Information Theory: Final Exam, January 16, 2014

Totally, there are 50 points in this exam. Each point will earn you 2 credits. So there will be totally 100 credits.

1. (15 points: Each subproblem costs 3 points; one for true/false answer, and two for the proof/counterexample) Decide whether each of the following statements is true or false. Prove the validity of those that are true and give counterexamples or arguments based on known facts to disprove those that are false.

   (a) There exists a discrete memoryless channel with a binary input alphabet and a ternary output alphabet such that its capacity is equal to C = 1.5 bits/channel use.

   (b) Let X be a Gaussian distributed random variable with zero mean and variance $\frac{1}{2\pi e}$, and let Y be a random variable that is uniformly distributed over the interval $[0, 1]$. Then
   $$h(2X) > h(2Y).$$

   (c) If X and Y are real-valued independent random variables, then $h(X+Y) \geq h(X+c)$ for any real-valued constant c.
   Hint: $h(X + c) = h(X)$.

   (d) Consider a discrete-time (additive) memoryless Gaussian channel with input power constraint P and noise power (variance) N. Let $C(P)$ denote the channel capacity in bits/channel use. Then $C(P) \geq \frac{1}{2\ln(2)} \frac{P}{P+N}$.
   Hint: $C(P) = \frac{1}{2} \log_2 \left(1 + \frac{P}{N}\right)$.

   (e) Among all continuous random variables admitting a probability density function on the interval $[0, \infty)$ with finite mean $\mu > 0$, the exponential random variable with probability density function
   $$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$
   and parameter $\lambda = \frac{1}{\mu}$ maximizes differential entropy.
   Hint: Denote by X and Y the random variables respectively admitting $f(x)$ and $g(x)$ as their probability density functions on the interval $[0, \infty)$ with finite mean $\mu > 0$, and compute $h(X) - h(Y)$.

Solutions.

   (a) The statement is false because $I(X;Y) \leq \min\{H(X), H(Y)\} \leq \log_2(2) = 1$.

   (b) The statement is false. Note that $\text{Var}[2X] = 4 \cdot \text{Var}[X] = \frac{2}{\pi e}$; hence, $h(2X) = \frac{1}{2}\log_2(2\pi e \text{Var}[2X]) = 1$, and $h(2Y) = \log_2(2 - 0) = 1$. So, $h(2X) = h(2Y)$.

   (c) The statement is true because
   $$\begin{aligned} h(X+Y) &\geq h(X+Y|Y) \\ &= h(X) \\ &= h(X+c) \end{aligned}$$

(d) The statement is true. The capacity of a discrete-time (additive) memoryless Gaussian channel with input power constraint P and noise power (variance) N is $\frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$. Hence, the question is the same as asking whether

$$\frac{1}{2}\log_2\left(1+\frac{P}{N}\right) \geq \frac{1}{2\ln(2)}\frac{P}{P+N},$$

or equivalently,

$$\ln\left(\frac{P+N}{N}\right) \geq 1-\frac{N}{P+N},$$

which is true because $\ln(x) \geq 1-\frac{1}{x}$ for $x>0$.

(e) The statement is true. Denote by X and Y the random variables respectively admitting $f(x)$ and $g(x)$ as their probability density functions on the interval $[0,\infty)$ with finite mean $\mu>0$.

$$\begin{aligned}
h(X)-h(Y) &= \int_0^\infty f(x)\log_2\frac{1}{f(x)}dx - \int_0^\infty g(x)\log_2\frac{1}{g(x)}dx \\
&= \int_0^\infty f(x)\log_2\frac{1}{\lambda e^{-\lambda x}}dx - \int_0^\infty g(x)\log_2\frac{1}{g(x)}dx \\
&= \int_0^\infty g(x)\log_2\frac{1}{\lambda e^{-\lambda x}}dx - \int_0^\infty g(x)\log_2\frac{1}{g(x)}dx \qquad (1) \\
&= \int_0^\infty g(x)\log_2\frac{g(x)}{\lambda e^{-\lambda x}}dx \\
&= \int_0^\infty g(x)\log_2\frac{g(x)}{f(x)}dx \\
&= D(g(x)\|f(x)) \geq 0,
\end{aligned}$$

where (**??**) holds since

$$\int_0^\infty f(x)\log_2(\lambda e^{-\lambda x})dx = \int_0^\infty g(x)\log_2(\lambda e^{-\lambda x})dx.$$

2. (8 points: Each subproblem costs 4 points) Answer the following questions.

