HOMEWORK 3 (2023) — SOLUTIONS

JAIST — SCHOOL OF INFORMATION SCIENCE — I232 INFORMATION THEORY

1. What is the correct relationship, $=$, $\geq$, $\leq$ or ? (for unknown) for each pair below. Give reason with equations or phrase like “conditioning reduces entropy.”

   1. $I(\mathsf{X};\mathsf{Y})$ ______ 0.

      Solution: Mutual information is non-negative, so $I(\mathsf{X};\mathsf{Y}) \geq 0$.

   2. $H(\mathsf{X},\mathsf{Y})$______ $H(\mathsf{X}) + H(\mathsf{Y})$

      Solution: $\leq$ Chain rule of entropy and conditioning reduces entropy.

   3. $I(\mathsf{X};\mathsf{Y}) + H(\mathsf{X}|\mathsf{Y})$ ______ $H(\mathsf{X})$.

      Solution: Variation on the definition of mutual information, so $=$.

   4. $I(\mathsf{X};\mathsf{X})$______ $H(\mathsf{X})$.

      Solution: By definition of mutual information, $I(\mathsf{X};\mathsf{X}) = H(\mathsf{X}) - H(\mathsf{X}|\mathsf{X})$. $H(\mathsf{X}|\mathsf{X}) = 0$, so $I(\mathsf{X};\mathsf{X}) = \ H(\mathsf{X})$.

   5. $I(\mathsf{X};\mathsf{Y})$ ______ $H(\mathsf{X}) - H(g(\mathsf{Y})|\mathsf{Y})$.

      Solution: By conditional entropy of functions, $H(g(\mathsf{Y})|\mathsf{Y}) = 0$. From the definition of mutual information $I(\mathsf{X};\mathsf{Y}) = H(\mathsf{X}) - H(\mathsf{X}|\mathsf{Y})$, we have $I(\mathsf{X};\mathsf{Y}) \leq H(\mathsf{X})$.

   6. $H(\mathsf{X}|\mathsf{Y})$______ $H(\mathsf{X}) + H(\mathsf{Y})$

      Solution: Conditioning reduces entropy: $H(\mathsf{X}|\mathsf{Y}) \leq H(\mathsf{X})$ and non-negativity of entropy: $0 \leq H(\mathsf{Y})$, correct answer is $\leq$.

   7. $H(2\mathsf{X})$ ______ $H(\mathsf{X})$

      Solution: $2X$ is deterministic function of X, so $H(2\mathsf{X}) \ = \ H(\mathsf{X})$.

   8. $H(\mathsf{X}_2|\mathsf{X}_1)$______ $H(\mathsf{X}_2|\mathsf{X}_1,\mathsf{X}_0)$

      Solution: Conditioning reduces entropy: $H(\mathsf{X}_2|\mathsf{X}_1) \geq \ H(\mathsf{X}_2|\mathsf{X}_1,\mathsf{X}_0)$.

2. Consider jointly distributed X and Y with $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1\}$ with joint distribution given by:

$p_{\mathsf{XY}}(x,y)$	$y=0$	$y=1$
$x=0$	$\frac{1}{3}$	$\frac{1}{6}$
$x=1$	$\frac{1}{6}$	$\frac{1}{3}$

   Let Z be a new random variable $\mathsf{Z} = \mathsf{X} + \mathsf{Y}$. Here “+” means real addition so $\mathcal{Z} = \{0,1,2\}$.

(a) Find $H(\mathsf{X})$, $H(\mathsf{Y})$ and $H(\mathsf{X},\mathsf{Y})$

(b) Find the joint distribution $p_{\mathsf{XYZ}}(x,y,z)$.

(c) Find $I(\mathsf{X},\mathsf{Y};\mathsf{Z})$.

(d) Find $I(\mathsf{X};\mathsf{Z})$.

