An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

July 18, 2016

Lecture #1B: Measure of Information

Outline of the lecture

- Introduction

Outline of the lecture

- Introduction
- Measures of information

Outline of the lecture

- Introduction
- Measures of information
- Definition of entropy, conditional entropy, relative entropy

Outline of the lecture

- Introduction
- Measures of information
- Definition of entropy, conditional entropy, relative entropy
- IT-inequality

Outline of the lecture

- Introduction
- Measures of information
- Definition of entropy, conditional entropy, relative entropy
- IT-inequality
- Properties of entropy

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy, m

Outline of the lecture

- Introduction
- Measures of information
- Definition of entropy, conditional entropy, relative entropy
- IT-inequality
- Properties of entropy
- Chain rules of entropy, mutual information

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Outline

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy, m

What to expect in this lecture

- How to quantify information?

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

What to expect in this lecture

- How to quantify information?
- Basic properties of entropy, relative entropy, mutual information

Outline

Hartley's Measure

- Information provided by observation of discrete random variable X is

$$I(X) = \log_b L$$

  where L is the number of possible values of X.

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Hartley's Measure

- Information provided by observation of discrete random variable X is

$$I(X) = \log_b L$$

  where L is the number of possible values of X.
- Question: What's wrong with Hartley's measure of information?

Shannon's Measure

- If i-th possible value of X has probability p_i, then the Hartley information $\log(1/p_i)$ should be weighted by p_i to give

$$-\sum_{i=1}^{L} p_i \log p_i$$

amount of information provided by X. This is Shannon's measure.

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Shannon's Measure

- If i-th possible value of X has probability p_i, then the Hartley information $\log(1/p_i)$ should be weighted by p_i to give

$$-\sum_{i=1}^{L} p_i \log p_i$$

amount of information provided by X. This is Shannon's measure.
- If f is any real-valued function, then support of f is defined as the subset of its domain where f takes on non-zero values, and is denoted by $supp(f)$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Shannon's Measure

- If i-th possible value of X has probability p_i, then the Hartley information $\log(1/p_i)$ should be weighted by p_i to give

$$-\sum_{i=1}^{L} p_i \log p_i$$

amount of information provided by X. This is Shannon's measure.
- If f is any real-valued function, then support of f is defined as the subset of its domain where f takes on non-zero values, and is denoted by $supp(f)$
- The uncertainty (entropy) of a discrete random variable X is given by

$$\begin{aligned} H(X) &= -\sum_{x \in supp(P_X)} P_X(x) \log_b P_X(x) \\ &= E[-\log P_X(X)] \end{aligned}$$

Outline

Entropy

- The joint entropy of discrete random variables X, and Y is given by

$$\begin{aligned} H(XY) &= E[-\log P_{XY}(X,Y)] \\ &= -\sum_{(x,y)\in supp(P_{XY})} P_{XY}(x,y)\log P_{XY}(x,y) \end{aligned}$$

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Entropy

- The joint entropy of discrete random variables X, and Y is given by

$$\begin{aligned} H(XY) &= E[-\log P_{XY}(X,Y)] \\ &= -\sum_{(x,y)\in supp(P_{XY})} P_{XY}(x,y)\log P_{XY}(x,y) \end{aligned}$$

- (Example) Suppose that X has two possible values x_1, x_2 and that $P_X(x_1) = p$ so that $P_X(x_2) = 1 - p$. Then the uncertainity of X in bits, provided that $0 < p < 1$

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p).$$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Entropy

- The joint entropy of discrete random variables X, and Y is given by

$$
\begin{aligned}
H(XY) &= E[-\log P_{XY}(X,Y)] \\
&= -\sum_{(x,y)\in supp(P_{XY})} P_{XY}(x,y)\log P_{XY}(x,y)
\end{aligned}
$$

