CSIE 7624: Homework 2
Week 3-5: The Asymptotic Equipartition Property and Entropy Rates, and Channel and Channel Capacity

Due on April 9, 2018

JL Wu 14:20

D06922002 Kuan-Lun Wang

Contents

Problem 1 1
- (a) 1
- (b) 1
- (c) 2

Problem 2 2
- (a) 2
- (b) 3
- (c) 3
- (d) 3

Problem 3 4

Problem 4 4
- (a) 4
- (b) 5
- (c) 5

Problem 5 5
- (a) 5
- (b) 6

Problem 6 6

Problem 7 7
- (a) 7
- (b) 8
- (c) 8

Problem 8 8
- (a) 9
- (b) 9
- (c) 10

Problem 9 10

Problem 10 11

(a) 12
(b) 12
(c) 12

Problem 1

Markov's inequality and Chebyshev's inequality

(a)

(Markov's inequality) For any non-negative random variable X and any $t > 0$, show that

$$\Pr\{X \geq t\} \leq \frac{\mathrm{E}[X]}{t}.$$

Exhibit a random variable that achieves this inequality with equality.

Denote the distribution of X by F, and give $t \in \mathbb{R}^+ \cup \{0\}$. Then we have:

$$\begin{aligned}
\mathrm{E}[X] &= \int_0^\infty x \,\mathrm{d}\, F(x) \\
&= \int_0^t x \,\mathrm{d}\, F(x) + \int_t^\infty x \,\mathrm{d}\, F(x) \\
&\geq \int_t^\infty x \,\mathrm{d}\, F(x) \\
&\geq \int_t^\infty t \,\mathrm{d}\, F(x) \\
&\geq t \int_t^\infty 1 \,\mathrm{d}\, F(x) \\
&= t \Pr\{X \geq t\};
\end{aligned}$$

(a)

that is, $\Pr\{X \geq t\} \leq \mathrm{E}[X]/t$.

(b)

(Chebyshev's inequality.) Let Y be a random variable with mean μ and variance σ^2. By letting $X = (Y - \mu)^2$, show that for any $\epsilon > 0$,

$$\Pr\{|Y - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}.$$

By Markov's inequality, we have:

$$\mathrm{E}[X] = \mathrm{E}[(Y - \mu)^2] = \sigma^2$$

(b)

implies

$$\begin{aligned}\Pr\{|Y-\mu| \geq \epsilon\} &= \Pr\{|Y-\mu|^2 \geq \epsilon^2\} \\ &= \Pr\{X \geq \epsilon^2\} \\ &\leq \frac{\mathrm{E}[X]}{\epsilon^2} \\ &= \frac{\sigma^2}{\epsilon^2}.\end{aligned}$$

(c)

(The weak law of large numbers.) Let $Z_1, Z_2, \ldots, Z_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Let $\overline{Z}_n = (1/n)\sum_{i=1}^n Z_i$ be the sample mean. Show that

$$\Pr\{|\overline{Z}_n - \mu| > \epsilon\} \leq \frac{\sigma^2}{n\epsilon^2}.$$

Thus $\Pr\{|\overline{Z}_n - \mu| > \epsilon\} \to 0$ as $n \to \infty$. This is known as the weak law of large numbers.

By the basic properties of mean and variance, we have:

$$\mathrm{E}[\overline{Z}_n] = \frac{\sum_{i=1}^n \mathrm{E}[Z_i]}{n} = \mu$$

and

$$\mathrm{Var}\{\overline{Z}_n\} = \frac{\sum_{i=1}^n \mathrm{Var}\{Z_i\}}{n^2} = \frac{\sigma^2}{n}.$$

So, by Chebyshev's inequality, we have:

$$\Pr\{|\overline{Z}_n - \mu| > \epsilon\} \leq \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Problem 2

AEP. Let X_i be i.i.d. $\sim p(x)$, $x \in \{1, 2, \ldots, m\}$. Let $\mu = \mathrm{E}[X]$ and $H = -\sum p(x)\log p(x)$. Let $A^n = \{x^n \in \chi^n : |-(1/n)\log p(x^n) - H| \leq \epsilon\}$. Let $B^n = \{x^n \in \chi^n : |(1/n)\sum_i^n X_i - \mu| \leq \epsilon\}$.

(a)

Does $\Pr\{X^n \in A^n\} \to 1$?

