2.2 Shannon’s Information Measures
Shannon’s Information Measures

• Entropy
• Conditional entropy
• Mutual information
• Conditional mutual information
Shannon’s Information Measures

- Entropy
Shannon’s Information Measures

- Entropy
- Conditional entropy
Shannon’s Information Measures

- Entropy
- Conditional entropy
- Mutual information
Shannon’s Information Measures

- Entropy
- Conditional entropy
- Mutual information
- Conditional mutual information
**Definition 2.13** The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = - \sum_x p(x) \log p(x).$$
Definition 2.13  The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = - \sum_x p(x) \log p(x).$$

- Convention: summation is taken over $S_X$. 

- When the base of the logarithm is $\beta$, write $H(\beta)(X)$ as $H(\beta)(X)$.

- Entropy measures the uncertainty of a discrete random variable.

- The unit for entropy is bit if $\beta = 2$; nat if $\beta = e$;dit if $\beta = D$. 

- A bit in information theory is different from a bit in computer science.
**Definition 2.13** The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = - \sum_x p(x) \log p(x).$$

- Convention: summation is taken over $S_X$.
- When the base of the logarithm is $\alpha$, write $H(X)$ as $H_\alpha(X)$. 

A bit in information theory is different from a bit in computer science.
Definition 2.13  The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = - \sum_x p(x) \log p(x).$$

- Convention: summation is taken over $S_X$.
- When the base of the logarithm is $\alpha$, write $H(X)$ as $H_\alpha(X)$. 

A bit in information theory is different from a bit in computer science.
**Definition 2.13** The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = - \sum_{x} p(x) \log p(x).$$

- Convention: summation is taken over $\mathcal{S}_X$.
- When the base of the logarithm is $\alpha$, write $H(X)$ as $H_{\alpha}(X)$.
Definition 2.13  The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = -\sum_{x} p(x) \log p(x).$$

- Convention: summation is taken over $S_X$.
- When the base of the logarithm is $\alpha$, write $H(X)$ as $H_\alpha(X)$.
- Entropy measures the uncertainty of a discrete random variable.
Definition 2.13  The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = -\sum_x p(x) \log p(x).$$

- Convention: summation is taken over $\mathcal{S}_X$.
- When the base of the logarithm is $\alpha$, write $H(X)$ as $H_\alpha(X)$.
- Entropy measures the uncertainty of a discrete random variable.
- The unit for entropy is
  
  \begin{align*}
  \text{bit} & \quad \text{if } \alpha = 2 \\
  \text{nat} & \quad \text{if } \alpha = e \\
  D\text{-it} & \quad \text{if } \alpha = D
  \end{align*}
Definition 2.13  The entropy $H(X)$ of a random variable $X$ is defined as

$$H(X) = -\sum_x p(x) \log p(x).$$

- Convention: summation is taken over $\mathcal{S}_X$.
- When the base of the logarithm is $\alpha$, write $H(X)$ as $H_\alpha(X)$.
- Entropy measures the uncertainty of a discrete random variable.
- The unit for entropy is
  
  bit \quad \text{if } \alpha = 2 \\
  \text{nat} \quad \text{if } \alpha = e \\
  D\text{-it} \quad \text{if } \alpha = D

- A bit in information theory is different from a bit in computer science.
Remark $H(X)$ depends only on the distribution of $X$ but not on the actual values taken by $X$, hence also write $H(p_X)$. 
Remark  $H(X)$ depends only on the distribution of $X$ but not on the actual values taken by $X$, hence also write $H(p_X)$.

Example  Let $X$ and $Y$ be random variables with $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and let

$$p_X(0) = 0.3, \ p_X(1) = 0.7$$

and

$$p_Y(0) = 0.7, \ p_Y(1) = 0.3.$$ 

Although $p_X \neq p_Y$, $H(X) = H(Y)$. 
**Remark**  $H(X)$ depends only on the distribution of $X$ but not on the actual values taken by $X$, hence also write $H(p_X)$.

**Example**  Let $X$ and $Y$ be random variables with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and let

\[ p_X(0) = 0.3, \quad p_X(1) = 0.7 \]

and

\[ p_Y(0) = 0.7, \quad p_Y(1) = 0.3. \]

Although $p_X \neq p_Y$, $H(X) = H(Y)$. 

**Remark** $H(X)$ depends only on the distribution of $X$ but not on the actual values taken by $X$, hence also write $H(p_X)$.

**Example** Let $X$ and $Y$ be random variables with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and let

$$p_X(0) = 0.3, \quad p_X(1) = 0.7$$

and

$$p_Y(0) = 0.7, \quad p_Y(1) = 0.3.$$ 

Although $p_X \neq p_Y$, $H(X) = H(Y)$. 
Remark  $H(X)$ depends only on the distribution of $X$ but not on the actual values taken by $X$, hence also write $H(p_X)$.

