2005 Spring Open Book Midterm for Information Theory

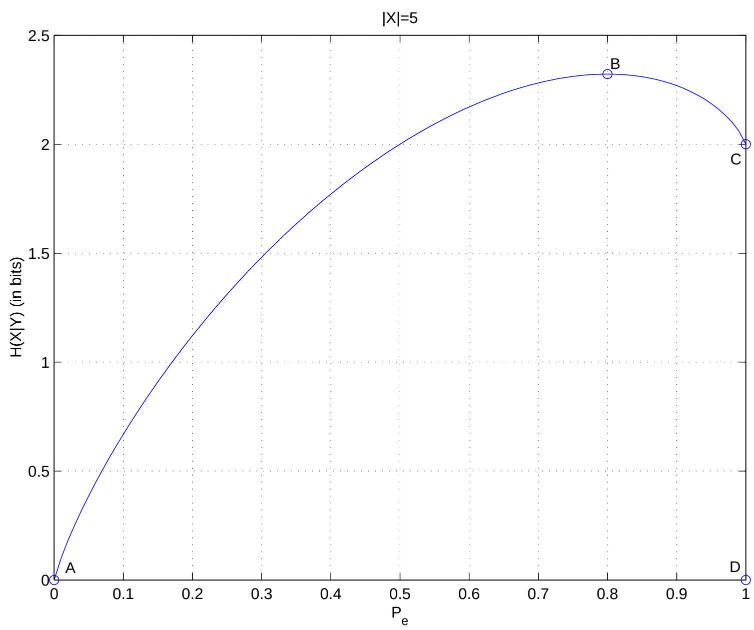
1. Assume that the alphabets for random variables X and Y are both $\{1,2,3,4,5\}$. Let $\hat{x} = g(y)$ be an estimate of x from observing y. Define the probability of estimating error as:

$$P_e = \Pr\{g(Y) \neq X\}.$$

Then, the Fano's inequality gives bounds for P_e as:

$$H_b(P_e) + 2P_e \geq H(X|Y),$$

where $H_b(p) = p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$ is the binary entropy function. The curve for $H_b(P_e) + 2P_e = H(X|Y)$ is plotted below.



(a) (8 pt) Point A on the above figure shows that if $H(X|Y) = 0$, zero estimation error, namely, $P_e = 0$, can be achieved. In this case, characterize the distribution $P_{X|Y}$. Also, give an estimator $g(\cdot)$ that achieves $P_e = 0$. (Hint: Think of what kind of statistical relation between X and Y can render $H(X|Y) = 0$.)

(b) (8 pt) Point B on the above figure indicates that when $H(X|Y) = \log_2(5)$, the estimation error can only be equal to 0.8. In this case, characterize the distributions $P_{X|Y}$ and P_X. Prove that at $H(X|Y) = \log_2(5)$, all estimators yield $P_e = 0.8$. (Hint: Think of what kind of statistical relation between X and Y can render $H(X|Y) = \log_2(5)$.)

(c) (4 pt) Point C on the above figure hints that when $H(X|Y) = 2$, the estimation error can be as worse as 1. Give an estimator $g(\cdot)$ that leads to $P_e = 1$, if $P_{X|Y}(x|y) = 1/4$ for $x \neq y$, and $P_{X|Y}(x|y) = 0$ for $x = y$. (Hint: The answer is apparent, isn't it?)

(d) (4 pt) Similarly, point D on the above figure hints that when $H(X|Y) = 0$, the estimation error can be as worse as 1. Give an estimator $g(\cdot)$ that leads to $P_e = 1$ at $H(X|Y) = 0$. (Hint: The answer is apparent, isn't it?)

Answer:

(a) $H(X|Y) = 0$ means that X is deterministic given Y, namely, $P_{X|Y}(x|y)$ is either 1 or 0. Therefore, the choice that $g(y)$ is equal to the x for which $P_{X|Y}(x|y) = 1$ can achieve $P_e = 0$.

