JAIST — School of Information Science — I232E Information Theory

Midterm Exam 2022

Instructor: Brian M. Kurkoski and Lei Liu

May 20, 2022 13:30–15:10

SOLUTIONS

Exam Policy. This is a *closed book* exam. You may use:

- One page of notes, A4-sized paper, double-sided OK.

You may not use anything else:

- No printed materials, including books, lecture notes and slides
- No notes (except as above)
- No internet-connected devices
- No calculators. Answers such as “log 3” are acceptable.

Question	Points	Score
1	25	
2	20	
3	25	
4	30	
Total:	100	

This exam has 10 pages.

1. What is the correct relationship, $=$, $\geq$, $\leq$ or ? (for unknown) for each pair below.

   (a) (2 points) $I(\mathsf{X};\mathsf{Y})$ ______ 0.

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

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

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

   (c) (2 points) $I(\mathsf{X};\mathsf{Y}) + H(\mathsf{X}|\mathsf{Y})$ ______ $H(\mathsf{X})$.

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

   (d) (2 points) $I(\mathsf{X};\mathsf{X})$______ $H(\mathsf{X})$.

   Solution: $I(\mathsf{X};\mathsf{X}) = H(\mathsf{X})$ since $H(\mathsf{X}|\mathsf{X}) = 0$.

   (e) (2 points) $I(\mathsf{X};\mathsf{Y})$ ______ $H(\mathsf{X}) - H(g(\mathsf{Y})|\mathsf{Y})$.

   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 $I(\mathsf{X};\mathsf{Y}) \leq H(\mathsf{X})$.

   (f) (2 points) $H(\mathsf{X}|\mathsf{Y})$______ $H(\mathsf{X}) + H(\mathsf{Y})$

   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 $\leq$.

   (g) (2 points) $H(2\mathsf{X})$ ______ $H(\mathsf{X})$

   Solution: For a discrete random variable X, the probability vectors of $2\mathsf{X}$ and X are the same, so correct answer is $H(2\mathsf{X}) = H(\mathsf{X})$.

   (h) (2 points) $H(\mathsf{X}^2)$ ______ $H(\mathsf{X})$

   Solution: The correct answer is $H(\mathsf{X}^2) \leq H(\mathsf{X})$, because

$$\begin{aligned} H(\mathsf{X}^2, \mathsf{X}^2) &= H(\mathsf{X}^2) + \underbrace{H(\mathsf{X}|\mathsf{X}^2)}_{\geq 0} \\ &= H(\mathsf{X}) + \underbrace{H(\mathsf{X}^2|\mathsf{X})}_{=0}. \end{aligned}$$

   (i) (2 points) $H(\mathsf{X}_2|\mathsf{X}_1)$______ $H(\mathsf{X}_2|\mathsf{X}_1, \mathsf{X}_0)$

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

   (j) (3 points) $H(\mathsf{X},\mathsf{Y}) + I(\mathsf{X};\mathsf{Y})$______ $H(\mathsf{X}) + H(\mathsf{Y})$

   Solution:

$$\begin{aligned} H(\mathsf{X},\mathsf{Y}) + I(\mathsf{X};\mathsf{Y}) &= \underbrace{H(\mathsf{X}) + H(\mathsf{Y}|\mathsf{X})}_{=H(\mathsf{X},\mathsf{Y})\text{ (chain rule)}} + \underbrace{H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X})}_{I(\mathsf{X};\mathsf{Y})} \\ &= I(\mathsf{X}) + H(\mathsf{Y}). \end{aligned}$$

(k) (4 points) $H(\mathsf{X},\mathsf{Y})$______ $H(\mathsf{X}|\mathsf{Y}) + H(\mathsf{Y}|\mathsf{X}) + I(\mathsf{X};\mathsf{Y})$