Example  Let $X$ and $Y$ be random variables with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and let

$$p_X(0) = 0.3, \ p_X(1) = 0.7$$

and

$$p_Y(0) = 0.7, \ p_Y(1) = 0.3.$$ 

Although $p_X \neq p_Y$, $H(X) = H(Y)$. 
**Remark**  $H(X)$ depends only on the distribution of $X$ but not on the actual values taken by $X$, hence also write $H(p_X)$.

**Example**  Let $X$ and $Y$ be random variables with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and let

$$p_X(0) = 0.3, \ p_X(1) = 0.7$$

and

$$p_Y(0) = 0.7, \ p_Y(1) = 0.3.$$  

Although $p_X \neq p_Y$, $H(X) = H(Y)$. 
Remark  \( H(X) \) depends only on the distribution of \( X \) but not on the actual values taken by \( X \), hence also write \( H(p_X) \).

Example  Let \( X \) and \( Y \) be random variables with \( \mathcal{X} = \mathcal{Y} = \{0, 1\} \), and let

\[
p_X(0) = 0.3, \quad p_X(1) = 0.7
\]

and

\[
p_Y(0) = 0.7, \quad p_Y(1) = 0.3.
\]

Although \( p_X \neq p_Y \), \( H(X) = H(Y) \).
Entropy as Expectation

- **Convention**

  \[ Eg(X) = \sum_x p(x)g(x) \]

  where summation is over \( S_X \).

See Problem 5 for details.
Entropy as Expectation

- **Convention**
  \[ Eg(X) = \sum_x p(x)g(x) \]
  where summation is over \( S_X \).

- **Linearity**
  \[ E[f(X) + g(X)] = Ef(X) + Eg(X) \]
  See Problem 5 for details.
Entropy as Expectation

- **Convention**
  \[ Eg(X) = \sum_x p(x)g(x) \]
  where summation is over \( S_X \).

- **Linearity**
  \[ E[f(X) + g(X)] = Ef(X) + Eg(X) \]
  See Problem 5 for details.

- **Can write**
  \[ H(X) = -E \log p(X) = -\sum_x p(x) \log p(x) \]
Entropy as Expectation

- Convention

\[ Eg(X) = \sum_x p(x)g(x) \]

where summation is over \( S_X \).

- Linearity

\[ E[f(X) + g(X)] = Ef(X) + Eg(X) \]

See Problem 5 for details.

- Can write

\[ H(X) = -E \log p(X) = -\sum_x p(x) \log p(x) \]
Entropy as Expectation

- **Convention**
  \[ E g(X) = \sum_x p(x) g(x) \]
  where summation is over \( S_X \).

- **Linearity**
  \[ E[f(X) + g(X)] = Ef(X) + Eg(X) \]
  See Problem 5 for details.

- **Can write**
  \[ H(X) = -E \log p(X) = -\sum_x p(x) \log p(x) \]
Entropy as Expectation

- Convention
  \[ Eg(X) = \sum_x p(x)g(x) \]
  where summation is over \( S_X \).

- Linearity
  \[ E[f(X) + g(X)] = Ef(X) + Eg(X) \]
  See Problem 5 for details.

- Can write
  \[ H(X) = -E \log p(X) = -\sum_x p(x) \log p(x) \]

- In probability theory, when \( Eg(X) \) is considered, usually \( g(x) \) depends only on the value of \( x \) but not on \( p(x) \).
Binary Entropy Function

- For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$
Binary Entropy Function

• For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hôpital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$
Binary Entropy Function

• For $0 \leq \gamma \leq 1$, define the binary entropy function

\[ h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma) \]

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

\[ \lim_{a \to 0} a \log a = 0. \]
Binary Entropy Function

- For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$
Binary Entropy Function

- For $0 \leq \gamma \leq 1$, define the binary entropy function

$$ h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma) $$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$ \lim_{a \to 0} a \log a = 0. $$
Binary Entropy Function

- For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hôpital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$
Binary Entropy Function

• For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$
Binary Entropy Function

• For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$
Binary Entropy Function

- For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hôpital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$
Binary Entropy Function

- For $0 \leq \gamma \leq 1$, define the binary entropy function
  \[ h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma) \]
  with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,
  \[ \lim_{a \to 0} a \log a = 0. \]

- For $X \sim \{\gamma, 1 - \gamma\}$,
  \[ H(X) = h_b(\gamma). \]
Binary Entropy Function

