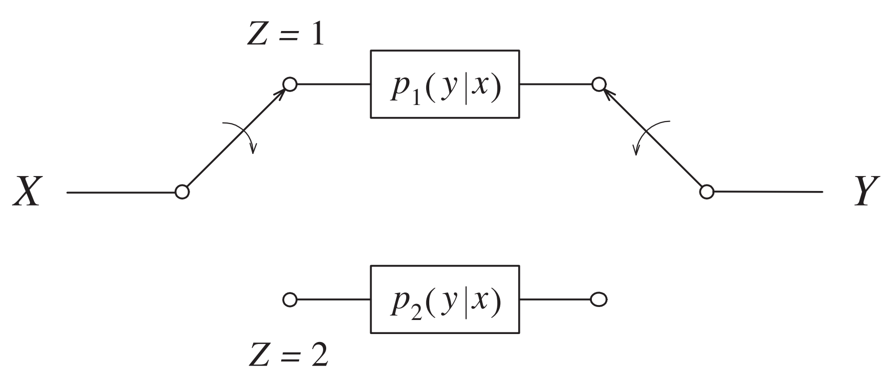
$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

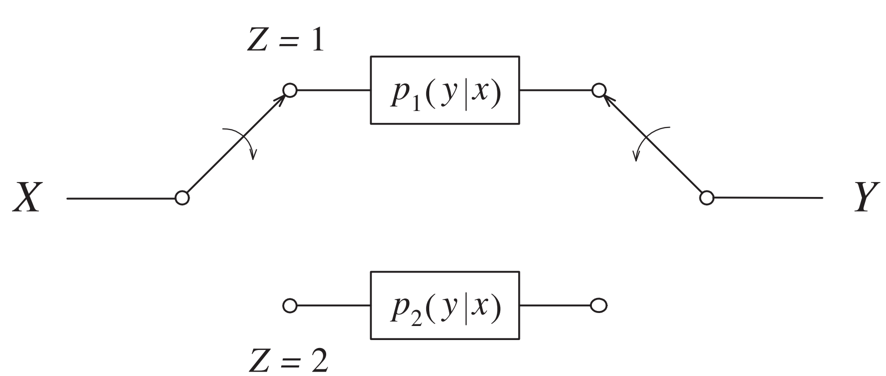
Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

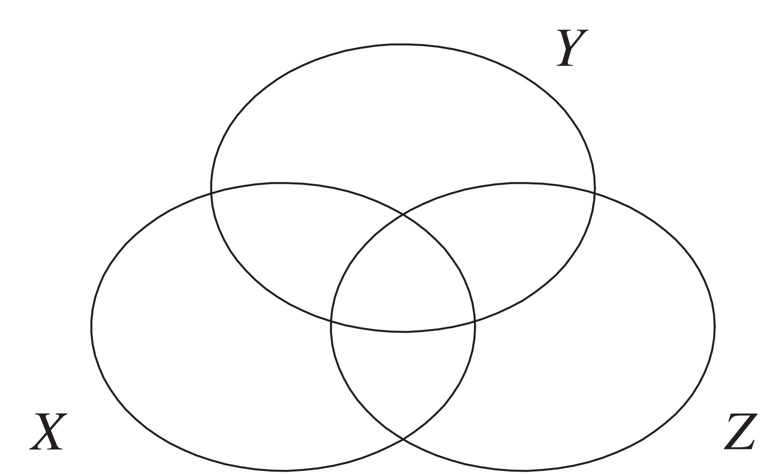
1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let





Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

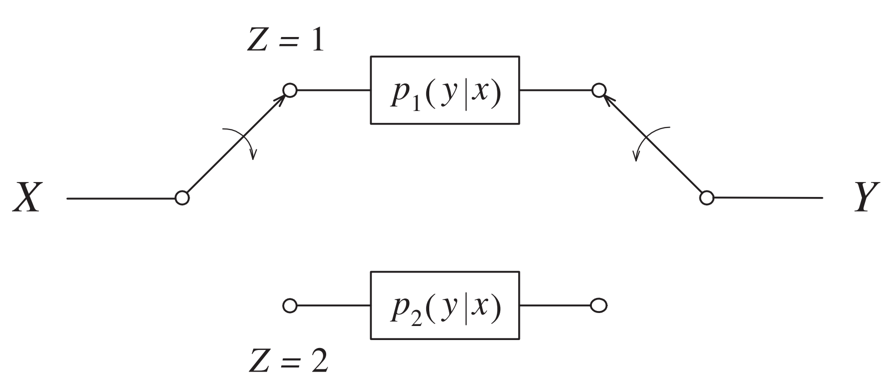
1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

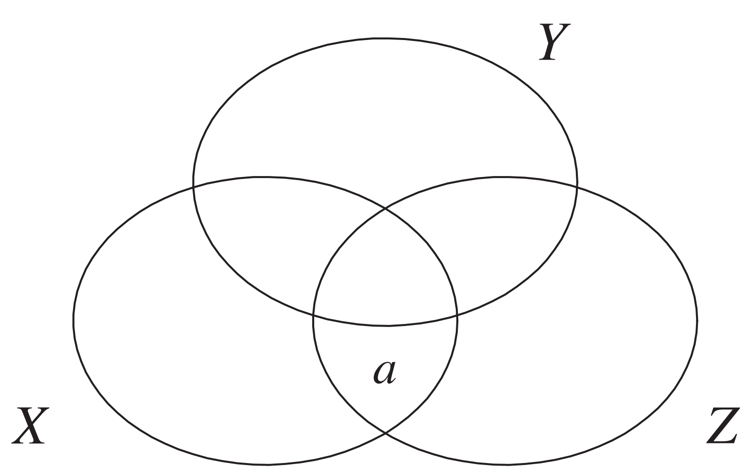
2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

$$I(X; Z|Y) = a \geq 0.$$





Example 3.13 (Convexity of Mutual Information)
Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

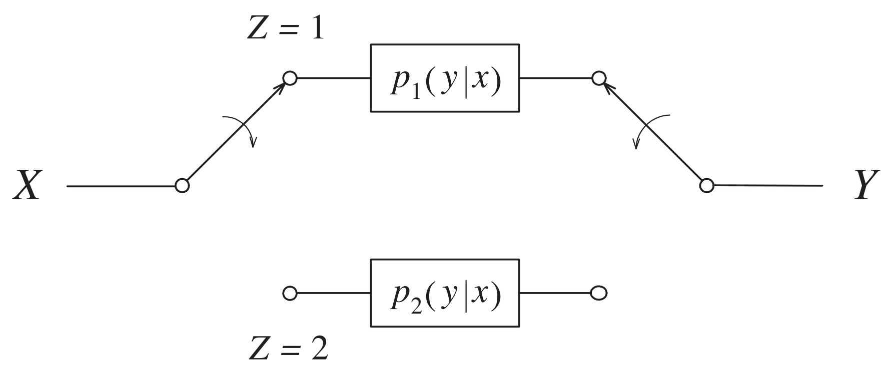
$$I(X;Z) = 0.$$

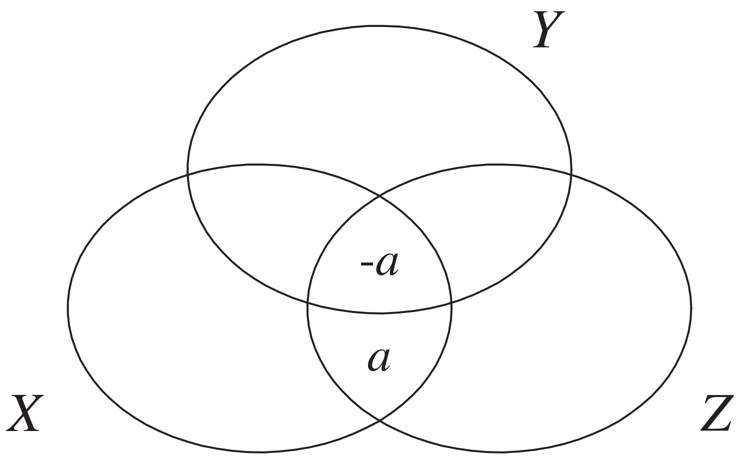
3. In the information diagram for X, Y, and Z, let

$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$





Example 3.13 (Convexity of Mutual Information)
Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

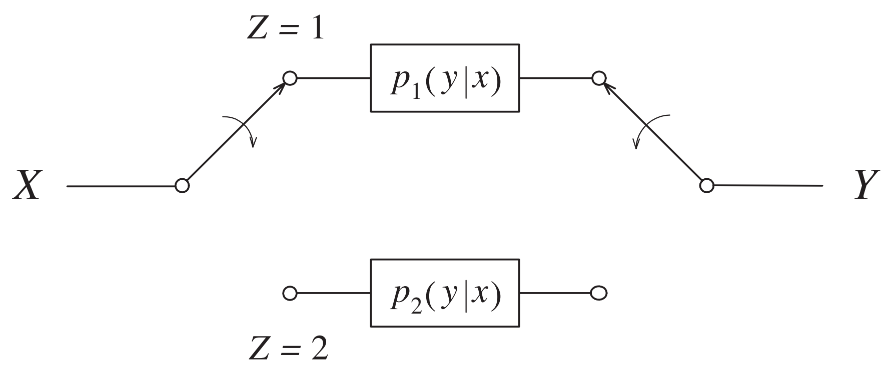
3. In the information diagram for X, Y, and Z, let

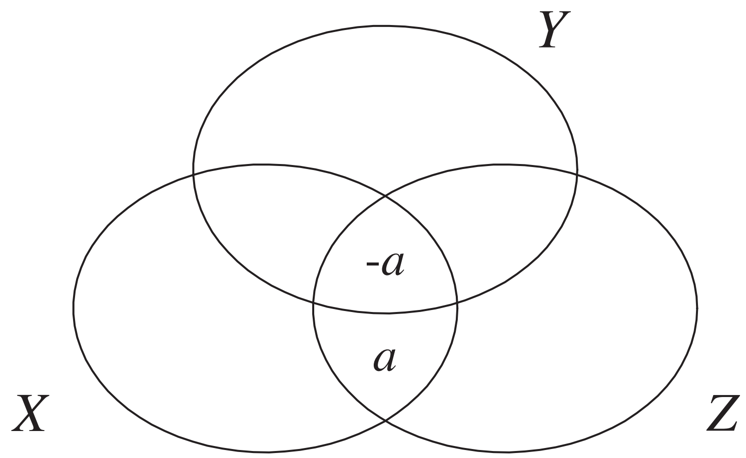
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

because $I(X;Z) = 0$.





Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

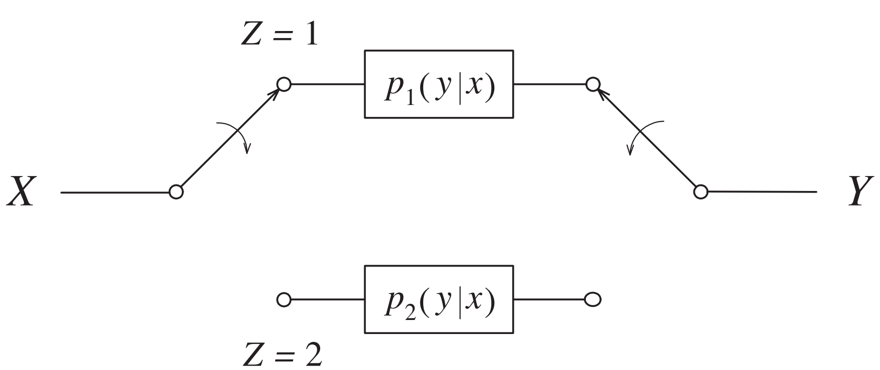
$$I(X; Z|Y) = a \geq 0.$$

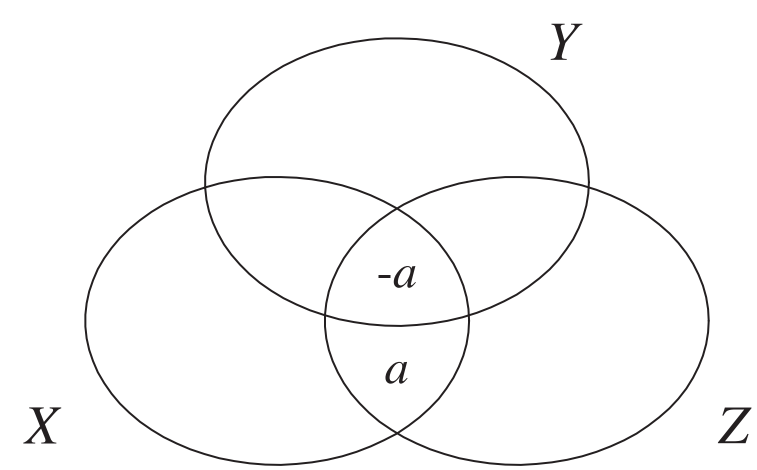
Then

$$I(X; Y; Z) = -a$$

because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.





Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

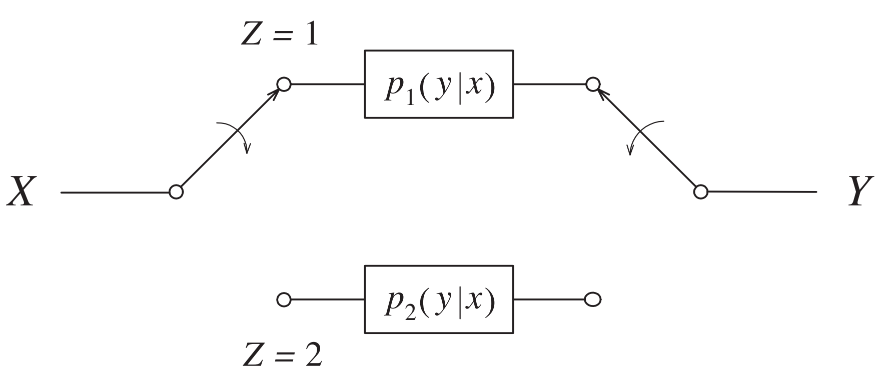
$$I(X;Z|Y) = a \geq 0.$$

Then

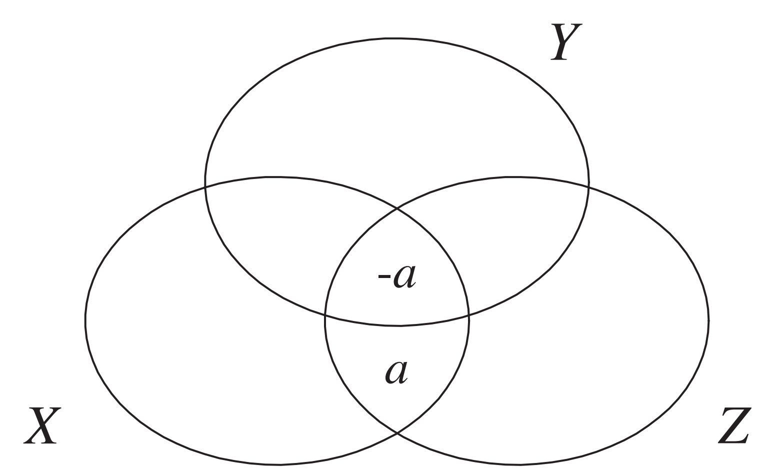
$$I(X;Y;Z) = -a$$

because $I(X;Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

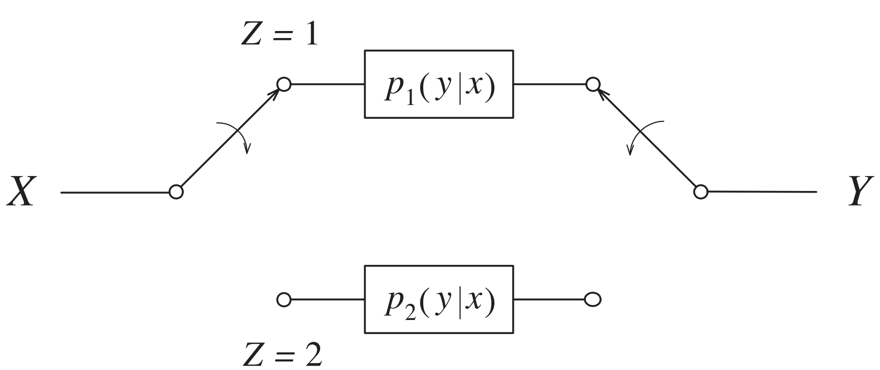
$$I(X; Z|Y) = a \geq 0.$$

Then

$$I(X; Y; Z) = -a$$

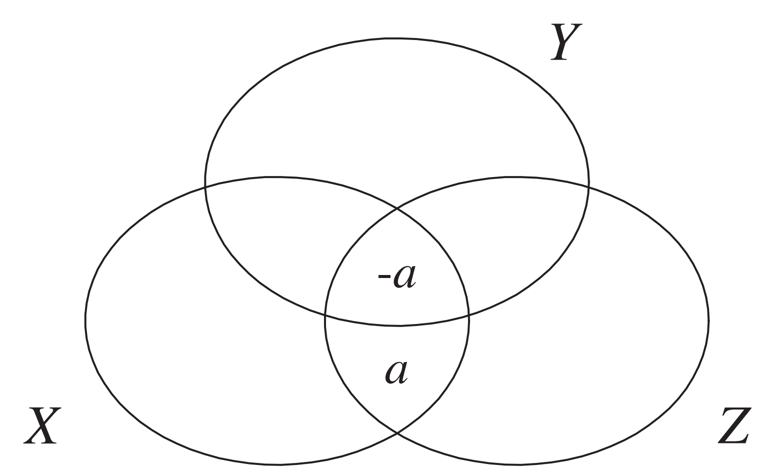
because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$I(X; Y)$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

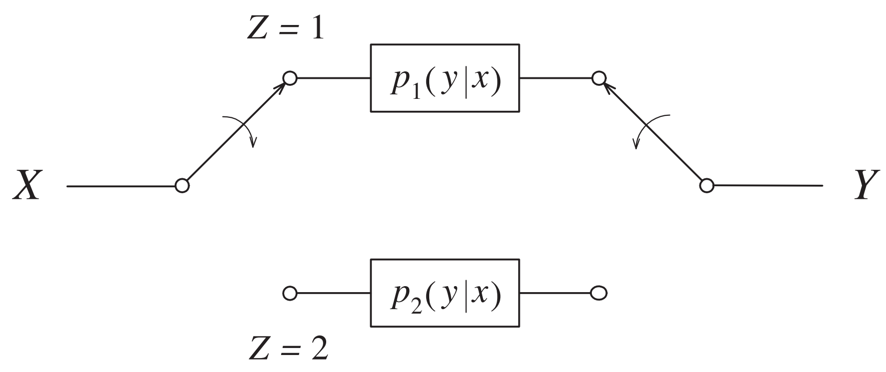
$$I(X; Z|Y) = a \geq 0.$$

Then

$$I(X; Y; Z) = -a$$

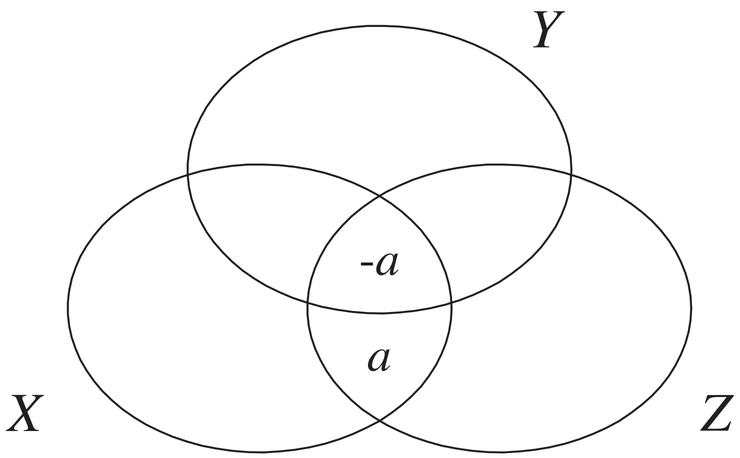
because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$I(X; Y) = {\color{blue}I(X; Y|Z)} + {\color{red}I(X; Y; Z)}$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

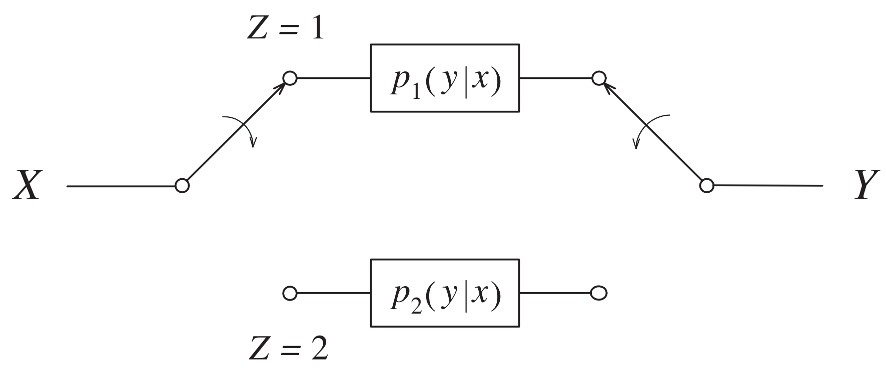
$$I(X; Z|Y) = a \geq 0.$$

Then

$$I(X; Y; Z) = -a$$

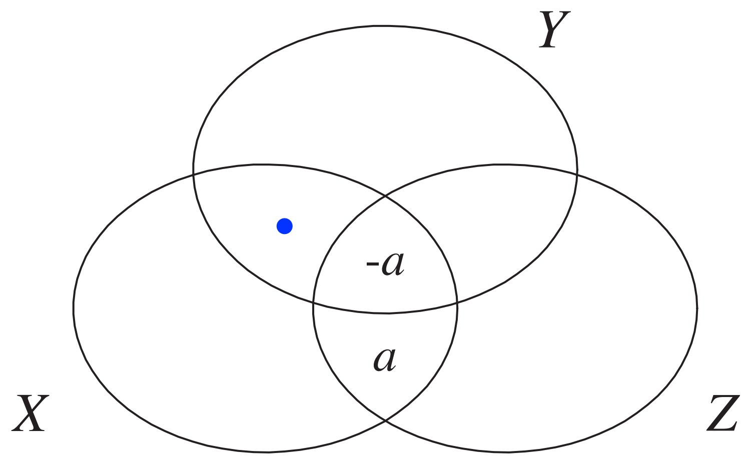
because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$I(X; Y) = I(X; Y|Z) + I(X; Y; Z)$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

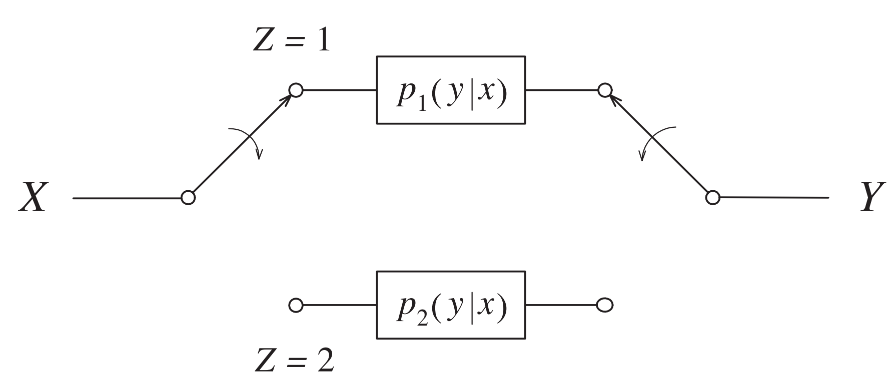
$$I(X; Z|Y) = a \geq 0.$$

Then

$$I(X; Y; Z) = -a$$

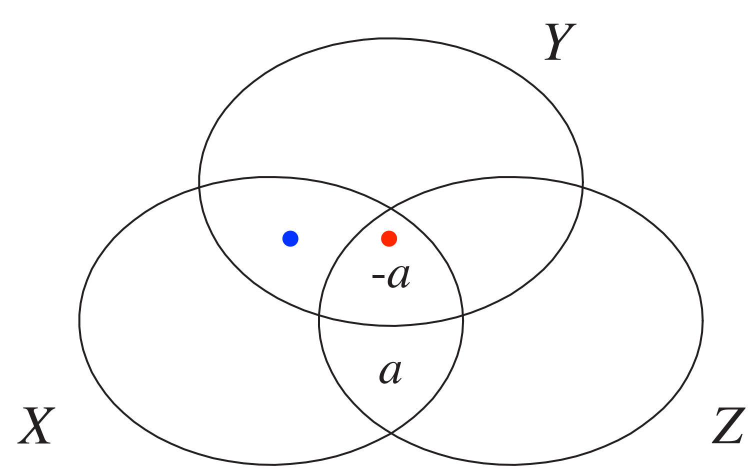
because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$I(X; Y) = I(X; Y|Z) + I(X; Y; Z)$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

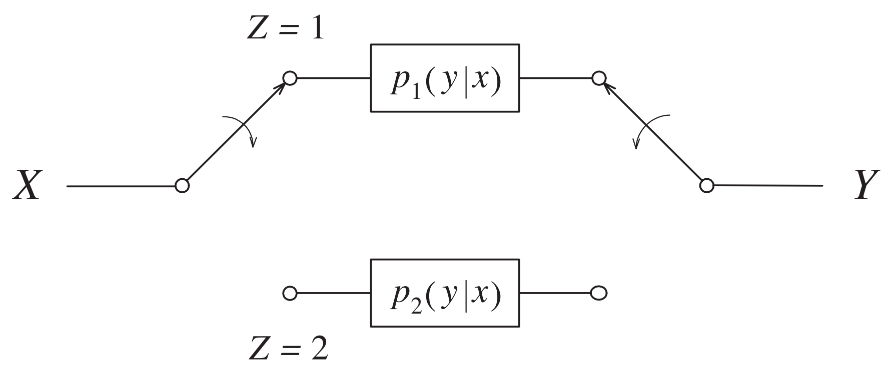
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

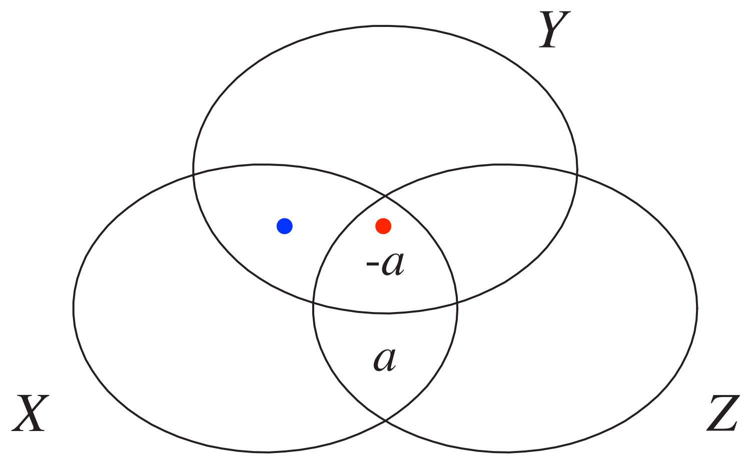
because $I(X;Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} &I(X;Y) \\ &\quad = \quad I(X;Y|Z) + I(X;Y;Z) \\ &\quad \leq \quad I(X;Y|Z) \end{aligned}$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

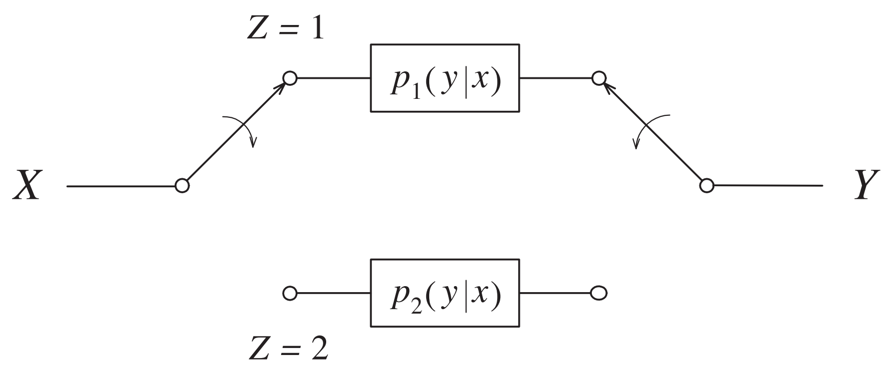
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

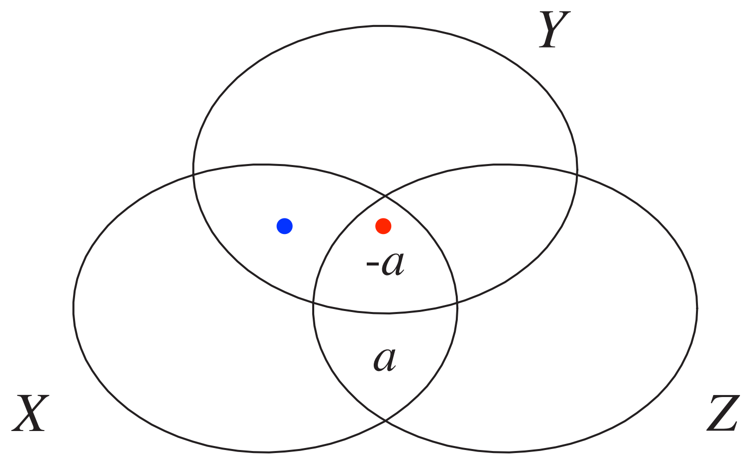
because $I(X;Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} I(X;Y) &= I(X;Y|Z) + I(X;Y;Z) \\ &\leq I(X;Y|Z) \\ &= \Pr\{Z = 1\}I(X;Y|Z = 1) \\ &\quad +\Pr\{Z = 2\}I(X;Y|Z = 2) \end{aligned}$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

