4F5 Advanced Communications and Coding
Michealmas 2015

Examples Paper I

Question 1

Inequalities: Which of the following inequalities are generally $\geq, =, \leq$? Label each with $\geq$, $=$, or $\leq$.

(a) $H(5X)$ vs. $H(X)$

(b) $H(X_1|X_0)$ vs $H(X_1|X_0, X_2)$

(c) $H(X, Y)$ vs. $H(X) + H(Y)$

(d) $I(g(X); Y)$ vs. $I(X; Y)$

Question 2

Entropy of Functions: Let X be a random variable with pmf P_X and let $Z = g(X)$, for some function g. Show that

$$H(X) \geq H(Z).$$

When does equality hold? In particular, does it hold when $Z = 2^X$? When $Z = \cos X$?

(Hint: Expand $H(X, Z)$ in two different ways. Alternatively, you could express the pmf of Z in terms of P_X, and use it in the formula for $H(X)$.)

Question 3

Discrete Entropies: Let X and Y be two independent, integer valued random variables. Let X be uniformly distributed over $\{1, 2, \ldots, 8\}$, and let $\Pr(Y = k) = 2^{-k},\ k = 1, 2, \ldots$.

(a) Find $H(X)$.

(b) Find $H(Y)$. The following expressions may be useful. For $0 < r < 1$:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \qquad \sum_{n=0}^{\infty} n r^n = \frac{r}{(1-r)^2}$$

(c) Find $H(X + Y, X - Y)$. (There are two ways to do this—one is tedious, the other elegant. Q.2 may be useful.)

Question 4

The Value of a Question: Let random variable X with pmf P take values in $\{1, \ldots, m\}$. We are given a set $S \subseteq \{1, \ldots, m\}$. We ask whether $X \in S$ and receive the answer

$$Y = \begin{cases} 1 & \text{if } X \in S \\ 0 & \text{if } X \notin S. \end{cases}$$

Suppose that $\Pr(X \in S) = \sum_{x: x \in S} P(x) = \alpha$.

(a) Find the decrease in uncertainty $H(X) - H(X|Y)$.

(b) Apparently, any set with a given α is as good as any other. What value of α yields the maximum decrease in uncertainty? What does this tell you about the most informative yes-no question(s)?

Question 5

Coin Flips: A fair coin is flipped until the first head occurs. Let X denote the number of flips required.

(a) Find the entropy $H(X)$ in bits. (Related to Q.3(b))

(b) A random variable X is generated according to this distribution. Find an "efficient" sequence of yes-no questions of the form "Is X contained in the set S?" Compare $H(X)$ to the expected number of questions required to determine X. (Hint: Use Q.4(b))

Question 6

AEP and Compression: A discrete source emits a sequence of i.i.d binary digits with the distribution $P(1) = 0.005, P(0) = 0.995$. The digits are taken one hundred at a time, and a binary codeword is provided for each sequence of 100 digits containing three or fewer 1's.

(a) Assuming that all codewords are the same length, find the minimum length (in bits) required to provide codewords for all sequences with three or fewer ones.

(b) Calculate the probability of observing a source sequence for which no codeword has been assigned.

Question 7

Asymptotic behaviour of products: Let

$$X = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ 2 & \text{with probability } \frac{1}{4} \\ 3 & \text{with probability } \frac{1}{4} \end{cases}$$

Let $X_1, X_2, \ldots$ be drawn i.i.d. according to this distribution. Find the limiting behaviour of the product

$$(X_1 X_2 \ldots X_n)^{\frac{1}{n}}.$$

Question 8

AEP and Relative Entropy: Let $X_1, X_2, \ldots$ be independent and identically distributed random variables drawn according to the probability mass function $P(x), x \in \{1, \ldots, m\}$. Thus $Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(x_i)$. We know that

$$-\frac{1}{n} \log P(X_1, \ldots, X_n) \to H(X)$$

in probability. Let Q be another pmf on $\{1, \ldots, m\}$ and define $Q(X_1, \ldots, X_n) = \prod_{i=1}^{n} Q(X_i)$.

(a) Evaluate $\lim_{n\to\infty} -\frac{1}{n} \log Q(X_1, \ldots, X_n)$ where $X_1, X_2, \ldots$ are i.i.d according to P.

(b) Now evaluate the limit of the log-likelihood ratio $\frac{1}{n} \log \frac{Q(X_1,\ldots,X_n)}{P(X_1,\ldots,X_n)}$ when $X_1, X_2 \ldots$ are i.i.d. according to P.