- For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$ 

- For $X \sim \{\gamma, 1 - \gamma\}$,

$$H(X) = h_b(\gamma).$$
Binary Entropy Function

• For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$

• For $X \sim \{\gamma, 1 - \gamma\}$,

$$H(X) = h_b(\gamma).$$
For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$ 

For $X \sim \{\gamma, 1 - \gamma\}$,

$$H(X) = h_b(\gamma).$$
Binary Entropy Function

• For $0 \leq \gamma \leq 1$, define the binary entropy function

$$h_b(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$$

with the convention $0 \log 0 = 0$, as by L’Hopital’s rule,

$$\lim_{a \to 0} a \log a = 0.$$

• For $X \sim \{\gamma, 1 - \gamma\}$,

$$H(X) = h_b(\gamma).$$

• $h_b(\gamma)$ achieves the maximum value 1 when $\gamma = \frac{1}{2}$. 
The diagram shows a plot of $h_b(\gamma)$ against $\gamma$. The curve reaches its peak at $\gamma = 0.5$ and asymptotically approaches 1 as $\gamma$ increases from 0 to 1.
Interpretation

Consider tossing a coin with

\[ p(H) = \gamma \quad \text{and} \quad p(T) = 1 - \gamma. \]

Then \( h_b(\gamma) \) measures the amount of uncertainty in the outcome of the toss.
Consider tossing a coin with

\[ p(H) = \gamma \quad \text{and} \quad p(T) = 1 - \gamma. \]

Then \( h_b(\gamma) \) measures the amount of uncertainty in the outcome of the toss.

- When \( \gamma = 0 \) or \( 1 \), the coin is deterministic and \( h_b(\gamma) = 0 \). This is consistent with our intuition because for such cases we need 0 bits to convey the outcome.

Interpretation
Consider tossing a coin with

\[ p(H) = \gamma \quad \text{and} \quad p(T) = 1 - \gamma. \]

Then \( h_b(\gamma) \) measures the amount of uncertainty in the outcome of the toss.

- When \( \gamma = 0 \) or \( 1 \), the coin is deterministic and \( h_b(\gamma) = 0 \). This is consistent with our intuition because for such cases we need 0 bits to convey the outcome.

- When \( \gamma = 0.5 \), the coin is fair and \( h_b(\gamma) = 1 \). This is consistent with our intuition because we need 1 bit to convey the outcome.
Interpretation

Consider tossing a coin with

\[ p(H) = \gamma \quad \text{and} \quad p(T) = 1 - \gamma. \]

Then \( h_b(\gamma) \) measures the amount of uncertainty in the outcome of the toss.

- When \( \gamma = 0 \) or \( 1 \), the coin is \textit{deterministic} and \( h_b(\gamma) = 0 \). This is consistent with our intuition because for such cases we need 0 bits to convey the outcome.

- When \( \gamma = 0.5 \), the coin is \textit{fair} and \( h_b(\gamma) = 1 \). This is consistent with our intuition because we need 1 bit to convey the outcome.

- When \( \gamma \not\in \{0, 0.5, 1\} \), \( 0 < h_b(\gamma) < 1 \), i.e., the uncertainty about the outcome is somewhere between 0 and 1 bit.
Interpretation

Consider tossing a coin with

\[ p(H) = \gamma \quad \text{and} \quad p(T) = 1 - \gamma. \]

Then \( h_b(\gamma) \) measures the amount of uncertainty in the outcome of the toss.

- When \( \gamma = 0 \) or 1, the coin is \textit{deterministic} and \( h_b(\gamma) = 0 \). This is consistent with our intuition because for such cases we need 0 bits to convey the outcome.

- When \( \gamma = 0.5 \), the coin is \textit{fair} and \( h_b(\gamma) = 1 \). This is consistent with our intuition because we need 1 bit to convey the outcome.

- When \( \gamma \not\in \{0, 0.5, 1\} \), \( 0 < h_b(\gamma) < 1 \), i.e., the uncertainty about the outcome is somewhere between 0 and 1 bit.

- This interpretation will be justified in terms of the source coding theorem.
**Definition 2.14** The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = - \sum_{x,y} p(x, y) \log p(x, y) = -E \log p(X, Y).$$
**Definition 2.14** The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y) = -E \log p(X, Y).$$
Definition 2.14  The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = - \sum_{x,y} p(x, y) \log p(x, y) = -E \log p(X, Y).$$
**Definition 2.14** The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = - \sum_{x, y} p(x, y) \log p(x, y) = -E \log p(X, Y).$$
**Definition 2.14** The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = -\sum_{x, y} p(x, y) \log p(x, y) = -E \log p(X, Y).$$
**Definition 2.14** The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = - \sum_{x,y} p(x, y) \log p(x, y) = - \mathbb{E} \log p(X, Y).$$