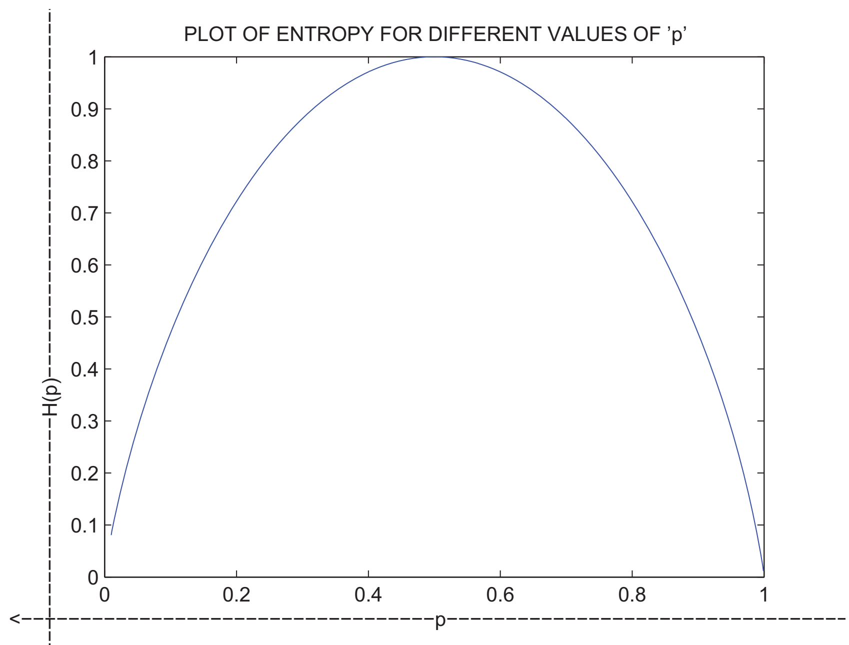
- (Example) Suppose that X has two possible values x_1, x_2 and that $P_X(x_1) = p$ so that $P_X(x_2) = 1 - p$. Then the uncertainity of X in bits, provided that $0 < p < 1$

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p).$$

- The above expression is known as binary entropy function and is denoted by $H(p)$.

Entropy

Figure of binary entropy function:



Outline

IT-inequality

- For a positive real number r.

$$\log r \leq (r-1)\log e$$

with equality if and only if $r = 1$.

IT-inequality

- For a positive real number r.

$$\log r \leq (r-1)\log e$$

with equality if and only if $r = 1$.

- Proof: The graphs of $\ln r$ and of $r-1$ coincide at $r = 1$. But

$$\begin{aligned} d(\ln r)/dr &= 1/r > 1; r < 1 \\ &= 1/r < 1; r > 1 \end{aligned}$$

so the graphs can never cross.

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy, r

IT-inequality

- For a positive real number r.

$$\log r \leq (r-1)\log e$$

with equality if and only if $r = 1$.

- Proof: The graphs of $\ln r$ and of $r-1$ coincide at $r = 1$. But

$$\begin{aligned} d(\ln r)/dr &= 1/r > 1; r < 1 \\ &= 1/r < 1; r > 1 \end{aligned}$$

so the graphs can never cross.

- Thus $\ln r \leq (r-1)$ with equality if and only if r = 1.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

IT-inequality

- For a positive real number r.

$$\log r \leq (r-1)\log e$$

with equality if and only if $r = 1$.

- Proof: The graphs of $\ln r$ and of $r-1$ coincide at $r = 1$. But

$$\begin{aligned} d(\ln r)/dr &= 1/r > 1; r < 1 \\ &= 1/r < 1; r > 1 \end{aligned}$$

so the graphs can never cross.

- Thus $\ln r \leq (r-1)$ with equality if and only if r = 1.
- Multiplying both sides of this inequality by $\log e$ and noting that $\log r = (\ln r)(\log e)$ gives the desired inequality.

Outline

Properties of entropy

- If the discrete random variable has L possible values then

$$0 \leq H(X) \leq \log L$$

with equality on the left if and only if $P_X(x) = 1$ for some x, and equality on the right if and only if $P_X(x) = 1/L$ for all x.

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Properties of entropy

- If the discrete random variable has L possible values then

$$0 \leq H(X) \leq \log L$$

with equality on the left if and only if $P_X(x) = 1$ for some x, and equality on the right if and only if $P_X(x) = 1/L$ for all x.
- Proof: If $x \in \text{supp}\,(P_X)$, then

$$\begin{aligned} -P_X(x)\log P_X(x) &= 0 \quad \text{for} \quad P_X(x) = 1 \\ -P_X(x)\log P_X(x) &> 0 \quad \text{for} \quad 0 < P_X(x) < 1 \end{aligned}$$

Properties of entropy

- If the discrete random variable has L possible values then

$$0 \leq H(X) \leq \log L$$

with equality on the left if and only if $P_X(x) = 1$ for some x, and equality on the right if and only if $P_X(x) = 1/L$ for all x.