Give $\epsilon > 0$. Then, by Archimedean property, there exists a natural number N such that $1/(N+1) < \epsilon < 1/N$. By Theorem 3.1.2, there is a natural number M such that

$$\Pr\left\{A_{1/N}^{(n)}\right\} > 1 - \frac{1}{N}$$

for all $n \geq M$. So,

$$\Pr\{X^n \in A^n\} \geq \Pr\left\{A_{1/N}^{(n)}\right\} > 1 - \frac{1}{N} > 1 - \epsilon.$$

Therefore we hold $\Pr\{X^n \in A^n\} \to 1$.

(a)

(b)

Does $\Pr\{X^n \in A^n \cap B^n\} \to 1$?

Recall $\epsilon > 0$. By the solution of above problem, there is a natural number N_1 such that $\Pr\{A^n\} > 1-\epsilon/2$. By problem 1, there is a natural number $N_2 > N_1$ such that $\Pr\{B^n\} > 1-\sigma^2/(n\epsilon^2) > 1-\epsilon/2$ for all $n > N_2$. Thus, we have:

$$\begin{aligned}\Pr(\{X^n \in A^n \cap B^n\} &= \Pr(\{X^n \in A^n\} + \Pr\{X^n \in A^n\} - \Pr(\{X^n \in A^n \cup B^n\} \\ &\geq \Pr(\{X^n \in A^n\} + \Pr\{X^n \in A^n\} - 1 \\ &> 1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 \\ &= 1 - \epsilon.\end{aligned}$$

Therefore we hold $\Pr\{X^n \in A^n \cap B^n\} \to 1$.

(b)

(c)

Show $|A^n \cap B^n| \leq 2^{n(H+\epsilon)}$, for all n.

By the definition of A^n, we have:

$$2^{-(H-\epsilon)} \leq p(x) \leq 2^{-(H+\epsilon)} \tag{1}$$

implies

$$1 = \sum_{x^n \in \mathcal{X}^n} p(x^n) \geq \sum_{x^n \in A^n} p(x)^n = \sum_{x^n \in A^n} 2^{-n(H-\epsilon)} = 2^{-n(H-\epsilon)}|A^n|;$$

that is, $|A^n| \leq 2^{n(H-\epsilon)} \leq 2^{n(H+\epsilon)}$. Therefore $|A^n \cap B^n| \leq 2^{n(H+\epsilon)}$ because that $A^n \cap B^n \subset A^n$.

(c)

(d)

Show $|A^n \cap B^n| \geq (1/2)2^{n(H-\epsilon)}$, for n sufficiently large.

By the solution (b) of this problem, there is a natural number N such that $\Pr\{X^n \in A^n \cap B^n\} > 1/2$.

By inequality (1), we have:

$$\frac{1}{2} \leq \sum_{x^n \in A^n \cap B^n} p(x^n) \leq \sum_{x^n \in A^n \cap B^n} 2^{-n(H+\epsilon)} = |A^n \cap B^n| 2^{-n(H+\epsilon)}$$

for all $n > N$; that is, $|A^n \cap B^n| \geq (1/2)2^{n(H-\epsilon)}$, for n sufficiently large.

(d)

Problem 3

Random box size. An n-dimensional rectangular box with sides $X_1, X_2, X_3, \ldots, X_n$ is to be constructed. The volume is $V_n = \prod_{i=1}^n X_i$. The edge length ℓ of a n-cube with the same volume as the random box is $\ell = V_n^{1/n}$. Let $X_1, X_2, \ldots$ be i.i.d. uniform random variables over the unit interval $[0,1]$. Find $\lim_{n\to\infty} V_n^{1/n}$ and compare to $\mathrm{E}[V_n]^{1/n}$. Clearly, the expected edge length does not capture the idea of the volume of the box. The geometric mean, rather than the arithmetic mean, characterizes the behavior of products.

