EECS 126 Probability and Random Processes — University of California, Berkeley: Fall 2015
Kannan Ramchandran — September 22, 2015

Midterm Exam 1 Solution

Last name	First name	SID

Name of student on your left:
Name of student on your right:

- DO NOT open the exam until instructed to do so.
- The total number of points is **110**, but a score of $\geq$ **100** is considered perfect.
- You have 10 minutes to read this exam without writing anything and 90 minutes to work on the problems.
- Box your final answers.
- Partial credit will not be given to answers that have no proper reasoning.
- **Remember to write your name and SID on the top left corner of every sheet of paper.**
- **Do not write on the reverse sides of the pages.**
- All eletronic devices must be turned off. Textbooks, computers, calculators, etc. are prohibited.
- No form of collaboration between students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.
- **You must include explanations to receive credit.**

Problem	Part	Max	Points	Problem	Part	Max	Points
1	(a)	8		2		20	
	(b)	8		3		25	
	(c)	8		4		25	
	(d)	8					
	(e)	8					
		40					
Total						110	

Cheat sheet

1. Discrete Random Variables

1) Geometric with parameter $p \in [0,1]$:
$P(X = n) = (1-p)^{n-1}p,\ n \geq 1$
$E[X] = \frac{1}{p},\ \text{var}(X) = \frac{1-p}{p^2}$

2) Binomial with parameters N and p:
$P(X = n) = \binom{N}{n} p^n (1-p)^{N-n},\ n = 0, \ldots, N,$ where $\binom{N}{n} = \frac{N!}{(N-n)!n!}$
$E[X] = Np,\ \text{var}(X) = Np(1-p)$

3) Poission with parameter λ:
$P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda},\ n \geq 0$
$E[X] = \lambda,\ \text{var}(X) = \lambda$

2. Continuous Random Variables

1) Uniformly distributed in $[a, b]$, for some $a < b$:
$f_X(x) = \frac{1}{b-a} 1\{a \leq x \leq b\}$
$E[X] = \frac{a+b}{2},\ \text{var}(X) = \frac{(b-a)^2}{12}$

2) Exponentially distributed with rate $\lambda > 0$:
$f_X(x) = \lambda e^{-\lambda x} 1\{x \geq 0\}$
$E[X] = \frac{1}{\lambda}, \text{var}(X) = \frac{1}{\lambda^2}$

3) Gaussian, or normal, with mean μ and variance σ^2:
$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$
$E[X] = \mu,\ \text{var}(X) = \sigma^2$

Problem 1. **Short questions: 8 points each.**

(a) Evaluate the following statements with *True* or *False*. Give one-line explanations to get credit.

(i) If two random variables are independent, they are uncorrelated.

Solution: True.

If X and Y are independent, $E[XY] = E[X]E[Y]$. Then they are uncorrelated.

(ii) Disjoint events are independent.

Solution: False.

If A and B are disjoint, $P(A \cap B) = 0 \neq P(A)P(B)$.

(iii) If events A and B are independent, then conditioned on any other event C, A and B are still independent.

Solution: False.

A, B: outcomes of two independent coin flips; C: there is only one Head in $\{A, B\}$.

(iv) The variance of a sum of uncorrelated random variables is the sum of their variances.

Solution: True.

When finding the variance of the sum, cross terms are 0 for uncorrelated random variables.

(b) You and your friend have created a new language in order to text message with each other. The language consists of an alphabet of 7 letters: $\{A, B, C, D, E, F, G\}$. Unfortunately, your friend has very strict parents who cap the amount of data she receives over text messaging. You decide to use Huffman coding to compress your messages as much as possible. Assume that the frequency of letters in your language is according to the following table:

$P(X)$	X
0.15	A
0.05	B
0.2	C
0.05	D
0.05	E
0.15	F
0.35	G

Now, instead of using binary codewords, we will use *ternary* codewords (e.g. $\{0, 1, 2\}$). Construct a ternary Huffman code for this alphabet and find the expected length of a codeword.

Solution: The Huffman tree is shown in Figure 1.