- Proof: If x $\in$ supp (P_X), then

$$\begin{aligned} -P_X(x)\log P_X(x) &= 0 \quad \text{for} \quad P_X(x) = 1 \\ -P_X(x)\log P_X(x) &> 0 \quad \text{for} \quad 0 < P_X(x) < 1 \end{aligned}$$

- $H(X) = 0$ if and only if $P_X(x)$ equals 1 for every $x \in supp(P_X)$, but there can be only one such x.

Properties of entropy

- We use IT-inequality to prove the right inequality

$$\begin{aligned} H(X) - \log L &= \left[-\sum_{x\in supp(P_X)} P_X(x)\log P_X(x)\right] - \log L \\ &= \sum_{x\in supp(P_X)} P_X(x)\left[\log\frac{1}{P(x)} - \log L\right] \\ &= \sum_{x\in supp(P_X)} P_X(x)\log\frac{1}{LP(x)} \\ &\leq \sum_{x\in supp(P_X)} P_X(x)\left[\frac{1}{LP(x)} - 1\right]\log e \\ &= \left[\sum_{x\in supp(P_X)} \frac{1}{L} - \sum_{x\in supp(P_X)} P_X(x)\right]\log e \\ &\leq (1-1)\log e = 0 \end{aligned}$$

Properties of entropy

- $H_b(X) = (\log_b a) H_a(X)$.

Properties of entropy

- $H_b(X) = (\log_b a) H_a(X)$.
- Proof:

$$\log_b p = \log_b a \log_a p$$

Properties of entropy

- $H_b(X) = (\log_b a) H_a(X)$.
- Proof:

$$\log_b p = \log_b a \log_a p$$

- Hence,

$$\begin{aligned} -p\log_b p &= -p\log_b a \log_a p \\ \sum_x -p\log_b p &= (\log_b a)\sum_x -p\log_a p \\ H_b(X) &= (\log_b a) H_a(X) \end{aligned}$$

Conditional Entropy

Definition

- The conditional entropy of the discrete random variable X, given that the event Y = y occurs, is given by

$$\begin{aligned} H(X|Y=y) &= -\sum_{x\in supp(P_{X|Y}(x|y))} P(x|y)\log P(x|y) \\ &= E[-\log P_{X|Y}(X|Y)|Y=y]. \end{aligned}$$

Conditional Entropy

Definition

- The conditional entropy of the discrete random variable X, given that the event Y = y occurs, is given by

$$\begin{aligned} H(X|Y=y) &= -\sum_{x\in supp(P_{X|Y}(x|y))} P(x|y)\log P(x|y) \\ &= E[-\log P_{X|Y}(X|Y)|Y=y]. \end{aligned}$$

- The conditional entropy of the discrete random variable X given the discrete random variable Y is given by

$$\begin{aligned} H(X|Y) &= \sum_{y\in P_Y} P_Y(y)H(X|Y=y) \\ &= E[-\log P_{X|Y}(X|Y)] \end{aligned}$$

Relative entropy

- If X and $\hat{X}$ are discrete random variables with the same set of possible values then the information divergence between P_X and$P_{\hat{X}}$ is the quantity

$$D(P_X||P_{\hat{X}}) = \sum_{x\in supp(P_X)} P_X(x)\log\frac{P_X(x)}{P_{\hat{X}(x)}}$$

Relative entropy

- If X and $\hat{X}$ are discrete random variables with the same set of possible values then the information divergence between P_X and $P_{\hat{X}}$ is the quantity

$$D(P_X||P_{\hat{X}}) = \sum_{x \in supp(P_X)} P_X(x) \log \frac{P_X(x)}{P_{\hat{X}(x)}}$$