First, we show $V_n^{1/n}$ converges to $1/e$ in probability, and $\mathrm{E}[V_n]^{1/n}$ converges to $1/2$. By the definition of X_i and Khinchine's Theorem, we have:

$$\operatorname*{plim}_{n\to\infty} \frac{\sum_{i=n}^{n} \log X_i}{n} = \mathrm{E}[\log X_1] = \int_0^1 \log X = -\log e$$

implies

$$\operatorname*{plim}_{n\to\infty} V_n^{1/n} = 2^{\operatorname{plim}_{n\to\infty}(1/n)\sum_{i=1}^n \log X_i} = 2^{-\log e} = \frac{1}{e}.$$

Since X_is are i.i.d.,

$$\mathrm{E}\left[V_n\right]^{1/n} = \mathrm{E}\left[\prod_{i=1}^n X_i\right]^{1/n} = \left(\prod_{i=1}^n \mathrm{E}\left[X_i\right]\right)^{1/n} = \left(\prod_{i=1}^n \frac{1}{2}\right)^{1/n} = \frac{1}{2}$$

for all n; that is, $\operatorname{plim}_{m\to\infty} V_m^{1/m} < \mathrm{E}\left[V_n\right]^{1/n}$ for all n.

Problem 4

Proof of Theorem 3.3.1. This problem shows that the size of the smallest "probable" set is about 2^{nH}. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim p(x)$. Let $B_\delta^{(n)} \subset \chi^n$ such that $\Pr(B_\delta^{(n)}) > 1-\delta$. Fix $\epsilon < 1/2$.

(a)

Given any two sets A, B such that $\Pr(A) > 1-\epsilon_1$ and $\Pr(A) > 1-\epsilon_1$ and $\Pr(B) > 1-\epsilon_2$, show that $\Pr(A\cap B) > 1-\epsilon_1-\epsilon_2$. Hence $\Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) \geq 1-\epsilon-\delta$.

Denote the complement of A by A^c. Then, we have:

$$\begin{aligned}
\Pr(A\cap B) &= 1 - \Pr(A^c \cap B^c) \\
&= 1 - (\Pr(A^c) + \Pr(B^c) - \Pr(A^c \cap B^c)) \\
&\geq 1 - \Pr(A^c) - \Pr(B^c) \\
&\geq 1 - (1-\epsilon_1) - (1-\epsilon_2) \\
&= 1 - \epsilon_1 - \epsilon_2.
\end{aligned}$$

Therefore $\Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) \geq 1 - \epsilon - \delta$ if we take $\epsilon_1 = \epsilon$, $A = A_\epsilon^{(n)}$, $\epsilon_2 = \delta$, and $B = B_\delta^{(n)}$. (a)

(b)

Justify the steps in the chain of inequalities

$$
\begin{aligned}
1 - \epsilon - \delta &\leq \Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) && (2)\\
&= \sum_{A_\epsilon^{(n)} \cap B_\delta^{(n)}} p(x^n) && (3)\\
&\leq \sum_{A_\epsilon^{(n)} \cap B_\delta^{(n)}} 2^{-n(H-\epsilon)} && (4)\\
&= \left|A_\epsilon^{(n)} \cap B_\delta^{(n)}\right| 2^{-n(H-\epsilon)} && (5)\\
&\leq \left|B_\delta^{(n)}\right| 2^{-n(H-\epsilon)} && (6)
\end{aligned}
$$

By the solution of above problem (a), inequalities (2) and (3) are clearly. By inequality (1) in Problem 2 (c), we have inequalities (4) and (5). Therefore inequality (6) must be true. (b)

(c)

Complete the proof of the theorem.

By Theorem 3.1.2, there is a natural number N such that $\Pr\{A_\epsilon^{(n)}\} > 1 - \delta'$. By inequality (6), we have:

$$1 - \delta' - \delta \leq \left|B_\delta^{(n)}\right| 2^{-n(H-\delta')}$$

implies

$$\frac{1}{n} \log \left|B_\delta^{(n)}\right| \leq H - \delta' + \log(1 - \delta - \delta') < H - \delta' + \log 1 = H - \delta'.$$

Therefore the proof is completed. (c)

Problem 5

Monotonicity of entropy per element. For a stationary stochastic process $X_1, X_2, \ldots, X_n$, show that

(a)

$H(X_1, X_2, \ldots, X_n)/n \leq H(X_1, X_2, \ldots, X_{n-1})/(n-1)$.

