Solutions to Final Examinations

1. *(20 points)* **Cookies.**
   Let

$$V_n = \prod_{i=1}^{n} X_i,$$

where X_i are i.i.d.

$$X_i = \begin{cases} 1/8, & \text{probability } 1/2, \\ 1/2, & \text{probability } 1/2. \end{cases}$$

Presumably, X_i is the fraction remaining after a single mouse bite.

(a) Let

$$V_n' = \alpha^n.$$

Find the value of α such that V_n and V_n' decrease at the same rate.

For parts (b) and (c), we mix V_n and V_n' as follows. Let

$$Y_i = \lambda\alpha + (1-\lambda)X_i,$$

where $\lambda \in (0,1)$. Let

$$V_n'' = \prod_{i=1}^{n} Y_i.$$

(b) Is the growth rate of V_n'' larger or smaller than $\log \alpha$?

(c) What is the growth rate of V_n'' for $\lambda = 1/2$?

Solution: Cookies.

(a) Since

$$\frac{1}{n} \log V_n \to E \log X_1 = -2 \quad \text{w.p.1},$$

we need $\alpha = 2^{-2} = 1/4$ to have the same growth (or decay) rate.

(b) The growth rate of V_n'' is larger than $\log \alpha$. Indeed, by Jensen's inequality,

$$\log(\lambda\alpha + (1-\lambda)X_i) \geq \lambda \log \alpha + (1-\lambda) \log X_i,$$

so that

$$E \log Y_i \geq \lambda \log \alpha + (1-\lambda) E \log X_i = \log \alpha.$$

(c) Since

$$Y_i = \begin{cases} 3/16, & \text{probability } 1/2, \\ 3/8, & \text{probability } 1/2, \end{cases}$$

the growth rate is given by

$$E \log Y_i = \log \left(\frac{3}{8\sqrt{2}} \right),$$

which is larger than $\log(1/4)$.

2. *(20 points)* **Huffman code.**
Find the binary Huffman encoding for

$$X \sim \mathbf{p} = \left(\frac{19}{40}, \frac{8}{40}, \frac{3}{40}, \frac{3}{40}, \frac{3}{40}, \frac{2}{40}, \frac{2}{40} \right).$$

Solution: Huffman code.

Codeword								
1	x_1	19	19	19	19	19	21	40
01	x_2	8	8	8	8	13	19	
0001	x_3	3	4	6	7	8		
0010	x_4	3	3	4	6			
0011	x_5	3	3	3				
00001	x_6	2	3					
00000	x_7	2						

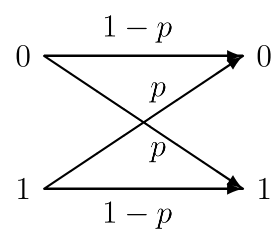
3. *(20 points)* **Good codes.**
Which of the following codes are possible Huffman codes?

   (a) {0,00,01}
   (b) {0,10,11}
   (c) {0,10}

Solution: Good codes.

Only (b) can be a Huffman code; it represents a complete binary tree. (a) is not prefix free; (c) can be improved by replacing the codeword 10 with 1.

4. *(20 points)* **Errors and erasures.**
Consider a binary symmetric channel (BSC) with crossover probability p.

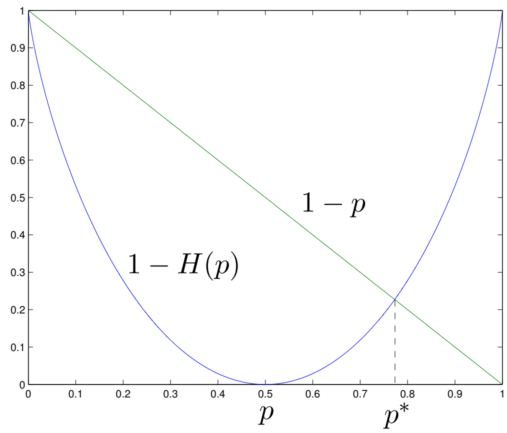


A helpful genie who knows the locations of all bit flips offers to convert flipped bits into erasures. In other words, the genie can transform the BSC into a binary erasure channel. Would you use his power? Be specific.

Solution: Errors and erasures.

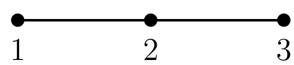
Although it is very tempting to accept the genie's offer, on a second thought, one realizes that it is disadvantageous to convert the bit flips into erasures when p is large. For example, when $p = 1$, the original BSC is noiseless, while the "helpful" genie will erase every single bit coming out from the channel.