**Definition 2.15** For random variables $X$ and $Y$, the conditional entropy of $Y$ given $X$ is defined as

$$H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) = - \mathbb{E} \log p(Y|X).$$
**Definition 2.14** The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y) = -E \log p(X, Y).$$

**Definition 2.15** For random variables $X$ and $Y$, the conditional entropy of $Y$ given $X$ is defined as

$$H(Y|X) = -\sum_{x,y} p(x, y) \log p(y|x) = -E \log p(Y|X).$$
**Definition 2.14** The joint entropy $H(X, Y)$ of a pair of random variables $X$ and $Y$ is defined as

$$H(X, Y) = - \sum_{x,y} p(x, y) \log p(x, y) = -E \log p(X, Y).$$

**Definition 2.15** For random variables $X$ and $Y$, the conditional entropy of $Y$ given $X$ is defined as

$$H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) = -E \log p(Y|X).$$
Write
• Write

\[ H(Y|X) = -\sum_{x,y} p(x, y) \log p(y|x) \]
• Write

\[ H(Y|X) = - \sum_{x,y} p(x,y) \log p(y|x) \]
• Write

\[ H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) \]

\[ = - \sum_x \sum_y p(x)p(y|x) \log p(y|x) \]
• Write

\[ H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) \]

\[ = - \sum_x \sum_y p(x)p(y|x) \log p(y|x) \]

\[ = \sum_x p(x) \left[ - \sum_y p(y|x) \log p(y|x) \right] \]
\* Write

\[
H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x)
\]

\[
= - \sum_x \sum_y p(x)p(y|x) \log p(y|x)
\]

\[
= \sum_x p(x) \left[ - \sum_y p(y|x) \log p(y|x) \right]
\]

\* The inner sum is the entropy of $Y$ conditioning on a fixed $x \in S_X$. 
Write

\[ H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) \]

\[ = - \sum_{x} \sum_{y} p(x)p(y|x) \log p(y|x) \]

\[ = \sum_{x} p(x) \left[ - \sum_{y} p(y|x) \log p(y|x) \right] \]

- The inner sum is the entropy of \( Y \) conditioning on a fixed \( x \in \mathcal{S}_X \).
• Write

\[ H(Y|X) = -\sum_{x,y} p(x, y) \log p(y|x) \]

\[ = -\sum_{x} \sum_{y} p(x)p(y|x) \log p(y|x) \]

\[ = \sum_{x} p(x) \left[ -\sum_{y} p(y|x) \log p(y|x) \right] \]

• The inner sum is the entropy of \( Y \) conditioning on a fixed \( x \in \mathcal{S}_X \).
Write

\[ H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) \]

\[ = - \sum_x \sum_y p(x)p(y|x) \log p(y|x) \]

\[ = \sum_x p(x) \left[ - \sum_y p(y|x) \log p(y|x) \right] \]

- The inner sum is the entropy of \( Y \) conditioning on a fixed \( x \in S_X \).
• Write

\[ H(Y|X) = -\sum_{x,y} p(x, y) \log p(y|x) \]

\[ = -\sum_x \sum_y p(x)p(y|x) \log p(y|x) \]

\[ = \sum_x p(x) \left[ -\sum_y p(y|x) \log p(y|x) \right] \]

• The inner sum is the entropy of \( Y \) conditioning on a fixed \( x \in \mathcal{S}_X \).

• Denoting the inner sum by \( H(Y|X = x) \), we have

\[ H(Y|X) = \sum_x p(x)H(Y|X = x) \]
• Write

\[ H(Y|X) = -\sum_{x,y} p(x, y) \log p(y|x) \]

\[ = -\sum_{x} \sum_{y} p(x)p(y|x) \log p(y|x) \]

\[ = \sum_{x} p(x) \left[ -\sum_{y} p(y|x) \log p(y|x) \right] \]

• The inner sum is the entropy of \( Y \) conditioning on a fixed \( x \in \mathcal{S}_X \).

• Denoting the inner sum by \( H(Y|X = x) \), we have

\[ H(Y|X) = \sum_{x} p(x) H(Y|X = x) \]
• Write

\[ H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) \]

\[ = - \sum_x \sum_y p(x)p(y|x) \log p(y|x) \]

\[ = \sum_x p(x) \left[ - \sum_y p(y|x) \log p(y|x) \right] \]

\[ = \sum_x p(x)H(Y|X = x) \]

• The inner sum is the entropy of \( Y \) conditioning on a fixed \( x \in S_X \).