By entropy chain rule, Theorem 2.5.1, we have:

$$
\begin{aligned}
\frac{H(X_1, \ldots, X_n)}{n} &= \frac{\sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_1)}{n} \times \frac{n-1}{n-1}\\
&= \frac{\sum_{i=1}^{n-1} H(X_i | X_{i-1}, \ldots, X_1)}{n-1} + \frac{-\sum_{i=1}^{n-1} H(X_i | X_{i-1}, \ldots, X_1)}{n(n-1)}.
\end{aligned}
$$

(a)

So,

$$\frac{H(X_1,\ldots,X_n)}{n} \le \frac{\sum_{i=1}^{n-1} H(X_i|X_{i-1},\ldots,X_1)}{n-1} = \frac{H(X_1,\ldots,X_{n-1})}{n-1}.$$

(b)

$H(X_1, X_2, \ldots, X_n)/n \ge H(X_n|X_{n-1}, \ldots, X_1)$.

Without loss of generality, we consider $X_0, \ldots, X_{-n+2}$ are available. By the definition of stationary, we have:

$$\begin{aligned} H(X_n|X_{n-1},\ldots,X_1) &= \frac{nH(X_n|X_{n-1},\ldots,X_1)}{n} \\ &= \frac{\sum_{i=0}^{n-1} H(X_{n-i}|X_{n-i-1},\ldots,X_{1-i})}{n} \\ &\le \frac{\sum_{i=1}^{n-1} H(X_{n-i}|X_{n-i-1},\ldots,X_{1-i})}{n} \\ &\le \frac{\sum_{i=1}^{n-1} H(X_i|X_{i-1},\ldots,X_1)}{n} \\ &= \frac{H(X_1,\ldots,X_n)}{n}. \end{aligned}$$

Therefore the proof is completed.

Problem 6

Maximum entropy process. A discrete memory-less source has the alphabet $\{1, 2\}$, where the symbol 1 has duration 1 and the symbol 2 has duration 2. The probabilities of 1 and 2 are p_1 and p_2, respectively. Find the value of p_1 that maximizes the source entropy per unit time $H(\chi) = H(X)/\operatorname{E}[T]$. What is the maximum value $H(\chi)$?

By definitions of entropy and T, we have:

$$H(p_1) = -p_1 \log p_1 - (1-p_1)\log(1-p_1)$$

and

$$\operatorname{E}[T] = 1 \cdot p_1 + 2 \cdot (1-p_1) = 2 - p_1;$$

that is,

$$H(\chi) = \frac{H(X)}{\operatorname{E}[T]} = \frac{-p_1 \log p_1 - (1-p_1)\log(1-p_1)}{2-p_1}.$$

Thus,

$$\frac{\mathrm{d}\,H(\chi)}{\mathrm{d}\,p_1} = \frac{\log(1-p_1) - 2\log p_1}{(2-p_1)^2} \text{ and } \frac{\mathrm{d}^2\,H(\chi)}{\mathrm{d}\,p_1^2} = -\frac{(2-p_1)^2 - 2(1-p_1)p_1\log(1-p_1) + 4(1-p_1)p_1\log p_1}{(2-p_1)^3(1-p_1)p_1}$$

Problem 6 continued on next page...

imply

$$\left.\frac{\mathrm{d}\, H(\chi)}{\mathrm{d}\, p_1}\right|_{p_1=p_{10}} = 0 \text{ and } \left.\frac{\mathrm{d}^2\, H(\chi)}{\mathrm{d}\, p_1^2}\right|_{p_1=p_{10}} = -\frac{(15+7\sqrt{5})\log e}{10} < 0,$$

where $p_{10} = (-1 \pm \sqrt{5})/2$, respectively. Since p_1 must between 0 and 1, the maximum point and value for entropy $H(\chi)$ are $(-1+\sqrt{5})/2$ and $\left(5+\sqrt{5}\right)\left(\sinh^{-1}\left(\frac{11}{2}\right) + 2\sqrt{5}\tanh^{-1}\left(2-\sqrt{5}\right)\right)\log e/20$, respectively.

Problem 7

Entropy rate of a dog looking for a bone. A dog walks on the integers, possibly reversing direction at each step with probability $p = 0.1$. Let $X_0 = 0$. The first step is equally likely to be positive or negative. A typical walk might look like this:

$$(X_0, X_1, \ldots) = (0, -1, -2, -3, -4, -3, -2, -1, 0, 1, \ldots).$$

(a)

Find $H(X_1, X_2, \ldots, X_n)$.