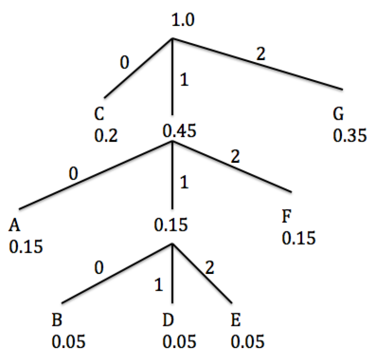

Figure 1: Huffman tree

The codewords are shown below:

X	codeword
A	10
B	110
C	0
D	111
E	112
F	12
G	2

The expected length of a codeword is

$$2 \times 0.15 + 3 \times 0.05 + 1 \times 0.2 + 3 \times 0.05 + 3 \times 0.05 + 2 \times 0.15 + 1 \times 0.35 = 1.6$$

(c) Consider a regular hexagon whose side length is 1. Five points A, B, C, D and E are chosen on the perimeter of the hexagon randomly as follows. First, we pick one of the 6 sides of the hexagon uniformly at random. Then, the position of point A is chosen uniformly at random on the picked side. We then pick one of the remaining 5 sides of the hexagon and decide the position of point B uniformly at random on the chosen side. We repeat this procedure until we locate all five points. Then, we draw edges between (A, B), (B, C), (C, D), (D, E) and (E, A). The figure formed by the 5 edges can be a "star" as in Figure 2(a), or not a "star" as in Figure 2(b). What is the probability that the formed figure is a "star"?

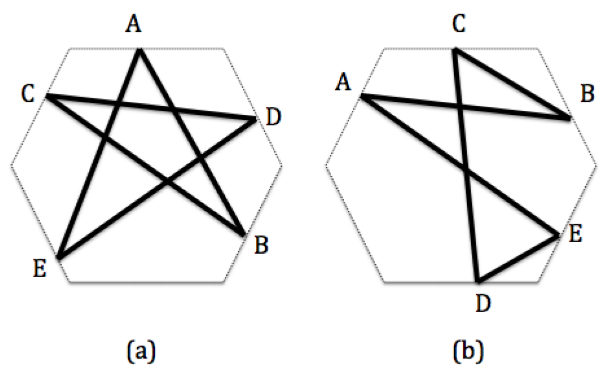

Figure 2: Examples.

Solution: Fix the position of A, there are 4! possible permutations of the other points. Among these permutations, only 2 of them are "stars", as shown in Figure 3. Therefore, the probability is

$$\frac{2}{4!} = \frac{1}{12}.$$

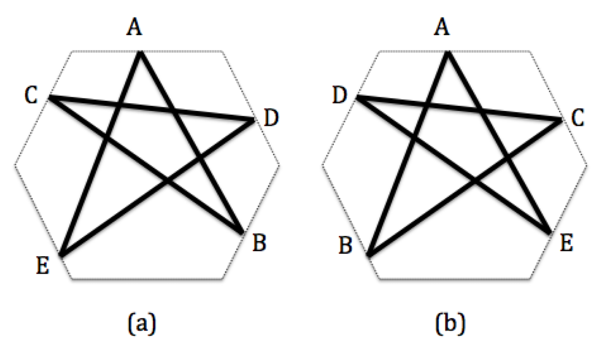

Figure 3: Stars.

(d) 8 couples enter a casino. After two hours, 8 of the original 16 people remain (the rest have left). Each person decides whether or not to leave independently of others' decisions. Find the expected number of couples still in the casino at the end of two hours.

Solution: Conditioned on the number of people remaining, all possible outcomes (of which 8 people remain) have the same probability, i.e., $\frac{1}{\binom{16}{8}}$. Using this, the probability of any particular couple remaining is obtained by counting the number of ways to choose 8 people from 16 people such that the particular couple is chosen:

$$p_c = \frac{\binom{14}{6}}{\binom{16}{8}} = \frac{8 \times 7}{16 \times 15} = \frac{7}{30}.$$

Now, consider a new random variable X_i that takes value of 1 if the ith couple remains, and takes 0 otherwise. The expected value of this random variable $E[X_i] = p_c$. Let X be the total number of couples remaining, then

$$X = X_1 + X_2 + \cdots + X_8.$$

Using linearity of expectation, we conclude

$$E[X] = 8 \times \frac{7}{30} = \frac{28}{15} \simeq 1.87$$

(e) A fair hundred-sided die (with values $0, 1, 2, \ldots, 99$) is rolled independently 50 times, with outcomes $X_1, \ldots, X_{50}$. Let Z be the empirical mean of the 50 rolls, i.e., $Z = \frac{1}{50}\sum_{i=1}^{50} X_i$, and Y be the maximum value attained over the 50 rolls, i.e. $Y = \max\{X_1, X_2, \ldots, X_{50}\}$. Find $E[Z|Y]$ as a function of Y.