(a) Let X and Y be two Gaussian random variables with $E[X]=E[Y]=0$, $E[X^2]=\sigma_X^2$ and $E[Y^2]=\sigma_Y^2$. Find $D(X\|Y)$ in terms of σ_X^2 and σ_Y^2.

(b) Let X and Z be two real-valued random variables both with zero-mean and the same variance: $E[X^2]=E[Z^2]=\sigma_X^2$. Assuming that X is Gaussian, show that

$$D(Z\|Y) \geq D(X\|Y)$$

where Y is a Gaussian random variable with zero-mean and variance σ_Y^2.
Hint: $h(X) \geq h(Z)$.

Solutions.

(a)

$$\begin{aligned} D(X\|Y) &= \int_{-\infty}^{\infty} f_X(x)\log_2\frac{f_X(x)}{f_Y(x)}dx \\ &= -h(X)+\int_{-\infty}^{\infty} f_X(x)\log_2\frac{1}{f_Y(x)}dx \\ &= -\frac{1}{2}\log_2(2\pi e\sigma_X^2)+\int_{-\infty}^{\infty} f_X(x)\left[\frac{1}{2}\log_2(2\pi\sigma_Y^2)+\frac{x^2}{2\sigma_Y^2}\log_2(e)\right]dx \\ &= -\frac{1}{2}\log_2(2\pi e\sigma_X^2)+\left[\frac{1}{2}\log_2(2\pi\sigma_Y^2)+\frac{\sigma_X^2}{2\sigma_Y^2}\log_2(e)\right] \\ &= \frac{1}{2}\log_2\frac{\sigma_Y^2}{e\sigma_X^2}+\frac{\sigma_X^2}{2\sigma_Y^2}\log_2(e) \end{aligned}$$

(b)

$$\begin{aligned} &D(Z\|Y)-D(X\|Y) \\ &= \int_{-\infty}^{\infty} f_Z(t)\log_2\frac{f_Z(t)}{f_Y(t)}dt-\int_{-\infty}^{\infty} f_X(t)\log_2\frac{f_X(t)}{f_Y(t)}dt \\ &= -h(Z)+\int_{-\infty}^{\infty} f_Z(t)\log_2\frac{1}{f_Y(t)}dt+h(X)-\int_{-\infty}^{\infty} f_X(t)\log_2\frac{1}{f_Y(t)}dt \\ &= -h(Z)+\left[\frac{1}{2}\log_2(2\pi\sigma_Y^2)+\frac{\sigma_X^2}{2\sigma_Y^2}\log_2(e)\right]+h(X)-\left[\frac{1}{2}\log_2(2\pi\sigma_Y^2)+\frac{\sigma_X^2}{2\sigma_Y^2}\log_2(e)\right] \\ &= h(X)-h(Z)\geq 0. \end{aligned}$$

3. (11 points: Each subproblems cost 4 points except the first one that only costs 3 points)

   (a) Consider a discrete-time memoryless source $\{U_n\}_{n=1}^{\infty}$ with alphabet $\mathcal{U}=\{a,b,c,d\}$ and distribution $\Pr[U=a]=\Pr[U=b]=\frac{1}{4}$ and $\Pr[U=c]=\frac{1}{8}$. Compute the source entropy rate $H(\mathcal{U})$ in bits/source symbol.

   (b) Consider a binary-input ternary-output discrete memoryless channel with the following transition matrix

$$Q=\begin{bmatrix} p_1 & p_2 & 1-p_1-p_2 \\ p_1 & 1-p_1-p_2 & p_2 \end{bmatrix},$$

   where $0\leq p_1\leq 1$ and $0\leq p_2\leq 1$. Find the capacity (in bits/channel symbol) of this channel in terms of p_1 and p_2.