By the definition of the walk, we have:

$$\Pr(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0 = 0) = \Pr(X_{n+1} = x_{n+1} | X_n = x_n)$$

and

$$\Pr(X_{n+1} = x_{n+1} | X_n = x_n) = \begin{cases} 0.9, & \text{if } x_{n+1} = x_n; \\ 0.1, & \text{if } x_{n+1} = x_n - 1, \end{cases}$$

for all natural number $n > 1$; that is, the process is a stationary Markov process on $(1, \infty)$. Then, by the entropy chain rule,

(a)

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= \sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_1) \\ &= H(X_1) + \sum_{i=2}^{n} H(X_i | X_{i-1}, \ldots, X_1) \\ &= H(0.5) + \sum_{i=2}^{n} H(X_i | X_{i-1}) \\ &= 1 + \sum_{i=2}^{n} \sum_{j=0}^{i-1} \binom{i-1}{j} 0.9^{i-j-1} 0.1^j H(X_i | X_{i-1} = -j) \\ &= 1 + (n-1) \sum_{j=0}^{i-1} \binom{i-1}{j} 0.9^{i-j-1} 0.1^j H(0.1) \end{aligned}$$

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= 1 + (n-1)H(0.1)\sum_{j=0}^{i-1}\binom{i-1}{j}0.9^{i-j-1}0.1^j \\ &= 1 + (n-1)\left(1 + \log 5 - \frac{9\log 3}{5}\right) \\ &= \left(1 + \log 5 - \frac{9\log 3}{5}\right)n - \log 5 + \frac{9\log 3}{5}. \end{aligned}$$

(b)

Find the entropy rate of the dog.

By the definition of entropy rate, we have:

$$\begin{aligned} H'(\chi) &= \lim_{n\to\infty} H(X_1, \ldots, H_n) \\ &= \lim_{n\to\infty}\left(1 + \log 5 - \frac{9\log 3}{5}\right)n - \log 5 + \frac{9\log 3}{5} \\ &= 1 + \log 5 - \frac{9\log 3}{5}. \end{aligned}$$

(c)

What is the expected number of steps the dog takes before reversing direction?

Let R be the number of steps the dog taken between reversals. By the definition of R, we have:

$$\mathrm{E}[R] = \sum_{i=1}^{\infty} 0.9^{i-1}0.1R = 10.$$

Problem 8

Random walk on graph. Consider a random walk on the following graph:

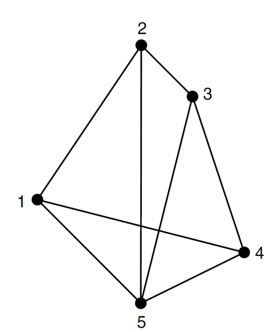


Figure 1: Random Walk on Graph

 Problem 8 continued on next page...

(a)

Calculate stationary distribution.

By the definition of adjacency matrix M_{adj} on the undirected graph, we have:

$$M_{adj} = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

implies the chain is Markov and the probability transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix}.$$

(a)

Take

$$P^{(\infty)} = \left(P_{ij}^{(\infty)}\right) = \lim_{n\to\infty} P^n = \begin{pmatrix} \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{4} \\ \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{4} \\ \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{4} \\ \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{4} \\ \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{4} \end{pmatrix}.$$

Since $P_{ij}^{(\infty)}$, for all $i \neq j$, are nonzero, all of states are not transient or null recurrent; that is, the Markov chain is stationary. So, by the stationary distribution definition of it, the distribution

$$\pi = (\pi_i) = \left(P_{ij}^{(\infty)}\right) = \begin{pmatrix} \frac{3}{16} \\ \frac{3}{16} \\ \frac{3}{16} \\ \frac{3}{16} \\ \frac{1}{4} \end{pmatrix}.$$

(b)

What is the entropy rate:

By Theorem 4.2.1 and 4.2.4, we have:

$$\begin{aligned} H'(\chi) &= \lim_{n\to\infty} H(X_n|X_{n-1},\ldots,X_1) \\ &= H(\pi) \\ &= -\sum_{i=1}^{5}\sum_{j=1}^{5} \pi_i P_{ij}\log P_{ij} \\ &= \frac{3\log 3}{4} + \frac{1}{2}. \end{aligned}$$

(c)

Find the mutual information $I(X_{n+1;X_n})$ assuming the process is stationary.