The capacity $C_1(p)$ of the binary symmetric channel with crossover probability p is $1-H(p)$ while the capacity $C_2(p)$ of the binary erasure channel with erasure probability p is $1 - p$. One would convert the BSC into a BEC only if $C_1(p) \leq C_2(p)$, that is, $p \leq p^* = .7729$. (See Figure 1.)



5. *(40 points)* **Random walks.**
Consider the following graph with three nodes:

$\{X_i\}$



(a) What is the entropy rate $H(\mathcal{X})$ of the random walk $\{X_i\}_{i=1}^{\infty}$ on this graph?

Now consider a derived process

$$Y_i = \begin{cases} 0, & \text{if } X_i = 1 \text{ or } 3, \\ 1, & \text{if } X_i = 2. \end{cases}$$

(b) Is it Markov?

(c) Find the entropy rate $H(\mathcal{Y})$ of $\{Y_i\}_{i=1}^{\infty}$.

Now consider another derived process

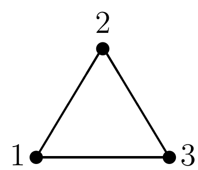
$$Z_i = \begin{cases} 0, & \text{if } X_i = 1 \text{ or } 2, \\ 1, & \text{if } X_i = 3. \end{cases}$$

(d) Is it Markov?

(e) Find the entropy rate $H(\mathcal{Z})$ of $\{Z_i\}_{i=1}^{\infty}$.

For parts (f), (g), and (h), consider the following graph with three nodes:

$\{U_i\}$



(f) What is the entropy rate $H(\mathcal{U})$ of the random walk $\{U_i\}_{i=1}^{\infty}$ on this graph?

Now consider a derived process

$$V_i = \begin{cases} 0, & \text{if } U_i = 1 \text{ or } 2, \\ 1, & \text{if } U_i = 3. \end{cases}$$

(g) Is it Markov?

(h) Find the entropy rate $H(\mathcal{V})$ of $\{V_i\}_{i=1}^{\infty}$.

Solution: Random walks.

(a) It is easy to see that the stationary distribution is given by $\mu = (1/4, 1/2, 1/4)$. The entropy rate is $\sum_j H(X_{n+1}|X_n = j)\mu_j = 1/2$.

(b) Yes, it is Markov. If $Y_n = 0$, then $Y_{n+1} = 1$ w.p.1, and vice versa.

(c) Since the process evolves deterministically, the entropy rate $H(\mathcal{Y})$ is 0.

(d) No, it is not Markov. For example, it is easy to check that $P(Z_{n+1} = 1|Z_n = 0, Z_{n-1} = 1) = 1/2$, while $P(Z_{n+1} = 1|Z_n = 0) = 2/3$.

(e) Although the process is not Markov, as in Problem 6 in midterm, knowing $(X_1, Z_1, \ldots, Z_{n-1})$ is equivalent to knowing $(X_1, \ldots, X_{n-1})$. Thus we have

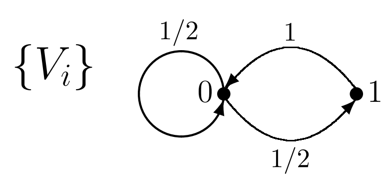
$$H(Z_n|X_1, Z^{n-1}) = H(Z_n|X^{n-1}) = H(Z_n|X_{n-1}) = 1/2,$$

and hence

$$H(\mathcal{Z}) = \lim_{n\to\infty} H(Z_n|X_1, Z^{n-1}) = 1/2.$$

(f) Given U_n, U_{n+1} takes two values with equal probability. Hence, $H(\mathcal{U}) = 1$.

(g) Yes, it is Markov with the following transition probability:



(h) The stationary distribution is $\mu = (2/3, 1/3)$, so that

$$H(\mathcal{V}) = \sum_j H(V_{n+1}|V_n = j)\mu_j = 2/3.$$

6. *(20 points)* **Code constraint.**
What is the capacity of a BSC(p) under the constraint that each of the codewords has a proportion of 1's less than or equal to α, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i(w) \leq \alpha, \qquad \text{for } w \in \{1, 2, \ldots, 2^{nR}\}.$$

(Pay attention when $\alpha > 1/2$.)

Solution: Code constraint.