• Denoting the inner sum by \( H(Y|X = x) \), we have

\[ H(Y|X) = \sum_x p(x)H(Y|X = x) \]
Similarly,

\[ H(Y|X, Z) = \sum_z p(z) H(Y|X, Z = z), \]
• Similarly,

\[ H(Y|X, Z) = \sum_z p(z) H(Y|X, Z = z), \]
• Similarly,

\[ H(Y|X, Z) = \sum_z p(z)H(Y|X, Z = z), \]
• Similarly,

\[ H(Y|X, Z) = \sum_z p(z)H(Y|X, Z = z), \]

where

\[ H(Y|X, Z = z) = -\sum_{x, y} p(x, y|z) \log p(y|x, z). \]
• Similarly,

\[
H(Y|X, Z) = \sum_z p(z) H(Y|X, Z = z),
\]

where

\[
H(Y|X, Z = z) = -\sum_{x,y} p(x, y|z) \log p(y|x, z).
\]
Similarly,

\[ H(Y|X, Z) = \sum_z p(z) H(Y|X, Z = z), \]

where

\[ H(Y|X, Z = z) = -\sum_{x,y} p(x, y|z) \log p(y|x, z). \]
• Similarly,

\[ H(Y|X, Z) = \sum_z p(z) H(Y|X, Z = z), \]

where

\[ H(Y|X, Z = z) = -\sum_{x, y} p(x, y|z) \log p(y|x, z). \]
• Similarly,

\[ H(Y|X, Z) = \sum_z p(z) H(Y|X, Z = z), \]

where

\[ H(Y|X, Z = z) = -\sum_{x,y} p(x,y|z) \log p(y|x, z). \]
• Similarly,

\[ H(Y|X, Z) = \sum_z p(z)H(Y|X, Z = z), \]

where

\[ H(Y|X, Z = z) = - \sum_{x, y} p(x, y|z) \log p(y|x, z). \]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]

Proof
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X | Y). \]

Proof

Consider

\[ H(X, Y) = -E \log p(X, Y) \]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]

Proof

Consider

\[ H(X, Y) = -E \log p(X, Y) \]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]

Proof

Consider

\[ H(X, Y) = -E \log p(X, Y) \]
\[ = -E \log [p(X)p(Y|X)] \]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]

Proof

Consider

\[ H(X, Y) = -E \log p(X, Y) \]

\[ = -E \log[p(X)p(Y|X)] \]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X | Y). \]

Proof

Consider

\[
H(X, Y) = -E \log p(X, Y) \\
= -E \log [p(X)p(Y | X)] \\
= -E \log p(X) - E \log p(Y | X)
\]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]

Proof

Consider

\[
H(X, Y) &= -E \log p(X, Y) \\
&= -E \log [p(X)p(Y|X)] \\
&= -E \log p(X) - E \log p(Y|X)
\]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]

Proof

Consider

\[
\begin{align*}
H(X, Y) &= -E \log p(X, Y) \\
&= -E \log [p(X)p(Y|X)] \\
&= -E \log p(X) - E \log p(Y|X) \\
&= H(X) + H(Y|X).
\end{align*}
\]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X | Y). \]

Proof

Consider

\[ H(X, Y) = -E \log p(X, Y) \]

\[ = -E \log[p(X)p(Y|X)] \]

\[ = -E \log p(X) - E \log p(Y|X) \]

\[ = H(X) + H(Y|X). \]
Proposition 2.16

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

and

\[ H(X, Y) = H(Y) + H(X|Y). \]

Proof

Consider

\[
H(X, Y) = -E \log p(X, Y) = -E \log[p(X)p(Y|X)] = -E \log p(X) - E \log p(Y|X) = H(X) + H(Y|X).
\]
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E \log \frac{p(X, Y)}{p(X)p(Y)}.$$
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \mathbb{E} \log \frac{p(X,Y)}{p(X)p(Y)}.$$
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

\[
I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = E \log \frac{p(X,Y)}{p(X)p(Y)}.
\]
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

\[
I(X; Y) = \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = \mathbb{E} \log \frac{p(X, Y)}{p(X)p(Y)}.
\]
Definition 2.17  For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

\[
I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E \log \frac{p(X, Y)}{p(X)p(Y)}.
\]
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E \log \frac{p(X, Y)}{p(X)p(Y)}.$$ 

**Remark** $I(X; Y)$ is symmetrical in $X$ and $Y$. 
Definition 2.17 For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

$$I(X; Y) = \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E \log \frac{p(X, Y)}{p(X)p(Y)}.$$ 

Remark $I(X; Y)$ is symmetrical in $X$ and $Y$.