Solution:

$$\begin{aligned} H(\mathsf{X},\mathsf{Y}) &= H(\mathsf{X}) + H(\mathsf{Y}|\mathsf{X}) \\ &= \underbrace{H(\mathsf{X}) - H(\mathsf{X}|\mathsf{Y})}_{=I(\mathsf{X};\mathsf{Y})} + H(\mathsf{X}|\mathsf{Y}) + H(\mathsf{Y}|\mathsf{X}) \\ &= I(\mathsf{X};\mathsf{Y}) + H(\mathsf{X}|\mathsf{Y}) + H(\mathsf{Y}\|\mathsf{X}). \end{aligned}$$

2. Let X be defined on $\mathcal{X} = \{-1, 0, 1\}$ with $p_{\mathsf{X}}(x) = [\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$. Let $g(x) = x^2$ and let $\mathsf{Y} = g(\mathsf{X})$, so that $\mathcal{Y} = \{0, 1\}$ and $p_{\mathsf{Y}}(0) = 1/2$ and $p_{\mathsf{Y}}(1) = 1/2$.
   (a) (4 points) Compute $E[\mathsf{X}]$ and $E[g(\mathsf{X})]$.
   (b) (4 points) What is $H(\mathsf{Y}|\mathsf{X})$? It is easily found without computations.
   (c) (4 points) Find $p_{\mathsf{XY}}(x, y)$. Find $p_{\mathsf{X}|\mathsf{Y}}(x|y)$.
   (d) (4 points) Compute $H(\mathsf{X}|\mathsf{Y} = 0)$ and $H(\mathsf{X}|\mathsf{Y} = 1)$.
   (e) (4 points) Compute $H(\mathsf{X}|\mathsf{Y})$.

Solution: (a)

$$\begin{aligned} E[\mathsf{X}] &= \sum_{x\in\mathcal{X}} x p_{\mathsf{X}}(x) \\ &= (-1)\frac{1}{4} + 0\frac{1}{2} + 1\frac{1}{4} = 0 \\ E[g(\mathsf{X})] &= \sum_{x\in\mathcal{X}} g(x) p_{\mathsf{X}}(x) \\ &= (-1)^2\frac{1}{4} + 0\frac{1}{2} + 1^2\frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

(b) $H(\mathsf{Y}|\mathsf{X}) = 0$ because if you know X, you can always compute Y, that is $H(g(\mathsf{X})|\mathsf{X}) = 0$.

(c) Note it is easy to find $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ from the problem statement:

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

since $y = x^2$. Since $p_{\mathsf{XY}}(x, y) = p_{\mathsf{Y}|\mathsf{X}}(y|x)p_{\mathsf{X}}(x)$,

$$p_{\mathsf{XY}}(x, y) = \begin{bmatrix} 0 & 1/4 \\ 1/2 & 0 \\ 0 & 1/4 \end{bmatrix},$$

Since $p_{\mathsf{Y}|\mathsf{X}}(y|x)$, $p_{\mathsf{X}}(x)$ and $p_{\mathsf{Y}}(y)$ are all known, use Bayes rule to find $p_{\mathsf{X}|\mathsf{Y}}(x|y)$:

$$\begin{aligned} p_{\mathsf{X}|\mathsf{Y}}(x|y) &= \frac{p_{\mathsf{Y}|\mathsf{X}}(y|x)p_{\mathsf{X}}(x)}{p_{\mathsf{Y}}(y)} \\ p_{\mathsf{X}|\mathsf{Y}}(-1|0) &= \frac{p_{\mathsf{Y}|\mathsf{X}}(0|-1)p_{\mathsf{X}}(-1)}{p_{\mathsf{Y}}(0)} = \frac{0\cdot 1/4}{1/2} = 0 \\ p_{\mathsf{X}|\mathsf{Y}}(0|0) &= \frac{p_{\mathsf{Y}|\mathsf{X}}(0|0)p_{\mathsf{X}}(0)}{p_{\mathsf{Y}}(0)} = \frac{1\cdot 1/2}{1/2} = 1 \\ p_{\mathsf{X}|\mathsf{Y}}(-1|1) &= \frac{p_{\mathsf{Y}|\mathsf{X}}(1|-1)p_{\mathsf{X}}(-1)}{p_{\mathsf{Y}}(1)} = \frac{1\cdot 1/4}{1/2} = \frac{1}{2}\cdots \\ \cdots\text{ so that } p_{\mathsf{X}|\mathsf{Y}}(x|y) &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