Using the similar argument for the capacity of Gaussian channels under the power constraint P, we find that the capacity C of a BSC(p) under the proportion constraint α is

$$C = \max_{p(x): EX \leq \alpha} I(X;Y).$$

Now under the Bernoulli(π) input distribution with $\pi \leq \alpha$, we have

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(Z|X) \\ &= H(Y) - H(Z) \\ &= H(\pi * p) - H(p), \end{aligned} \tag{1}$$

where $\pi * p = (1-\pi)p + \pi(1-p)$. (Breaking $I(X;Y) = H(X) - H(X|Y) = H(X) - H(Z|Y)$ is way more complicated since Z and Y are correlated.) Now when $\alpha > 1/2$, we have

$$\max_{\pi} H(\pi * p) - H(p) = 1 - H(p),$$

with the capacity-achieving $\pi^* = 1/2$. On the other hand, when $\alpha \leq 1/2$, $\pi^* = \alpha$ achieves the maximum of (1); hence

$$C = H(\alpha * p) - H(p).$$

7. *(20 points)* **Typicality.**
Let (X,Y) have joint probability mass function $p(x,y)$ given as

X \ Y	0	1
0	.1	.3
1	.4	.2

(a) Find $H(X), H(Y)$, and $I(X;Y)$. (Don't bother to compute the actual numerical values.)

(b) Suppose $\{X_i\}$ is independent and identically distributed (i.i.d.) according to Bern(.4), $\{Y_i\}$ is i.i.d. Bern(1/2), and X^n and Y^n are independent. Find (to first order in the exponent) the probability that (X^n, Y^n) is jointly typical (with respect to the joint distribution $p(x,y)$.

Solution: Typicality.

(a)

$$\begin{aligned} H(X) &= H(.4), \\ H(Y) &= H(1/2) = 1, \\ I(X;Y) &= H(Y) - H(Y|X) = 1 - .4H(1/4) - .6H(1/3). \end{aligned}$$

(b) From the joint AEP, the probability (X^n, Y^n) is jointly typical w.r.t. $p(x,y)$ is $\doteq 2^{-n(I(X;Y)\pm\epsilon)}$.

8. *(20 points)* **Partition.**
Let (X, Y) denote height and weight. Let $[Y]$ be Y rounded off to the nearest pound.

(a) Which is greater $I(X;Y)$ or $I(X;[Y])$?

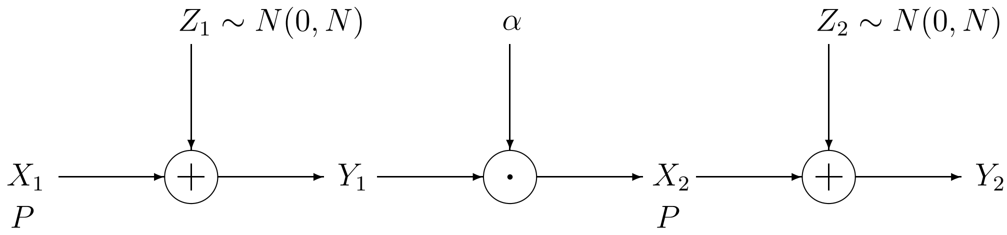
(b) Why?

Solution: Partition.

(a) $I(X;Y) \geq I(X;[Y])$.

(b) Data processing inequality.

9. *(20 points)* **Amplify and forward.**
We cascade two Gaussian channels by feeding the (scaled) output of the first channel into the second.



Thus noises Z_1 and Z_2 are independent and identically distributed according to $N(0, N)$,

$$EX_1^2 = EX_2^2 = P,$$

$$Y_1 = X_1 + Z_1,$$
$$Y_2 = X_2 + Z_2,$$

and

$$X_2 = \alpha Y_1,$$

where the scaling factor α is chosen to satisfy the power constraint $EX_2^2 = P$.

(a) *(5 points)* What scaling factor α satisfies the power constraint?

(b) *(10 points)* Find

$$C = \max_{p(x_1)} I(X_1; Y_2).$$

(c) *(5 points)* Is the cascade capacity C greater or less than $\frac{1}{2}\log\left(1 + \frac{P}{N}\right)$?

Solution: Amplify and forward.

(a) We want $\alpha^2 EY_1^2 = \alpha^2(P + N) = P$. Hence $\alpha = \sqrt{\frac{P}{P+N}}$.

(b) Since $Y_2 = X_2 + Z_2 = \alpha Y_1 + Z_2 = \alpha X_1 + (\alpha Z_1 + Z_2)$, the channel from X_1 to Y_2 is a Gaussian channel with signal-to-noise ratio $\alpha^2 P : (\alpha^2 N + N)$. Hence, the capacity is

$$C = \frac{1}{2}\log\left(1 + \frac{\alpha^2 P}{(\alpha^2+1)N}\right) = \frac{1}{2}\log\left(1 + \frac{P^2}{(2P+N)N}\right) = \frac{1}{2}\log\left(\frac{(P+N)^2}{(2P+N)N}\right).$$

(c) The cascade capacity C is less than $\frac{1}{2}\log\left(1 + \frac{P}{N}\right)$, which can be shown by data processing inequality. Adding an extra noise wouldn't increase the capacity.