   Hint: Both $\frac{1}{p_1}Q_1=\begin{bmatrix}1\\1\end{bmatrix}$ and $\frac{1}{1-p_1}Q_2=\begin{bmatrix}\frac{p_2}{1-p_1} & \frac{1-p_1-p_2}{1-p_1}\\ \frac{1-p_1-p_2}{1-p_1} & \frac{p_2}{1-p_1}\end{bmatrix}$ are transition matrices of weakly symmetric channels.

   (c) Assume that we wish to encode the above source and transmit the result over the above channel with $p_2=0$. For what values of p_1 can we ensure the existence of a source-channel code with transmission rate of 0.25 source symbols/channel symbol that provides arbitrarily good asymptotic performance (i.e., arbitrarily low probability of decoding error) ?

Hint: Here, 0.25 source symbols/channel symbol implies that one source symbol can be transmitted by four channel symbols, or $T_s = 4T_c$, where T_s and T_c represent the durations per source symbol and per channel symbol, respectively.

Hint: Information Transmission Theorem or Joint Source-Channel Coding Theorem.

Solutions.

(a) It is a memoryless source; hence, the entropy rate is equal to the marginal entropy.

$$\begin{aligned}
H(\mathcal{U}) &= \lim_{n\to\infty} \frac{1}{n} H(U^n) = H(U) \\
&= \frac{1}{4}\log_2 \frac{1}{\frac{1}{4}} + \frac{1}{4}\log_2 \frac{1}{\frac{1}{4}} + \frac{1}{8}\log_2 \frac{1}{\frac{1}{8}} + \frac{3}{8}\log_2 \frac{1}{\frac{3}{8}} \\
&= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{9}{8} - \frac{3}{8}\log_2(3) \\
&= \frac{5}{2} - \frac{3}{8}\log_2(3).
\end{aligned}$$

(b) Both

$$\frac{1}{p_1}Q_1 = \frac{1}{p_1}\begin{bmatrix} p_1 \\ p_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\frac{1}{1-p_1}Q_2 = \frac{1}{1-p_1}\begin{bmatrix} p_2 & 1-p_1-p_2 \\ 1-p_1-p_2 & p_2 \end{bmatrix} = \begin{bmatrix} \frac{p_2}{1-p_1} & \frac{1-p_1-p_2}{1-p_1} \\ \frac{1-p_1-p_2}{1-p_1} & \frac{p_2}{1-p_1} \end{bmatrix}$$

are transition matrices of weakly symmetric channels. Their corresponding channel capacities are respectively $C_1 = 0$ and $C_2 = 1 - h_b\left(\frac{1-p_1-p_2}{1-p_1}\right)$. Hence,

$$C = p_1C_1 + (1-p_1)C_2 = (1-p_1)\left(1 - h_b\left(\frac{1-p_1-p_2}{1-p_1}\right)\right).$$

(c) We first note that $p_2 = 0$ implies $(1-p_1)\left(1 - h_b\left(\frac{1-p_1-p_2}{1-p_1}\right)\right) = (1-p_1)$. Hence, if

$$\frac{\frac{5}{2} - \frac{3}{8}\log_2(3)}{T_s} \frac{\text{bits/source symbol}}{\text{seconds/source symbol}} < \frac{1-p_1}{T_c} \frac{\text{bits/channel symbol}}{\text{seconds/channel symbol}},$$

we can ensure the existence of a source-channel code with transmission rate of 0.25 source symbols/channel symbols that provides arbitrarily good asymptotic performance. The answer to this question is accordingly $0 \le p_1 < \frac{3}{8} + \frac{3}{32}\log_2(3)$ (≈ 0.5236).