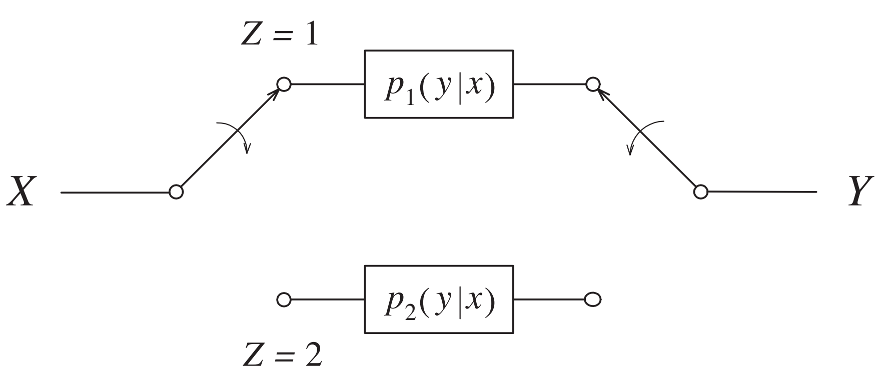
$$I(X; Z|Y) = a \geq 0.$$

Then

$$I(X; Y; Z) = -a$$

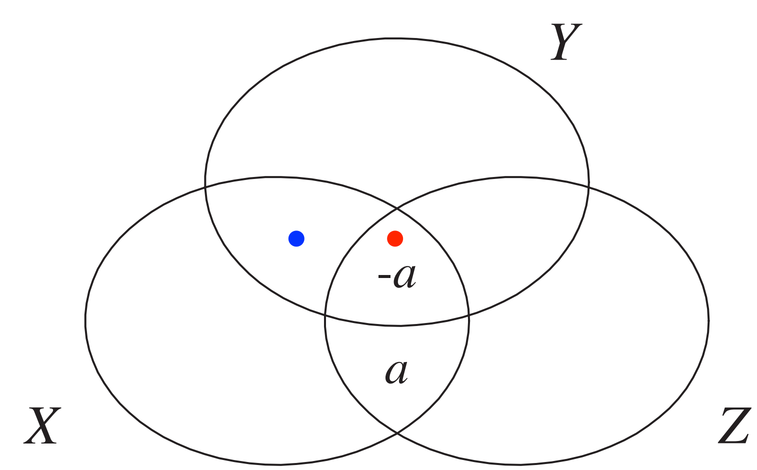
because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} I(X; Y) &= I(X; Y|Z) + I(X; Y; Z) \\ &\leq I(X; Y|Z) \\ &= \Pr\{Z = 1\}I(X; Y|Z = 1) \\ &\quad + \Pr\{Z = 2\}I(X; Y|Z = 2) \end{aligned}$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

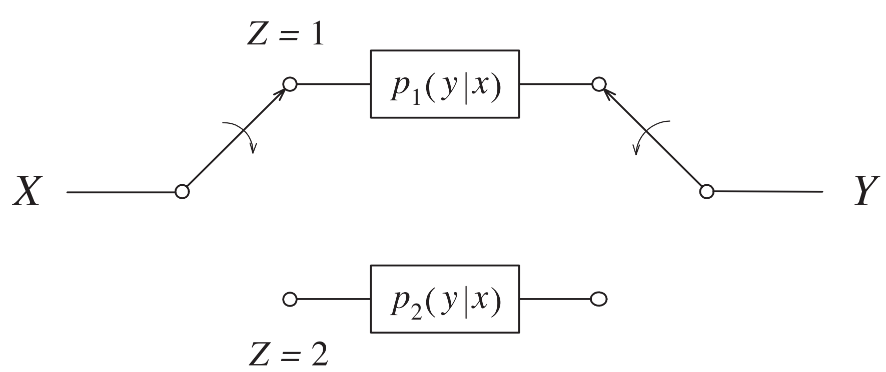
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

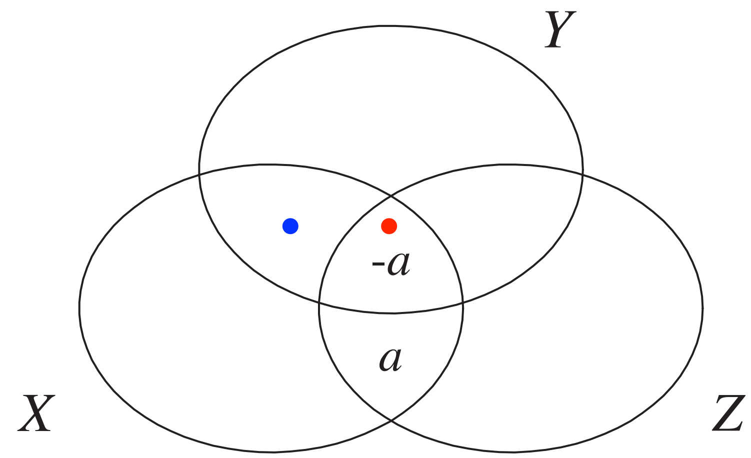
because $I(X;Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} & I(X;Y) \\ &= I(X;Y|Z) + I(X;Y;Z) \\ &\leq I(X;Y|Z) \\ &= \Pr\{Z=1\} I(X;Y|Z=1) \\ &\quad + \Pr\{Z=2\} I(X;Y|Z=2) \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)), \end{aligned}$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.
2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

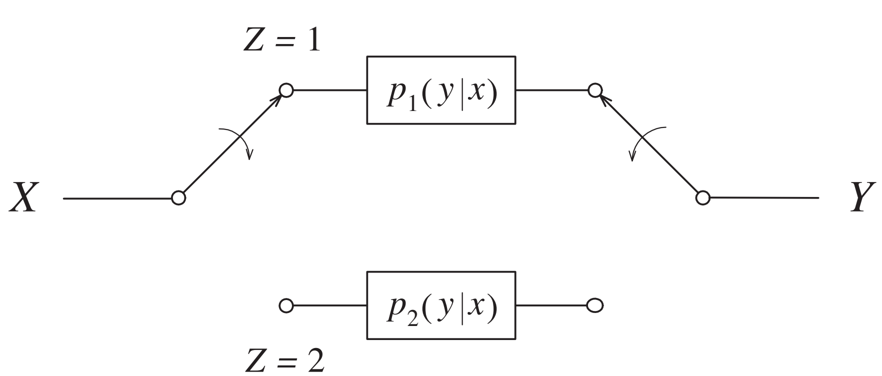
$$I(X; Z|Y) = a \geq 0.$$

Then

$$I(X; Y; Z) = -a$$

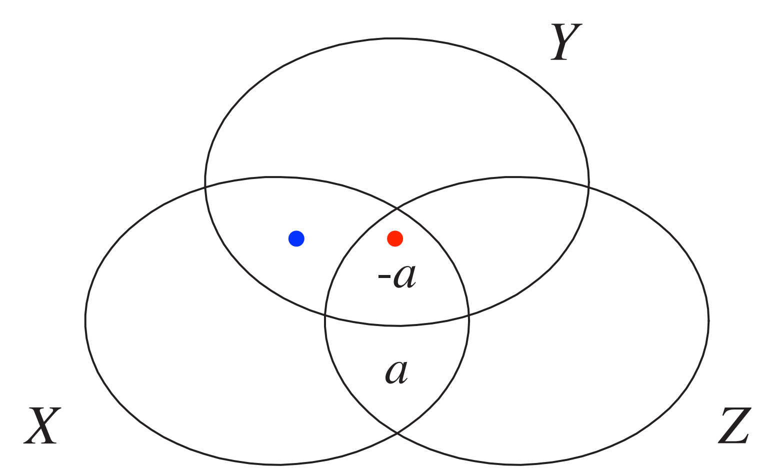
because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} & I(X; Y) \\ &= I(X; Y|Z) + I(X; Y; Z) \\ &\leq I(X; Y|Z) \\ &= \Pr\{Z = 1\}I(X; Y|Z = 1) \\ &\quad + \Pr\{Z = 2\}I(X; Y|Z = 2) \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)), \end{aligned}$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X; Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

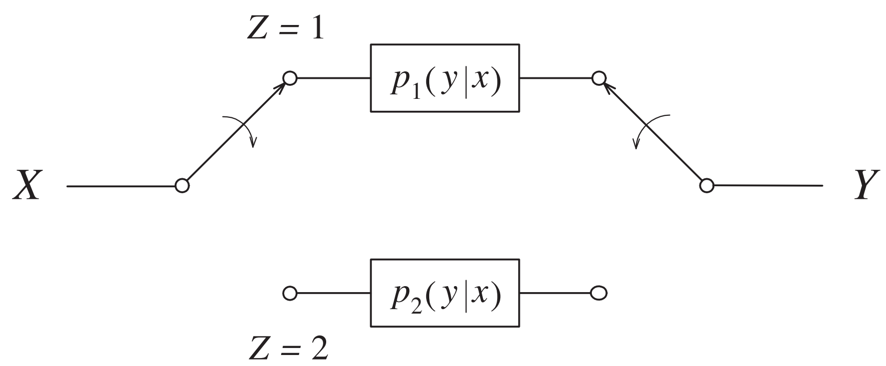
$$I(X; Z|Y) = a \geq 0.$$

Then

$$I(X; Y; Z) = -a$$

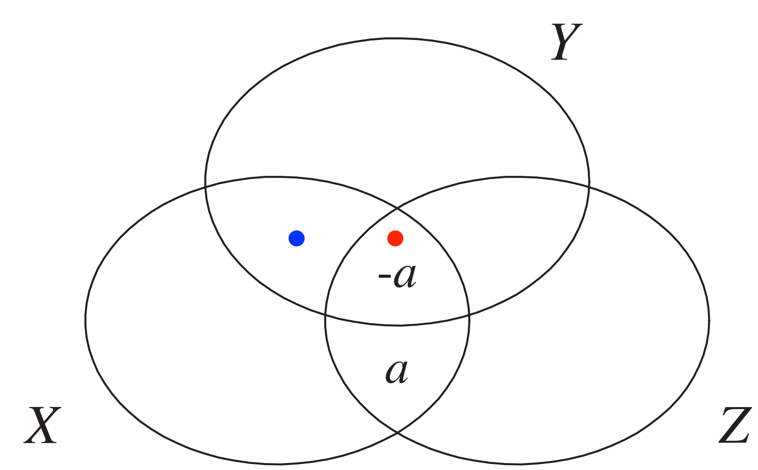
because $I(X; Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} I(X; Y) &= I(X; Y|Z) + I(X; Y; Z) \\ &\leq I(X; Y|Z) \\ &= \Pr\{Z = 1\}I(X; Y|Z = 1) \\ &\quad + \Pr\{Z = 2\}I(X; Y|Z = 2) \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)), \end{aligned}$$



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

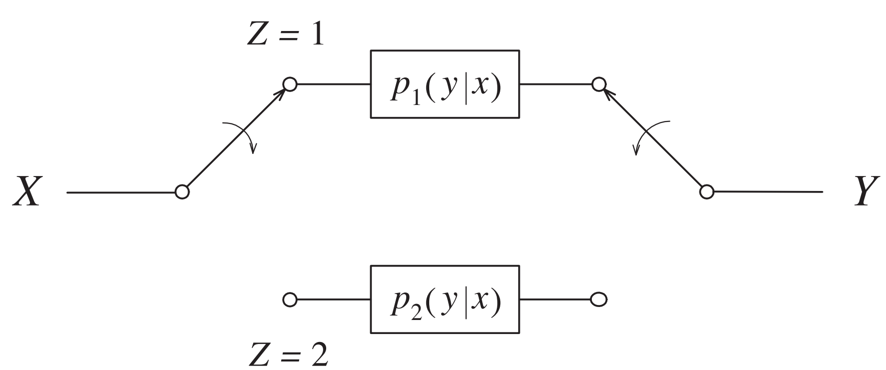
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

because $I(X;Z) = 0$.

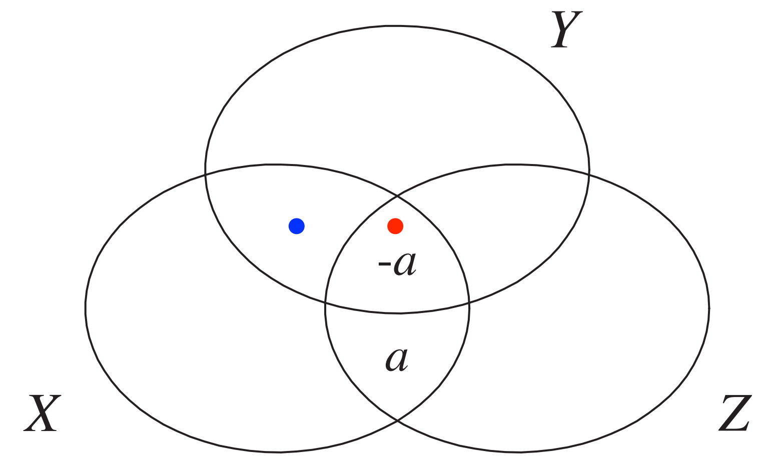
4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} I(X;Y) &= I(X;Y|Z) + I(X;Y;Z) \\ &\leq I(X;Y|Z) \\ &= \Pr\{Z=1\}\underline{I(X;Y|Z=1)} \\ &\quad +\Pr\{Z=2\}I(X;Y|Z=2) \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)), \end{aligned}$$

where $I(p(x), p_1(y|x))$ is the mutual information between X and Y when the switch is up, and $I(p(x), p_2(y|x))$ is the mutual information between X and Y when the switch is down. This shows that $I(X;Y)$ is a convex functional of $p(y|x)$.



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.
2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

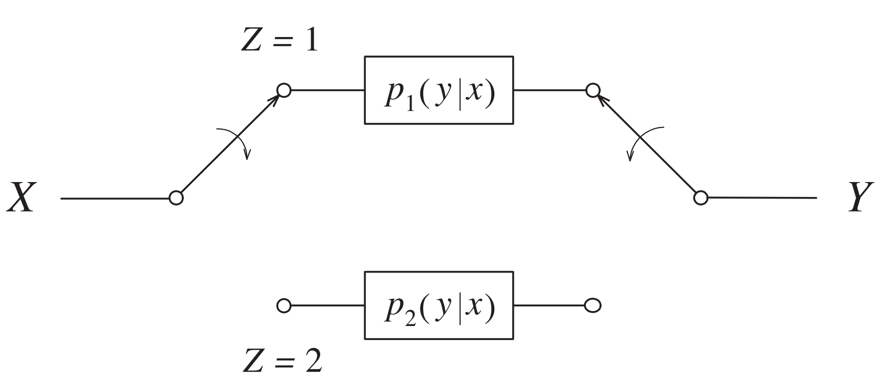
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

because $I(X;Z) = 0$.

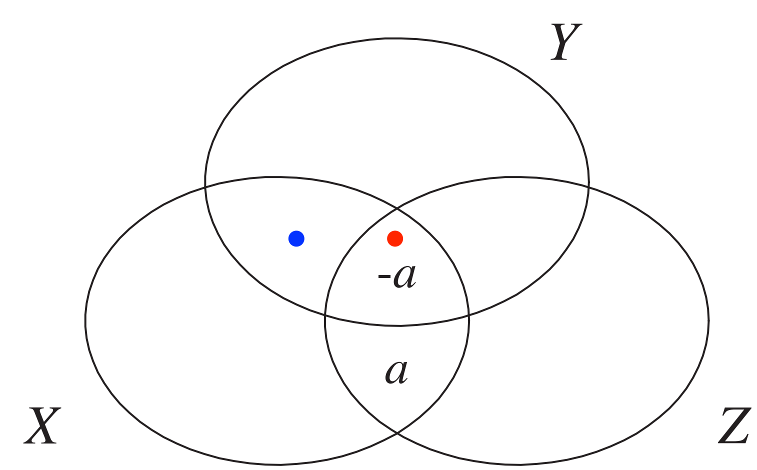
4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned}
I(X;Y) &= I(X;Y|Z) + I(X;Y;Z) \\
&\leq I(X;Y|Z) \\
&= \Pr\{Z=1\}I(X;Y|Z=1) \\
&\quad + \Pr\{Z=2\}I(X;Y|Z=2) \\
&= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)),
\end{aligned}$$

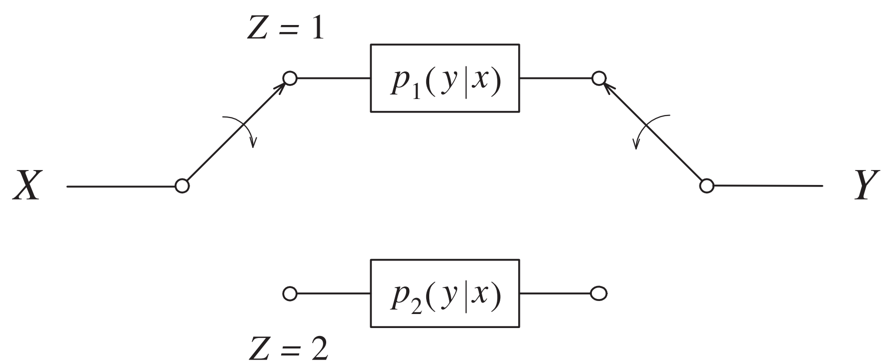
where $I(p(x), p_1(y|x))$ is the mutual information between X and Y when the switch is up, and $I(p(x), p_2(y|x))$ is the mutual information between X and Y when the switch is down. This shows that $I(X;Y)$ is a convex functional of $p(y|x)$.



Example 3.13 (Convexity of Mutual Information)
Let
$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

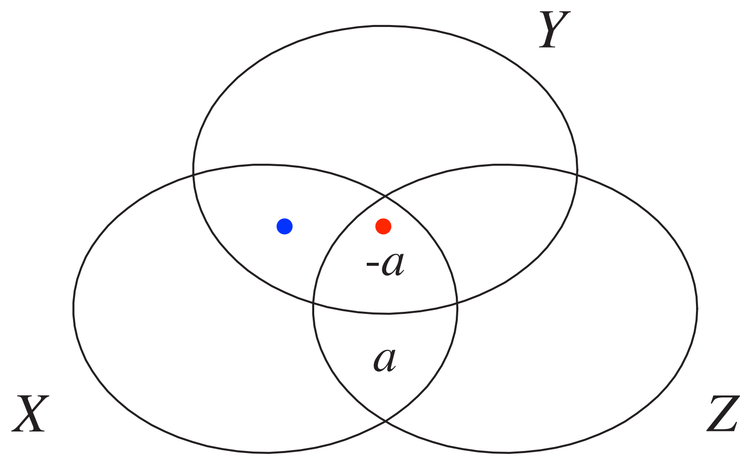
Show that for fixed $p(x)$, $I(X; Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.
2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,
$$I(X; Z) = 0.$$
3. In the information diagram for X, Y, and Z, let
$$I(X; Z|Y) = a \geq 0.$$
Then
$$I(X; Y; Z) = -a$$
because $I(X; Z) = 0$.
4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then
$$\begin{aligned} I(X; Y) &= I(X; Y|Z) + I(X; Y; Z) \\ &\leq I(X; Y|Z) \\ &= \Pr\{Z = 1\}I(X; Y|Z = 1) \\ &\quad +\Pr\{Z = 2\}\underline{I(X; Y|Z = 2)} \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)), \end{aligned}$$

where $I(p(x), p_1(y|x))$ is the mutual information between X and Y when the switch is up, and $I(p(x), p_2(y|x))$ is the mutual information between X and Y when the switch is down. This shows that $I(X; Y)$ is a convex functional of $p(y|x)$.



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.
2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

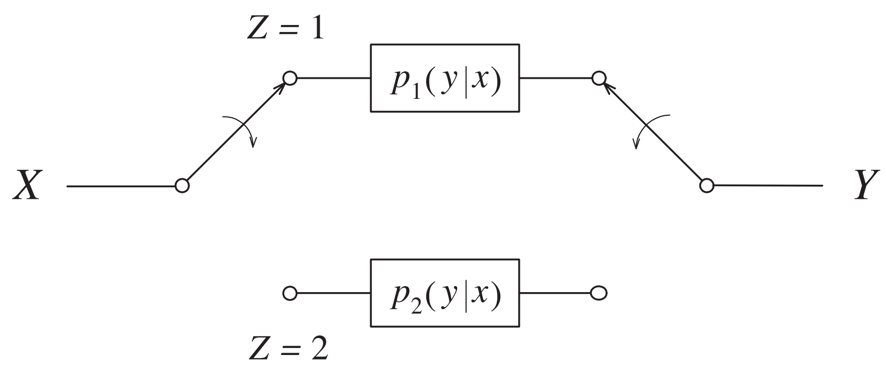
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

because $I(X;Z) = 0$.

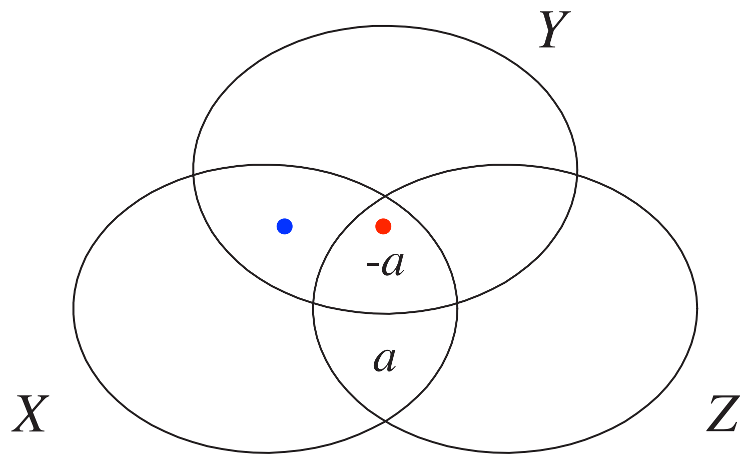
4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} I(X;Y) &= I(X;Y|Z) + I(X;Y;Z) \\ &\leq I(X;Y|Z) \\ &= \Pr\{Z=1\}I(X;Y|Z=1) \\ &\quad +\Pr\{Z=2\}\underline{I(X;Y|Z=2)} \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda}\underline{I(p(x), p_2(y|x))}, \end{aligned}$$

where $I(p(x), p_1(y|x))$ is the mutual information between X and Y when the switch is up, and $I(p(x), p_2(y|x))$ is the mutual information between X and Y when the switch is down. This shows that $I(X;Y)$ is a convex functional of $p(y|x)$.



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

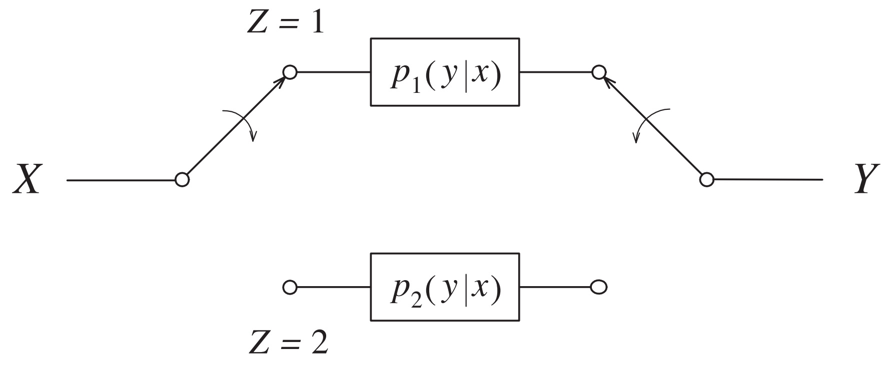
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

because $I(X;Z) = 0$.

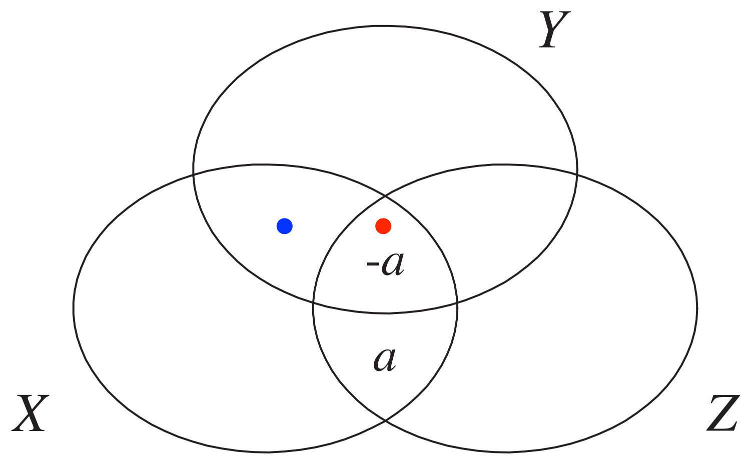
4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.



5. Then

$$\begin{aligned} I(X;Y) &= I(X;Y|Z) + I(X;Y;Z) \\ &\leq I(X;Y|Z) \\ &= \Pr\{Z = 1\}I(X;Y|Z = 1) \\ &\quad + \Pr\{Z = 2\}I(X;Y|Z = 2) \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)), \end{aligned}$$

where $I(p(x), p_1(y|x))$ is the mutual information between X and Y when the switch is up, and $I(p(x), p_2(y|x))$ is the mutual information between X and Y when the switch is down. This shows that $I(X;Y)$ is a convex functional of $p(y|x)$.



Example 3.13 (Convexity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$.

1. Let $p_1(y|x)$ and $p_2(y|x)$ be 2 transition matrices representing 2 channels.

2. Consider the system as shown in which the position of the switch is determined by a random variable Z as in the last example, where Z is independent of X, i.e.,

$$I(X;Z) = 0.$$

3. In the information diagram for X, Y, and Z, let

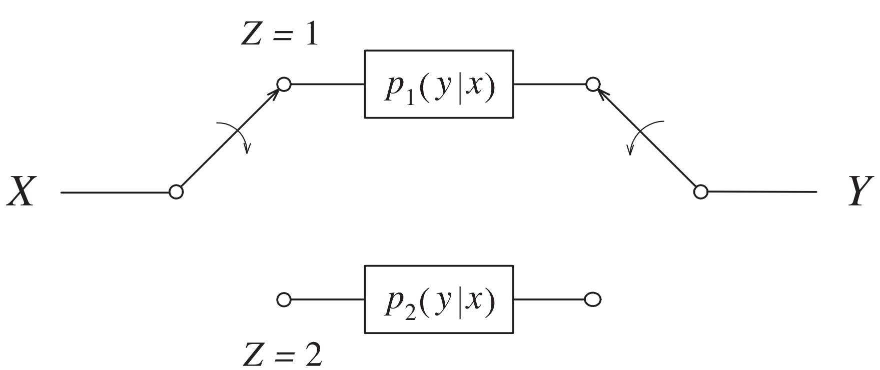
$$I(X;Z|Y) = a \geq 0.$$

Then

$$I(X;Y;Z) = -a$$

because $I(X;Z) = 0$.

4. Recall that $\Pr\{Z = 1\} = \lambda$ and $\Pr\{Z = 2\} = \bar{\lambda}$.

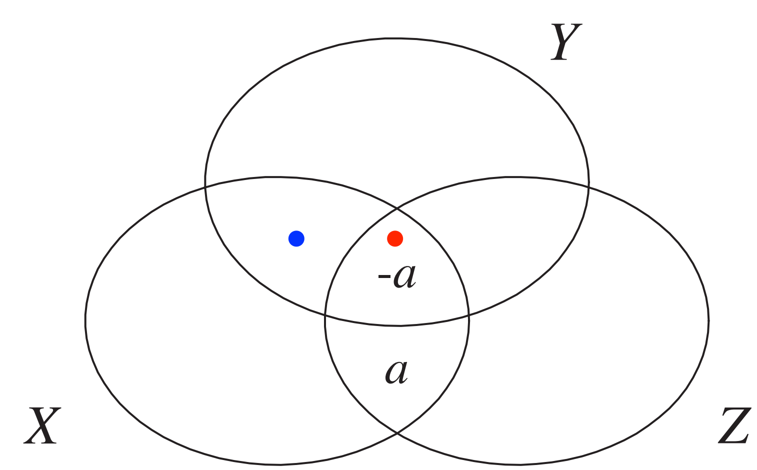


5. Then

$$\begin{aligned} I(X;Y) &= I(X;Y|Z) + I(X;Y;Z) \\ &\leq I(X;Y|Z) \\ &= \Pr\{Z=1\}I(X;Y|Z=1) \\ &\quad +\Pr\{Z=2\}I(X;Y|Z=2) \\ &= \lambda I(p(x), p_1(y|x)) + \bar{\lambda} I(p(x), p_2(y|x)), \end{aligned}$$

where $I(p(x), p_1(y|x))$ is the mutual information between X and Y when the switch is up, and $I(p(x), p_2(y|x))$ is the mutual information between X and Y when the switch is down. This shows that $I(X;Y)$ is a convex functional of $p(y|x)$.

Interpretation For a fixed input distribution $p(x)$, the mutual information between the input and the output of the system as shown, which is obtained by mixing 2 channels $p_1(y|x)$ and $p_2(y|x)$, is at most the mixture of the 2 mutual informations corresponding to $p_1(y|x)$ and $p_2(y|x)$, respectively.



Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

Example 3.14 (Concavity of Mutual Information)
Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

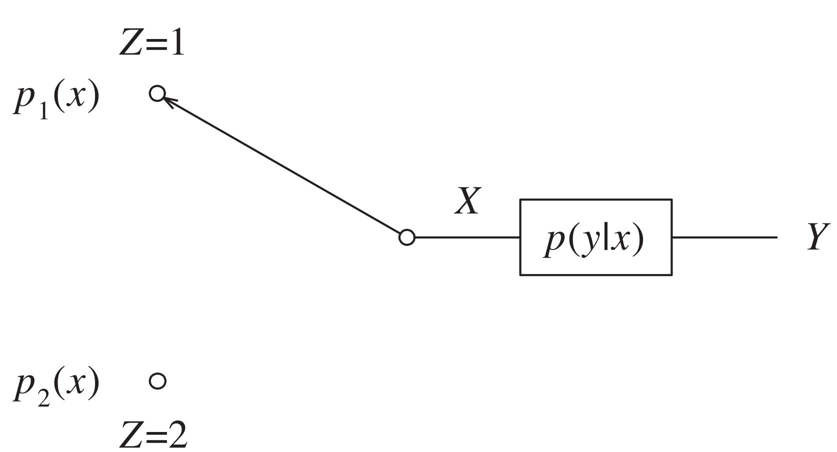
Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

Example 3.14 (Concavity of Mutual Information)
Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.



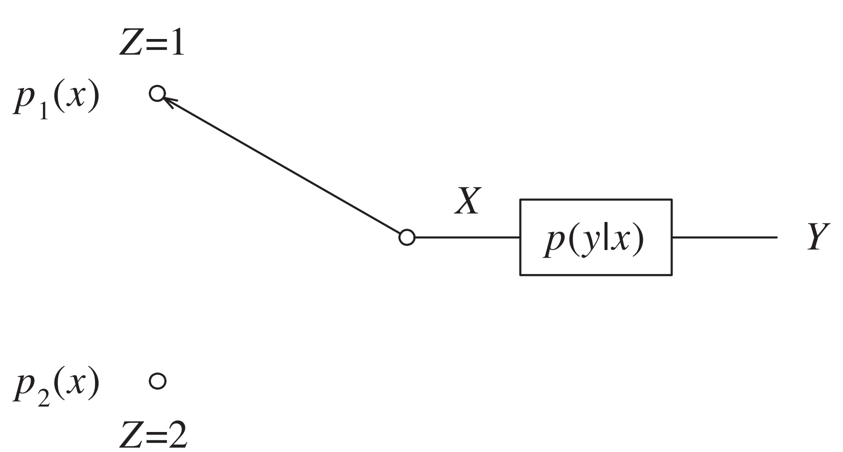
Example 3.14 (Concavity of Mutual Information)
Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \rightarrow X \rightarrow Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



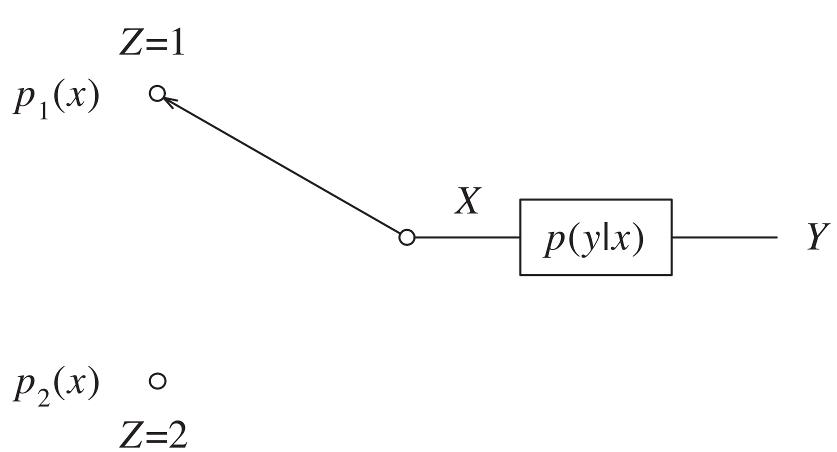
Example 3.14 (Concavity of Mutual Information) Let

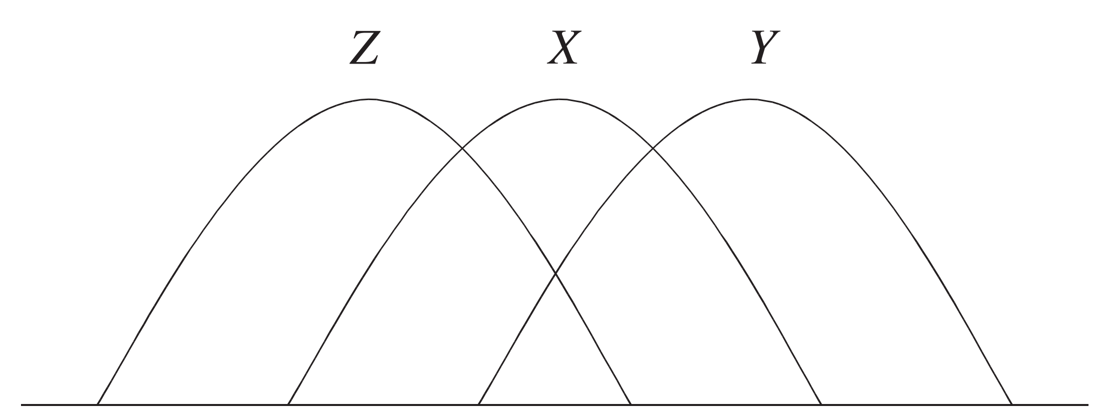
$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.





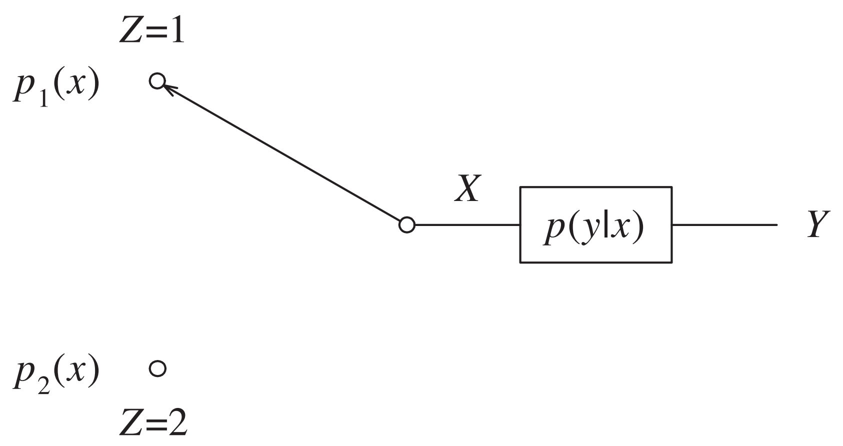
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

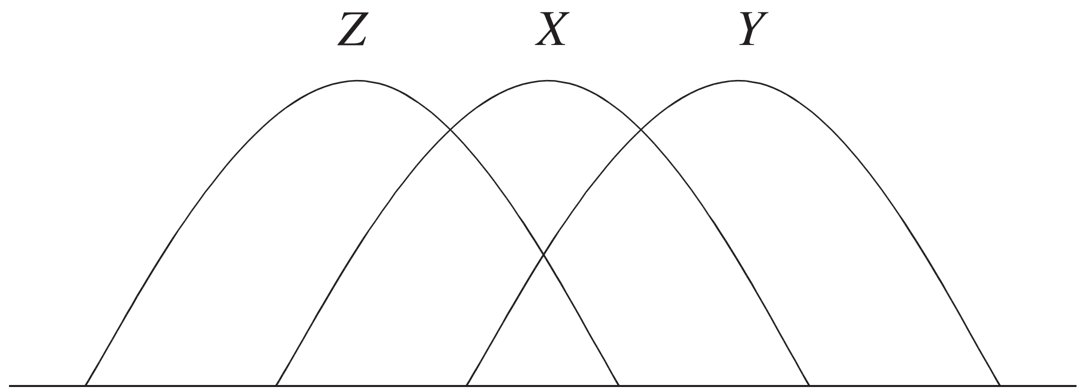
Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that



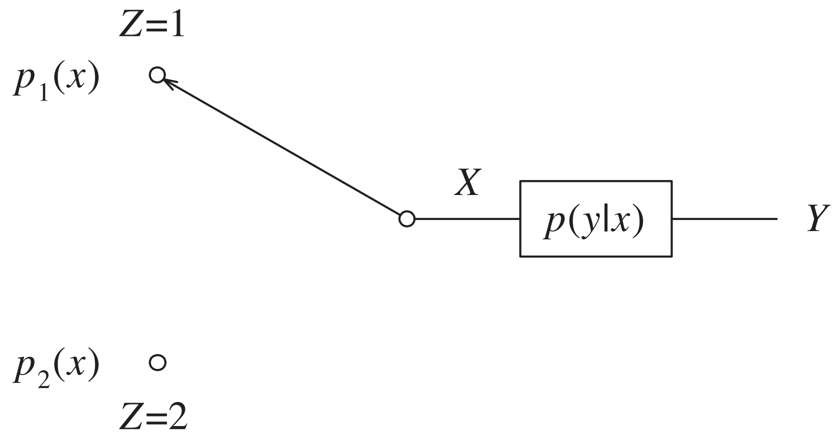
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

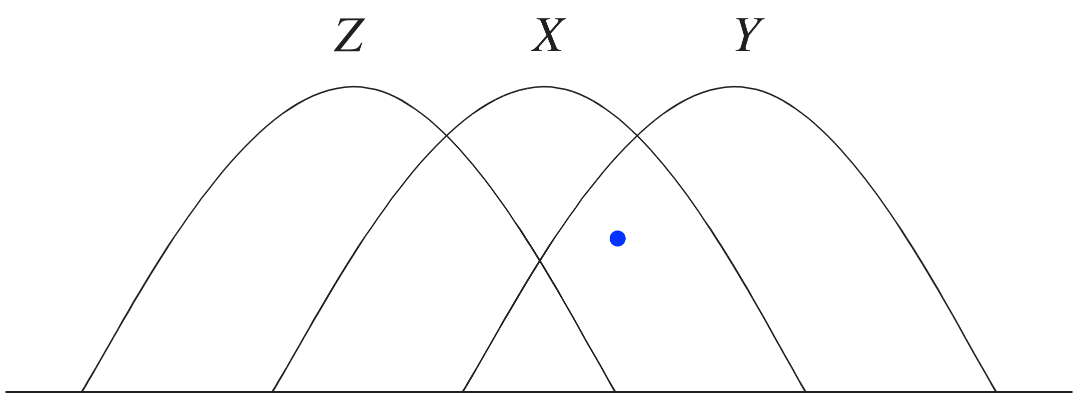
Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that



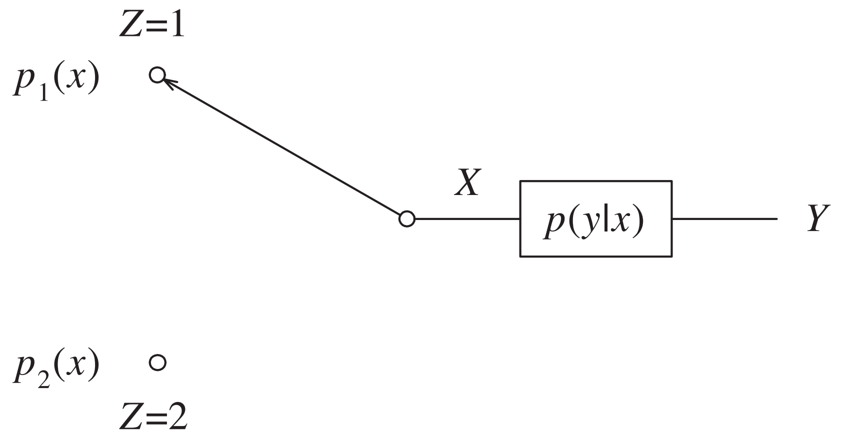
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

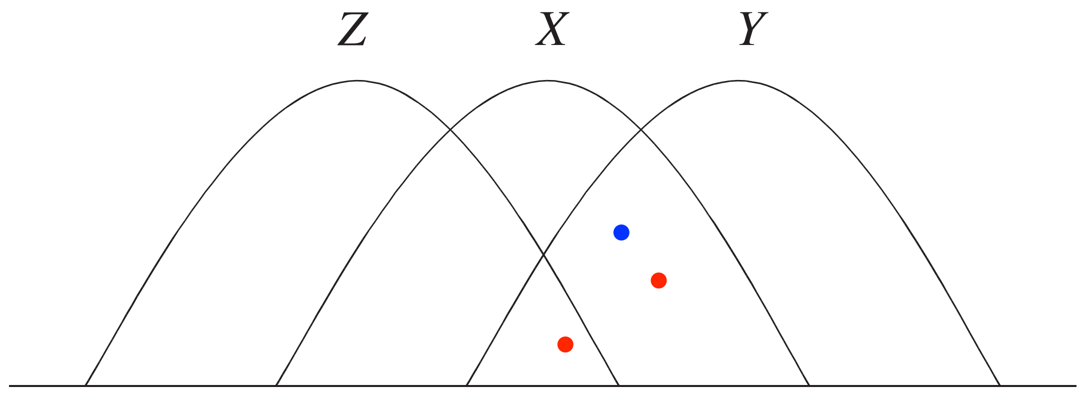
Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that



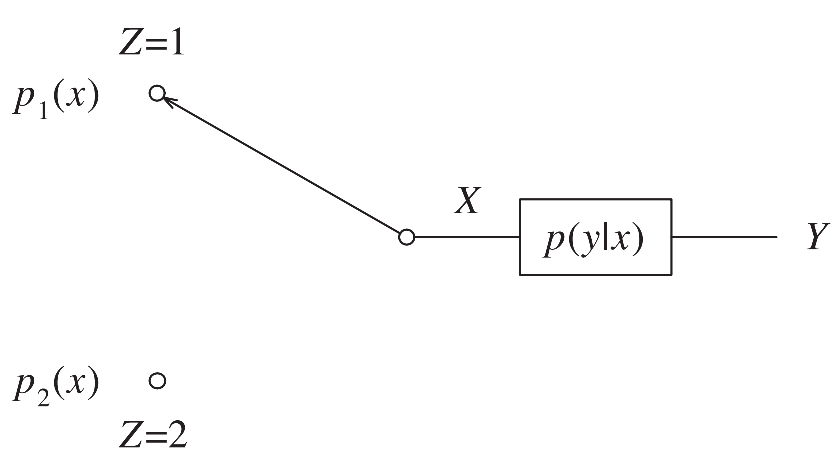
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

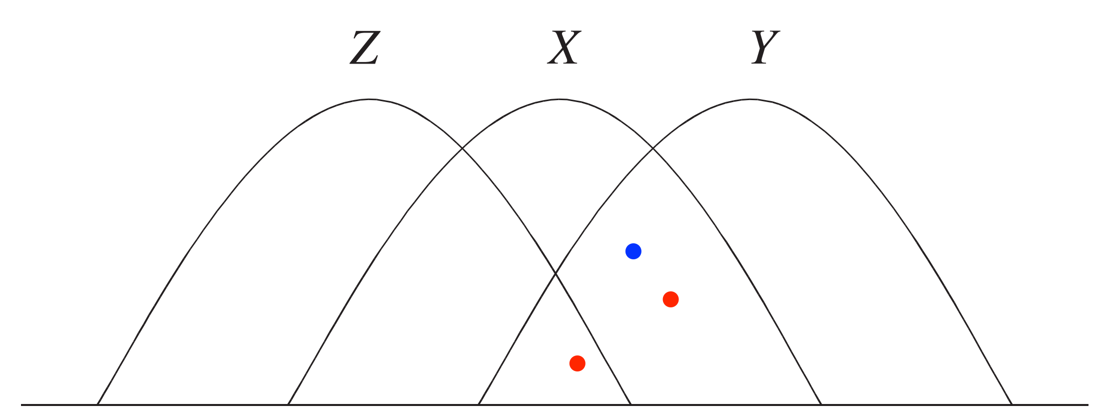
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$I(X; Y)$$



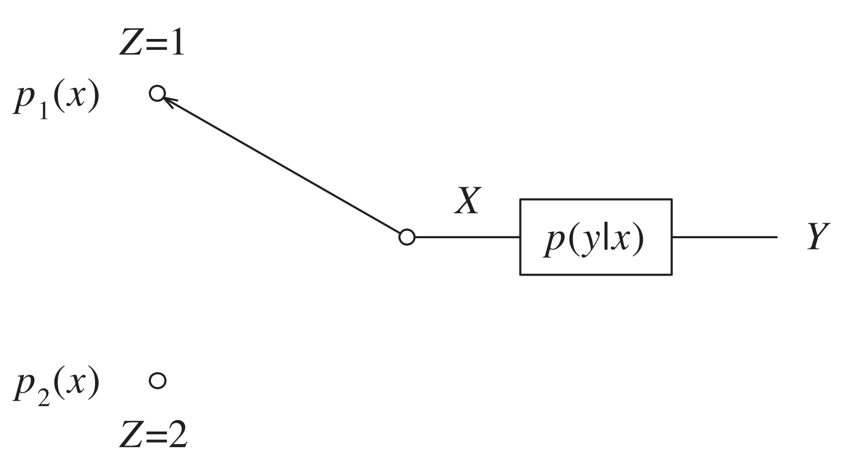
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

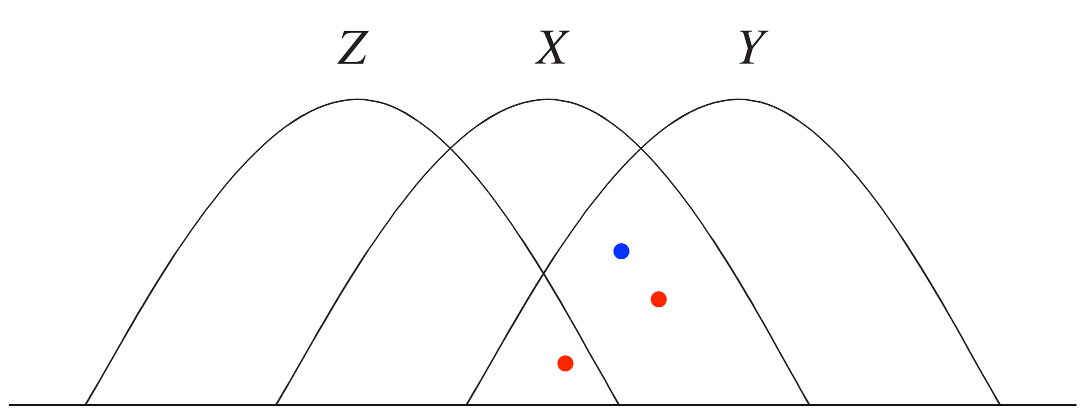
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X;Y) \\ \geq \quad & I(X;Y|Z) \end{aligned}$$



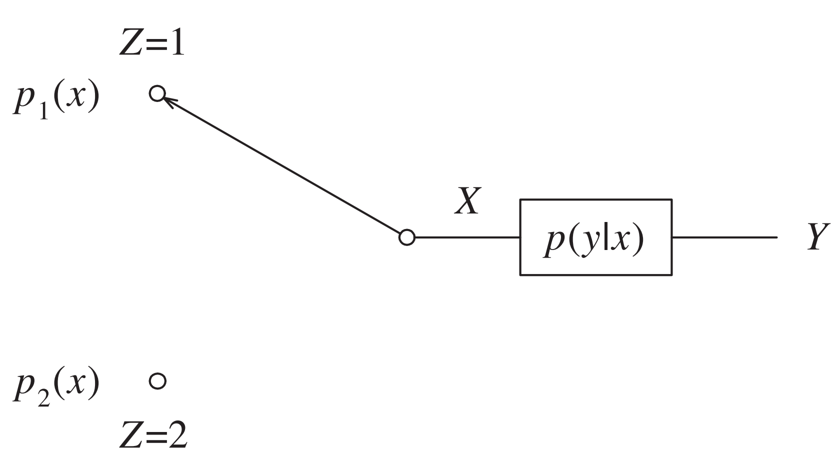
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

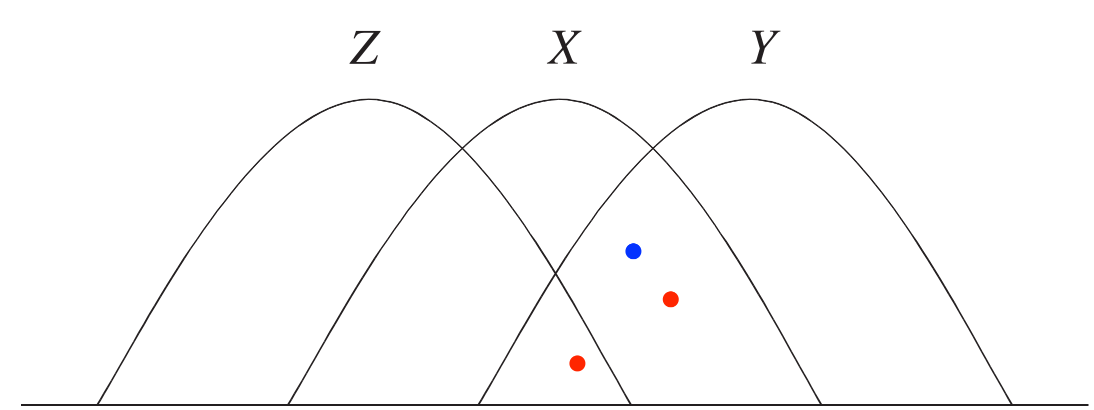
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X;Y) \\ & \quad \geq \quad I(X;Y|Z) \\ & \quad = \quad \Pr\{Z=1\}I(X;Y|Z=1) \\ & \qquad\quad + \Pr\{Z=2\}I(X;Y|Z=2) \end{aligned}$$



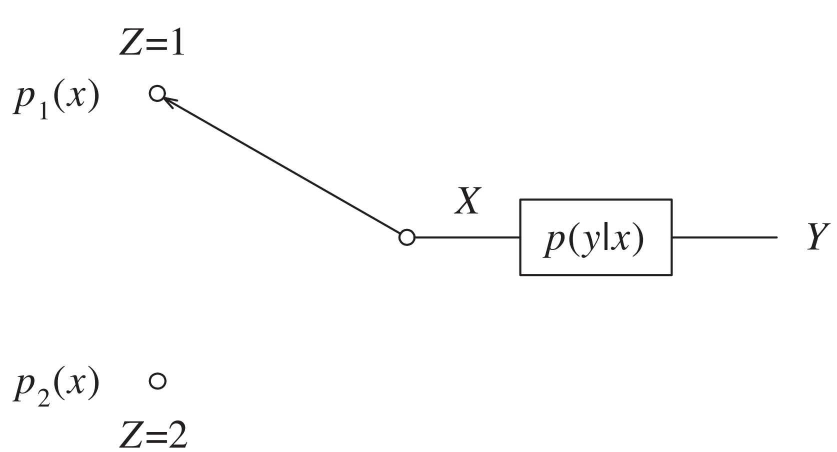
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

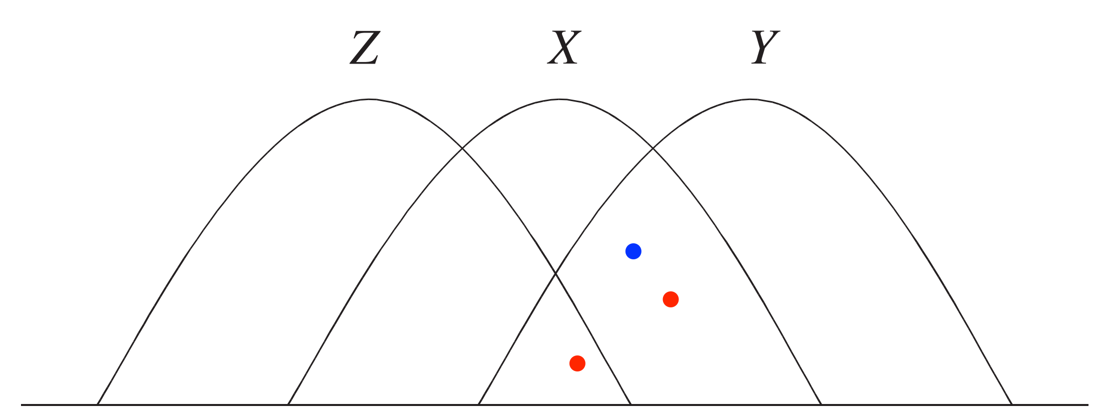
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X; Y) \\ & \quad \geq I(X; Y|Z) \\ & \quad = \underline{\Pr\{Z = 1\}} I(X; Y|Z = 1) \\ & \quad\quad + \Pr\{Z = 2\} I(X; Y|Z = 2) \end{aligned}$$



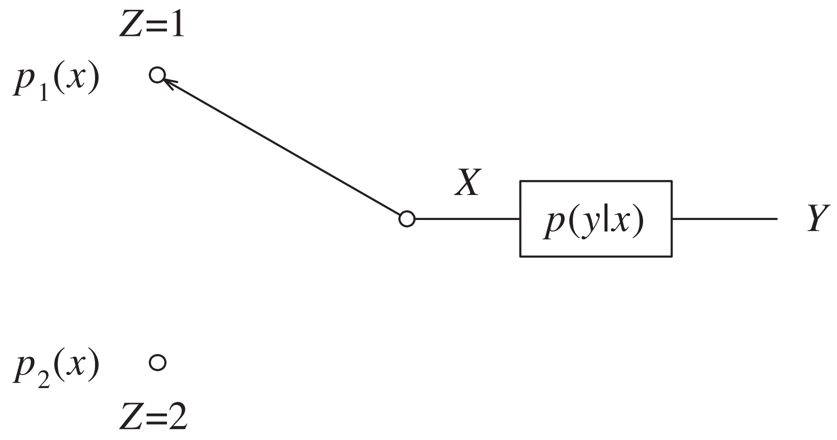
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

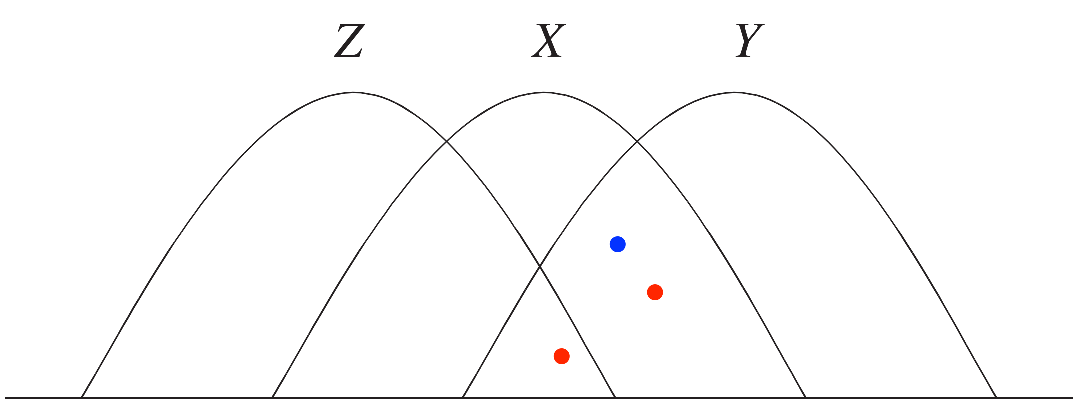
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned}
& I(X;Y) \\
& \quad \geq \quad I(X;Y|Z) \\
& \quad = \quad \Pr\{Z=1\} I(X;Y|Z=1) \\
& \qquad\quad + \Pr\{Z=2\} I(X;Y|Z=2) \\
& \quad = \quad \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)).
\end{aligned}$$



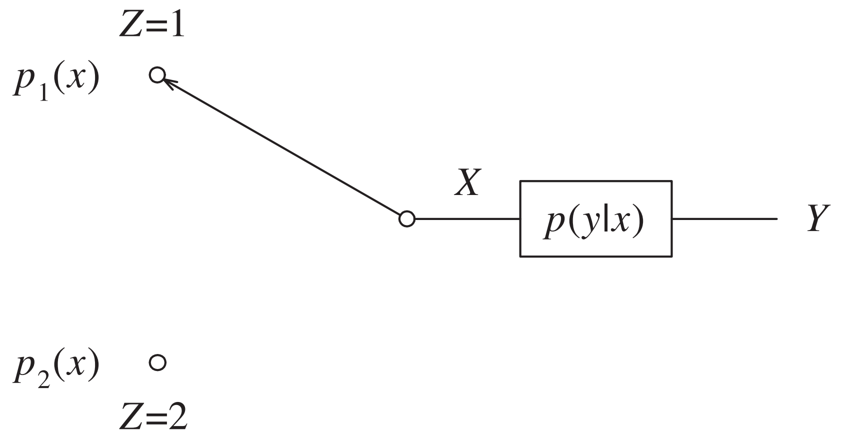
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