Remark Alternatively, we can write

$$I(X; Y) = \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = \sum_{x, y} p(x, y) \log \frac{p(x|y)}{p(x)} = E \log \frac{p(X|Y)}{p(X)}.$$ 

However, it is not apparent from this form that $I(X; Y)$ is symmetrical in $X$ and $Y.$
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E \log \frac{p(X, Y)}{p(X)p(Y)}.$$ 

**Remark** $I(X; Y)$ is symmetrical in $X$ and $Y$.

**Remark** Alternatively, we can write

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = \sum_{x,y} p(x, y) \log \frac{p(x|y)}{p(x)} = E \log \frac{p(X|Y)}{p(X)}.$$ 

However, it is not apparent from this form that $I(X; Y)$ is symmetrical in $X$ and $Y$. 
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = E \log \frac{p(X, Y)}{p(X)p(Y)}.$$ 

**Remark** $I(X; Y)$ is symmetrical in $X$ and $Y$.

**Remark** Alternatively, we can write

$$I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = \sum_{x,y} p(x, y) \log \frac{p(x|y)}{p(x)} = E \log \frac{p(X|Y)}{p(X)}.$$ 

However, it is not apparent from this form that $I(X; Y)$ is symmetrical in $X$ and $Y$. 
**Definition 2.17** For random variables $X$ and $Y$, the mutual information between $X$ and $Y$ is defined as

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = E \log \frac{p(X,Y)}{p(X)p(Y)}.$$ 

**Remark** $I(X;Y)$ is symmetrical in $X$ and $Y$.

**Remark** Alternatively, we can write

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} = E \log \frac{p(X|Y)}{p(X)}.$$ 

However, it is not apparent from this form that $I(X;Y)$ is symmetrical in $X$ and $Y$. 
Proposition 2.18 The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

Proof
Proposition 2.18  The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

Proof

\[ I(X; Y) = E \log \frac{p(X, Y)}{p(X)p(Y)} \]
**Proposition 2.18** The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

**Proof**

$$I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)}.$$
Proposition 2.18  The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

Proof

\[
I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)}
\]

\[
= E \log \frac{p(X)}{p(X)p(X)}
\]
Proposition 2.18  The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

Proof

\[
I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)}
\]

\[
= E \log \frac{p(X)}{p(X)p(X)}
\]

\[
I(X; Y) = E \log \frac{p(X, Y)}{p(X)p(Y)}
\]
Proposition 2.18  The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

Proof

$$I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)} = E \log \frac{p(X)}{p(X)p(X)} = H(X).$$

Remark

The entropy of $X$ is sometimes called the self-information of $X$. 

$I(X; Y) = E \log \frac{p(X, Y)}{p(X)p(Y)}$
Proposition 2.18 The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

Proof

\[
I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)}
\]

\[
= E \log \frac{p(X)}{p(X)p(X)}
\]

\[
= -E \log p(X)
\]
**Proposition 2.18** The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

**Proof**

\[
I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)}
\]

\[
= E \log \frac{p(X)}{p(X)p(X)}
\]

\[
= E \log \frac{p(X)}{p(X)}
\]

\[
= -E \log p(X)
\]

\[
= H(X).
\]
Proposition 2.18  The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

Proof

$$I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)}$$

$$= E \log \frac{p(X)}{p(X)p(X)}$$

$$= -E \log p(X)$$

$$= H(X).$$
**Proposition 2.18** The mutual information between a random variable $X$ and itself is equal to the entropy of $X$, i.e., $I(X; X) = H(X)$.

**Proof**

$$I(X; X) = E \log \frac{p(X, X)}{p(X)p(X)}$$

$$= E \log \frac{p(X)}{p(X)p(X)}$$

$$= -E \log p(X)$$

$$= H(X).$$

**Remark** The entropy of $X$ is sometimes called the *self-information* of $X$. 

$$I(X; Y) = E \log \frac{p(X, Y)}{p(X)p(Y)}$$
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]
\[ I(X;Y) = H(Y) - H(Y|X), \]
\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]

provided that all the entropies and conditional entropies are finite.

(Exercise)

Remark

\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]

is analogous to

\[ \mu(A \setminus B) = \mu(A) + \mu(B) - \mu(A \cap B), \]

where \( \mu \) is a set-additive function and \( A \) and \( B \) are sets.
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]

provided that all the entropies and conditional entropies are finite.

(Exercise) Remark

\[ I(X;Y) = H(X) + H(Y) - H(X,Y) \]

is analogous to

\[ \mu(A \setminus B) = \mu(A) + \mu(B) - \mu(A \setminus \overline{B}) \]

where \( \mu \) is a set-additive function and \( A \) and \( B \) are sets.
Proposition 2.19

\[
I(X; Y) = H(X) - H(X|Y),
\]
\[
I(X; Y) = H(Y) - H(Y|X),
\]

provided that all the entropies and conditional entropies are finite.