- Example: Suppose that X has L possible values,i.e; $\#X = L$. Let $\hat{X}$ have the probability distribution $P_{\hat{X}}(x) = 1/L$ for all $x \in X$.Then

$$\begin{aligned} D(P_X||P_{\hat{X}}) &= E[\log \frac{P_X(X)}{P_{\hat{X}}(\hat{X})}] \\ &= E[\log \frac{P_X(X)}{1/L}] \\ &= \log L - E[-\log P_X(X)] \\ &= \log L - H(X) \end{aligned}$$

Divergence Inequality

- Divergence Inequality For any two probability distributions p and q on X,

$$D(p||q) \geq 0$$

Divergence Inequality

- Divergence Inequality For any two probability distributions p and q on X,

$$D(p||q) \geq 0$$

- Proof: For $q(x) > 0$

$$\begin{aligned}
-D(p||q) &= \sum_{x \in supp(P_X)} p(x) \log \frac{q(x)}{p(x)} \\
&\leq \sum_{x \in supp(P_X)} p(x) \left(\frac{q(x)}{p(x)} - 1 \right) \log e \\
&= (\sum_{x \in supp(P_X)} q(x) - \sum_{x \in supp(P_X)} p(x)) \log e \\
&\leq (1-1) \log e \\
&= 0
\end{aligned}$$

Outline

Chain rules

- Entropy

$$H(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} H(X_i/X_{i-1}, \cdots, X_1)$$

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Chain rules

- Entropy

$$H(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} H(X_i/X_{i-1}, \cdots, X_1)$$

- Proof: We can write $P_{X_1 X_2, \cdots X_N}(X_1, X_2, \cdots, X_N)$ as

$$P_{X_1 X_2, \cdots X_N}(X_1, X_2, \cdots, X_N) = \prod_{i=1}^{n} P(X_i/X_{i-1}, \cdots, X_1)$$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Chain rules

- Entropy

$$H(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} H(X_i/X_{i-1}, \cdots, X_1)$$

- Proof: We can write $P_{X_1X_2,\cdots X_N}(X_1, X_2, \cdots, X_N)$ as

$$P_{X_1X_2,\cdots X_N}(X_1, X_2, \cdots, X_N) = \prod_{i=1}^{n} P(X_i/X_{i-1}, \cdots, X_1)$$

- Thus

$$\begin{aligned} H(X_1X_2\cdots X_N) &= E[-\log P_{X_1X_2,\cdots X_N}(X_1, X_2, \cdots, X_N)], \\ &= \sum_{i=1}^{n} H(X_i/X_{i-1}, \cdots, X_1) \end{aligned}$$

Mutual information

- The mutual information between the discrete random variables X and Y is the quantity

$$\begin{aligned} I(X;Y) &= \sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \\ &= \sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X/Y}(x|y)}{P_X(x)} \\ &= -\sum_{x,y} P_{X,Y}(x,y) \log P_X(x) + \sum_{x,y} P_{X,Y}(x,y) \log P_{X/Y}(x|y) \\ &= -\sum_{x} P_X(x) \log P_X(x) - (-\sum_{x,y} P_{X,Y}(x,y) \log P_{X/Y}(x|y)) \\ &= H(X) - H(X|Y) \end{aligned}$$

Mutual information

- The mutual information between the discrete random variables X and Y is the quantity

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

Mutual information

- The mutual information between the discrete random variables X and Y is the quantity

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

- We know that

$$\begin{aligned} H(XY) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y) \end{aligned}$$

Mutual information

- The mutual information between the discrete random variables X and Y is the quantity

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

- We know that

$$\begin{aligned} H(XY) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y) \end{aligned}$$

- This implies that

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

Properties of entropy

- For any two discrete random variables X and Y,

$$H(X|Y) \leq H(X)$$

with equality if and only if X and Y are independent random variables.

Properties of entropy

- For any two discrete random variables X and Y,

$$H(X|Y) \leq H(X)$$

with equality if and only if X and Y are independent random variables.

- Proof: We know that

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

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Properties of entropy

- For any two discrete random variables X and Y,

$$H(X|Y) \leq H(X)$$

with equality if and only if X and Y are independent random variables.

- Proof: We know that

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

- Also, $I(X;Y) = D(P_{X,Y}(x,y)||P_X(x)P_Y(y)) \geq 0$.

Properties of entropy

- For any two discrete random variables X and Y,

$$H(X|Y) \leq H(X)$$

with equality if and only if X and Y are independent random variables.