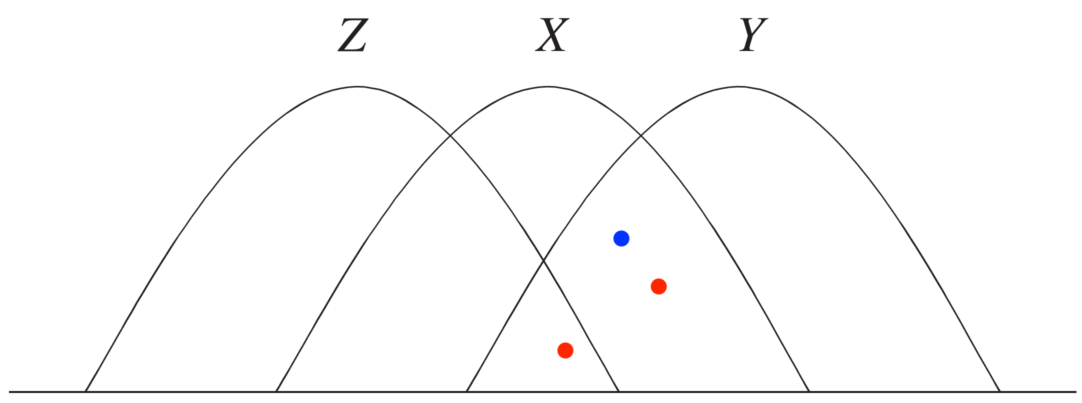
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \rightarrow X \rightarrow Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} I(X; Y) & \\ \geq \quad & I(X; Y|Z) \\ = \quad & \Pr\{Z = 1\}\underline{I(X; Y|Z = 1)} \\ & + \Pr\{Z = 2\}I(X; Y|Z = 2) \\ = \quad & \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$



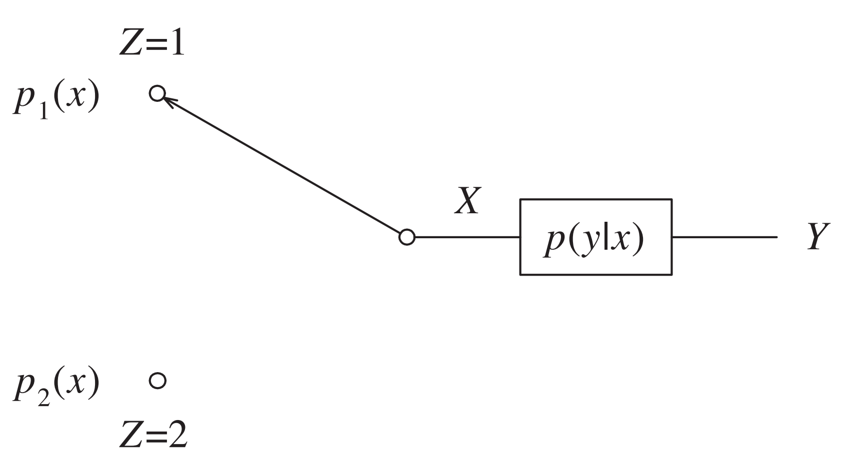
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

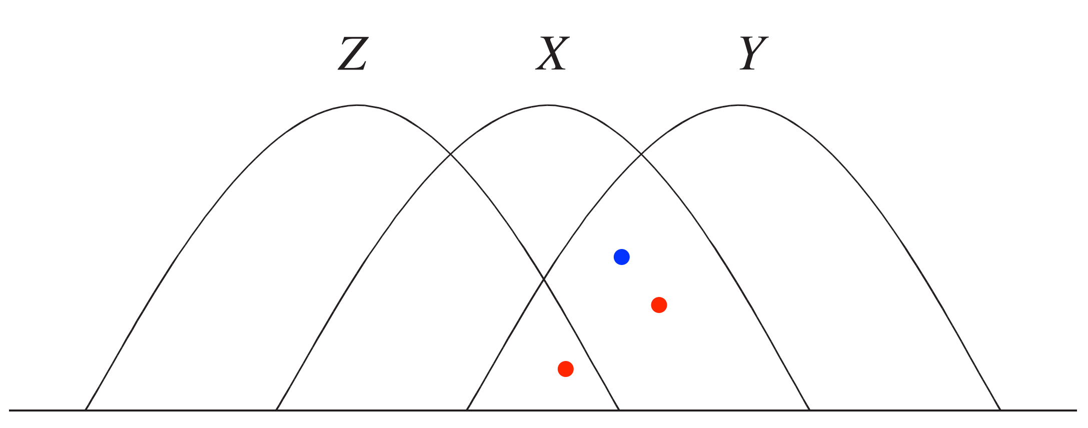
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X; Y) \\ & \quad \geq \quad I(X; Y|Z) \\ & \quad = \quad \Pr\{Z = 1\} I(X; Y|Z = 1) \\ & \quad \quad \quad + \Pr\{Z = 2\} I(X; Y|Z = 2) \\ & \quad = \quad \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$



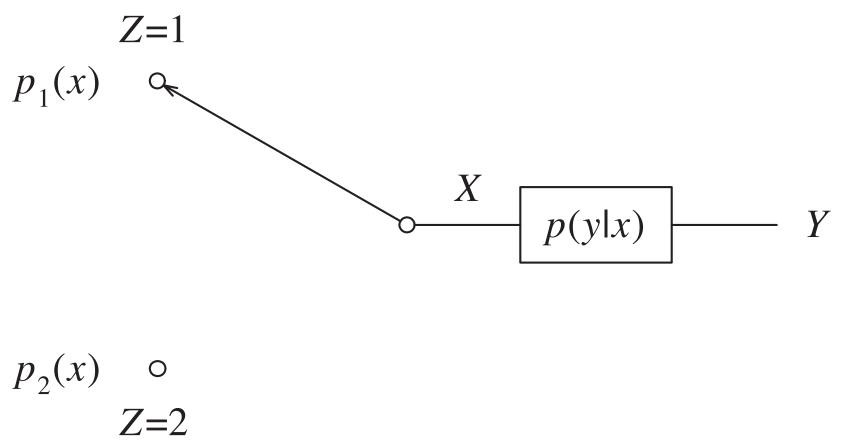
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

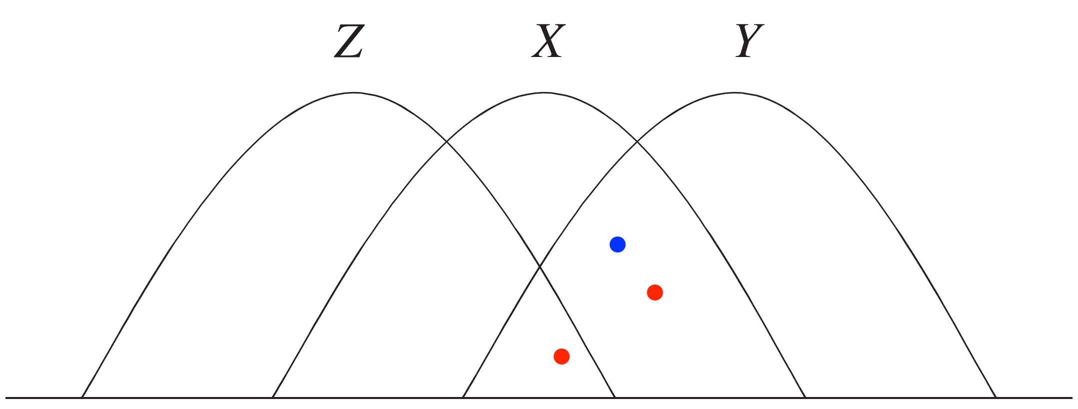
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \rightarrow X \rightarrow Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X;Y) \\ & \quad \geq \quad I(X;Y|Z) \\ & \quad = \quad \Pr\{Z=1\} I(X;Y|Z=1) \\ & \qquad\quad + \Pr\{Z=2\} I(X;Y|Z=2) \\ & \quad = \quad \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$



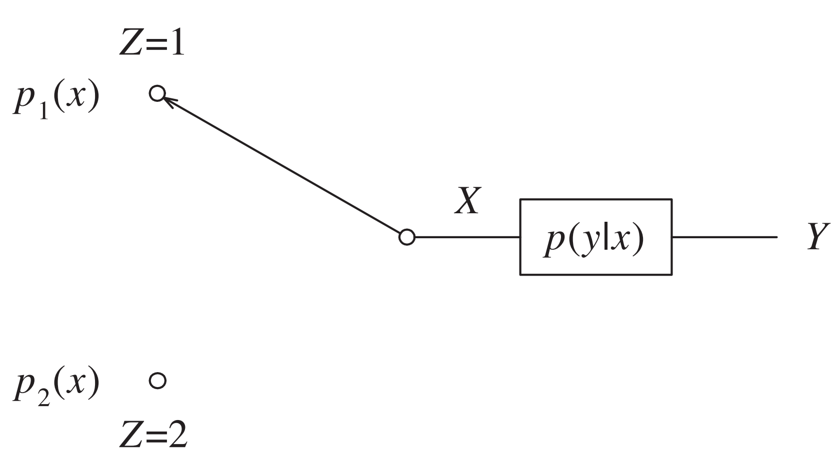
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

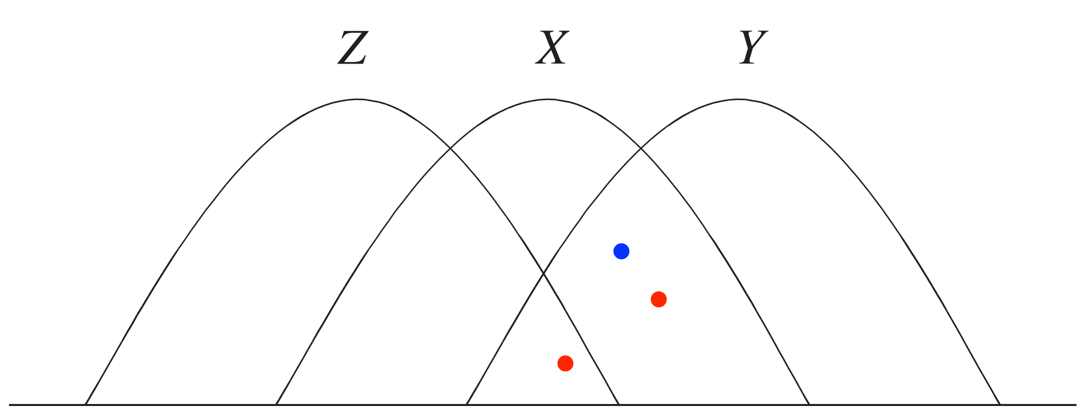
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X;Y) \\ \geq \quad & I(X;Y|Z) \\ = \quad & \Pr\{Z=1\} I(X;Y|Z=1) \\ & + \Pr\{Z=2\} I(X;Y|Z=2) \\ = \quad & \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$



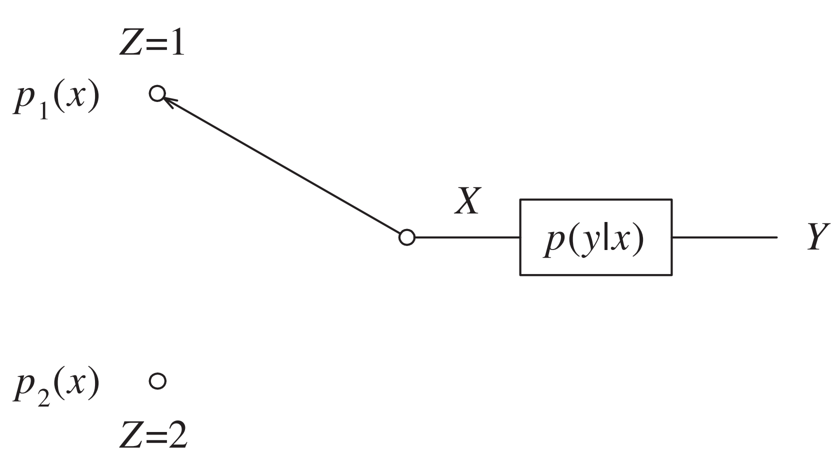
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

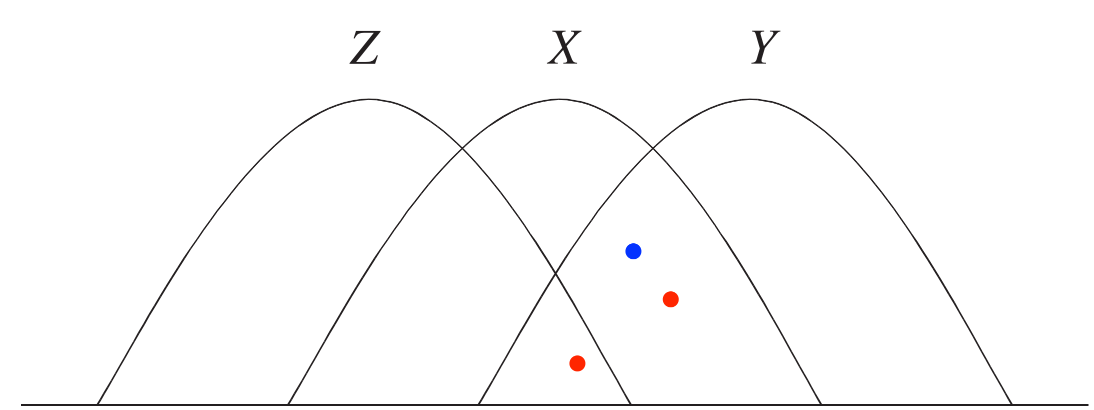
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X; Y) \\ \geq \quad & I(X; Y|Z) \\ = \quad & \Pr\{Z = 1\} I(X; Y|Z = 1) \\ & + \Pr\{Z = 2\} I(X; Y|Z = 2) \\ = \quad & \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$



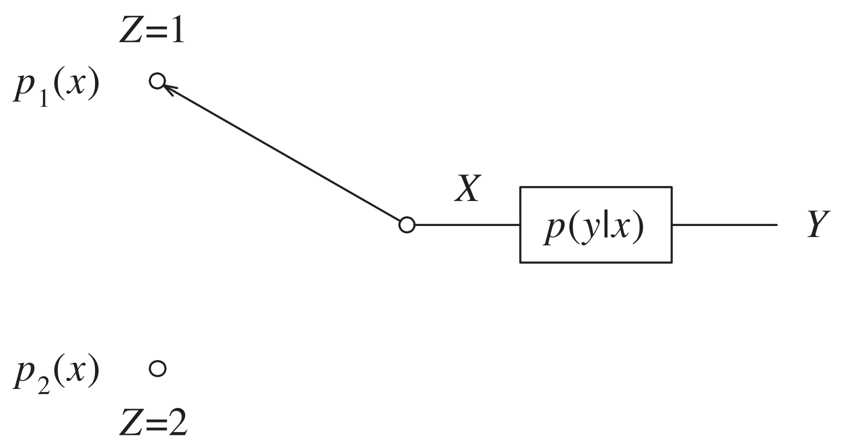
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

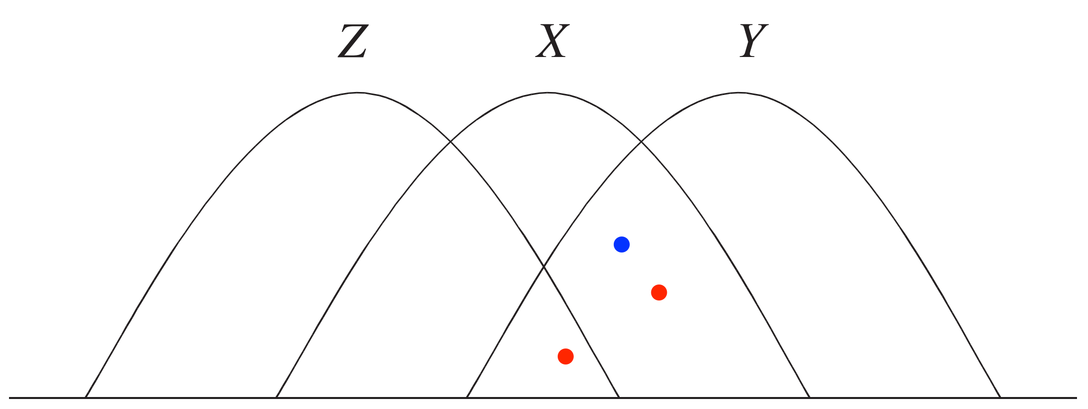
1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X;Y) \\ \geq \quad & I(X;Y|Z) \\ = \quad & \Pr\{Z=1\}I(X;Y|Z=1) \\ & + \Pr\{Z=2\}I(X;Y|Z=2) \\ = \quad & \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$



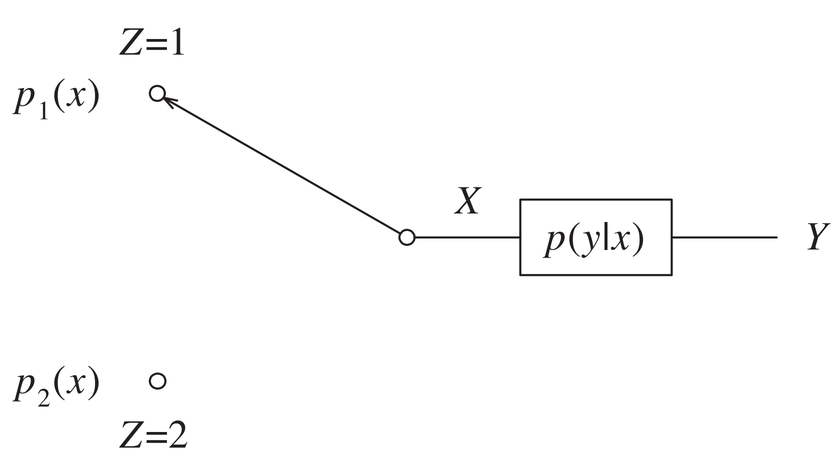
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

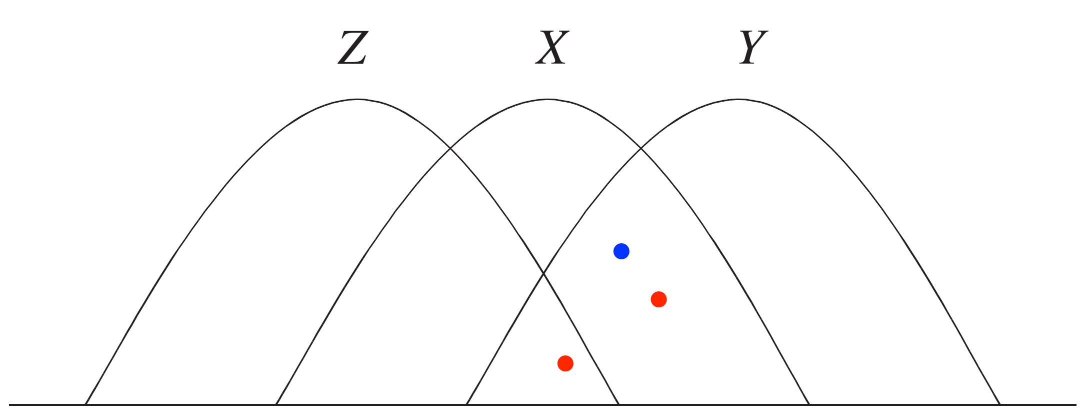
2. In this setup, when X is given, Y is independent of Z, or $Z \rightarrow X \rightarrow Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.



3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X; Y) \\ & \quad \geq I(X; Y|Z) \\ & \quad = \Pr\{Z = 1\}I(X; Y|Z = 1) \\ & \quad\quad + \Pr\{Z = 2\}I(X; Y|Z = 2) \\ & \quad = \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$

This shows that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.



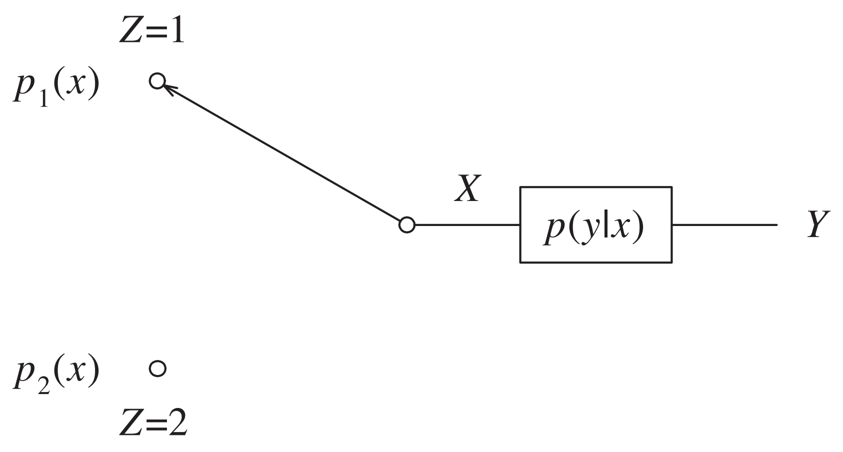
Example 3.14 (Concavity of Mutual Information) Let

$$(X, Y) \sim p(x, y) = p(x)p(y|x).$$

Show that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

1. Consider the system as shown, where the position of the switch is determined by a random variable Z as in the last example.

2. In this setup, when X is given, Y is independent of Z, or $Z \to X \to Y$ forms a Markov chain. Then μ^* is nonnegative, and the information diagram for X, Y, and Z is as shown.

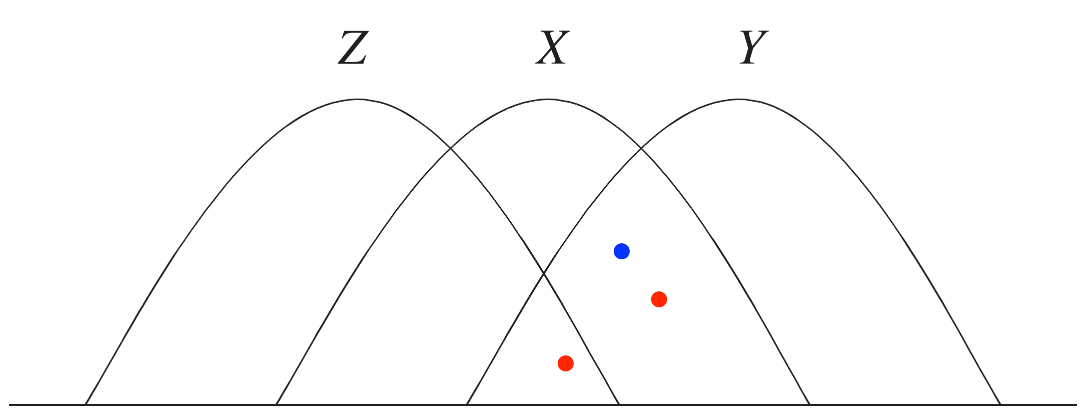


3. From the information diagram, since $\tilde{X} \cap \tilde{Y} - \tilde{Z}$ is a subset of $\tilde{X} \cap \tilde{Y}$ and μ^* is nonnegative, we immediately see that

$$\begin{aligned} & I(X; Y) \\ \geq \quad & I(X; Y|Z) \\ = \quad & \Pr\{Z = 1\} I(X; Y|Z = 1) \\ & + \Pr\{Z = 2\} I(X; Y|Z = 2) \\ = \quad & \lambda I(p_1(x), p(y|x)) + \bar{\lambda} I(p_2(x), p(y|x)). \end{aligned}$$

This shows that for fixed $p(y|x)$, $I(X; Y)$ is a concave functional of $p(x)$.

Interpretation For a fixed channel, by mixing the input distribution, the mutual information is at least equal to the mixture of the corresponding mutual informations.

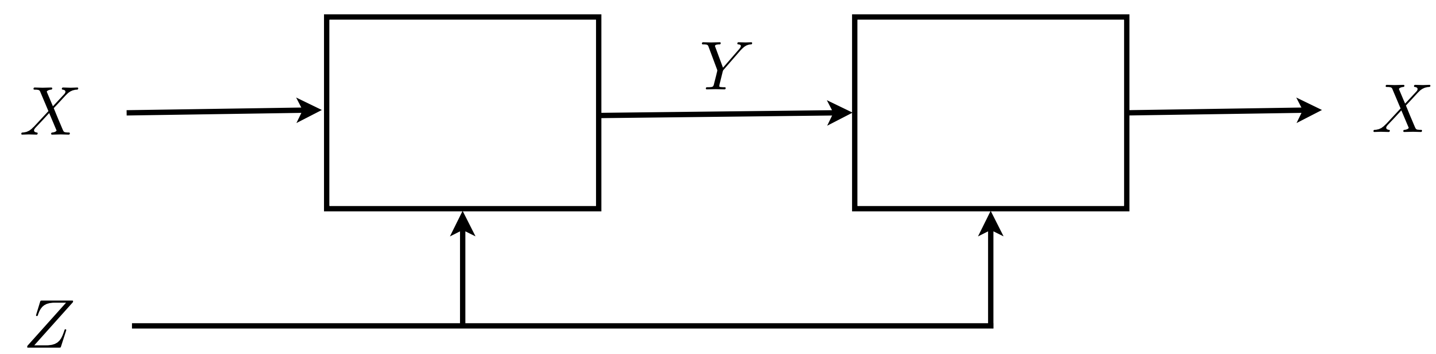


Shannon's Perfect Secrecy Theorem

X – plaintext

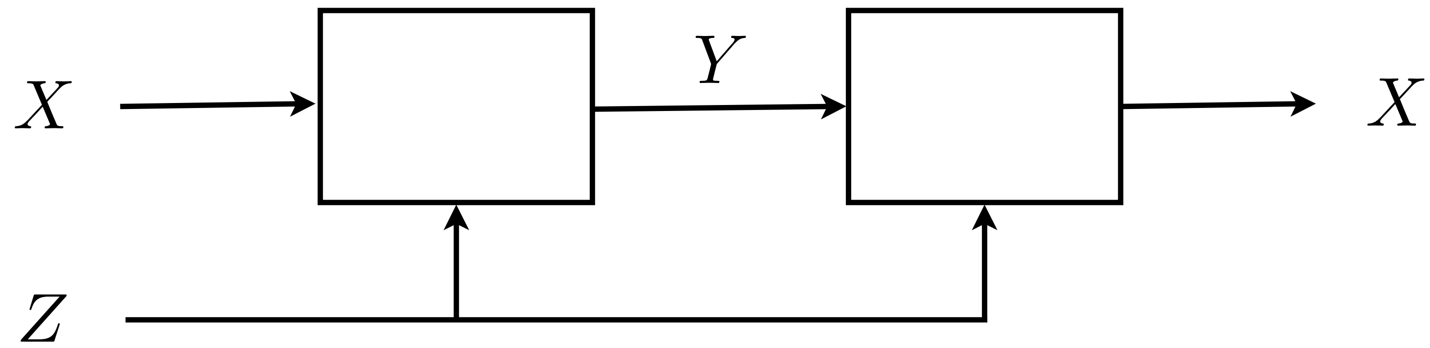
Y – ciphertext

Z – key



Shannon's Perfect Secrecy Theorem

X – plaintext
Y – ciphertext
Z – key



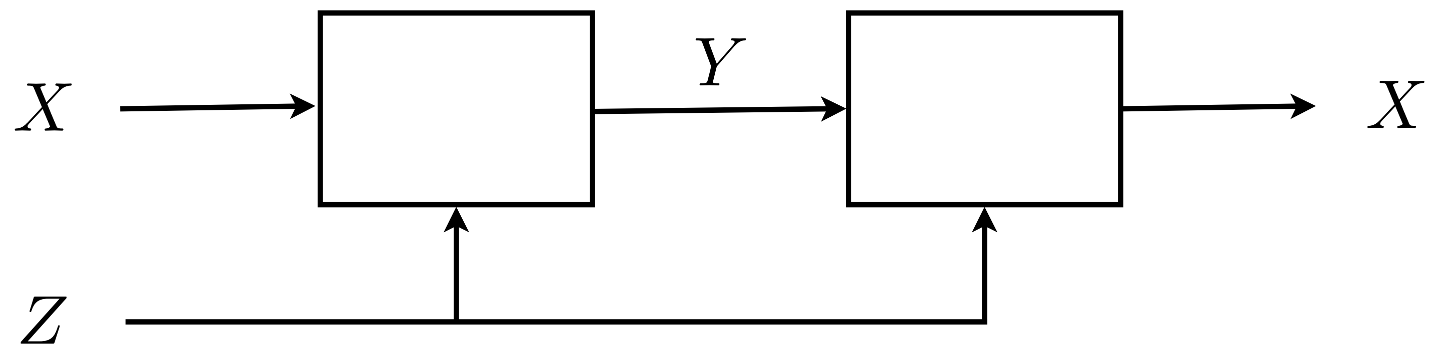
- Perfect Secrecy: $I(X;Y) = 0$

Shannon's Perfect Secrecy Theorem

X – plaintext

Y – ciphertext

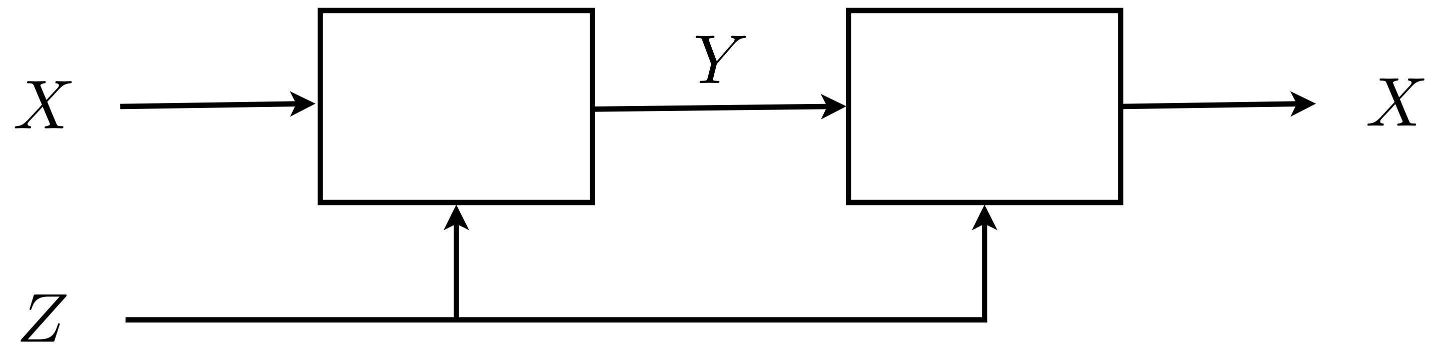
Z – key



- Perfect Secrecy: $I(X;Y) = 0$
- Decipherability: $H(X|Y,Z) = 0$

Shannon's Perfect Secrecy Theorem

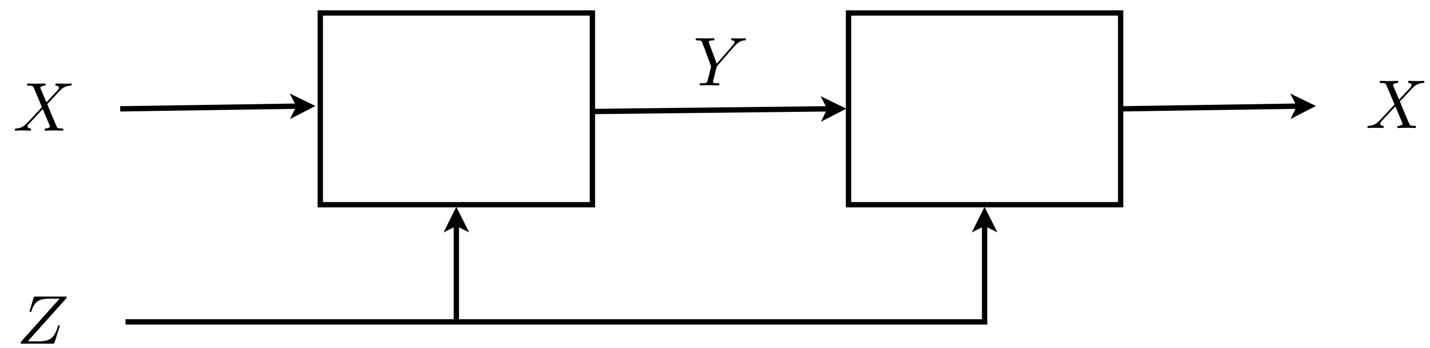
X – plaintext
Y – ciphertext
Z – key



- Perfect Secrecy: $I(X;Y) = 0$
- Decipherability: $H(X|Y,Z) = 0$
- These requirement implies $H(Z) \geq H(X)$, i.e., the length of the key is at least the same as the length of the plaintext.