4. (16 points: Each subproblems cost 4 points) Consider a discrete memoryless channel for which the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ are both the set of binary n-tuples; i.e., $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$ where $n \geq 2$ is a given integer. Except for the two cases $(x^n, y^n) = (0^n, 0^n)$ and $(x^n, y^n) = (1^n, 1^n)$, where 0^n and 1^n are the all-zero and all-one n-tuples, respectively, whenever an input x^n is transmitted over the channel, one of two binary n-tuples is received – either x^n or the binary n-tuple obtained by performing a single right-cyclic shift of x^n. More precisely, for any $x^n = (x_1, x_2, \cdots, x_n) \in \{0,1\}^n$ and any $y^n \in \{0,1\}^n$, the channel's conditional distribution $Q(y^n|x^n) := \Pr[Y^n = y^n | X^n = x^n]$ is given by

$$Q(y^n|x^n) = \begin{cases} 1-\epsilon, & \text{if } y^n = x^n \\ \epsilon, & \text{if } y_1 = x_n \text{ and } y_i = x_{i-1} \text{ for } i = 2, 3, \cdots, n \\ 0, & \text{otherwise} \end{cases}$$

if $(x^n, y^n) \neq (0^n, 0^n)$ and $(x^n, y^n) \neq (1^n, 1^n)$, where $0 \leq \epsilon \leq 1$. Furthermore, $Q(0^n|0^n) = Q(1^n|1^n) = 1$.

(a) Determine the channel transition probability matrix $[Q(y^n|x^n)]$ for the cases $n = 2$ and $n = 3$.

Hint:

$$[Q(y^2|x^2)] = \begin{bmatrix} Q(00|00) & Q(11|00) & Q(01|00) & Q(10|00) \\ Q(00|11) & Q(11|11) & Q(01|11) & Q(10|11) \\ Q(00|01) & Q(11|01) & Q(01|01) & Q(10|01) \\ Q(00|10) & Q(11|10) & Q(01|10) & Q(10|10) \end{bmatrix}$$

$$[Q(y^3|x^3)] = \begin{bmatrix} Q(000|000) & Q(111|000) & Q(001|000) & Q(100|000) & Q(010|000) & Q(011|000) & q(101|000) & Q(110|000) \\ Q(000|111) & Q(111|111) & Q(001|111) & Q(100|111) & Q(010|111) & Q(011|111) & q(101|111) & Q(110|111) \\ Q(000|001) & Q(111|001) & Q(001|001) & Q(100|001) & Q(010|001) & Q(011|001) & q(101|001) & Q(110|001) \\ Q(000|100) & Q(111|100) & Q(001|100) & Q(100|100) & Q(010|100) & Q(011|100) & q(101|100) & Q(110|100) \\ Q(000|010) & Q(111|010) & Q(001|010) & Q(100|010) & Q(010|010) & Q(011|010) & q(101|010) & Q(110|010) \\ Q(000|011) & Q(111|011) & Q(001|011) & Q(100|011) & Q(010|011) & Q(011|011) & q(101|011) & Q(110|011) \\ Q(000|101) & Q(111|101) & Q(001|101) & Q(100|101) & Q(010|101) & Q(011|101) & q(101|101) & Q(110|101) \\ Q(000|110) & Q(111|110) & Q(001|110) & Q(100|110) & Q(010|110) & Q(011|110) & q(101|110) & Q(110|110) \end{bmatrix}$$

(b) In which category for $n = 3$ does this channel belong to, (i) symmetric, (ii) weakly symmetric, (iii) quasi-symmetric, or (iv) none of the above? Justify your answer by a few words.

(c) We can actually sub-divide $[Q(y^2|x^2)]$ into two parallel channels respectively with inputs S, T and outputs U, V, where $S, U \in \{00, 11\}$ and $T, V \in \{01, 10\}$. By denoting

$$a_1 = P_{X^2}(00) + P_{X^2}(11), \quad a_2 = P_{X^2}(01) + P_{X^2}(10),$$

and

$$P_S(00) = \frac{P_{X^2}(00)}{a_1}, P_S(11) = \frac{P_{X^2}(11)}{a_1}, \quad P_T(01) = \frac{P_{X^2}(01)}{a_2}, P_T(10) = \frac{P_{X^2}(10)}{a_2},$$