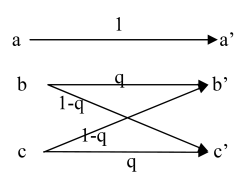
By the definition of mutual information and the equation (4.25) in the text, we have:

$$\begin{aligned} I(X_{n+1};X_n) &= H(X_{n+1}) - H(X_{n+1}|X_n) \\ &= H(\pi) - H(\chi) \\ &= -\frac{3\log 3}{4} + \frac{7}{2} - \frac{3\log 3}{4} - \frac{1}{2} \\ &= -\frac{3\log 3}{2} + 3. \end{aligned}$$

Problem 9

Please find the channel capacity of the following channel:

- $P(a) = \bar{P}$;
- $P(b) = P(c) = \bar{Q}$.



Channel Capacity $C =$?

Let X be discrete random variables with alphabets $\mathcal{X} = \{a, b, c\}$. By the definition, we have:

$$\Pr(Y = y|X = a') = \begin{cases} 1, & \text{if } y = a'; \\ 0, & \text{if } y = b'; \\ 0, & \text{if } y = c', \end{cases}$$

$$\Pr(Y = y|X = b') = \begin{cases} 0, & \text{if } y = a'; \\ q, & \text{if } y = b'; \\ 1-q, & \text{if } y = c', \end{cases}$$

and

Problem 9 continued on next page. . .

$$\Pr(Y=y|X=c') = \begin{cases} 0, & \text{if } y=a'; \\ 1-q, & \text{if } y=b'; \\ q, & \text{if } y=c', \end{cases}$$

Then,

$$\begin{aligned} H(Y|X) &= \textstyle\sum_{X\in\mathcal{X}} \Pr(X=X)H(Y|X=x) \\ &= \Pr(X=a)H(Y|X=a)+\Pr(X=b)H(Y|X=b)+\Pr(X=c)H(Y|X=c) \\ &= \bar{P}(-0\log 0 - 1\log 1 - 1\log 1) \\ &\quad +\bar{Q}(-0\log 0 - q\log q - (1-q)\log(1-q)) \\ &\quad +\bar{Q}(-0\log 0 - q\log q - (1-q)\log(1-q)) \\ &= 2\bar{Q}(-q\log q - (1-q)\log(1-q)). \end{aligned}$$

Moreover, by the definition, we have the following joint p.m.f. $p(x,y)$ of X and Y:

$Y \backslash X$	a	b	c
a'	$\bar{P}$	0	0
b'	0	$q\bar{Q}$	$(1-q)\bar{Q}$
c'	0	$(1-q)\bar{Q}$	$q\bar{Q}$

So,

$$\Pr(Y=y) = \begin{cases} \bar{P}, & \text{if } y=a'; \\ \bar{Q}, & \text{if } y=b'; \\ \bar{Q}, & \text{if } y=c' \end{cases}$$

implies

$$H(Y) = -\bar{P}\log\bar{P} - 2\bar{Q}\log\bar{Q}.$$

Therefore

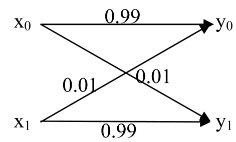
$$\begin{aligned} C &= \max_{p(x,y)} I(X;Y) \\ &= I(X;Y) \\ &= H(Y) - H(Y|X) \\ &= -\bar{P}\log\bar{P} - 2\bar{Q}\log\bar{Q} - 2\bar{Q}(-q\log q - (1-q)\log(1-q)). \end{aligned}$$

Problem 10

For the following BSC.

If $\Pr(X_0) = \Pr(X_1) = 1/2$.

Please calculate:



(a)

$\Pr(y_0)$ and $\Pr(y_1)$.

By the above figure, we have the joint p.m.f. of X and Y:

$Y \backslash X$	x_0	x_1
y_0	$\frac{99}{200}$	$\frac{1}{200}$
y_1	$\frac{1}{200}$	$\frac{99}{200}$

So, $\Pr(y_0) = \Pr(y_1) = 1/2$.

(b)

$H(Y)$

By the solution of above (a), $H(Y) = 1$.

(c)

$I(X;Y)$.

By the above table,

$$\begin{aligned} H(X,Y) &= -\frac{99}{200}\log\frac{99}{200} - \frac{1}{200}\log\frac{1}{200} - \frac{99}{200}\log\frac{99}{200} - \frac{1}{200}\log\frac{1}{200} \\ &= 3 - \frac{99\log 3}{50} + 2\log 5 - \frac{99\log 11}{100}. \end{aligned}$$

By the definition of mutual information,

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = -1 + \frac{99\log 3}{50} - 2\log 5 + \frac{99\log 11}{100}.$$