Solution: We know that there must exist j $(1 \leq j \leq 50)$ such that $X_j = Y$. For $i \neq j$, conditioned on Y, X_i's are i.i.d. uniformly distributed in the set $\{0, \ldots, Y\}$, and $E[X_i|Y] = \frac{Y}{2}$. Then we know

$$
\begin{aligned}
E[Z|Y] &= \frac{1}{50}[E[X_j|Y] + \sum_{i \neq j} E[X_i|Y]] \\
&= \frac{1}{50}[Y + \frac{49}{2}Y] \\
&= \frac{51}{100}Y.
\end{aligned}
$$

Problem 2. (20 points) Consider two continuous random variables X and Y that are uniformly distributed with their joint probability density function (PDF) equal to A over the shaded region as shown below.

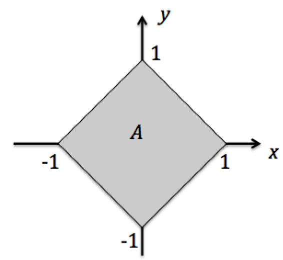

Figure 4: Joint PDF of X and Y

(a) What is A?

Solution: The area of the shaded square is 2. In order that the PDF is valid, the integration of the PDF over the shaded square should be one. Therefore, $A = 1/2$.

(b) Are X and Y independent? Are X and Y uncorrelated? Explain.

Solution: X and Y are not independent. To see this, think about $f_{X,Y}(0.75, 0.75)$, $f_X(0.75)$, and $f_Y(0.75)$. According to the joint PDF, $f_{X,Y}(0.75, 0.75) = 0$. However, as for the marginal distribution, we can see that $f_X(0.75) > 0$ and $f_Y(0.75) > 0$. Then we have $f_{X,Y}(0.75, 0.75) \neq f_X(0.75)f_Y(0.75)$. Therefore, X and Y are not independent.

Since the joint PDF is symmetrical with respect to the x-axis and y-axis, we know that $E[XY] = 0$, $E[X] = 0$, and $E[Y] = 0$. Then $\text{cov}(X, Y) = E[XY] - E[X]E[Y] = 0$. Therefore, X and Y are uncorrelated.

(c) Let $Z = |X| + |Y|$. Find the cumulative distribution function (CDF) of Z.

Solution: We want to find $F_Z(z) = P(Z \leq z)$. For $z < 0$, we have $P(Z \leq z) = 0$ and for $z > 1$, we have $P(Z \leq z) = 1$. Now consider $0 \leq z \leq 1$. We have

$$P(Z \leq z) = \int_{Z \leq z} f_{X,Y}(x,y)dxdy = \int_{|X|+|Y| \leq z} \frac{1}{2} dxdy = z^2.$$

Therefore, the CDF of Z is

$$F_Z(z) = P(Z \leq z) = \begin{cases} 0, & \text{if } z < 0 \\ z^2, & \text{if } 0 \leq z \leq 1 \\ 1, & \text{otherwise.} \end{cases}$$

Problem 3. (25 points) In an EE126 office hour, students bring either a difficult-to-answer question with probability $p = 0.2$ or an easy-to-answer question with probability $1 - p = 0.8$. A GSI takes a random amount of time to answer a question, with this time duration being exponentially distributed with rate $\mu_D = 1$ (question per minute)-where D denotes "difficult"-if the problem is difficult, and $\mu_E = 2$ (questions per minute)-where E denotes "easy"-if the problem is easy.

(a) Is the amount of time taken by the GSI to answer a random question exponentially distributed? Justify your answer.

Solution: The amount of time taken by the GSI to answer a random question is a mixture of two exponential distributions. Therefore, the time is not exponentially distributed.

One day, you visit an office hour and find the GSI answering the question of another student who came before you. You find out that the GSI has been busy with the other student's question for $T = \ln 4 \simeq 1.3863$ *(minutes) before you come. Assume that you are the only student waiting in the room.*

(b) Conditioned on the fact that the GSI has been busy with the other student's question for T minutes, let q be the conditional probability that the problem is difficult. Find the value of q.