Remark

\[
I(X; Y) = H(X) + H(Y) - H(X,Y)
\]
is analogous to

\[
\mu(A - B) = \mu(A) + \mu(B) - \mu(A \cap B),
\]
where \(\mu\) is a set-additive function and \(A\) and \(B\) are sets.
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]
\[ I(X;Y) = H(Y) - H(Y|X), \]

and

\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]
\[ I(X;Y) = H(Y) - H(Y|X), \]

and

\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]

provided that all the entropies and conditional entropies are finite. (Exercise)
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]
\[ I(X;Y) = H(Y) - H(Y|X), \]

and

\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]

provided that all the entropies and conditional entropies are finite. (Exercise)

Remark
Proposition 2.19

\[ I(X; Y) = H(X) - H(X|Y), \]
\[ I(X; Y) = H(Y) - H(Y|X), \]

and

\[ I(X; Y) = H(X) + H(Y) - H(X, Y), \]

provided that all the entropies and conditional entropies are finite. (Exercise)

Remark

\[ I(X; Y) = H(X) + H(Y) - H(X, Y) \]
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]
\[ I(X;Y) = H(Y) - H(Y|X), \]

and

\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]

provided that all the entropies and conditional entropies are finite. (Exercise)

Remark

\[ I(X;Y) = H(X) + H(Y) - H(X,Y) \]
Proposition 2.19

\[ I(X; Y) = H(X) - H(X|Y), \]
\[ I(X; Y) = H(Y) - H(Y|X), \]

and

\[ I(X; Y) = H(X) + H(Y) - H(X,Y), \]

provided that all the entropies and conditional entropies are finite. (Exercise)

Remark

\[ I(X; Y) = H(X) + H(Y) - H(X,Y) \]

is analogous to

\[ \mu(A \cap B) = \mu(A) + \mu(B) - \mu(A \cup B), \]
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]
\[ I(X;Y) = H(Y) - H(Y|X), \]

and

\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]

provided that all the entropies and conditional entropies are finite. (Exercise)

Remark

\[ I(X;Y) = H(X) + H(Y) - H(X,Y) \]

is analogous to

\[ \mu(A \cap B) = \mu(A) + \mu(B) - \mu(A \cup B), \]
Proposition 2.19

\[ I(X;Y) = H(X) - H(X|Y), \]
\[ I(X;Y) = H(Y) - H(Y|X), \]

and

\[ I(X;Y) = H(X) + H(Y) - H(X,Y), \]

provided that all the entropies and conditional entropies are finite. (Exercise)

Remark

\[ I(X;Y) = H(X) + H(Y) - H(X,Y) \]

is analogous to

\[ \mu(A \cap B) = \mu(A) + \mu(B) - \mu(A \cup B), \]

where \( \mu \) is a set-additive function and \( A \) and \( B \) are sets.
Information Diagram

$H(X, Y)$

$H(X|Y)$

$H(X)$

$I(X; Y)$

$H(Y|X)$

$H(Y)$
Information Diagram

$H(X,Y)$

$H(X|Y)$

$H(Y|X)$

$I(X;Y)$

$H(X)$

$H(Y)$
Information Diagram

\[ H(X, Y) \]

\[ H(X | Y) \]

\[ H(Y | X) \]

\[ H(X) \]

\[ H(Y) \]

\[ I(X; Y) \]
Information Diagram

$H(X,Y)$

$H(X|Y)$

$H(X)$

$I(X;Y)$

$H(Y|X)$

$H(Y)$
Information Diagram

\[ H(X,Y) \]

\[ H(X|Y) \]

\[ H(Y|X) \]

\[ H(X) \]

\[ H(Y) \]

\[ I(X;Y) \]
Information Diagram

$H(X, Y)$

$H(X|Y)$

$H(X)$

$I(X;Y)$

$H(Y)$

$H(Y|X)$
**Definition 2.20** For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

$$I(X; Y|Z) = \sum_{x, y, z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = E \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$
Definition 2.20  For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

$$I(X; Y \mid Z) = \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y \mid z)}{p(x \mid z)p(y \mid z)} = E \log \frac{p(X, Y \mid Z)}{p(X \mid Z)p(Y \mid Z)}.$$
Definition 2.20 For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

$$I(X; Y|Z) = \sum_{x, y, z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = E \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$
Definition 2.20 For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

$$I(X; Y|Z) = \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = E \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$

Remark $I(X; Y|Z)$ is symmetrical in $X$ and $Y$. 
Definition 2.20 For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

$$I(X; Y|Z) = \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = E \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$ 

Remark $I(X; Y|Z)$ is symmetrical in $X$ and $Y$.