If you are trying to resolve the hypothesis of whether the observed data was generated from P or from Q, this tells you that the odds of favouring Q are exponentially small when P is true.

Question 9

Conditional Mutual Information and the Chain Rule: Consider random variables X, Y, Z jointly distributed with pmf P_{XYZ}. The conditional mutual information $I(X;Y|Z)$ is defined as

$$I(X;Y|Z) := H(X|Z) - H(X|Y,Z).$$

(a) Show that $I(X;Y|Z)$ is equal to $D(P_{XY|Z}||P_{X|Z}P_{Y|Z})$, where

$$D(P_{XY|Z}||P_{X|Z}P_{Y|Z}) = \sum_{x,y,z} P_{XYZ}(x,y,z) \log \frac{P_{XY|Z}(x,y|z)}{P_{X|Z}(x|z)P_{Y|Z}(y|z)}.$$

Hence the conditional mutual information is always non-negative.

(b) *The Chain Rule for Information*: For any sequence of random variables $X_1, X_2, \ldots, X_n$ jointly distributed with Y, show that the mutual information can be expressed as

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_{i-1}, X_{i-2}, \ldots, X_1).$$

(Hint: Use the chain rule for entropy.)

Question 10

The Data Processing Inequality: Random variables X, Y, Z are said to form a *Markov chain* (denoted by $X - Y - Z$) if their joint pmf P_{XYZ} can be written as

$$P_{XYZ}(x,y,z) = P_X(x)P_{Y|X}(y|x)P_{Z|Y}(z|y) \quad \text{for all } x,y,z.$$

(a) Verify that if $X - Y - Z$, then X and Z are conditionally independent given Y, i.e.,

$$P_{XZ|Y}(x,z|y) = P_{X|Y}(x|y)P_{Z|Y}(z|y) \quad \text{for all } x,y,z.$$

Conclude that $X - Y - Z$ implies that $Z - Y - X$.

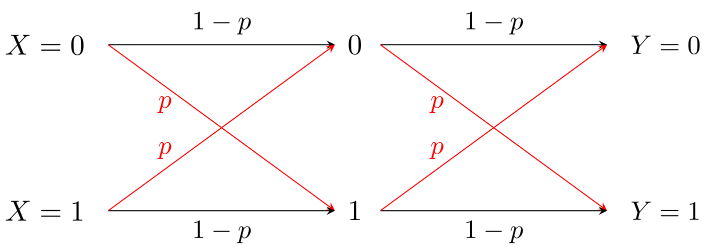
(b) Prove the following data processing inequality: If $X - Y - Z$, then

$$I(X;Y) \geq I(X;Z).$$

(Hint: Consider $I(X;Y,Z)$ and expand in two different ways using the chain rule for information.)

The data processing inequality is useful in many estimation problems. E.g., let Y be a noisy version of X, and $Z = g(Y)$ be an estimate of X based on observing only Y. The inequality says that functions of the data Y cannot increase the information about X.

Question 11



Cascade Channel: Consider the cascade of two independent binary symmetric channels, each with crossover probability $p < \frac{1}{2}$.

(a) What is the capacity of the cascade channel shown above?

(b) Show that the capacity of a cascade of m independent BSCs, each with crossover probability p, is given by

$$1 - H_2\left(\tfrac{1}{2}(1-(1-2p)^m\right)$$

where $H_2(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function. Observe that the capacity monotonically decreases as m increases and tends to 0. Is this consistent with what you'd expect?

(Hint: Use induction on m to compute the overall crossover probability.)

Question 12

Z-channel: The Z-channel has binary input and output alphabets and transition probabilities $P_{Y|X}$ given by the following matrix:

	$P_{Y\|X}$	Y = 0	1
X	0	1	0
	1	$\frac{1}{2}$	$\frac{1}{2}$

Find the capacity of the Z-channel and the maximising input distribution. (Sketching the input-output relationship for this channel should explain its name.)

Question 13

Modulo-addition channel: Consider the DMC defined by $Y = X + N (\text{mod } 11)$, where the input $X \in \{0, 1, 2, \ldots, 10\}$. The noise N is uniformly distributed in the set $\{1, 2, 3\}$, i.e., $Pr(N = 1) = Pr(N = 2) = Pr(N = 3) = \frac{1}{3}$. Assume that N is independent of X.

Find the capacity of this channel and the maximising input distribution.