Solution: Denote by X the random amount of time to answer a question and by Z the indicator of the event that the problem being answered is difficult. Then,

$$P(X > t | Z = 0) = e^{-\mu_E t}$$
$$P(X > t | Z = 1) = e^{-\mu_D t}$$

for $t \geq 0$. Thus,

$$P(X > t) = pe^{-\mu_D t} + (1 - p)e^{-\mu_E t} = 0.2e^{-t} + 0.8e^{-2t}.$$

As the GSI has been helping the other student for T time units, we know $X > T$. Thus,

$$q = P(Z = 1 | X > T) = \frac{P(Z = 1, X > T)}{P(X > T)} = \frac{pe^{-\mu_D T}}{pe^{-\mu_D T} + (1 - p)e^{-\mu_E T}}$$
$$= \frac{0.2e^{-T}}{0.2e^{-T} + 0.8e^{-2T}} = \frac{1}{1 + 4e^{-T}} = \frac{1}{1 + 1} = \frac{1}{2}.$$

(c) Conditioned on the information above, you want to find the expected amount of time you have to wait from the time you arrive until the other student's question is answered. Express your answer in terms of q of part (b).

Solution: Using the memoryless property of exponential random variables,

$$\begin{aligned}
&E[X-T|X>T] \\
=&P(Z=0|X>T)E[X-T|X>T,Z=0]+P(Z=1|X>T)E[X-T|X>T,Z=1] \\
=&(1-q)\frac{1}{\mu_E}+q\frac{1}{\mu_D}=\frac{1+q}{2}(=\frac{3}{4}).
\end{aligned}$$

The day before the final exam, two GSIs and the lecturer hold their office hours at the same time. The two GSIs share a room and the lecturer uses a different room. Both GSIs in the common room are busy helping one student each: both GSIs have been answering questions for T minutes, independently. There is no other student waiting in this room. The lecturer (in the other room) takes a random amount of time to answer a question, with this time duration being exponentially distributed with rate $\mu' = 6$ (questions per minute) regardless of the problem's difficulty. In the lecturer's room, there are two students including the one who is being helped.

(d) You want to minimize the expected waiting time by choosing the right room between the two. Which room should you join? (*Hint*: your answer should depend on the value of T.)

Solution: Let X_1 and X_2 be the amount of time that the two GSIs still need to take to answer their questions. The amount time to wait for the GSIs is $\min\{X_1, X_2\}$. Let X_3 be the amount of time that the lecturer need to take to finish the two students' questions. Let a be the probability that a particular GSI is answering a difficult question, i.e.,

$$a = \frac{0.2e^{-T}}{0.2e^{-T} + 0.8e^{-2T}} = \frac{1}{1+4Y},$$

where $Y = e^{-T}$. Thus,

$$E[\min\{X_1, X_2\}] = \frac{a^2}{2\mu_D} + \frac{2a(1-a)}{\mu_D + \mu_E} + \frac{(1-a)^2}{2\mu_E} = \frac{6a^2 + 8a(1-a) + 3(1-a)^2}{12},$$
$$E[X_3] = \frac{2}{\mu'} = \frac{1}{3}.$$

By equating two equations,

$$6a^2 + 8a(1-a) + 3(1-a)^2 = 4 \Rightarrow a = \sqrt{2} - 1 \Rightarrow Y = \frac{\sqrt{2}}{4}.$$

Thus, if $T > \ln(2\sqrt{2})$, you have to choose the GSI's room, and choose the lecturer's room if $T \leq \ln(2\sqrt{2})$.

Problem 4. (25 points) There are N $(N \geq 2)$ bins and two colors of balls, black and white. In each experiment, d $(1 \leq d < N)$ bins are chosen uniformly at random from the N bins. In each chosen bin, we place either a black ball or a white ball in that bin. The probability that we place a black ball is p, and the probability that we place a white ball is $1-p$. All placements are independent. This experiment is conducted K times independently. Let B_i and W_i be the number of black balls and white balls in the ith bin after the K experiments, respectively. An example of one experiment is shown in Figure 5.

To clarify, N and d are constants. In part (a), (b), and (c), K is a constant. In part (d), K is a random variable.

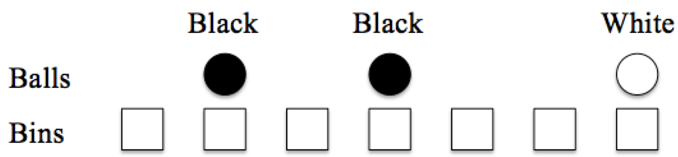

Figure 5: An example of one experiment with $N = 7$ bins and $d = 3$ balls.

(a) For the ith bin, what is the probability mass function (PMF) of B_i?

Solution: Since there are K independent experiments and in each experiment, the probability that a black ball is placed in the ith bin is $\frac{dp}{N}$. Then we know that $B_i \sim \text{Binomial}(K, \frac{dp}{N})$, and

$$P(B_i = x) = \binom{K}{x}(\frac{dp}{N})^x(1-\frac{dp}{N})^{K-x}.$$

(b) For the ith bin, what is the joint PMF of B_i and W_i, i.e., $P(B_i = x, W_i = y)$? Are B_i and W_i independent?