Solution:

(a) To calculate $H(\mathsf{X})$ and $H(\mathsf{Y})$, we first calculate $p_{\mathsf{X}}(x) = \sum_{y\in\mathcal{Y}} p_{\mathsf{XY}}(x,y)$ and $p_{\mathsf{Y}}(y) = \sum_{x\in\mathcal{X}} p_{\mathsf{XY}}(x,y)$. Then $H(\mathsf{X}) = h(\frac{1}{2}) = 1$ and $H(\mathsf{Y}) = h(\frac{1}{2}) = 1$

$$\begin{aligned} H(\mathsf{X},\mathsf{Y}) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{\mathsf{XY}}(x,y)\log p_{\mathsf{XY}}(x,y) = -\big(2\frac{1}{3}\log\frac{1}{3} + 2\frac{1}{6}\log\frac{1}{6}\big) \\ &= \frac{2}{3}\log 3 + \frac{1}{3}\log 6 = \frac{2}{3}\log 3 + \frac{1}{3}\log 3 + \frac{1}{3}\log 2 \\ &= \frac{1}{3} + \log 3 \quad \big(\text{Other answers: } h(\frac{1}{2}) + h(\frac{1}{3}) = \frac{1}{3}\log 54 \approx 1.9182\big) \end{aligned}$$

(b) One way to write a joint distribution:

$p_{\mathsf{XYZ}}(x,y,z)$	$z=0$	$z=1$	$z=2$
$x=0,y=0$	$\frac{1}{3}$	0	0
$x=0,y=1$	0	$\frac{1}{6}$	0
$x=1,y=0$	0	$\frac{1}{6}$	0
$x=1,y=1$	0	0	$\frac{1}{3}$

(c) $H(\mathsf{Z}|\mathsf{X},\mathsf{Y}) = 0$ because Z is a deterministic function of X and Y. $I(\mathsf{X},\mathsf{Y};\mathsf{Z}) = H(\mathsf{Z}) - H(\mathsf{Z}|\mathsf{X},\mathsf{Y}) = \log 3$.

(d) Two possible solutions are given:

$$\begin{aligned} I(\mathsf{X};\mathsf{Z}) &= I(\mathsf{X},\mathsf{Y};\mathsf{Z}) - I(\mathsf{Y};\mathsf{Z}|\mathsf{X}) && \text{Chain rule of mutual information} \\ &= \log 3 - (H(\mathsf{Y}|\mathsf{X}) - H(\mathsf{Y}|\mathsf{Z},\mathsf{X})) && \text{Using result of (c) and definition of mutual information} \\ &= \log 3 - h(\frac{1}{3}) + 0 && \text{Compute } H(\mathsf{Y}|\mathsf{X}) \text{ from } p_{\mathsf{Y}|\mathsf{X}}(y|x)\text{; given } \mathsf{Z} \text{ and } \mathsf{X}, \mathsf{Y} \text{ is known} \\ &= \frac{2}{3} \end{aligned}$$

$$\begin{aligned} I(\mathsf{X};\mathsf{Z}) &= H(\mathsf{Z}) - H(\mathsf{Z}|\mathsf{X}) && \text{Definition of mutual information} \\ &= \log 3 - H(\mathsf{Z}|\mathsf{X}) && \text{See part (b)} \\ &= \log 3 - \sum_{x\in\mathcal{X}} p_{\mathsf{X}}(x) H(\mathsf{Z}|\mathsf{X}=x) \\ &= \log 3 - h(\frac{1}{3}) = \frac{2}{3} && p_{\mathsf{Z}|\mathsf{X}}(0|0) = 2/3, p_{\mathsf{Z}|\mathsf{X}}(1|0) = 1/3 \text{ and } p_{\mathsf{X}}(x) = 1/2 \end{aligned}$$

3. Let X,Y and Z be jointly distributed variables such that $\mathsf{X}\to\mathsf{Y}\to\mathsf{Z}$ forms a Markov chain. Prove $H(\mathsf{X}|\mathsf{Z}) \geq H(\mathsf{X}|\mathsf{Y})$.

Solution: Since $\mathsf{X}\to\mathsf{Y}\to\mathsf{Z}$ forms a Markov chain, the data processing inequality holds.

$$\begin{aligned} I(\mathsf{X};\mathsf{Y}) &\geq I(\mathsf{X};\mathsf{Z}) && \text{data processing inequality} \\ H(\mathsf{X}) - H(\mathsf{X}|\mathsf{Y}) &\geq H(\mathsf{X}) - H(\mathsf{X}|\mathsf{Z}) && \text{definition of mutual information} \\ H(\mathsf{X}|\mathsf{Z}) &\geq H(\mathsf{X}|\mathsf{Y}) \end{aligned}$$