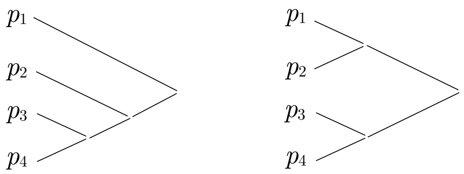
(d)

$$\begin{aligned} H(\mathsf{X}|\mathsf{Y}=0) &= -\sum_{x\in\mathcal{X}} p_{\mathsf{X}|\mathsf{Y}}(x|0)\log p_{\mathsf{X}|\mathsf{Y}}(x|0) = -0\log 0 - 1\log 1 - 0\log 0 = 0 \\ H(\mathsf{X}|\mathsf{Y}=1) &= -\sum_{x\in\mathcal{X}} p_{\mathsf{X}|\mathsf{Y}}(x|1)\log p_{\mathsf{X}|\mathsf{Y}}(x|1) = -\frac{1}{2}\log\frac{1}{2} - 0\log 0 - \frac{1}{2}\log\frac{1}{2} = 1 \text{ bit} \end{aligned}$$

(e)

$$H(\mathsf{X}|\mathsf{Y}) = p_{\mathsf{Y}}(0)H(\mathsf{X}|\mathsf{Y}=1) + p_{\mathsf{Y}}(1)H(\mathsf{X}|\mathsf{Y}=1) = \frac{1}{2}\cdot 0 + \frac{1}{2}\cdot 1 = \frac{1}{2} \text{ bit}$$

3. *Huffman code trees* Consider a source with $\mathcal{X} = \{1, 2, 3, 4\}$ and $p_1 > p_2 > p_3 > p_4$ and $p_1 + p_2 + p_3 + p_4 = 1$. There are only two possible binary Huffman codes for this source, with corresponding trees (A) and (B):



(A) (B)

(a) (5 points) Let $(p_1, p_2, p_3, p_4) = (0.39, 0.21, 0.2, 0.2)$. Give a binary Huffman code for this source. What is the expected codeword length?

(b) (5 points) Give a ternary Huffman code for the source in part (a). What is the expected codeword length?

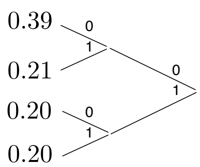
(c) (5 points) Give an inequality using p_1, p_3 and p_4 such that tree (A) is always obtained.

(d) (5 points) Show that if $p_1 > \frac{2}{5}$ then $p_3 + p_4 < \frac{2}{5}$.

(e) (5 points) Show that if $p_1 > \frac{2}{5}$, then the length of the corresponding Huffman codeword for "$x = 1$" is 1.

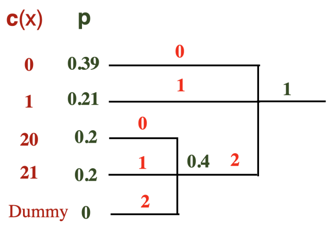
Solution:

(a) There are many possible codes. One code is below, but all codes have expected length 2.



$p_X(x)$	code
0.39	00
0.21	00
0.20	01
0.20	10

(b) There are many possible codes. One code is below.



$p_X(x)$	code
0.39	0
0.21	1
0.20	20
0.20	21

The expected length is $0.6 * 1 + 0.4 * 2 = 1.4$.

(c) Note that symbol 3 and symbol 4 are always combined first. Tree (B) will be obtained only when $p_3 + p_4$ is greater than p_1 (which is greater than p_2). So, tree (A) is obtained when $p_3 + p_4 \le p_1$.