Shannon's Perfect Secrecy Theorem

X – plaintext
Y – ciphertext
Z – key



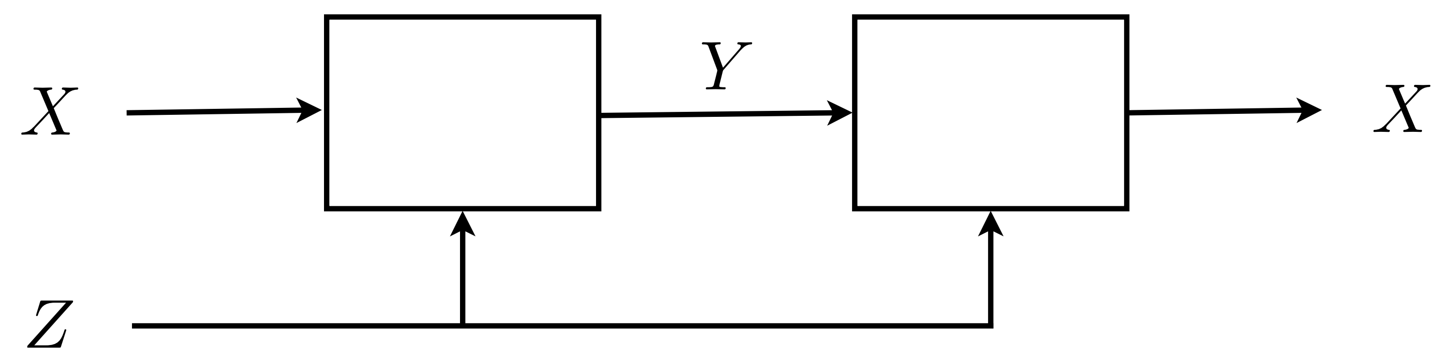
- Perfect Secrecy: $I(X;Y) = 0$
- Decipherability: $H(X|Y,Z) = 0$
- These requirement implies $H(Z) \geq H(X)$, i.e., the length of the key is at least the same as the length of the plaintext.
- Shannon (1949) gave a combinatorial proof.

Shannon's Perfect Secrecy Theorem

X – plaintext

Y – ciphertext

Z – key

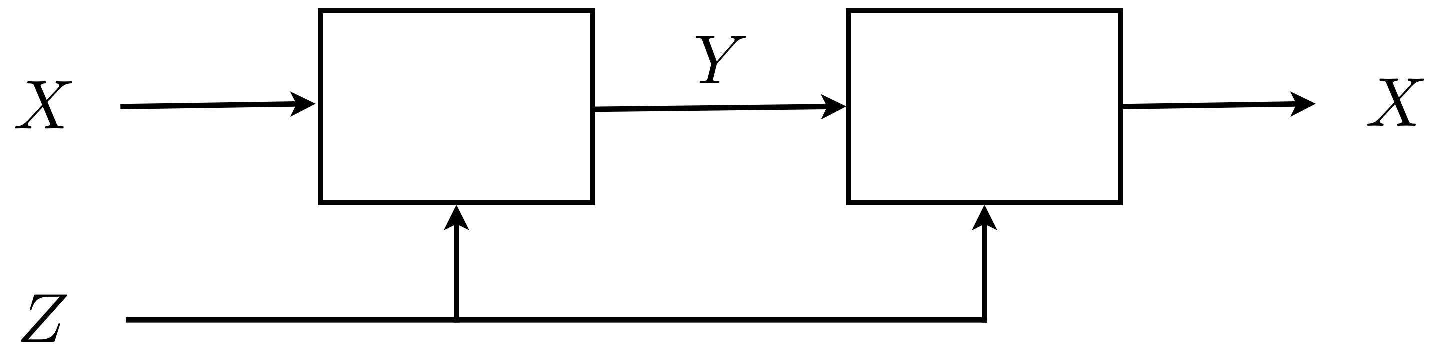


- Perfect Secrecy: $I(X;Y) = 0$
- Decipherability: $H(X|Y,Z) = 0$
- These requirement implies $H(Z) \geq H(X)$, i.e., the length of the key is at least the same as the length of the plaintext.
- Shannon (1949) gave a combinatorial proof.
- Can readily be proved by an information diagram.