Question 14

Joint AEP: Consider the following joint pmf P_{XY}:

	P_{XY}	Y = 0	1
X	0	p_{00}	p_{01}
	1	p_{10}	p_{11}

Suppose that pairs of random variables $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are generated i.i.d. according to P_{XY}. Let the number of occurrences of $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$ pairs in the observed sequence (x^n, y^n) be denoted by $n_{00}, n_{01}, n_{10}, n_{11}$, respectively.

(a) Show the following:

$$P_{XY}(x^n, y^n) = Pr(X^n = x^n, Y^n = y^n) = p_{00}^{n_{00}} \cdot p_{01}^{n_{01}} \cdot p_{10}^{n_{10}} \cdot p_{11}^{n_{11}}, \tag{1}$$

$$\hat{H}_{XY} := -\frac{1}{n}\log P_{XY}(x^n, y^n) = -\frac{1}{n}\sum_{i,j\in\{0,1\}} n_{ij}\log p_{ij}, \tag{2}$$

$$\hat{H}_X := -\frac{1}{n}\log P_X(x^n) = -\frac{1}{n}((n_{00}+n_{01})\log(p_{00}+p_{01}) + (n_{10}+n_{11})\log(p_{10}+p_{11})), \tag{3}$$

$$\hat{H}_Y := -\frac{1}{n}\log P_Y(y^n) = -\frac{1}{n}((n_{00}+n_{10})\log(p_{00}+p_{10}) + (n_{01}+n_{11})\log(p_{01}+p_{11})). \tag{4}$$

(b) For $p_{00} = 0.5, p_{11} = 0.3, p_{01} = p_{10} = 0.1$, calculate $H(X, Y), H(X)$, and $H(Y)$.

(c) Matlab Exercise: Generate n i.i.d pairs according to the pmf specified in part (b), for $n = 10, 100, 10000$. For each case, count the number of occurrences of each of the four binary pairs and use these to compute the *empirical* entropies $\hat{H}_{XY}, \hat{H}_X$, and $\hat{H}_Y$ according to the formulas above. Compare with the theoretical values obtained in part (b).

Question 15

Estimation and Fano's inequality: We are given the following joint distribution on (X, Y):

	P_{XY}	Y: a	b	c
	1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
X	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
	3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Let $\hat{X}(Y)$ be an estimator for X (based on Y) and let $P_e = \Pr(\hat{X}(Y) \neq X)$.

(a) Find the minimum probability of error estimator $\hat{X}(Y)$ and the associated P_e.

(b) Evaluate Fano's inequality for this problem and compare.

(c) **Extra**: Revisit the proof of Fano's inequality in Handout 5, and argue that the term $P_e \log|\mathcal{X}|$ can be replaced by $P_e \log(|\mathcal{X}| - 1)$. This gives a stronger version of Fano's inequality:

$$P_e \geq \frac{H(X|Y) - 1}{\log(|\mathcal{X}| - 1)}.$$

Comapre this stronger bound with the P_e of the estimator in Q. 15 (a).

Answers to Selected Questions

Q3. $H(X) = 3$, $H(Y) = 2$, $H(X + Y, X - Y) = 5$.

Q4. (a) $H_2(\alpha)$; (b) $\alpha = 0.5$

Q5. (a) $H(X) = 2$

Q6. (a) 18 bits; (b) 0.00167. Note that the entropy of the source is $H_2(0.005) = 0.0454$/sample, so the average code length of the proposed code is quite a bit larger than the entropy of a 100-digit sequence which is 4.54 bits.

Q7. $6^{1/4}$

Q8. (a) $D(P||Q) + H(P)$; (b) $-D(P||Q)$

Q12. The capacity of the Z-channel is 0.322 bits. The maximising distribution is $P_X(0) = \frac{3}{5}, P_X(1) = \frac{2}{5}$.

Q13. The capacity is $\log \frac{11}{3}$ bits.

Q14. (b) $H(X, Y) = 1.6855$ bits, $H(X) = H(Y) = 0.971$ bits.

Q15. (a) $P_e = \frac{1}{2}$; (b) $H(X|Y) = 1.5$ bits. Therefore, Fano's inequality yields

$$P_e \geq \frac{H(X|Y) - 1}{\log|\mathcal{X}|} = 0.316.$$

Extra: The stronger version of Fano's inequality yields $P_e \geq \frac{1}{2}$ for this problem. Therefore, the estimator in Q.15 (a) is as good as it gets (in terms of probability of error).