(d) Assume that $p_3 + p_4 \ge 2/5$.

(1) From $p_3 + p_4 \ge 2/5$ and $p_3 > p_4$, we have $p_3 > 1/5$. Since $p_2 > p_3$, we then have $p_2 > 1/5$.

(2) From $p_3 + p_4 \ge 2/5$, $p_1 > \frac{2}{5}$ and $p_1 + p_2 + p_3 + p_4 = 1$, we then have $p_2 = 1 - (p_2 + p_3 + p_4) < 1/5$, which contradicts the result in point (1).

Therefore, $p_3 + p_4 \ge 2/5$ does not hold. That is, $p_3 + p_4 < 2/5$.

(e) If $p_1 > \frac{2}{5}$, then by part (d), we have that $p_3 + p_4 < \frac{2}{5}$. Then, $p_1 > p_3 + p_4$, so by the condition in part (c), we must have Tree (A), which has the codeword for "$x = 1$" with length 1.

4. Consider a two-state Markov chain $\mathbf{X} = [\mathsf{X}_1, \mathsf{X}_2, \cdots]$ with probability transition matrix:

$\mathbf{P}_{\mathsf{X}_n\|\mathsf{X}_{n-1}}$	$x_{n-1} = 0$	$x_{n-1} = 1$
$x_n = 0$	$4/5$	$1/2$
$x_n = 1$	$1/5$	$1/2$

(a) (3 points) What is the stationary distribution $\mathbf{p}_\mathsf{X}$?

Solution:

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ \{\mathbf{P} - \mathbf{I}\}_{\backslash \mathbf{r}_1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1/5 & -1/2 \end{bmatrix},$$

$$\mathbf{p}_\mathsf{X} = \mathbf{Q}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{-1/2 - 1/5} \begin{bmatrix} -1/2 & -1 \\ -1/5 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5/7 \\ 2/7 \end{bmatrix} \approx \begin{bmatrix} 0.714 \\ 0.286 \end{bmatrix}.$$

(b) (3 points) What is the entropy rate $H(\mathcal{X})$?

Solution: $H(\mathcal{X}) = H(\mathsf{X}_n|\mathsf{X}_{n-1}) = \frac{5}{7}h(1/5) + \frac{2}{7}h(1/2) = \frac{5}{7}h(1/5) + \frac{2}{7} \approx 0.801$ bits.

(c) (3 points) What is the (single-variable) entropy $\lim_{n\to\infty} H(\mathsf{X}_n)$?

Solution: $\lim_{n\to\infty} H(\mathsf{X}_n) = h(\mathbf{p}_\mathsf{X}) = h(2/7) \approx 0.863$ bits.

(d) (3 points) Which has lower compression rate, compression using the Markov property, or single-variable compression?

Solution: Since $\lim_{n\to\infty} H(\mathsf{X}_n) > H(\mathcal{X})$, compression using the Markov property has lower compression rate .

Let $\mathbf{Y} = [\mathsf{Y}_1, \mathsf{Y}_2, \ldots]$ and $\mathsf{Y}_n = [\mathsf{X}_{2n-1}, \mathsf{X}_{2n}]$. Then $\mathsf{Y}_1 = [\mathsf{X}_1, \mathsf{X}_2], \mathsf{Y}_2 = [\mathsf{X}_3, \mathsf{X}_4], \ldots$ is a four-state Markov chain.

(e) (4 points) What is the probability transition matrix $\mathbf{P}_{\mathsf{Y}_n|\mathsf{X}_{2n-2}}$? What is the probability transition matrix $\mathbf{P}_{\mathsf{Y}_n|\mathsf{Y}_{n-1}}$?