Solution: We have

$$\begin{aligned} P(B_i = x, W_i = y) &= P(B_i = x, B_i + W_i = x + y) \\ &= P(B_i + W_i = x + y)P(B_i = x | B_i + W_i = x + y). \end{aligned}$$

Since $B_i + W_i$ is the total number of balls in a bin, we have $B_i + W_i \sim \text{Binomial}(K, \frac{d}{N})$. Conditioned on $B_i + W_i = x + y$, the distribution of B_i is also binomial, with parameters $x + y$ and p. Therefore we have

$$\begin{aligned} P(B_i = x, W_i = y) &= \binom{K}{x+y}(\frac{d}{N})^{x+y}(1-\frac{d}{N})^{K-(x+y)}\binom{x+y}{x}p^x(1-p)^y \\ &= \frac{K!}{(K-x-y)!x!y!}(\frac{dp}{N})^x(\frac{d(1-p)}{N})^y(1-\frac{d}{N})^{K-(x+y)}. \end{aligned}$$

B_i and W_i are not independent. According to part (a), $B_i \sim \text{Binomial}(K, \frac{dp}{N})$, and similarly, $W_i \sim \text{Binomial}(K, \frac{d(1-p)}{N})$. Then we can see that

$$P(B_i = x, W_i = y) \neq P(B_i = x)P(W_i = y).$$

(c) We call a bin a *mixed bin* if the bin contains both black and white balls. Suppose after K experiments, we find that the first bin is empty. Conditioned on this, what is the probability that the second bin is a mixed bin?

Solution: Let E_1 be the event that the first bin is empty, and M_2 be the event that the second bin is a mixed bin. We want to find $P(M_2|E_1)$. We have

$$\begin{aligned}
P(M_2|E_1) &= P(B_2 \geq 1, W_2 \geq 1|E_1) \\
&= 1 - P(B_2 = 0 \text{ or } W_2 = 0|E_1) \\
&= 1 - [P(B_2 = 0|E_1) + P(W_2 = 0|E_1) - P(B_2 = 0, W_2 = 0|E_1)].
\end{aligned}$$

Conditioned on the event that the first bin is empty, we know that in each experiment, the d balls are chosen uniformly at random from all the bins except the first bin. Therefore, we have

$$\begin{aligned}
P(B_2 = 0|E_1) &= (1 - \frac{dp}{N-1})^K \\
P(W_2 = 0|E_1) &= (1 - \frac{d(1-p)}{N-1})^K \\
P(B_2 = 0, W_2 = 0|E_1) &= (1 - \frac{d}{N-1})^K
\end{aligned}$$

Therefore, we know that

$$P(M_2|E_1) = 1 - (1 - \frac{dp}{N-1})^K - (1 - \frac{d(1-p)}{N-1})^K + (1 - \frac{d}{N-1})^K.$$

(d) Now we consider a special case of our experiments where $d = 1$. This means that in each experiment we choose a single bin uniformly at random and put either a black ball in that bin with probability p, or a white ball with probability $1 - p$. We keep repeating the experiments until all the bins are mixed bins. We want to find the expected number of experiments that we need to do, i.e., $E[K]$. But here, we ask for something simpler.

In order to find $E[K]$, we define $T_{n,m}$ as the expected number of remaining experiments that we still need to do when the number of bins without black balls is n and the number of bins without white balls is m. Consequently, $E[K] = T_{N,N}$. Find a recursive equation that describe the relationship between $T_{n,m}$, $T_{n-1,m}$, and $T_{n,m-1}$ $(n \geq 2, m \geq 2)$.

Solution: When the number of bins without black balls is n and the number of bins without white balls is m, we do one experiment, and then, three events can happen:

A: one bin among the n bins which do not contain black balls gets a black ball;
B: one bin among the m bins which do not contain white balls gets a white ball;
C: other cases.

We can see that the three events happen with probabilities $\frac{np}{N}$, $\frac{m(1-p)}{N}$, and $1 - \frac{np}{N} - \frac{m(1-p)}{N}$, respectively. Therefore, we know that

$$T_{n,m} = 1 + \frac{np}{N}T_{n-1,m} + \frac{m(1-p)}{N}T_{n,m-1} + (1 - \frac{np}{N} - \frac{m(1-p)}{N})T_{n,m}.$$

END OF THE EXAM.

Please check whether you have written your name and SID on every page.