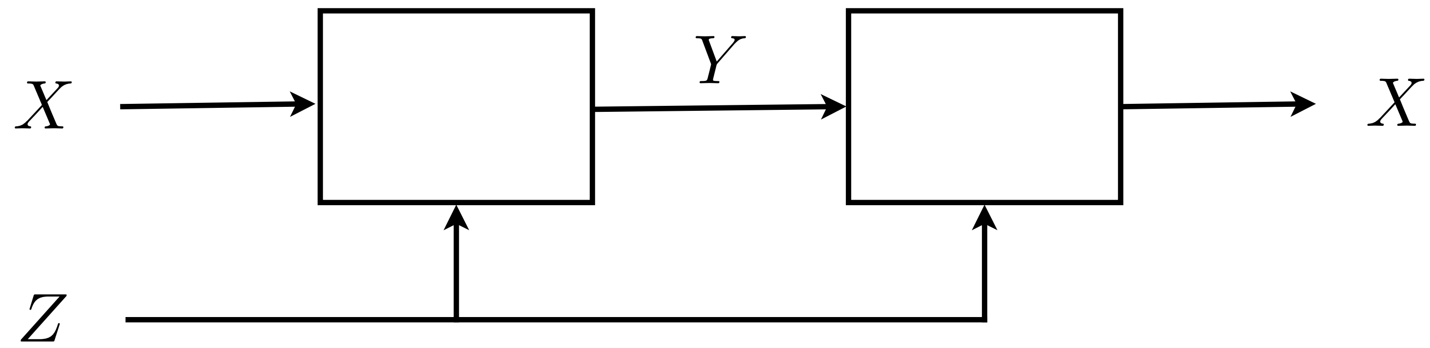
X – plaintext

Y – ciphertext

Z – key

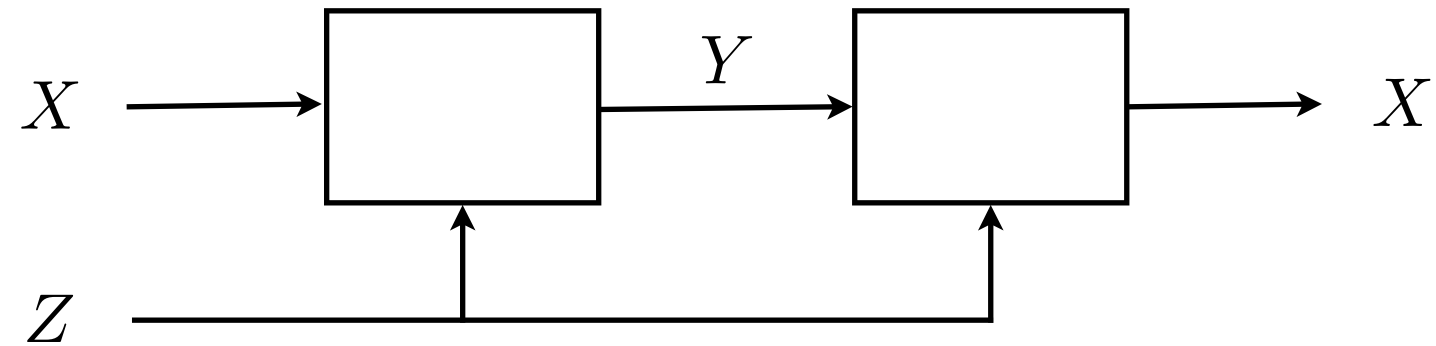


X – plaintext
Y – ciphertext
Z – key

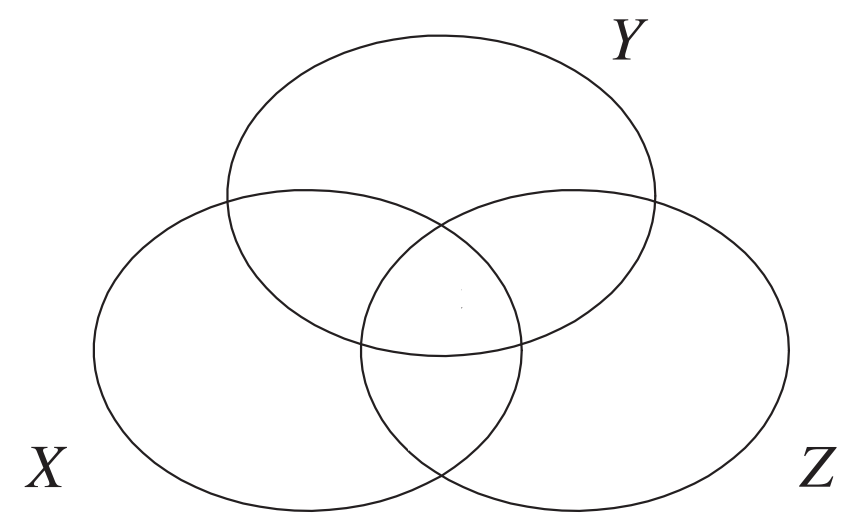


Perfect Secrecy $I(X;Y) = 0$

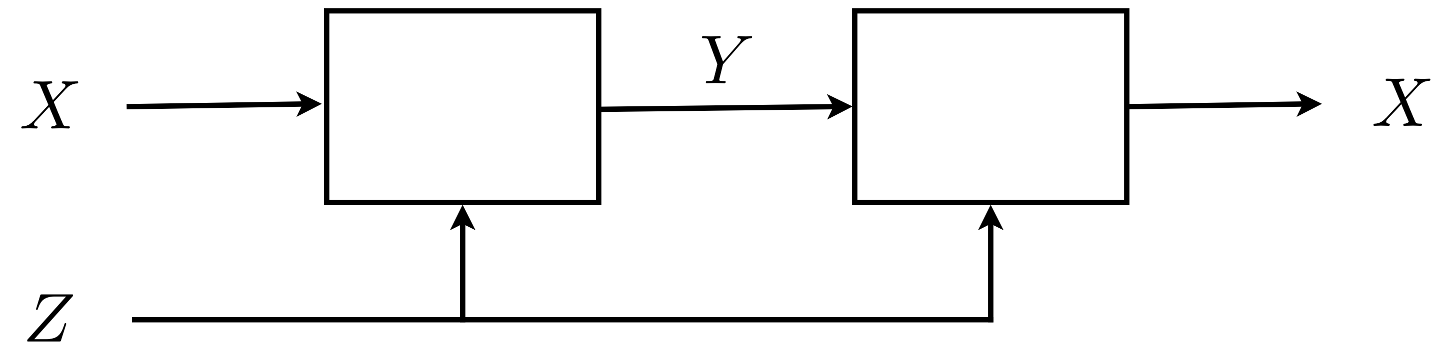
X	–	plaintext
Y	–	ciphertext
Z	–	key



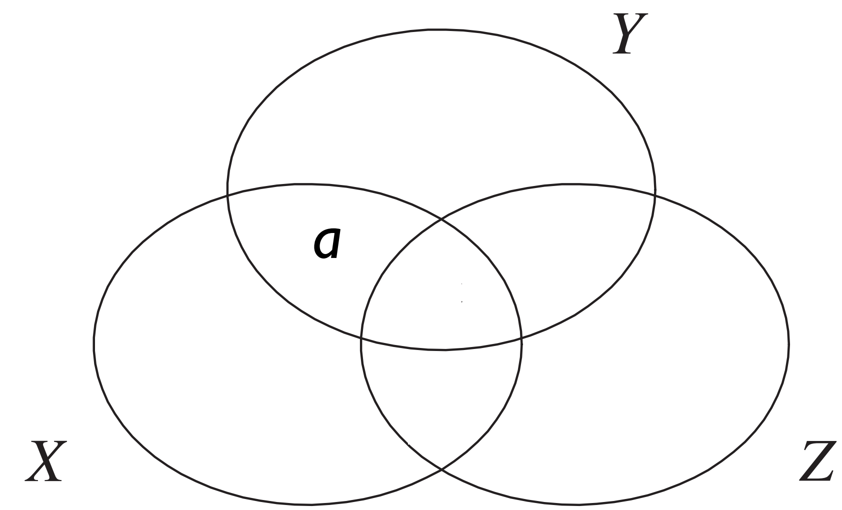
Perfect Secrecy $I(X;Y) = 0$



X	–	plaintext
Y	–	ciphertext
Z	–	key



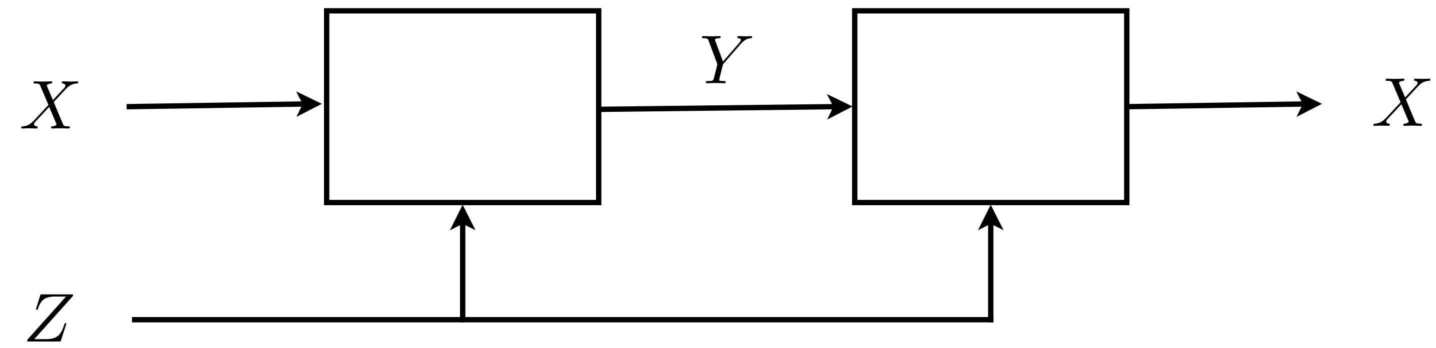
Perfect Secrecy $I(X;Y) = 0$



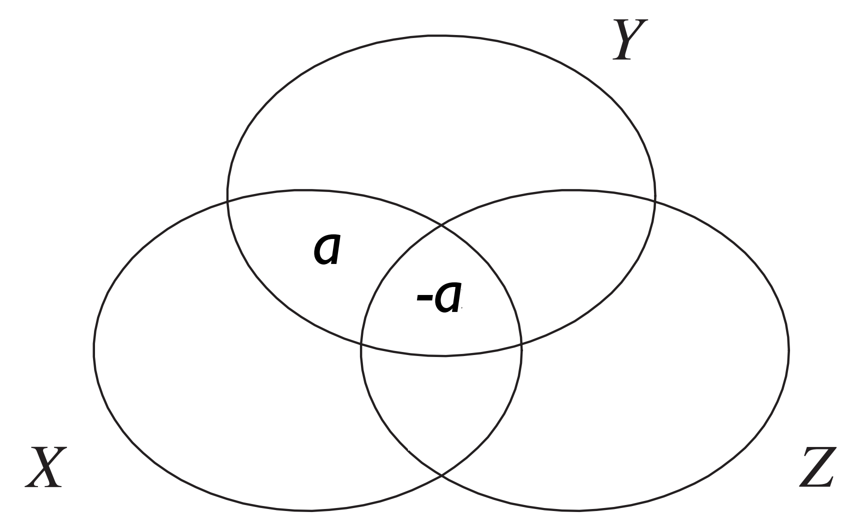
X – plaintext

Y – ciphertext

Z – key



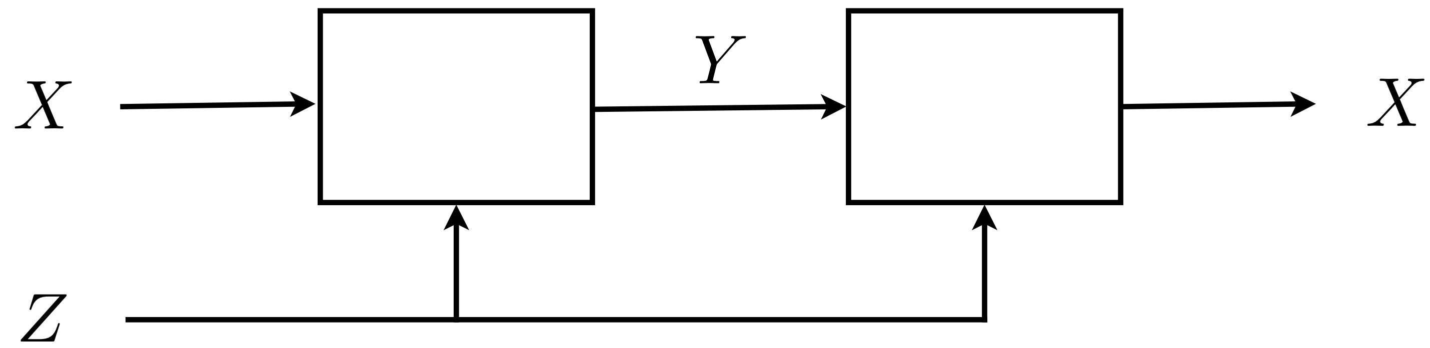
Perfect Secrecy $I(X;Y) = 0$



X – plaintext

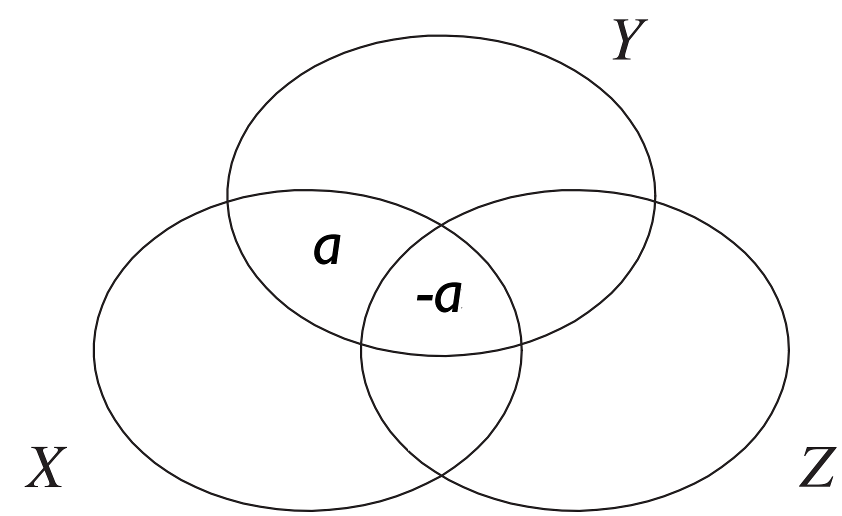
Y – ciphertext

Z – key

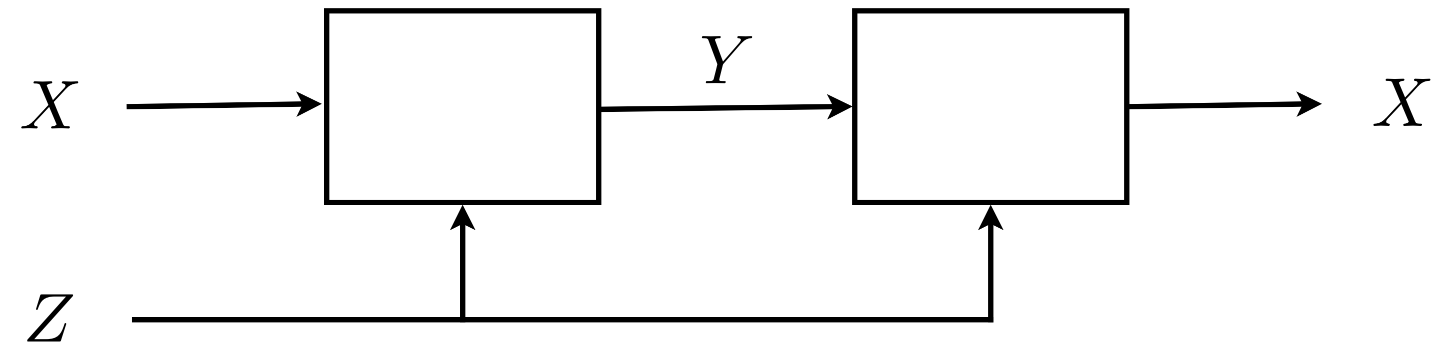


Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y, Z) = 0$

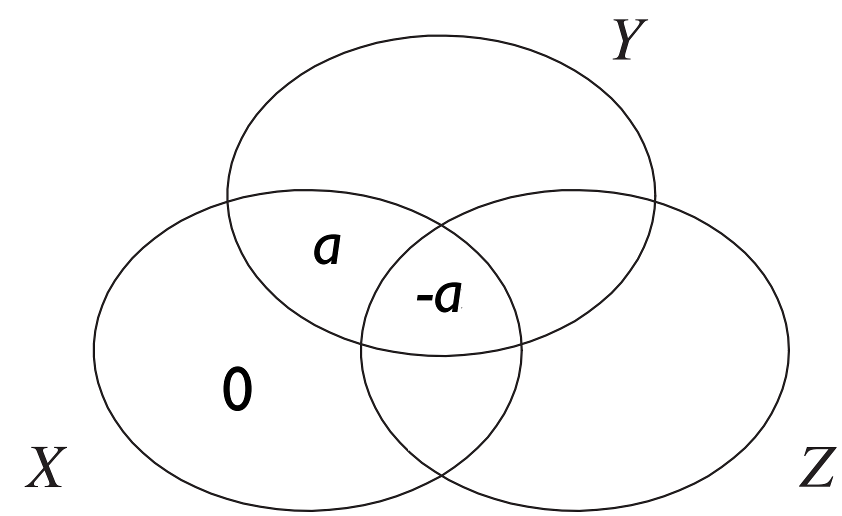


X – plaintext
Y – ciphertext
Z – key



Perfect Secrecy $I(X;Y)=0$

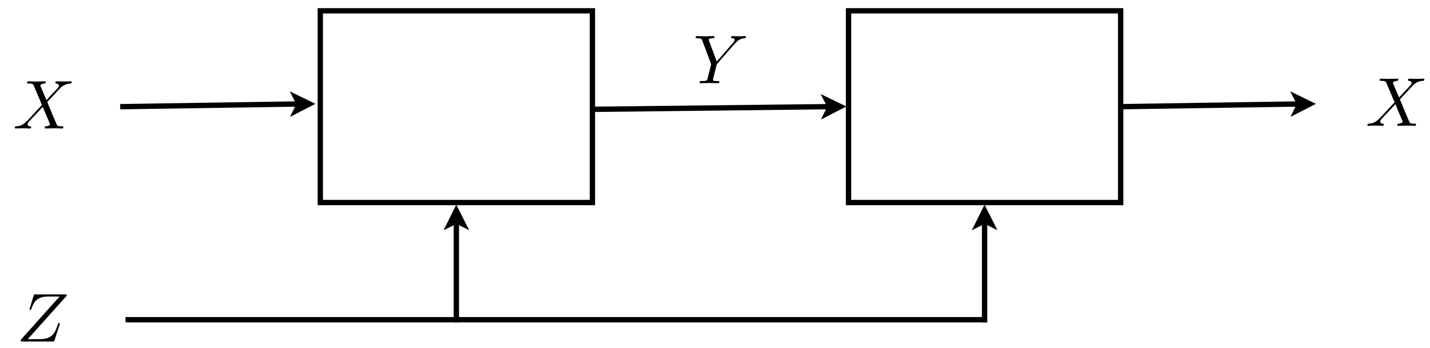
Decipherbility $H(X|Y,Z)=0$



X — plaintext

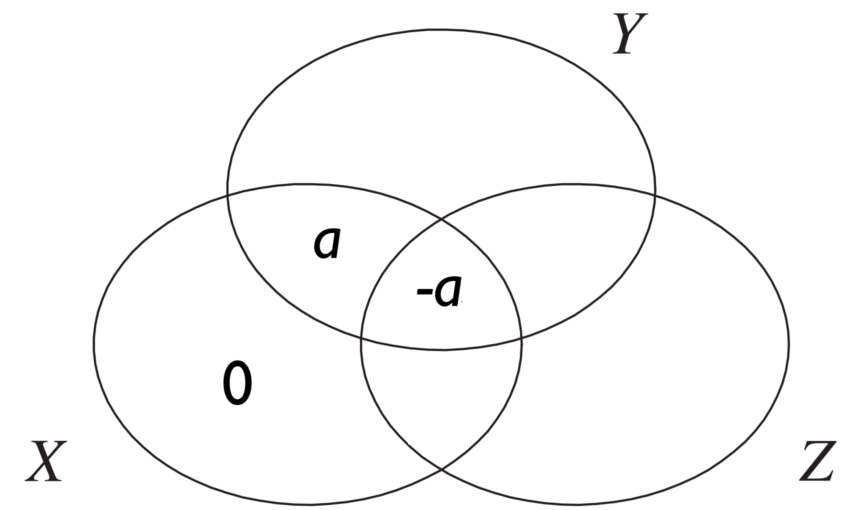
Y — ciphertext

Z — key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

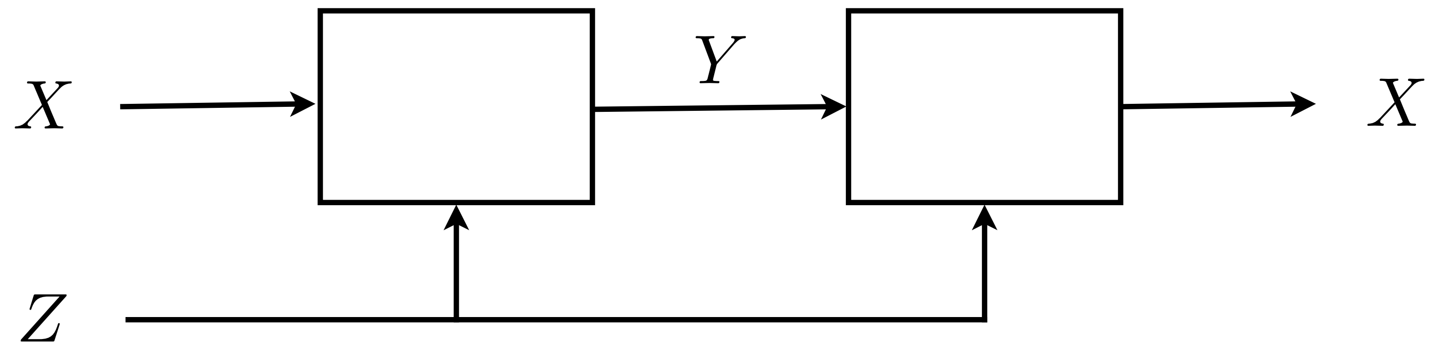


1. Since $I(Y;Z) \geq 0$, we have

X – plaintext

Y – ciphertext

Z – key

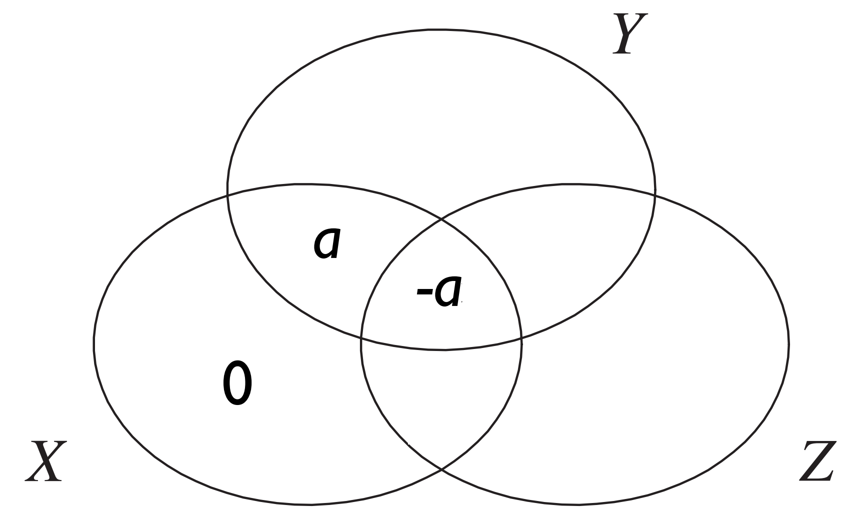


Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

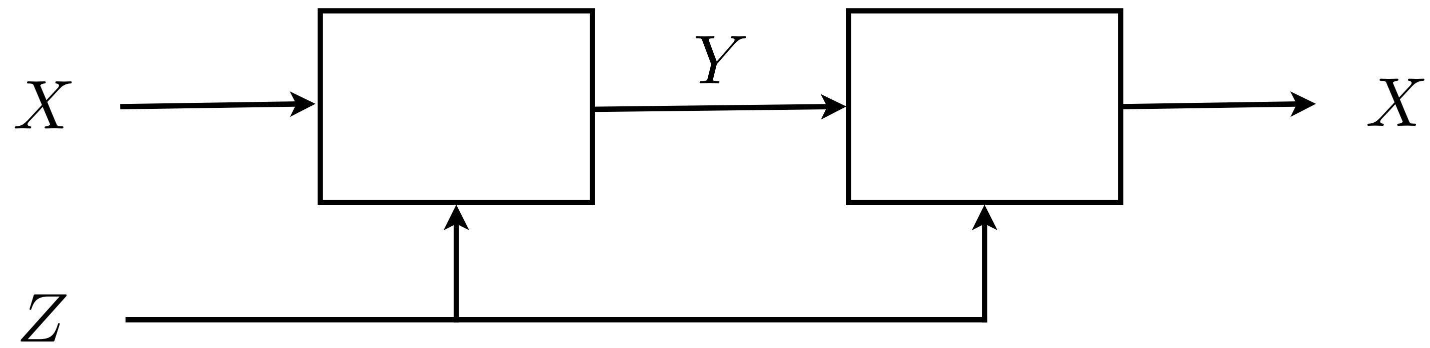
$$I(Y;Z|X) \geq a.$$



X — plaintext

Y — ciphertext

Z — key

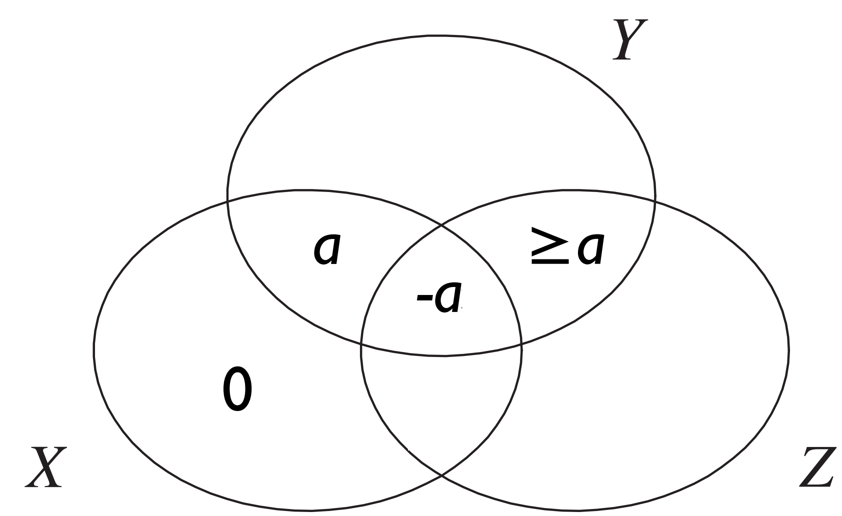


Perfect Secrecy $I(X;Y) = 0$

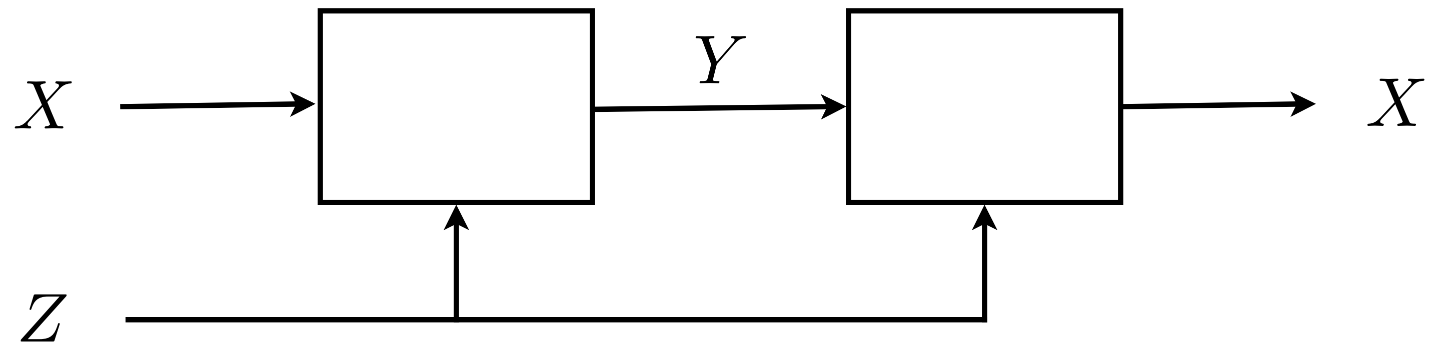
Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$



X – plaintext
Y – ciphertext
Z – key



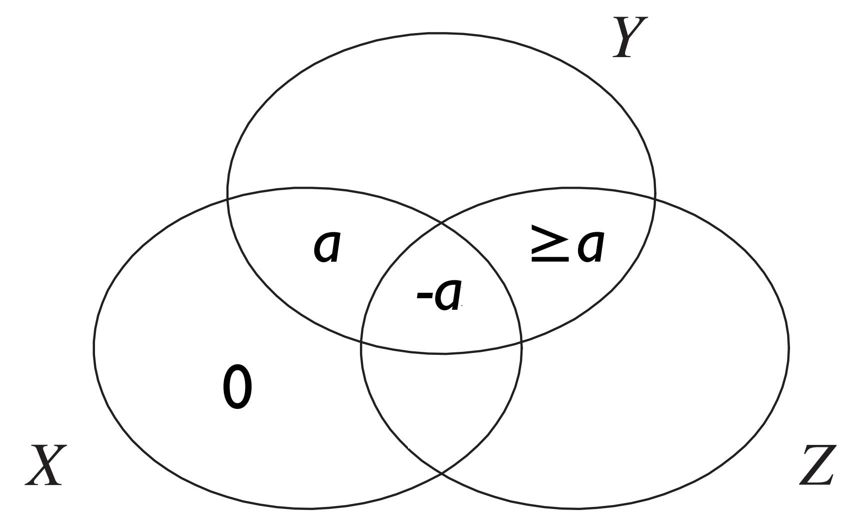
Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

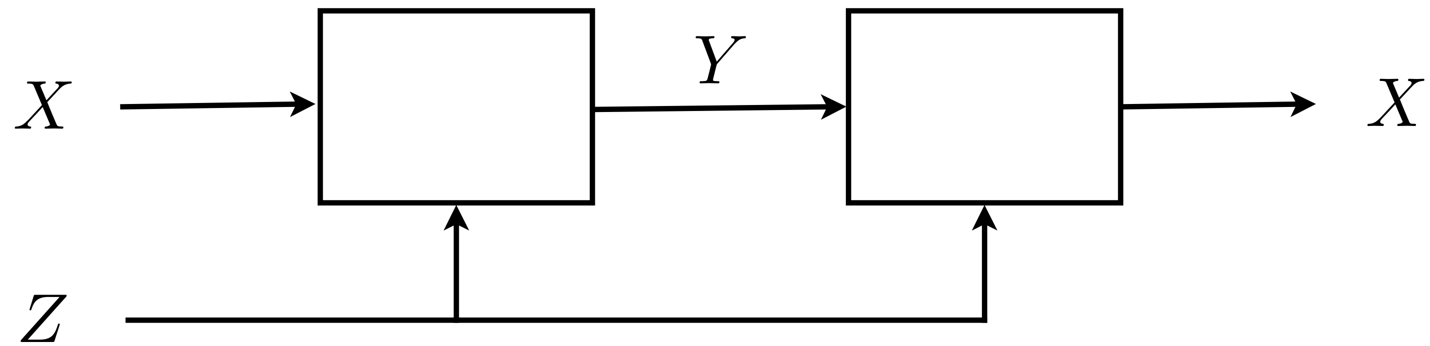
1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.



X – plaintext
Y – ciphertext
Z – key



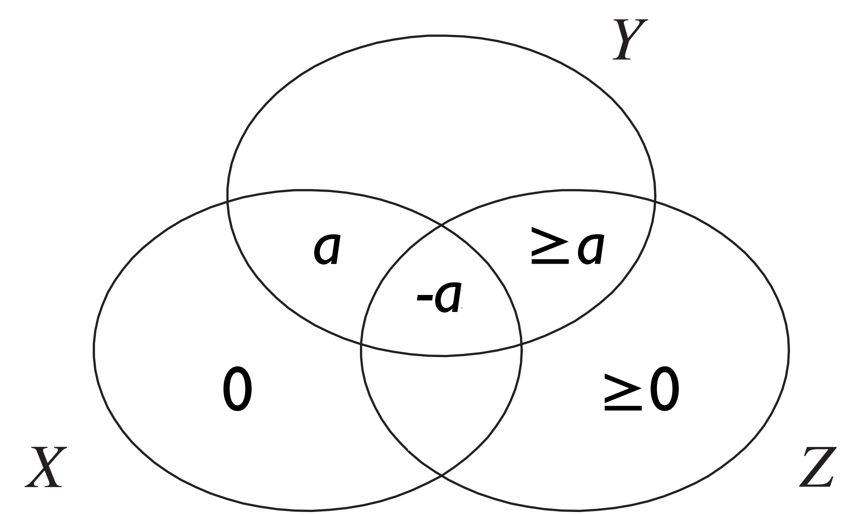
Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

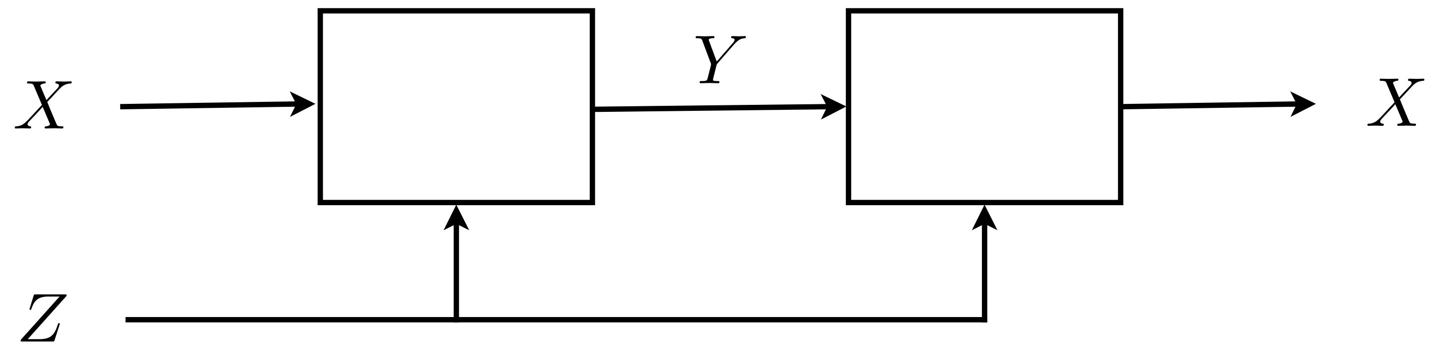
1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.



X — plaintext
Y — ciphertext
Z — key



Perfect Secrecy $I(X;Y) = 0$

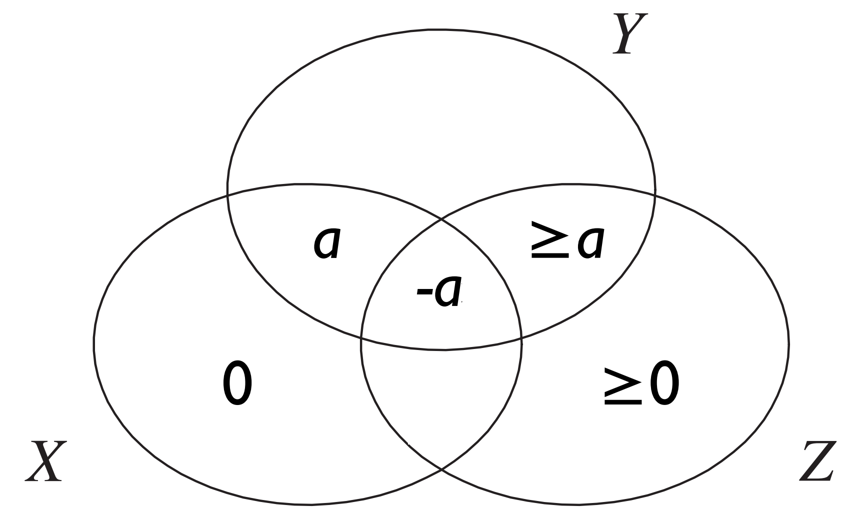
Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

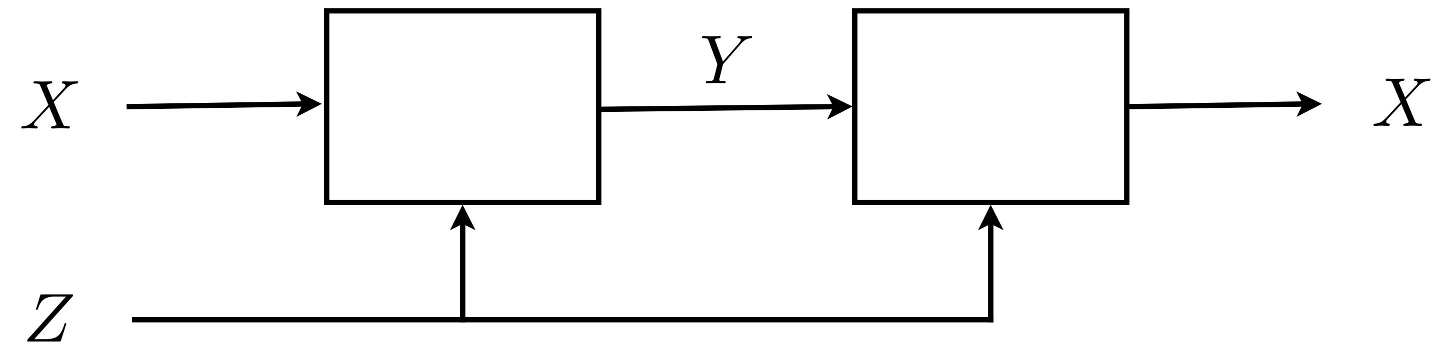
$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.

3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.



X – plaintext
Y – ciphertext
Z – key



Perfect Secrecy $I(X;Y) = 0$

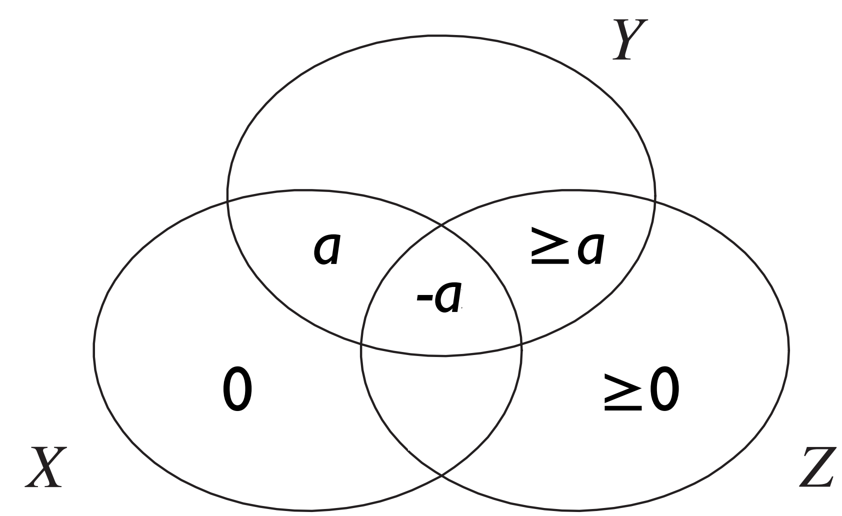
Decipherbility $H(X|Y,Z) = 0$

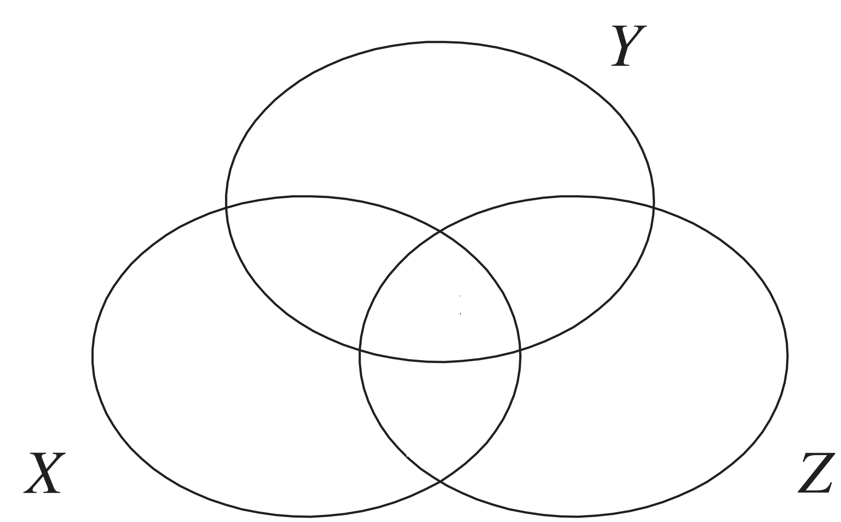
1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.

3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

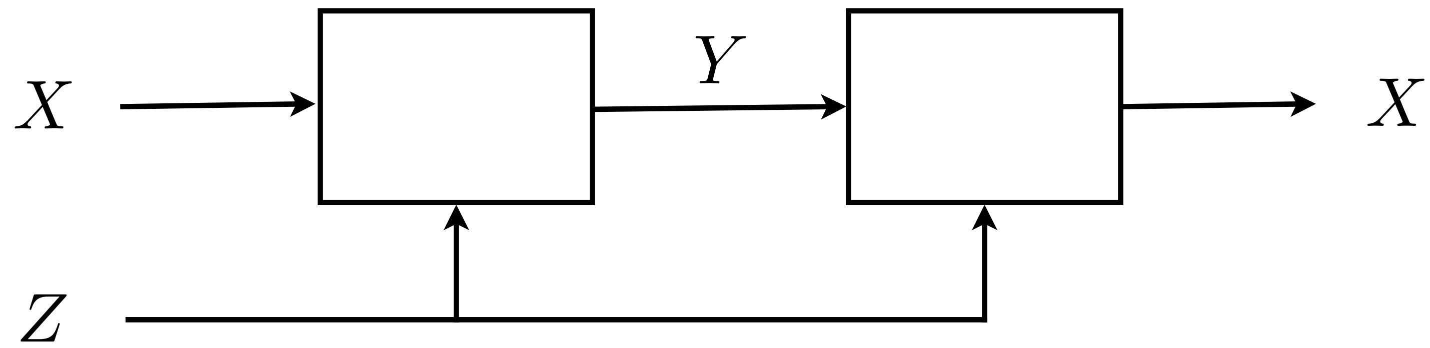




X — plaintext

Y — ciphertext

Z — key



Perfect Secrecy $I(X;Y) = 0$

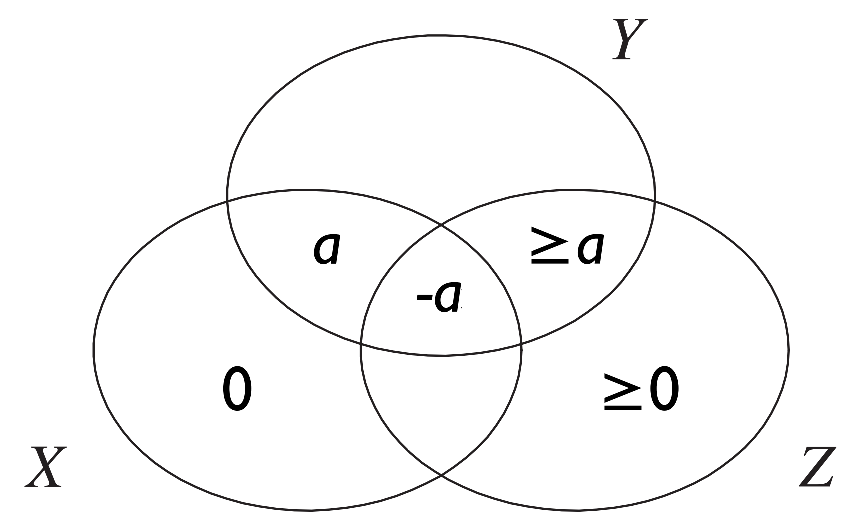
Decipherbility $H(X|Y,Z) = 0$

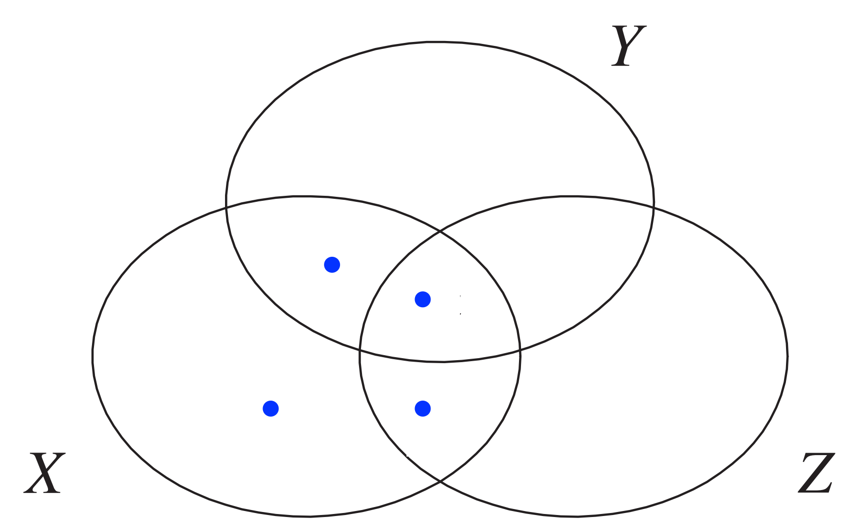
1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

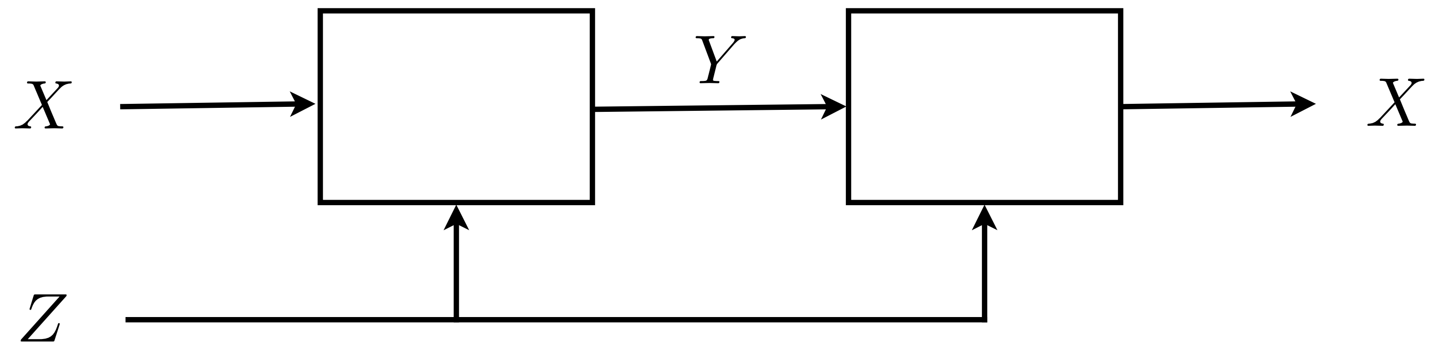
2. We also have $H(Z|X,Y) \geq 0$.

3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.





X — plaintext
Y — ciphertext
Z — key



Perfect Secrecy $I(X;Y) = 0$

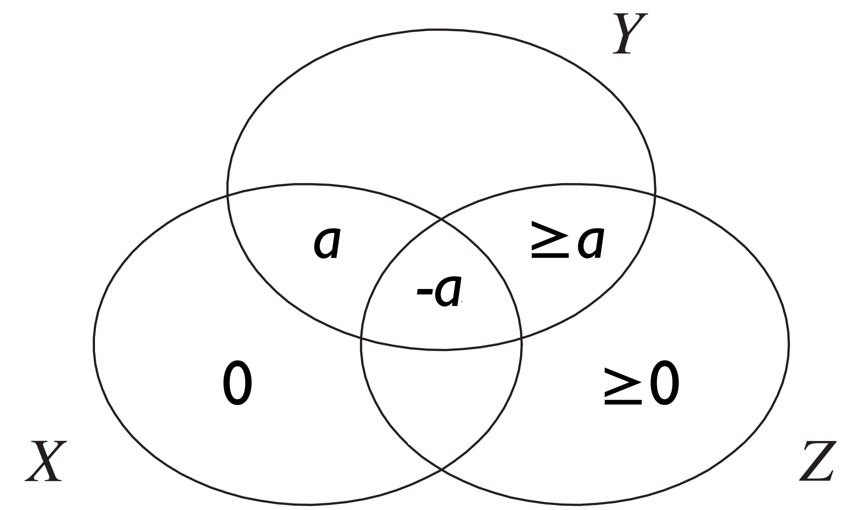
Decipherbility $H(X|Y,Z) = 0$

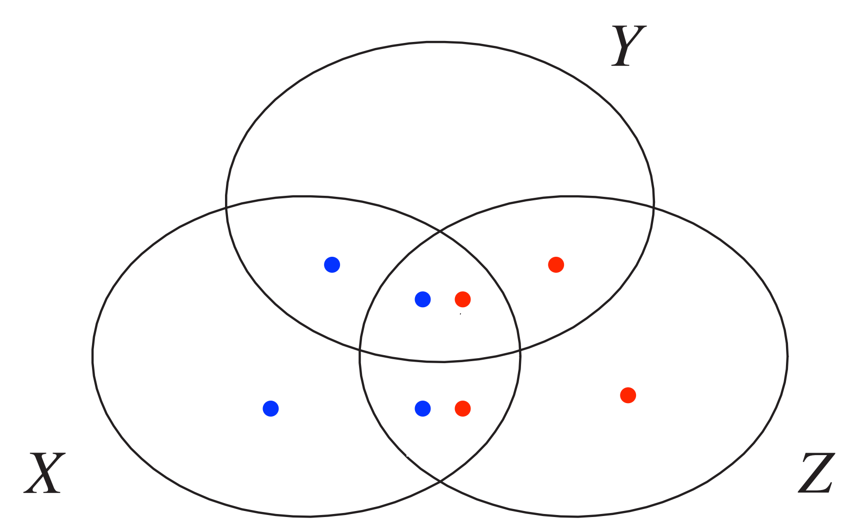
1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.