(b) Since $\log_2(5) = H(X|Y) \leq H(X) \leq \log_2(5)$, we have $H(X|Y) = H(X)$, which implies that X and Y are independent. Hence, $P_{X|Y}(x|y) = P_X(x)$. In addition, $H(X) = \log_2(5)$ implies that X is uniformly distributed, namely, $P_X(x) = 1/5$. This concludes that $P_{X|Y}(x|y) = 1/5$. Finally, any estimator $g(\cdot)$ gives $P_e = 0.8$, because

$$\begin{aligned}
\Pr\{g(Y) \neq X\} &= 1 - \sum_{x\in\mathcal{X}} P_X(x) \Pr\{g(Y) = x | X = x\} \\
&= 1 - \sum_{x\in\mathcal{X}} P_X(x) \Pr\{g(Y) = x\} \\
&= 1 - \frac{1}{5} \sum_{x\in\mathcal{X}} \Pr\{g(Y) = x\} \\
&= 1 - \frac{1}{5} = \frac{4}{5}.
\end{aligned}$$

(c) Let $g(y) = y$. Then, $\Pr\{g(Y) = X\} = \sum_{y\in\mathcal{Y}} P_Y(y) P_{X|Y}(y|y) = 0$.

(d) That $g(y) = x$ for any x with $P_{X|Y}(x|y) = 0$ satisfies the need.

2. (12 pt) In the second part of Theorem 3.22 (slide I:3-49), it is shown that there exists a prefix code with

$$\bar{\ell} = \sum_x P_x(x) \ell(\boldsymbol{c}_x) \leq H(X) + 1,$$

where $\boldsymbol{c}_x$ is the codeword for the source symbol x and $\ell(\boldsymbol{c}_x)$ is the length of codeword $\boldsymbol{c}_x$. Show that the upper bound can be improved to:

$$\bar{\ell} < H(X) + 1.$$

(Hint: Replace $\ell(\boldsymbol{c}_x) = \lfloor -\log_2 P_X(x) \rfloor + 1$ by a new assignment.)

Answer: We can modify the proof of Theorem 3.22 as follows.

Choose the codeword length for source symbol x as

$$\ell(\boldsymbol{c}_x) = \lceil -\log_2 P_X(x) \rceil. \tag{1}$$

Then

$$2^{-\ell(\boldsymbol{c}_x)} \leq P_X(x).$$

Summing both sides over all source symbols, we obtain

$$\sum_{x \in \mathcal{X}} 2^{-\ell(\boldsymbol{c}_x)} \leq 1,$$

which is exactly the Kraft inequality. On the other hand, (1) implies

$$\ell(\boldsymbol{c}_x) < -\log_2 P_X(x) + 1,$$

which in turn implies

$$\sum_{x \in \mathcal{X}} P_X(x)\ell(\boldsymbol{c}_x) < \sum_{x \in \mathcal{X}} [-P_X(x)\log_2 P_X(x)] + \sum_{x \in \mathcal{X}} P_X(x) = H(X) + 1.$$

3. Answer the following questions.

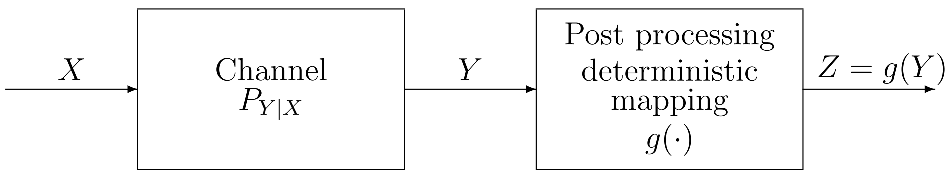
   (a) Let X_1, X_2, X_3, $\cdots$ be an i.i.d. discrete source with marginal alphabet $\{x_1,\ x_2,\ x_3,\ \cdots\}$, and assume that $P_X(x_i) > 0$ for every i.

      i. (8pt) Prove that the average codeword length of the single-letter binary Huffman code is equal to $H(X)$ if, and only if, $P_X(x_i)$ is equal to 2^{-n_i} for every i, where $\{n_i\}$ is a sequence of positive integers. (Hint: The if-part can be proved by the new bound in Problem 2, and the only-if-part can be proved by modifying the proof of Theorem 3.18 or slide I:3-39.)

      ii. (6pt) What is the sufficient and necessary condition under which the average codeword length of the single-letter ternary Huffman code equals $H(X)$? (Hint: You only need to write down the condition. No proof is necessary.)

      iii. (4pt) Prove that the average codeword length of the two-letter Huffman code cannot be equal to $H(X) + 1/2$ bits? (Hint: Use the new bound in Problem 2.)

   (b) (4 pt) Can the channel capacity between channel input X and channel output Z be strictly larger than the channel capacity between channel input X and channel output Y? Which lemma or theorem is your answer based on?