we obtain

$$\begin{aligned} I(X^2;Y^2) &= \sum_{x^2\in\{0,1\}^2}\sum_{y^2\in\{0,1\}^2} P_{X^2}(x^2)Q(y^2|x^2)\log_2\frac{Q(y^2|x^2)}{\sum_{\tilde{x}^2\in\{0,1\}^2}P_{X^2}(\tilde{x}^2)Q(y^2|\tilde{x}^2)} \\ &= \sum_{s\in\{00,11\}}\sum_{u\in\{00,11\}} a_1P_S(s)Q(u|s)\log_2\frac{Q(u|s)}{\sum_{\tilde{s}\in\{00,11\}}a_1P_S(\tilde{s})Q(u|\tilde{s})} \\ &+ \sum_{t\in\{01,10\}}\sum_{v\in\{01,10\}} a_2P_T(t)Q(v|t)\log_2\frac{Q(v|t)}{\sum_{\tilde{t}\in\{01,10\}}a_2P_T(\tilde{t})Q(v|\tilde{t})} \\ &= a_1I(S;U)+a_2I(T;V)-a_1\log_2(a_1)-a_2\log_2(a_2) \end{aligned}$$

Based the above derivation, determine the capacity of this channel for $n=2$ and $\epsilon=\frac{1}{2}$.
Hint: Here, $C=\max_{P_{X^2}} I(X^2;Y^2)$.

(d) Likewise, we can sub-divide $[Q(y^3|x^3)]$ into three parallel channels respectively with inputs S, T, W and outputs U, V, Z, where $S,U\in\{000,111\}$, $T,V\in\{001,100,010\}$, $W,Z\in\{011,101,110\}$. Show that the mutual information of this channel with $n=3$ can be represented as:

$$I(X^3;Y^3)=a_1I(S;U)+a_2I(T;V)+a_3I(W;Z)-a_1\log_2(a_1)-a_2\log_2(a_2)-a_3\log_2(a_3)$$

for some a_1,a_2,a_3 satisfying $a_1+a_2+a_3=1$.

Solutions.

(a)

$$[Q(y^2|x^2)]=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1-\epsilon&\epsilon\\0&0&\epsilon&1-\epsilon\end{bmatrix}$$

$$[Q(y^3|x^3)]=\begin{bmatrix}1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1-\epsilon&\epsilon&0&0&0&0\\0&0&0&1-\epsilon&\epsilon&0&0&0\\0&0&\epsilon&0&1-\epsilon&0&0&0\\0&0&0&0&0&1-\epsilon&\epsilon&0\\0&0&0&0&0&0&1-\epsilon&\epsilon\\0&0&0&0&0&\epsilon&0&1-\epsilon\end{bmatrix}$$

(b) *(iv)* none of the above. Explanation on why it is not symmetric, weakly symmetric, or quasi-symmetric is necessary.

(c)

$$
\begin{aligned}
C &= \max_{P_{X^2}} I(X^2;Y^2) \\
&= \max_{(a_1,a_2)\in[0,1]^2:a_1+a_2=1} \max_{P_S} \max_{P_T} [a_1 I(S;U) + a_2 I(T;V) - a_1 \log_2(a_1) - a_2 \log_2(a_2)] \\
&= \max_{(a_1,a_2)\in[0,1]^2:a_1+a_2=1} [a_1 C_1 + a_2 C_2 - a_1 \log_2(a_1) - a_2 \log_2(a_2)] \\
&= \max_{0\le p\le 1} [pC_1 + (1-p)C_2 - p\log_2(p) - (1-p)\log_2(1-p)] \\
&= p^* C_1 + (1-p^*)C_2 + h_b(p^*),
\end{aligned}
$$

where $C_1 = 1$, $C_2 = 1 - h_b(\epsilon)$, and $p^* = \frac{1}{1+\epsilon^\epsilon(1-\epsilon)^{1-\epsilon}}$. Hence, for $\epsilon = \frac{1}{2}$, we have $C_1 = 1$, $C_2 = 0$, $p^* = \frac{2}{3}$, and

$$
C = p^* C_1 + (1-p^*)C_2 + h_b(p^*) = \frac{2}{3} + h_b\left(\frac{2}{3}\right).
$$

(d) The procedure is similar what has been given in subproblem (c); hence, we omit it.