3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

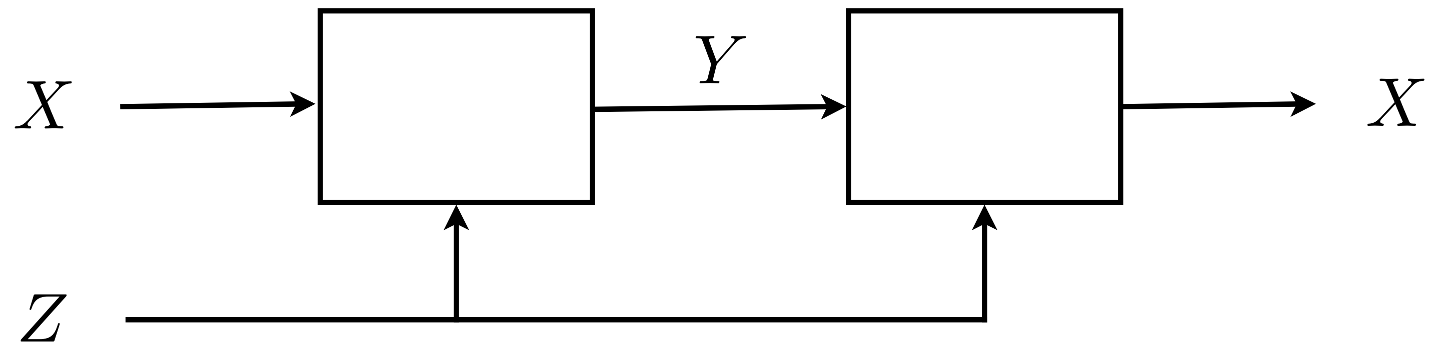




X – plaintext

Y – ciphertext

Z – key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

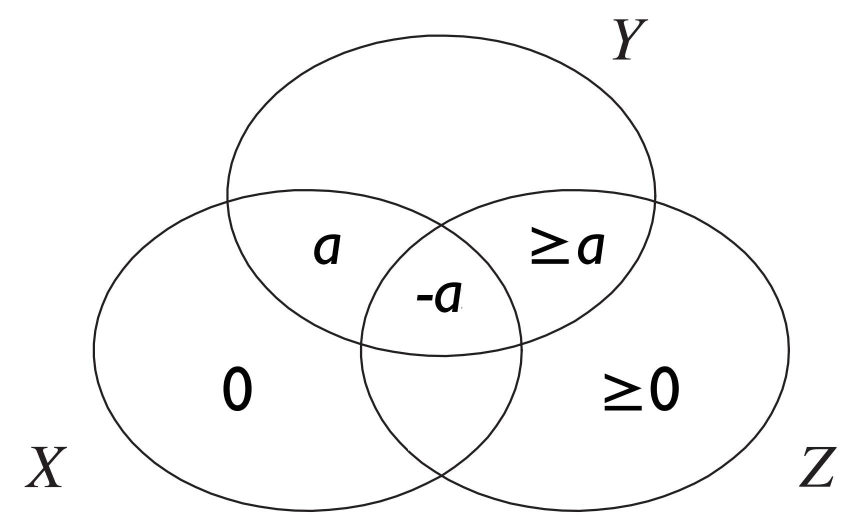
1. Since $I(Y;Z) \geq 0$, we have

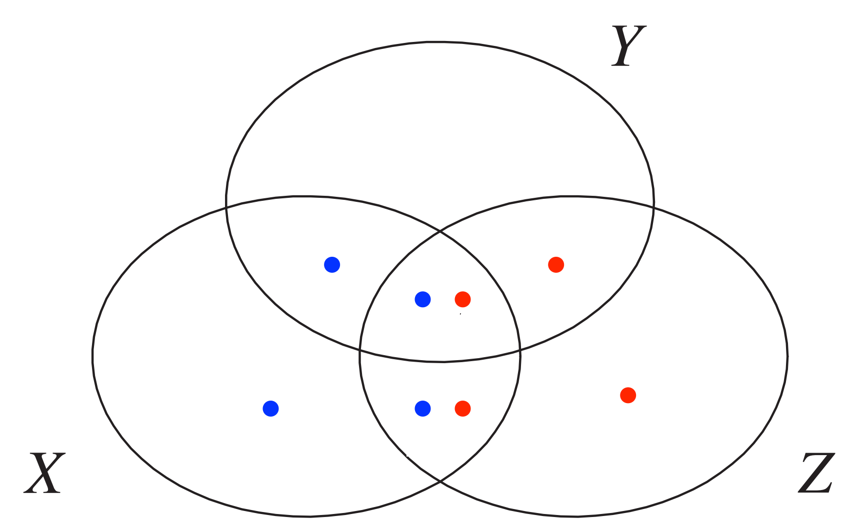
$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.

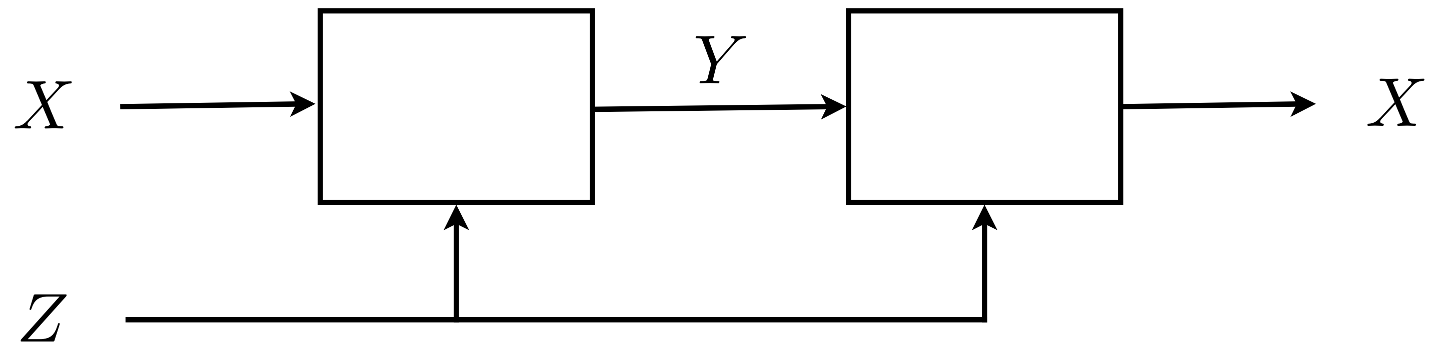
3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.





X — plaintext
Y — ciphertext
Z — key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

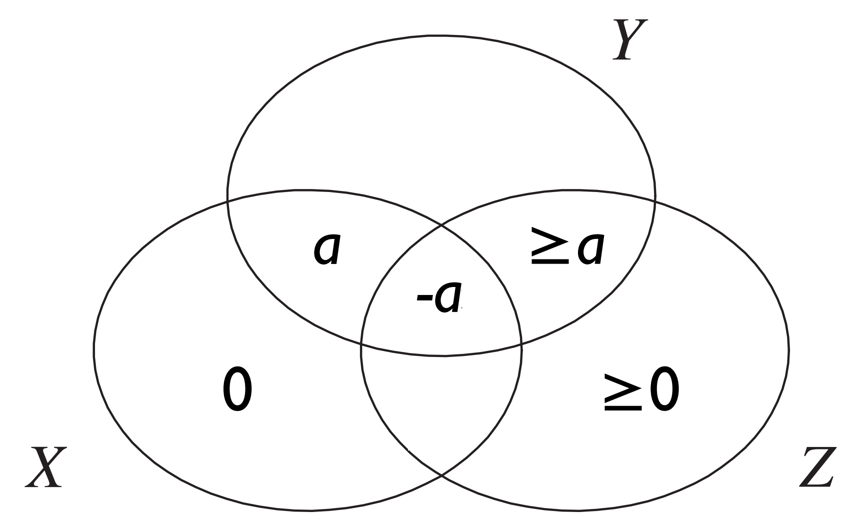
1. Since $I(Y;Z) \geq 0$, we have

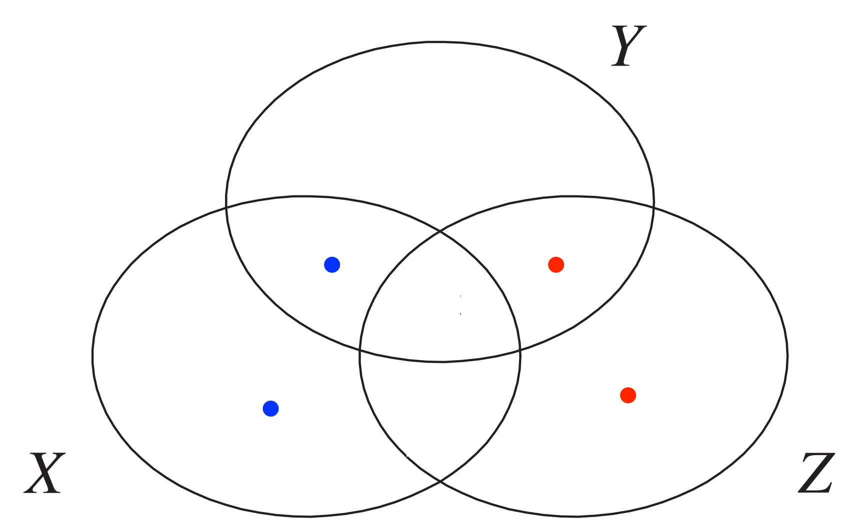
$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.

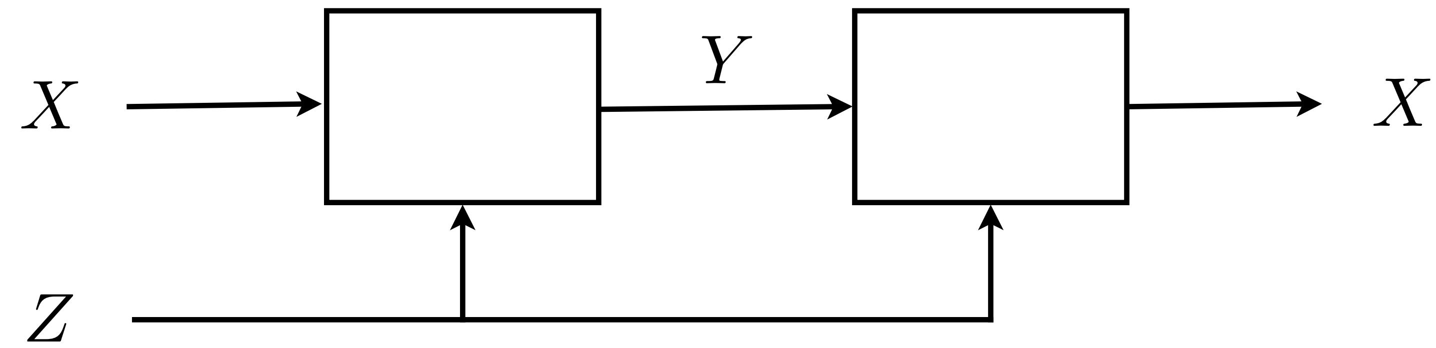
3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

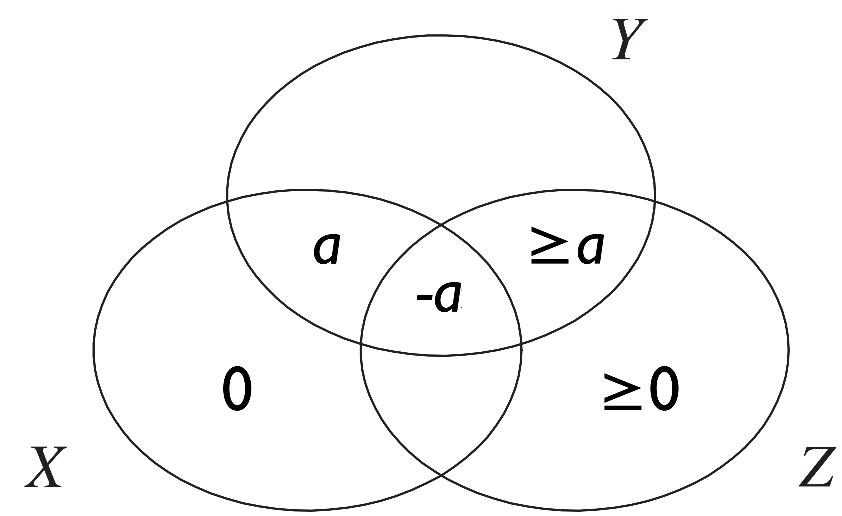
4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.





X — plaintext
Y — ciphertext
Z — key





Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

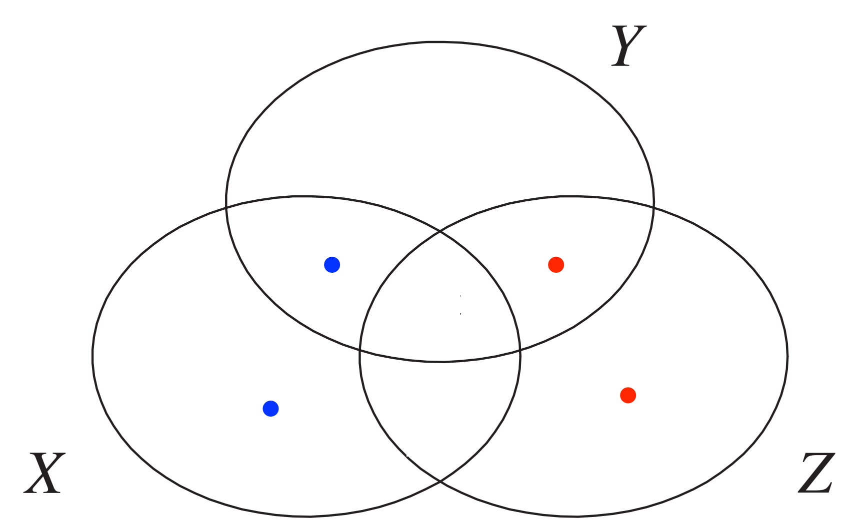
2. We also have $H(Z|X,Y) \geq 0$.

3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

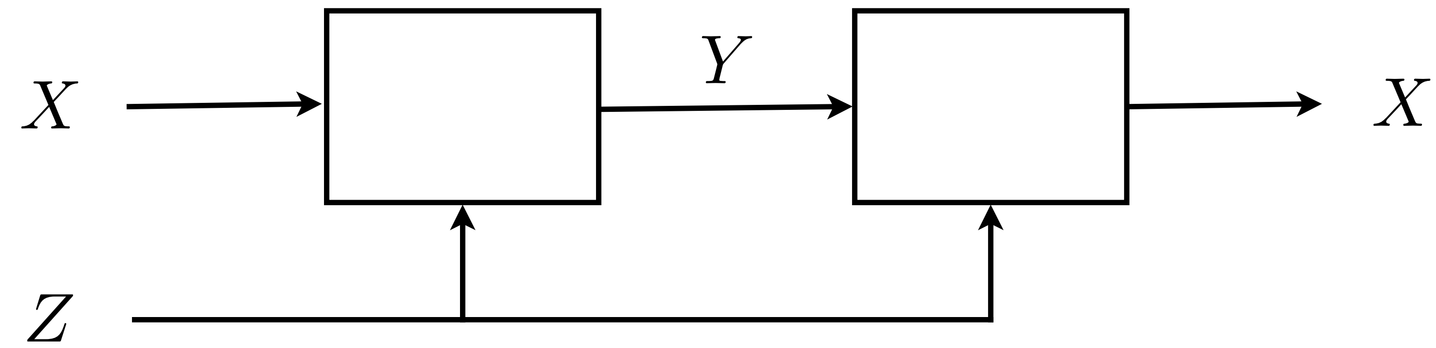
4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.

5. From the information diagram at the top, it is evident that

$$\mu^*(\tilde{Z} - \tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}).$$



X – plaintext
Y – ciphertext
Z – key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

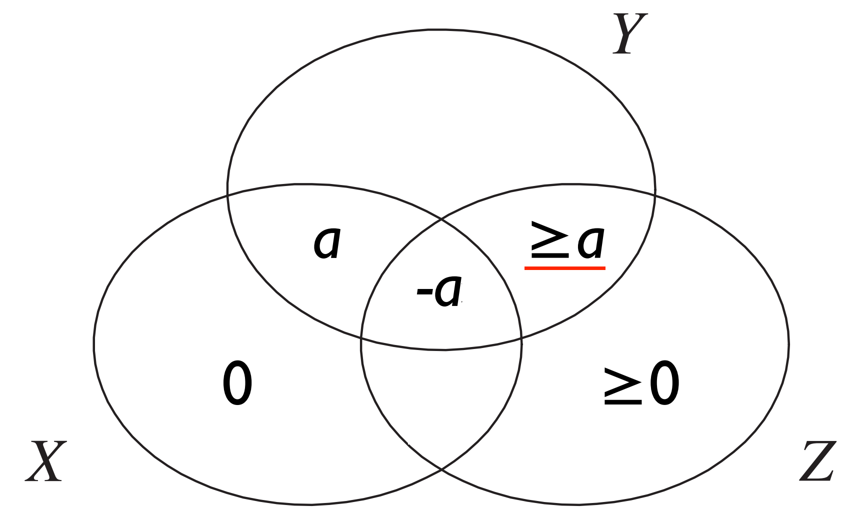
2. We also have $H(Z|X,Y) \geq 0$.

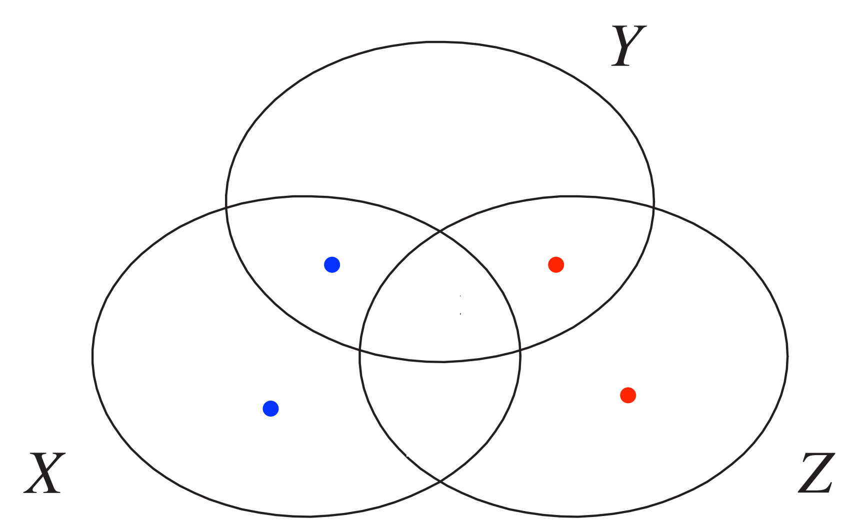
3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.

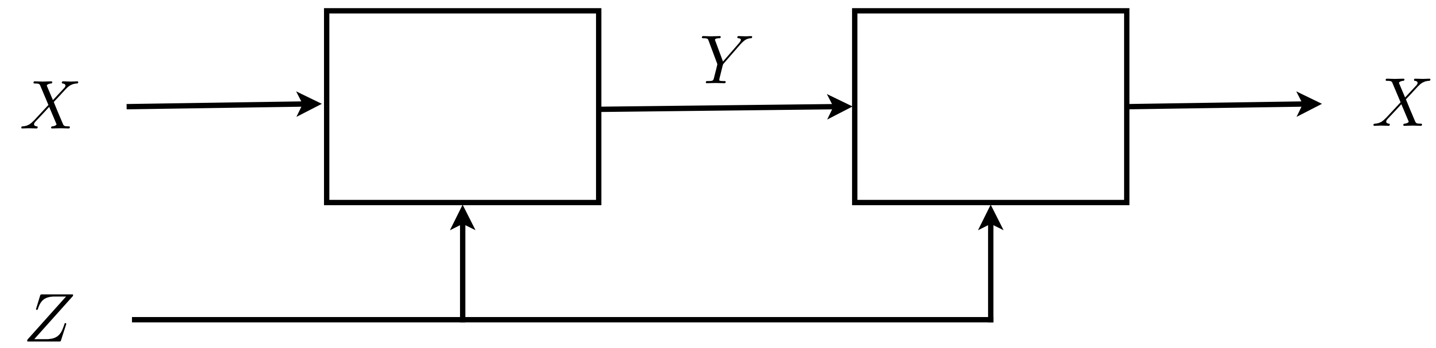
5. From the information diagram at the top, it is evident that

$$\mu^*(\tilde{Z} - \tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}).$$





X — plaintext
Y — ciphertext
Z — key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

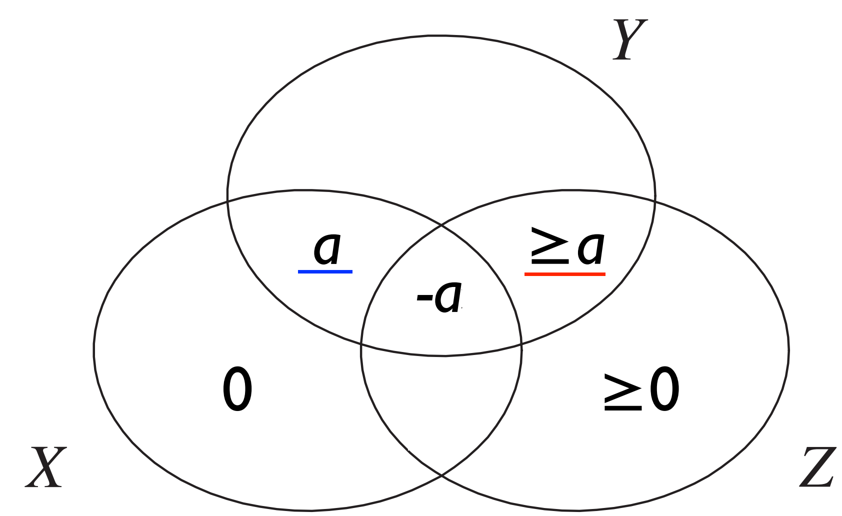
2. We also have $H(Z|X,Y) \geq 0$.

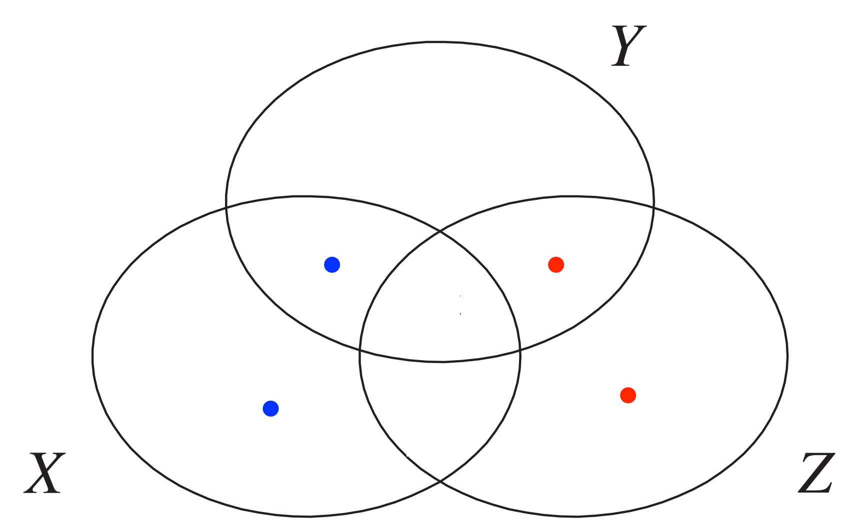
3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.

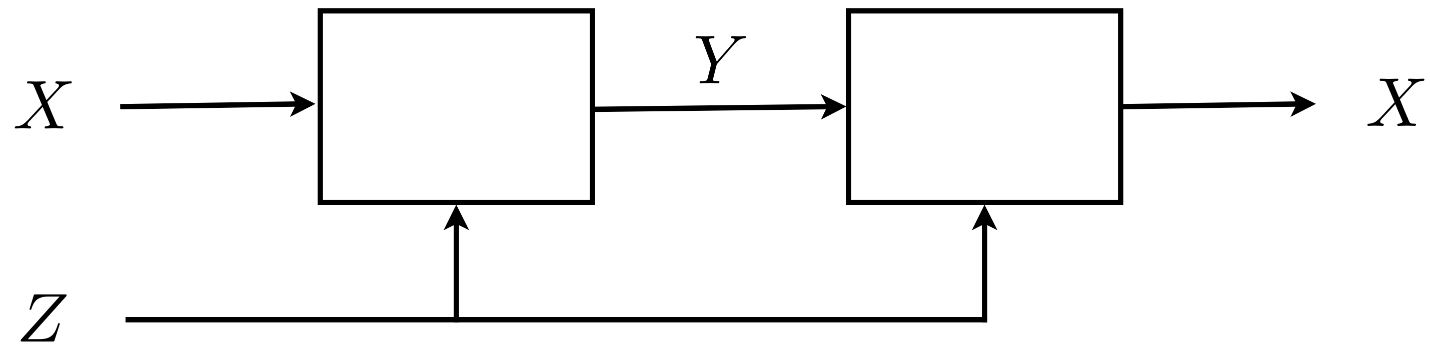
5. From the information diagram at the top, it is evident that

$$\mu^*(\tilde{Z} - \tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}).$$





X – plaintext
Y – ciphertext
Z – key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

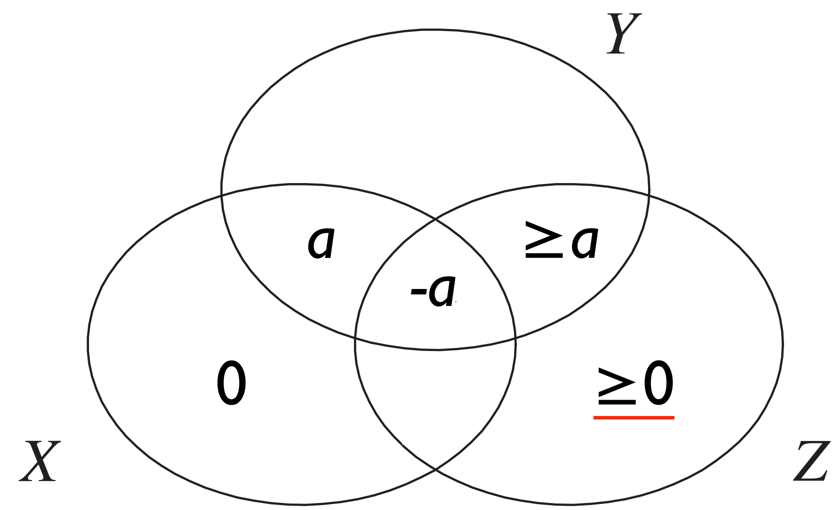
2. We also have $H(Z|X,Y) \geq 0$.

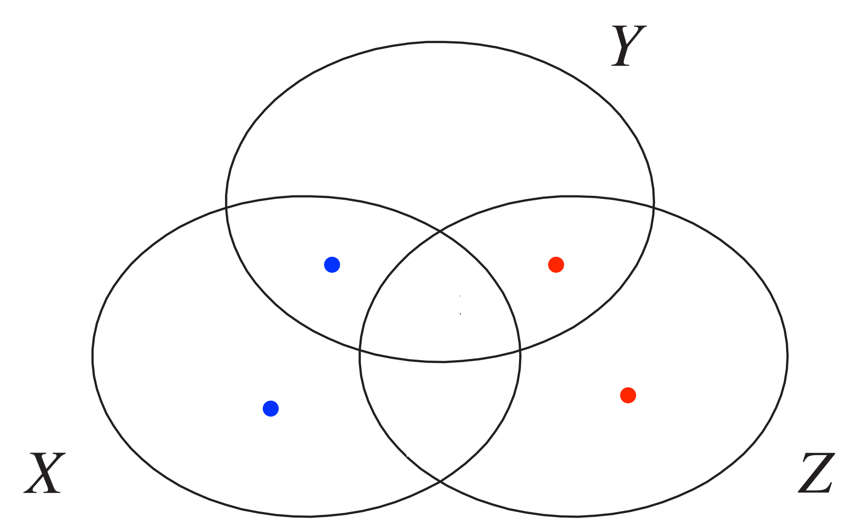
3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.

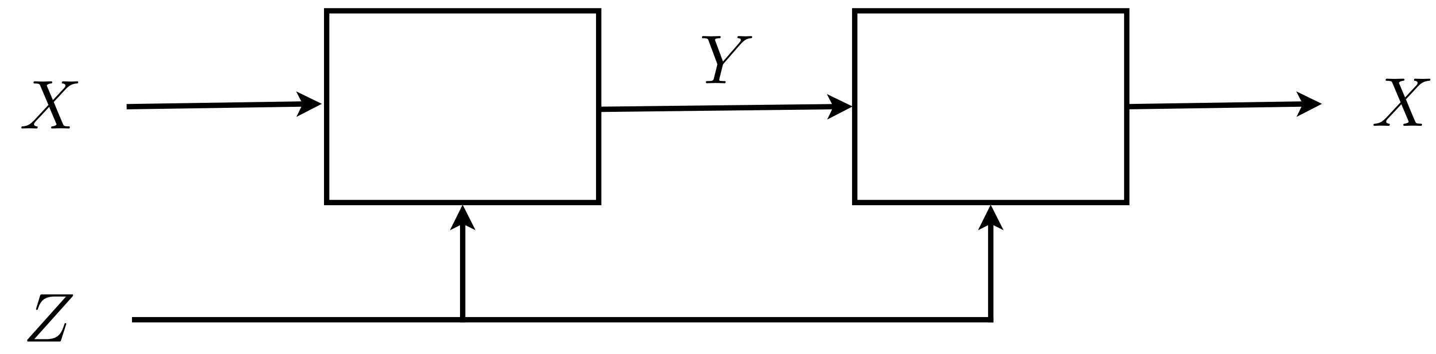
5. From the information diagram at the top, it is evident that

$$\mu^*(\tilde{Z} - \tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}).$$





X	–	plaintext
Y	–	ciphertext
Z	–	key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

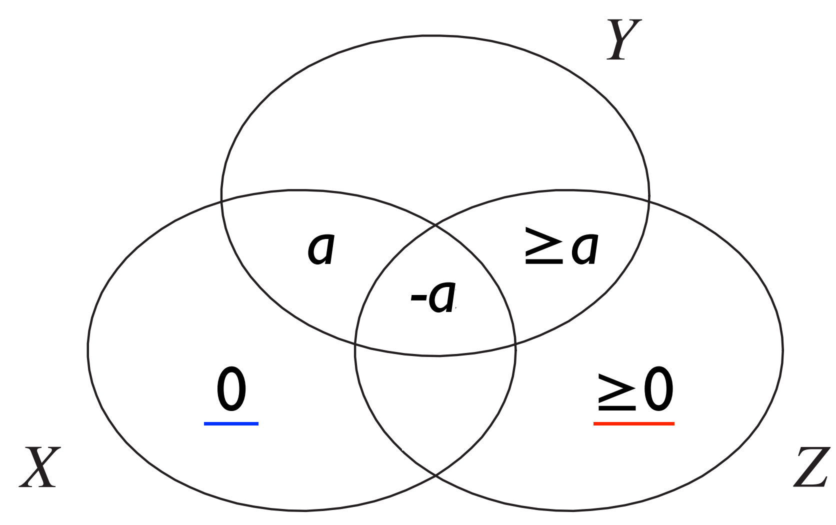
2. We also have $H(Z|X,Y) \geq 0$.