Answer:

(a) i. If $P_X(x_i)$ is 2^{-n_i} for every i, then the source entropy is equal to

$$H(X) = \sum_{x\in\mathcal{X}} P_X(x)\log_2 \frac{1}{P_X(x)} = \sum_{j=1}^{\infty} n_j 2^{-n_j}.$$

Problem 2 gives that we can take $\ell(\boldsymbol{c}_{x_j}) = \lceil -\log_2 P_X(x_j)\rceil = n_j$ to form an optimal variable-length code (for which there exists a Huffman code that performs as good.) Hence, $\bar{\ell} = \sum_{j=1}^{\infty} P_X(x_j)n_j = \sum_{j=1}^{\infty} n_j 2^{-n_j} = H(X)$.

On the contrary, if $\bar{\ell} = H(X)$, then the proof of Theorem 3.18 indicates that $\bar{\ell} = H(X)$ only when

$$P_X(x_j) = 2^{-\ell(\mathbf{c}_{x_j})}$$

and

$$\sum_{j=1}^{\infty} 2^{-\ell(\mathbf{c}_{x_j})} = 1, \tag{2}$$

for which (2) is always true since the probability mass sums to 1, namely,

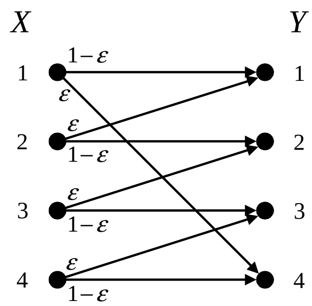
$$\sum_{j=1}^{\infty} 2^{-\ell(\mathbf{c}_{x_j})} = \sum_{j=1}^{\infty} P_X(x_j) = 1.$$

ii. $P_X(x_i)$ is equal to 3^{-n_i} for every i, where $\{n_i\}$ is a sequence of positive integers.

iii. From Problem 2, $\bar{\ell} < H(X^2) + 1 = 2H(X) + 1$ bits. Hence, the average codeword length $\frac{1}{2}\bar{\ell}$ must be strictly less than $H(X) + 1/2$.

(b) The answer is no by the data processing lemma.

4. Let the single-letter channel transition probability $P_{Y|X}$ of the discrete memoryless channel be defined as the following figure, where $0 < \epsilon < 0.5$.



(a) (4 pt) Is the channel a *weakly symmetric* channel? Is the channel a *symmetric* channel?

(b) (8 pt) Determine the channel capacity of this channel (in unit of bits). Also, indicate the input distribution that achieves the channel capacity. (Hint: You can directly apply the conclusions that we draw about the *weakly symmetric* and *symmetric* channels to obtain the answers.)

Answer:

(a) The transition matrix is given by

$$[P_{Y|X}] = \begin{bmatrix} 1-\epsilon & 0 & 0 & \epsilon \\ \epsilon & 1-\epsilon & 0 & 0 \\ 0 & \epsilon & 1-\epsilon & 0 \\ 0 & 0 & \epsilon & 1-\epsilon \end{bmatrix}.$$

Hence, the channel is not only *weakly symmetric* but also *symmetric.*

(b) For a symmetric channel, the capacity is achieved by uniform input. Also, the channel output is uniform due to uniform channel input. Therefore,

$$\begin{aligned} \max_{P_X} I(X;Y) &= \max_{P_X}[H(Y) - H(Y|X)] \\ &= \max_{P_X}[H(Y) - H_b(\epsilon)] \\ &= \max_{P_X}[H(Y)] - H_b(\epsilon) \\ &= \log_2(4) - H_b(\epsilon) \\ &= 2 - H_b(\epsilon) \text{ bits,} \end{aligned}$$

where $H_b(\epsilon) = \epsilon \log_2 \frac{1}{\epsilon} + (1-\epsilon)\log_2 \frac{1}{1-\epsilon}$ is the binary entropy function.

5. (a) (8 pt) Show that the exponential distribution has the largest differential entropy among all probability density functions (pdfs) with mean μ and continuous support $[0, \infty]$. (Hint: The pdf of the exponential distribution with mean μ is given by $p_X(x) = \frac{1}{\mu}\exp(-\frac{x}{\mu})$ for $x \geq 0$.)