Solution: Following the Markov property, we have

$$P_{\mathsf{Y}_n|\mathsf{X}_{2n-2}} = P_{\mathsf{X}_{2n-1},\mathsf{X}_{2n}|\mathsf{X}_{2n-2}} = P_{\mathsf{X}_{2n-1}|\mathsf{X}_{2n-2}} P_{\mathsf{X}_{2n}|\mathsf{X}_{2n-1}},$$

$\mathbf{P}_{\mathsf{Y}_n\|\mathsf{X}_{2n-2}}$	$x_{2n-2} = 0$	$x_{2n-2} = 1$
$y_n = 00$	$16/25$	$2/5$
$y_n = 01$	$4/25$	$1/10$
$y_n = 10$	$1/10$	$1/4$
$y_n = 11$	$1/10$	$1/4$

.

Similarly,

$$P_{\mathsf{Y}_n|\mathsf{Y}_{n-1}} = P_{\mathsf{X}_{2n-1},\mathsf{X}_{2n}|\mathsf{X}_{2n-3},\mathsf{X}_{2n-2}} = P_{\mathsf{X}_{2n-1},\mathsf{X}_{2n}|\mathsf{X}_{2n-2}} = P_{\mathsf{Y}_n|\mathsf{X}_{2n-2}},$$

$\mathbf{P}_{\mathsf{Y}_n\|\mathsf{Y}_{n-1}}$	$y_{n-1} = 00$	$y_{n-1} = 01$	$y_{n-1} = 10$	$y_{n-1} = 11$
$y_n = 00$	$16/25$	$2/5$	$16/25$	$2/5$
$y_n = 01$	$4/25$	$1/10$	$4/25$	$1/10$
$y_n = 10$	$1/10$	$1/4$	$1/10$	$1/4$
$y_n = 11$	$1/10$	$1/4$	$1/10$	$1/4$

.

(f) (2 points) What is the stationary distribution $\mathbf{p}_\mathsf{Y}$? (Hint: Matrix inverse is not necessary. You can utilize the stationary distribution $\mathbf{p}_\mathsf{X}$.)

Solution:

$$\mathbf{p}_\mathsf{Y} = \mathbf{P}_{\mathsf{Y}_n|\mathsf{X}_{2n-2}} * \mathbf{P}_\mathsf{X} = \begin{bmatrix} 4/7 \\ 1/7 \\ 1/7 \\ 1/7 \end{bmatrix}.$$

(g) (3 points) What is the entropy rate $H(\mathcal{Y})$?

Solution: $H(\mathcal{Y}) = 2H(\mathcal{X}) = \frac{10}{7}h(1/5) + \frac{4}{7} \approx 1.602$.

(h) (3 points) Which has lower compression rate per symbol, compression using the Markov property of $\mathbf{Y}$, or compression using the Markov property of $\mathbf{X}$?

Solution: Since $H(\mathcal{Y}) = 2H(\mathcal{X})$, both of them are optimal. Compression using the Markov property of $\mathbf{Y}$ has the same compression rate per symbol as that of $\mathbf{X}$.

(i) (3 points) What is the (two-variable) entropy $\lim_{n\to\infty} H(\mathsf{Y}_n)$?

Solution: $\lim_{n\to\infty} H(\mathsf{Y}_n) = h(\mathbf{p}_\mathsf{Y}) = -\frac{4}{7}\log_2\frac{4}{7} - 3 * \frac{1}{7}\log_2\frac{1}{7} = \log_2 7 - \dfrac{8}{7} \approx 1.665$ bits.

(j) (3 points) Which has lower compression rate per symbol, two-variable compression, or single-variable compression?

Solution: $\lim_{n\to\infty} H(\mathsf{Y}_n) \approx 1.665$ bits $< 2 * \lim_{n\to\infty} H(\mathsf{X}_n) \approx 1.726$ bits. Hence, two-variable compression has lower compression rate per symbol than single-variable compression.

Notes: $\log_2 5 \approx 2.3219, \log_2 7 \approx 2.8074$, and you may use the binary entropy function $h(p) \equiv -p\log p - (1-p)\log(1-p)$ for the questions above.