- Proof: We know that

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

- Also, $I(X;Y) = D(P_{X,Y}(x,y)||P_X(x)P_Y(y)) \geq 0$.
- From divergence inequality we get

$$I(X;Y) = H(X) - H(X|Y) \geq 0$$

with equality if $P_{X,Y}(x,y) = P_X(x)P_Y(y)$.

Chain rules

- Mutual information

$$I(X_1, X_2, \cdots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y/X_1, X_2, \cdots, X_{i-1})$$

Chain rules

- Mutual information

$$I(X_1, X_2, \cdots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y/X_1, X_2, \cdots, X_{i-1})$$

- Proof:

$$\begin{aligned} & I(X_1, X_2, \cdots, X_n; Y) \\ = & H(X_1, X_2, \cdots, X_n) - H(X_1, X_2, \cdots, X_n|Y) \\ = & \sum_{i=1}^{n} H(X_i|X_{i-1}, \cdots, X_1) - \sum_{i=1}^{n} H(X_i|X_{i-1}, \cdots, X_1, Y) \\ = & \sum_{i=1}^{n} I(X_i; Y|X_1, X_2, \cdots, X_{i-1}) \end{aligned}$$

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Chain rules

- Relative entropy

$$D(P_{X,Y}(x,y)||Q_{X,Y}(x,y)) = D(P_X(x)||Q_X(x)) + D(P_{Y/X}(y|x)||Q_{Y/X}(y|x)$$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Chain rules

- Relative entropy

$$D(P_{X,Y}(x,y)||Q_{X,Y}(x,y)) = D(P_X(x)||Q_X(x)) + D(P_{Y/X}(y|x)||Q_{Y/X}(y|x)$$

- Proof:

$$\begin{aligned}
& D(P_{X,Y}(x,y)||Q_{X,Y}(x,y)) \\
= & \sum_x \sum_y P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{Q_{X,Y}(x,y)} \\
= & \sum_x \sum_y P_{X,Y}(x,y) \log \frac{P_X(x)P_{Y/X}(y|x)}{Q_X(x)Q_{Y/X}(y|x)} \\
= & \sum_x \sum_y P_{X,Y}(x,y) \log \frac{P_X(x)}{Q_X(x)} + \sum_x \sum_y P_{X,Y}(x,y) \log \frac{P_{Y/X}(y|x)}{Q_{Y/X}(y|x)} \\
= & D(P_X(x)||Q_X(x)) + D(P_{Y/X}(y|x)||Q_{Y/X}(y|x))
\end{aligned}$$

Example

- A single unbiased die is tossed once. If the face of the die is 1, 2, 3, or 4, an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information about the face of the die by the number of heads obtained.

Example

- A single unbiased die is tossed once. If the face of the die is 1, 2, 3, or 4, an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information about the face of the die by the number of heads obtained.
- Let X be a random variable that denotes whether the face of the die is 1, 2, 3 or 4 or 5, 6.

Example

- A single unbiased die is tossed once. If the face of the die is 1, 2, 3, or 4, an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information about the face of the die by the number of heads obtained.
- Let X be a random variable that denotes whether the face of the die is 1, 2, 3 or 4 or 5, 6.
- Let Y be the random variable that denotes the number of heads.

Example

- A single unbiased die is tossed once. If the face of the die is 1, 2, 3, or 4, an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information about the face of the die by the number of heads obtained.
- Let X be a random variable that denotes whether the face of the die is 1, 2, 3 or 4 or 5, 6.
- Let Y be the random variable that denotes the number of heads.
- X takes two values x_1 when face of the die is 1, 2, 3, 4 and x_2 when face of the die is 5 or 6.

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Example

- A single unbiased die is tossed once. If the face of the die is 1, 2, 3, or 4, an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information about the face of the die by the number of heads obtained.
- Let X be a random variable that denotes whether the face of the die is 1, 2, 3 or 4 or 5, 6.
- Let Y be the random variable that denotes the number of heads.
- X takes two values x_1 when face of the die is 1, 2, 3, 4 and x_2 when face of the die is 5 or 6.
- Y takes three values y_0, y_1, y_2 depending upon whether there are no heads, one head or two heads.