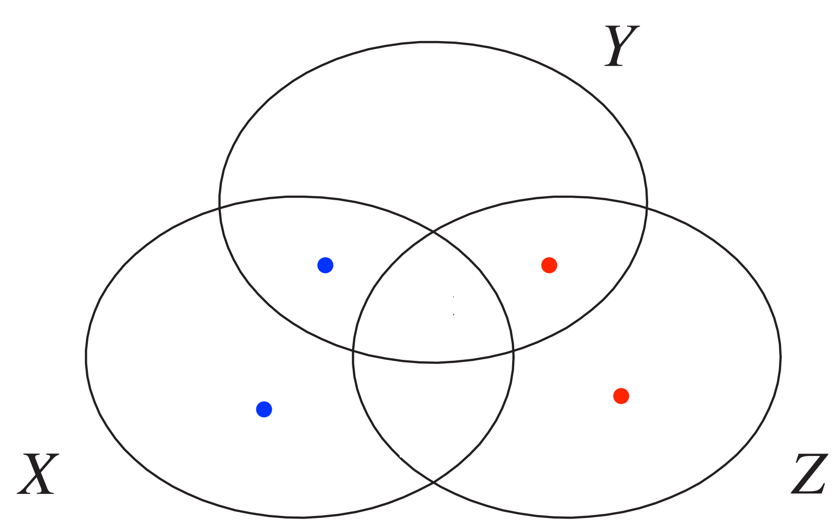
3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.

5. From the information diagram at the top, it is evident that

$$\mu^*(\tilde{Z} - \tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}).$$

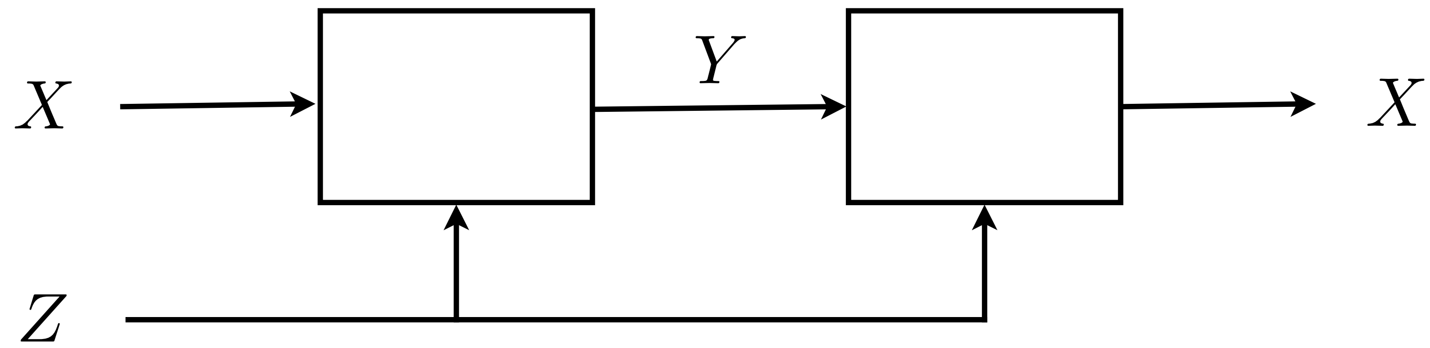




X – plaintext

Y – ciphertext

Z – key



Perfect Secrecy $I(X;Y) = 0$

Decipherbility $H(X|Y,Z) = 0$

1. Since $I(Y;Z) \geq 0$, we have

$$I(Y;Z|X) \geq a.$$

2. We also have $H(Z|X,Y) \geq 0$.

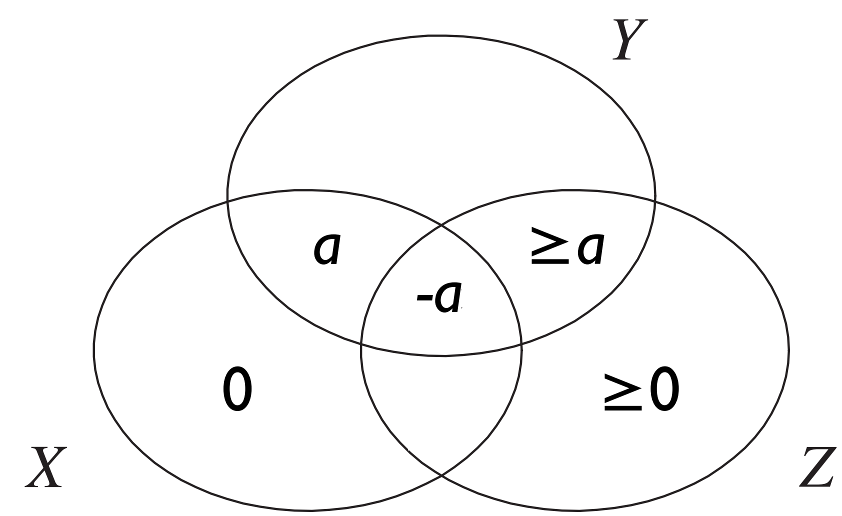
3. We need to show that $\mu^*(\tilde{Z}) \geq \mu^*(\tilde{X})$. Compare in the information diagram at the bottom the atoms of $\tilde{X}$ and $\tilde{Z}$.

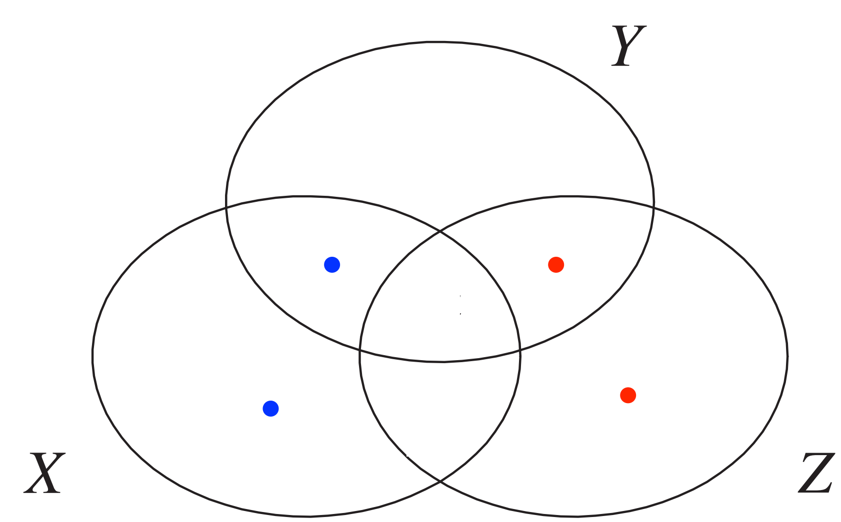
4. The atoms in $\tilde{X} \cap \tilde{Z}$ are common to both $\tilde{X}$ and $\tilde{Z}$. Their measures cancel with each other and so these atoms do not need to be considered. Only the atoms in $\tilde{X} - \tilde{Z}$ and $\tilde{Z} - \tilde{X}$ need to be compared.

5. From the information diagram at the top, it is evident that

$$\mu^*(\tilde{Z} - \tilde{X}) \geq \mu^*(\tilde{X} - \tilde{Z}).$$

6. Therefore we conclude that $H(Z) \geq H(X)$, as is to be shown.





Example 3.15 (Imperfect Secrecy Theorem) Let X be the plain text, Y be the cipher text, and Z be the key in a secret key cryptosystem. Since X can be recovered from Y and Z, we have

$$H(X|Y,Z) = 0.$$

Show that this constraint implies

$$I(X;Y) \geq H(X) - H(Z).$$

Exercise Study Example 3.15.

Example 3.15 (Imperfect Secrecy Theorem) Let X be the plain text, Y be the cipher text, and Z be the key in a secret key cryptosystem. Since X can be recovered from Y and Z, we have

$$H(X|Y, Z) = 0.$$

Show that this constraint implies

$$I(X; Y) \geq H(X) - H(Z).$$

Exercise Study Example 3.15.

Remark

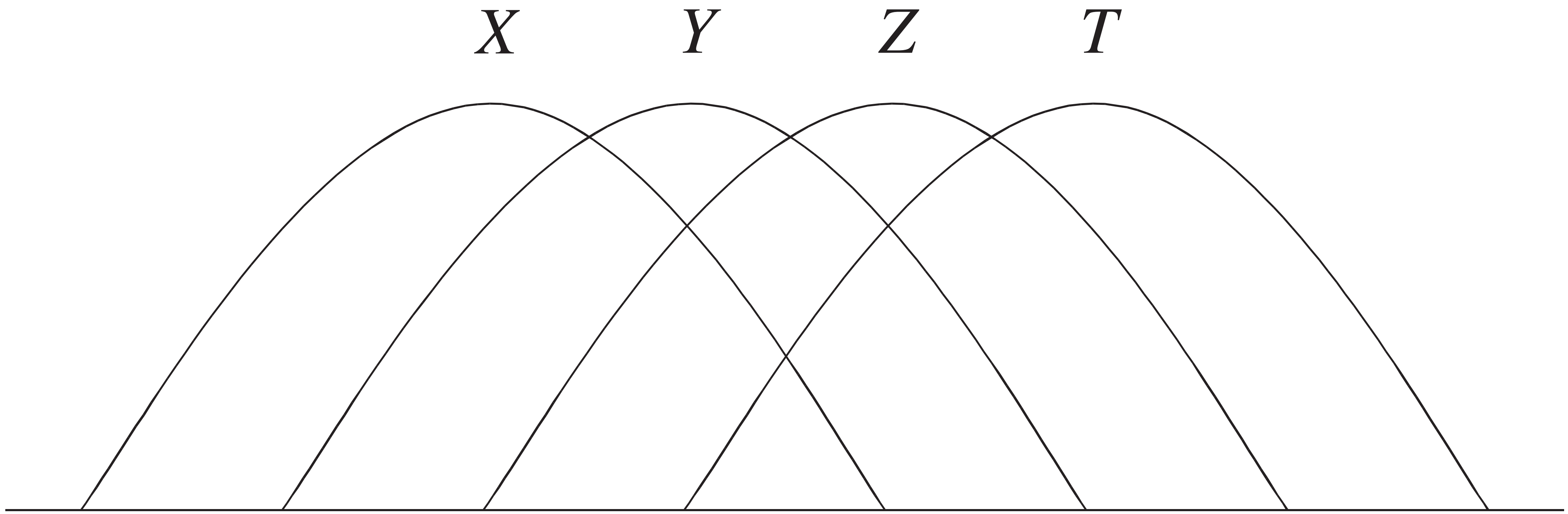
- $I(X; Y)$ measures the "leakage of information." When $I(X; Y) = 0$, it reduces Shannon's perfect secrecy theorem.

Example 3.17 (Data Processing Theorem) If $X \to Y \to Z \to T$, then

- $I(X;T) \leq I(Y;Z)$

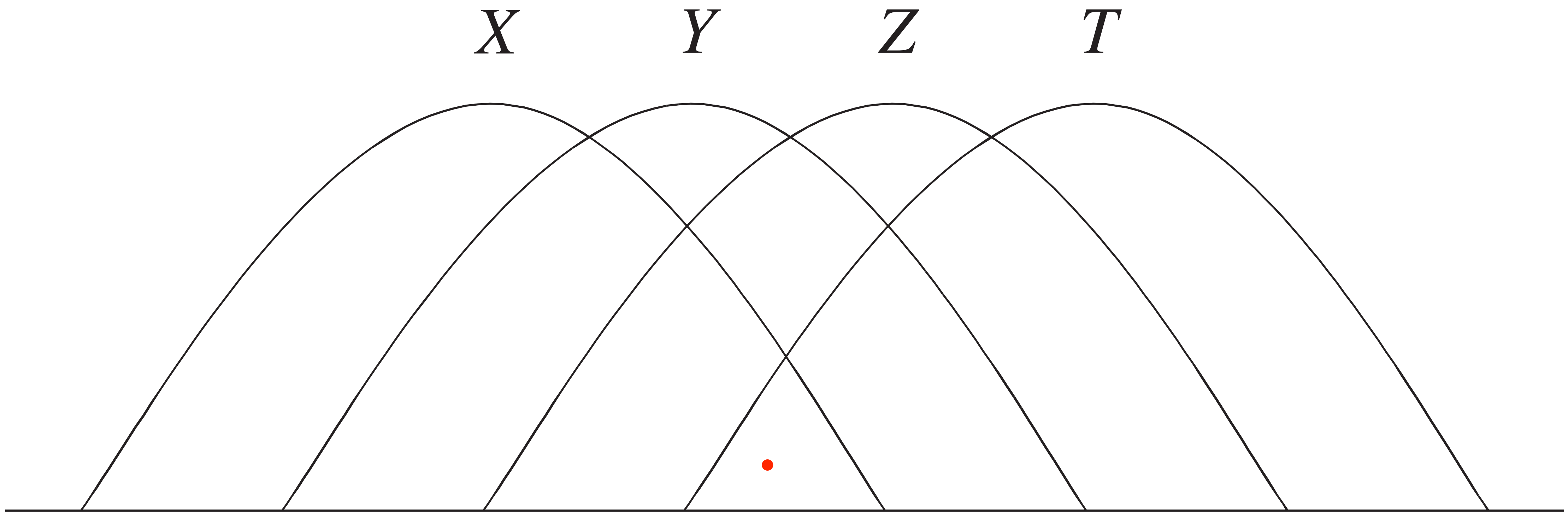
Example 3.17 (Data Processing Theorem) If $X \to Y \to Z \to T$, then

- $I(X;T) \leq I(Y;Z)$



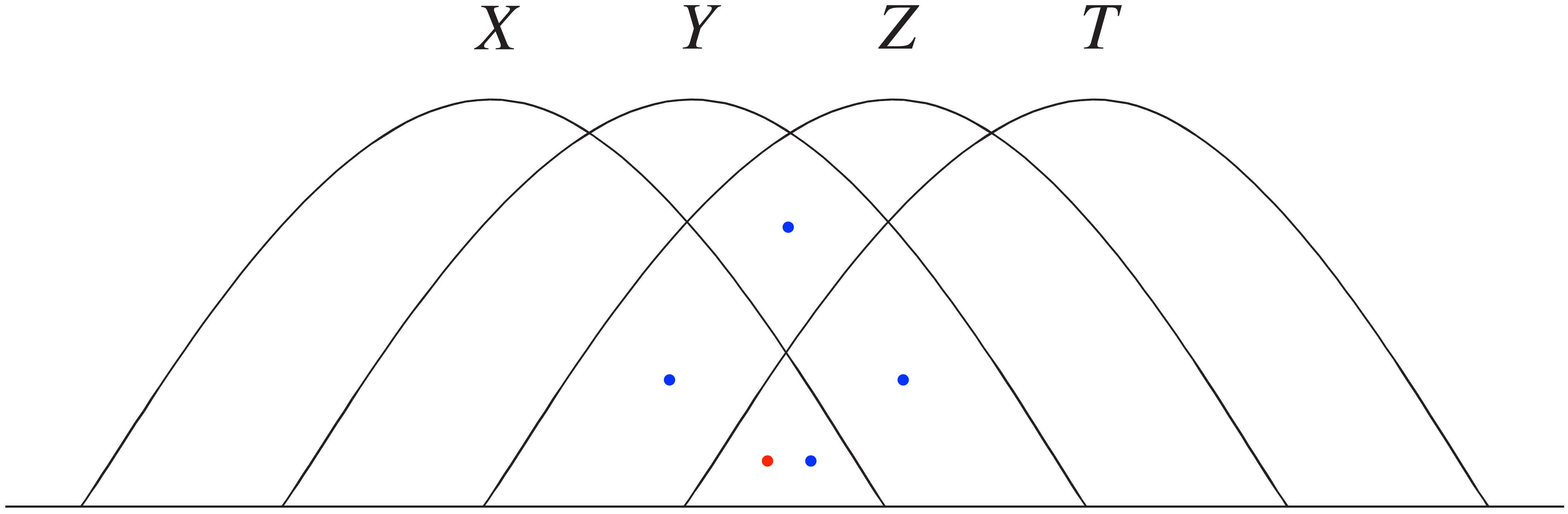
Example 3.17 (Data Processing Theorem) If $X \to Y \to Z \to T$, then

- $I(X;T) \leq I(Y;Z)$



Example 3.17 (Data Processing Theorem) If $X \to Y \to Z \to T$, then

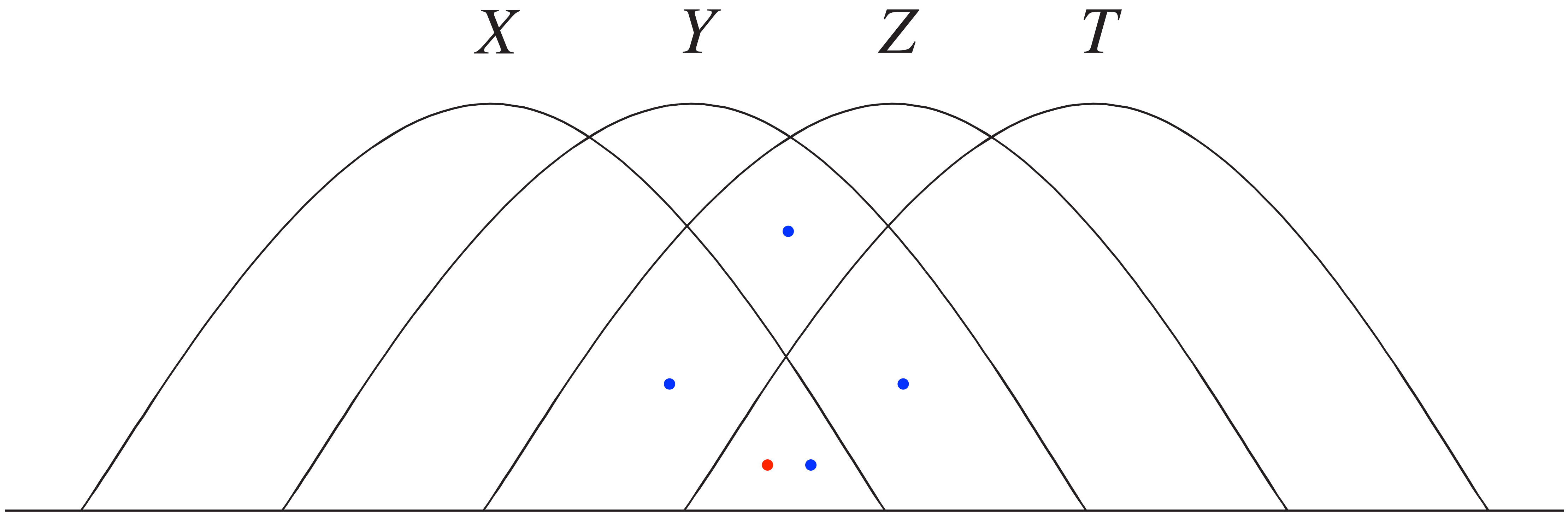
- $I(X;T) \le I(Y;Z)$



Example 3.17 (Data Processing Theorem) If $X \to Y \to Z \to T$, then

- $I(X;T) \le I(Y;Z)$
- in fact

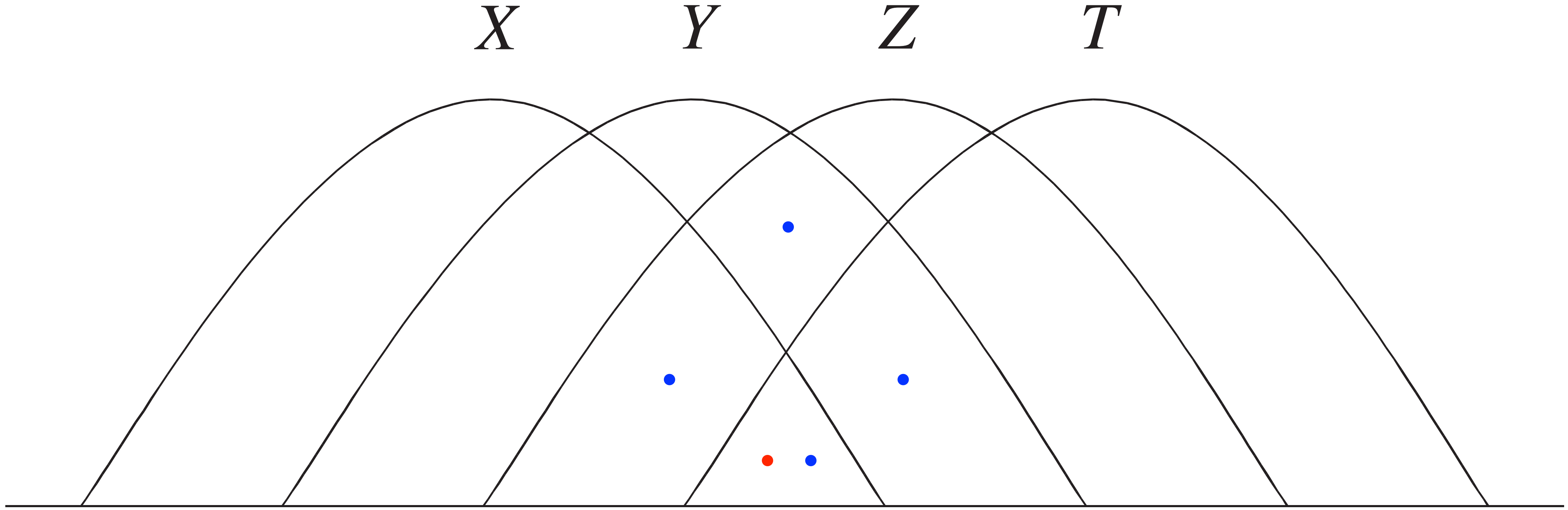
$$I(Y;Z) = I(X;T) + I(X;Z|T) + I(Y;T|X) + I(Y;Z|X,T)$$



Example 3.17 (Data Processing Theorem) If $X \to Y \to Z \to T$, then

- $I(X;T) \leq I(Y;Z)$
- in fact

$$I(Y;Z) = \underline{I(X;T)} + I(X;Z|T) + I(Y;T|X) + I(Y;Z|X,T)$$



Example 3.18 If $X \to Y \to Z \to T \to U$, then

$$H(Y) + H(T) = I(Z; X, Y, T, U) + I(X, Y; T, U) + H(Y|Z) + H(T|Z)$$

Example 3.18 If $X \to Y \to Z \to T \to U$, then

$$H(Y) + H(T) = I(Z; X, Y, T, U) + I(X, Y; T, U) + H(Y|Z) + H(T|Z)$$

Remarks

- Very difficult to discover without an information diagram.
- Instrumental in proving an outer bound for the multiple description problem.

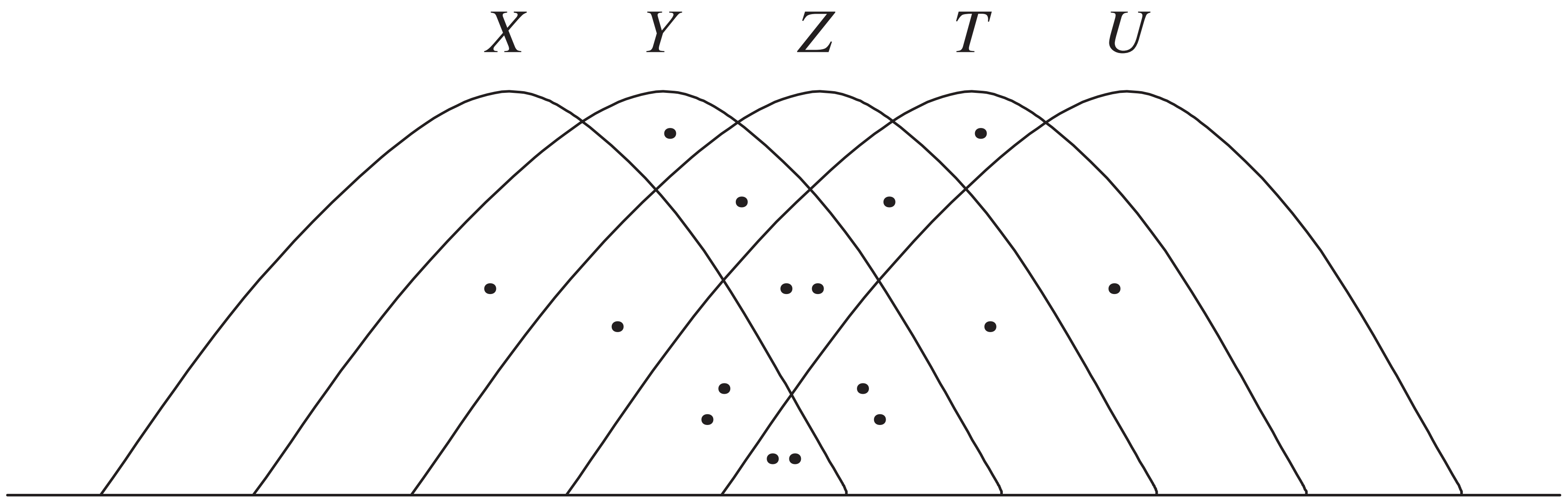
Example 3.18 If $X \to Y \to Z \to T \to U$, then

$$H(Y) + H(T) = I(Z; X, Y, T, U) + I(X, Y; T, U) + H(Y|Z) + H(T|Z)$$

Remarks

- Very difficult to discover without an information diagram.
- Instrumental in proving an outer bound for the multiple description problem.

Exercise Verify the following information diagram for the above equality.



Proving Information Inequalities

- Information inequalities that are implied by the basic inequalities are called Shannon-type inequalities.

Proving Information Inequalities

- Information inequalities that are implied by the basic inequalities are called Shannon-type inequalities.

- They can be proved by means of a linear program called ITIP (Information Theoretic Inequality Prover), developed on Matlab at CUHK (1996):

http://user-www.ie.cuhk.edu.hk/~ITIP/

Proving Information Inequalities

- Information inequalities that are implied by the basic inequalities are called Shannon-type inequalities.

- They can be proved by means of a linear program called ITIP (Information Theoretic Inequality Prover), developed on Matlab at CUHK (1996):

  http://user-www.ie.cuhk.edu.hk/∼ITIP/

- A version running on C called Xitip was developed at EPFL (2007):

  http://xitip.epfl.ch/

Proving Information Inequalities

- Information inequalities that are implied by the basic inequalities are called Shannon-type inequalities.

- They can be proved by means of a linear program called ITIP (Information Theoretic Inequality Prover), developed on Matlab at CUHK (1996):

  http://user-www.ie.cuhk.edu.hk/∼ITIP/

- A version running on C called Xitip was developed at EPFL (2007):

  http://xitip.epfl.ch/

- See Ch. 13 and 14 for discussion.

ITIP Examples

1.
```
>> ITIP(’H(XYZ) <= H(X) + H(Y) + H(Z)’)
True
```

2.
```
>> ITIP(’I(X;Z) = 0’,’I(X;Z|Y) = 0’,’I(X;Y) = 0’)
True
```

3.
```
>> ITIP(’X/Y/Z/T’, ’X/Y/Z’, ’Y/Z/T’)
Not provable by ITIP
```

4.
```
>> ITIP(’I(Z;U) - I(Z;U|X) - I(Z;U|Y) <=
 0.5 I(X;Y) + 0.25 I(X;ZU) + 0.25 I(Y;ZU)’)
Not provable by ITIP
```

ITIP Examples

1. ```
   >> ITIP(’H(XYZ) <= H(X) + H(Y) + H(Z)’)
   True
   ```

2. ```
   >> ITIP(’I(X;Z) = 0’,’I(X;Z|Y) = 0’,’I(X;Y) = 0’)
   True
   ```

3. ```
   >> ITIP(’X/Y/Z/T’, ’X/Y/Z’, ’Y/Z/T’)
   Not provable by ITIP
   ```

4. ```
   >> ITIP(’I(Z;U) - I(Z;U|X) - I(Z;U|Y) <=
    0.5 I(X;Y) + 0.25 I(X;ZU) + 0.25 I(Y;ZU)’)
   Not provable by ITIP
   ```

- #4 is a so-called non-Shannon-type inequality which is valid but not implied by the basic inequalities. See Ch. 15 for discussion.