(b) (8 pt) Of all pdfs with continuous support $[0, K]$, where K is finite and $K > 1$, which pdf has the largest differential entropy? (Hint: If p_X is the pdf that maximizes differential entropy among all pdfs with continuous support $[0, K]$, then $E[\log p_X(X)] = E[\log p_X(Y)]$ for any random variable Y with continuous support $[0, K]$.)

Answer:

(a) Let $p_Y(\cdot)$ be a pdf with mean μ and continuous support $[0,\infty)$. Then,

$$\begin{aligned}
h(p_X)-h(p_Y) &= \int_0^\infty p_X(x)\log\frac{1}{p_X(x)}dx-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_X(x)\left[\log(\mu)+\frac{1}{\mu}x\right]dx-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \left[\log(\mu)+\frac{1}{\mu}E[X]\right]-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \left[\log(\mu)+\frac{1}{\mu}E[Y]\right]-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_Y(y)\left[\log(\mu)+\frac{1}{\mu}y\right]dy-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_Y(y)\log\frac{1}{p_X(y)}dy-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_Y(y)\log\frac{p_Y(y)}{p_X(y)}dy\\
&= D(p_Y\|p_X)\ge 0,
\end{aligned}$$

with equality holds if, and only if, $p_Y\equiv p_X$.

(b) One choice of p_X that makes $E[\log p_X(X)]=E[\log p_X(Y)]$ is the uniform distribution over $[0,K]$, for which $p_X(x)=\frac{1}{K}$ for $x\in[0,K]$. Then,

$$\begin{aligned}
h(p_X)-h(p_Y) &= \int_0^\infty p_X(x)\log\frac{1}{p_X(x)}dx-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_X(x)\log(K)dx-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \log(K)-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_Y(y)\log(K)dy-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_Y(y)\log\frac{1}{p_X(y)}dy-\int_0^\infty p_Y(y)\log\frac{1}{p_Y(y)}dy\\
&= \int_0^\infty p_Y(y)\log\frac{p_Y(y)}{p_X(y)}dy\\
&= D(p_Y\|p_X)\ge 0,
\end{aligned}$$

with equality holds if, and only if, $p_Y\equiv p_X$.

6. Consider the 3-input 3-output memoryless additive Gaussian channel

$$\boldsymbol{Y}=\boldsymbol{X}+\boldsymbol{Z},$$

where $\boldsymbol{X} = [X_1, X_2, X_3]$, $\boldsymbol{Y} = [Y_1, Y_2, Y_3]$ and $\boldsymbol{Z} = [Z_1, Z_2, Z_3]$ are all 3-Dimensional real vectors. Assume that $\boldsymbol{X}$ is independent of $\boldsymbol{Z}$, and the input power constraint is S (i.e., $E(X_1^2 + X_2^2 + X_3^2) \leq S$). Also, assume that $\boldsymbol{Z}$ is Gaussian distributed with zero mean and covariance matrix $\mathbb{K}$, where

$$\mathbb{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}.$$

(a) (6 pt) Determine the capacity-cost function of the channel, if $\rho = 0$. (Hint: Directly apply Theorem 6.31 or slide I:6-55.)

(b) (8 pt) Determine the capacity-cost function of the channel, if $0 < \rho < 1$. (Hint: Directly apply Theorem 6.33 or slide I:6-60.)

Answer:

(a)

$$C(S) = \max_{\{P_{X^3} \,:\, E[X_1^2]+E[X_2^2]+E[X_3^2] \leq S\}} I(X^3; Y^3) = \sum_{i=1}^{3} \frac{1}{2} \log\left(1 + \frac{S}{3}\right).$$

(b) The eigenvalues of $\mathbb{K}$ are $\lambda_1 = 1$, $\lambda_2 = 1 - \rho$ and $\lambda_3 = 1 + \rho$. Hence,

$$C(S) = \sum_{i=1}^{3} \frac{1}{2} \log\left(1 + \frac{S_i}{\lambda_i}\right),$$

where

$$\begin{cases} S_1 = 0, & S_2 = S, & S_3 = 0, & \text{if } 0 \leq S < \rho; \\ S_1 = \dfrac{S-\rho}{2}, & S_2 = \dfrac{S+\rho}{2}, & S_3 = 0, & \text{if } \rho \leq S < 3\rho; \\ S_1 = \dfrac{S}{3}, & S_2 = \dfrac{S}{3} + \rho, & S_3 = \dfrac{S}{3} - \rho, & \text{if } S \geq 3\rho. \end{cases}$$