Example

- A single unbiased die is tossed once. If the face of the die is 1, 2, 3, or 4, an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information about the face of the die by the number of heads obtained.
- Let X be a random variable that denotes whether the face of the die is 1, 2, 3 or 4 or 5, 6.
- Let Y be the random variable that denotes the number of heads.
- X takes two values x_1 when face of the die is 1, 2, 3, 4 and x_2 when face of the die is 5 or 6.
- Y takes three values y_0, y_1, y_2 depending upon whether there are no heads, one head or two heads.
- Since the die is unbiased $P(x_1) = 2/3$ and $P(x_2) = 1/3$.

Example

- A single unbiased die is tossed once. If the face of the die is 1, 2, 3, or 4, an unbiased coin is tossed once. If the face of the die is 5 or 6, the coin is tossed twice. Find the information about the face of the die by the number of heads obtained.
- Let X be a random variable that denotes whether the face of the die is 1, 2, 3 or 4 or 5, 6.
- Let Y be the random variable that denotes the number of heads.
- X takes two values x_1 when face of the die is 1, 2, 3, 4 and x_2 when face of the die is 5 or 6.
- Y takes three values y_0, y_1, y_2 depending upon whether there are no heads, one head or two heads.
- Since the die is unbiased $P(x_1) = 2/3$ and $P(x_2) = 1/3$.
- Since the coin is also unbiased $P(y_0/x_1) = P(y_1/x_1) = 1/2$. Also $P(y_2/x_1) = 0$.

Example (contd.)

- Similarly, $P(y_0/x_2) = P(y_2/x_2) = 1/4$ and $P(y_1/x_2) = 1/2$.

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Example (contd.)

- Similarly, $P(y_0/x_2) = P(y_2/x_2) = 1/4$ and $P(y_1/x_2) = 1/2$.
- Thus we have $P(y_0) = 5/12, P(y_1) = 1/2, P(y_2) = 1/12$.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Example (contd.)

- Similarly, $P(y_0/x_2) = P(y_2/x_2) = 1/4$ and $P(y_1/x_2) = 1/2$.
- Thus we have $P(y_0) = 5/12, P(y_1) = 1/2, P(y_2) = 1/12$.
- He know that

$$H(Y) = H(5/12, 1/2, 1/12) = 1.325 \text{ bits}$$

Introduction Measure of Information Definition of entropy, conditional entropy, relative entropy IT-inequality Properties of entropy Chain rules of entropy,

Example (contd.)

- Similarly, $P(y_0/x_2) = P(y_2/x_2) = 1/4$ and $P(y_1/x_2) = 1/2$.
- Thus we have $P(y_0) = 5/12, P(y_1) = 1/2, P(y_2) = 1/12$.
- He know that

$$H(Y) = H(5/12, 1/2, 1/12) = 1.325 \text{ bits}$$

- Also

$$\begin{aligned} H(Y/X) &= P(x_1) \cdot H(Y/X = x_1) + P(x_2) \cdot H(Y/X = x_2) \\ &= 2/3 \cdot H(1/2, 1/2, 0) + 1/3 \cdot H(1/4, 1/2, 1/4) \\ &= 2/3 \cdot 1 + 1/3 \cdot 3/2 = 1.167 \text{ bits} \end{aligned}$$

Example (contd.)

- Similarly, $P(y_0/x_2) = P(y_2/x_2) = 1/4$ and $P(y_1/x_2) = 1/2$.
- Thus we have $P(y_0) = 5/12, P(y_1) = 1/2, P(y_2) = 1/12$.
- He know that

$$H(Y) = H(5/12, 1/2, 1/12) = 1.325 \text{ bits}$$

- Also

$$\begin{aligned} H(Y/X) &= P(x_1) \cdot H(Y/X = x_1) + P(x_2) \cdot H(Y/X = x_2) \\ &= 2/3 \cdot H(1/2, 1/2, 0) + 1/3 \cdot H(1/4, 1/2, 1/4) \\ &= 2/3 \cdot 1 + 1/3 \cdot 3/2 = 1.167 \text{ bits} \end{aligned}$$

- Thus we have

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y/X) \\ &= 1.325 - 1.167 = 0.158 \text{ bits} \end{aligned}$$