Similar to entropy, we have

$$I(X; Y|Z) = \sum_z p(z) I(X; Y|Z = z),$$
**Definition 2.20** For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

$$I(X; Y|Z) = \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = E \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$ 

**Remark** $I(X; Y|Z)$ is symmetrical in $X$ and $Y$.

Similar to entropy, we have

$$I(X; Y|Z) = \sum_z p(z) I(X; Y|Z = z),$$

where

$$I(X; Y|Z = z) = \sum_{x,y} p(x, y|z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)}.$$
Definition 2.20  For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

$$I(X; Y|Z) = \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = E \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.$$ 

Remark  $I(X; Y|Z)$ is symmetrical in $X$ and $Y$.

Similar to entropy, we have

$$I(X; Y|Z) = \sum_z p(z)I(X; Y|Z = z),$$

where

$$I(X; Y|Z = z) = \sum_{x,y} p(x, y|z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)}.$$
**Definition 2.20** For random variables $X$, $Y$ and $Z$, the mutual information between $X$ and $Y$ conditioning on $Z$ is defined as

\[
I(X; Y|Z) = \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = E \log \frac{p(X, Y|Z)}{p(X|Z)p(Y|Z)}.
\]

**Remark** $I(X; Y|Z)$ is symmetrical in $X$ and $Y$.

Similar to entropy, we have

\[
I(X; Y|Z) = \sum_z p(z)I(X; Y|Z = z),
\]

where

\[
I(X; Y|Z = z) = \sum_{x,y} p(x, y|z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)}.
\]
Proposition 2.21  The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X; X|Z) = H(X|Z)$. 
**Proposition 2.21** The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X; X|Z) = H(X|Z)$.
**Proposition 2.21** The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X; X|Z) = H(X|Z)$. (Proposition 2.18)
**Proposition 2.21** The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X;X|Z) = H(X|Z)$. 

(Proposition 2.19)
Proposition 2.21 The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X; X|Z) = H(X|Z)$.

Proposition 2.22

\[
I(X; Y|Z) = H(X|Z) - H(X|Y, Z),
\]
\[
I(X; Y|Z) = H(Y|Z) - H(Y|X, Z),
\]

and

\[
I(X; Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z),
\]

provided that all the conditional entropies are finite.
Proposition 2.21  The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X; X|Z) = H(X|Z)$.

Proposition 2.22

\[
I(X; Y \mid Z) = H(X \mid Z) - H(X|Y, Z), \\
I(X; Y \mid Z) = H(Y \mid Z) - H(Y|X, Z),
\]

and

\[
I(X; Y \mid Z) = H(X \mid Z) + H(Y \mid Z) - H(X, Y \mid Z),
\]

provided that all the conditional entropies are finite.
**Proposition 2.21** The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X; X|Z) = H(X|Z)$.

**Proposition 2.22**

\[
I(X; Y | Z) = H(X | Z) - H(X | Y, Z),
\]
\[
I(X; Y | Z) = H(Y | Z) - H(Y | X, Z),
\]

and

\[
I(X; Y | Z) = H(X | Z) + H(Y | Z) - H(X, Y | Z),
\]

provided that all the conditional entropies are finite. (Proposition 2.19)
Proposition 2.21  The mutual information between a random variable $X$ and itself conditioning on a random variable $Z$ is equal to the conditional entropy of $X$ given $Z$, i.e., $I(X; X|Z) = H(X|Z)$.

Proposition 2.22

\[
I(X; Y|Z) = H(X|Z) - H(X|Y, Z),
\]

\[
I(X; Y|Z) = H(Y|Z) - H(Y|X, Z),
\]

and

\[
I(X; Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z),
\]

provided that all the conditional entropies are finite.
Remark  All Shannon’s information measures are special cases of conditional mutual information. Let \( \Phi \) denote a random variable that takes a constant value. Then
Remark All Shannon’s information measures are special cases of conditional mutual information. Let $\Phi$ denote a random variable that takes a constant value. Then

$$H(X) = I(X; X|\Phi)$$
Remark  All Shannon’s information measures are special cases of conditional mutual information. Let $\Phi$ denote a random variable that takes a constant value. Then

\[
\begin{align*}
H(X) &= I(X; X|\Phi) \\
H(X|Z) &= I(X; X|Z)
\end{align*}
\]
**Remark**  All Shannon’s information measures are special cases of conditional mutual information. Let $\Phi$ denote a random variable that takes a constant value. Then

\[
\begin{align*}
H(X) &= I(X; X|\Phi) \\
H(X|Z) &= I(X; X|Z) \\
I(X; Y) &= I(X; Y|\Phi).
\end{align*}
\]