EECS 126 Probability and Random Processes University of California, Berkeley: Spring 2017
Kannan Ramchandran February 13, 2017

Midterm Exam 1 (**Solutions**)

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.
- Note that the test has **110** points. but a score ≥ 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.
- **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 electronic 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		15	
	(b)	10		3		25	
	(c)	8		4		25	
	(d)	9					
	(e)	10					
		45					
Total						110	

Cheat sheet

1. Discrete Random Variables

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

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

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

2. Continuous Random Variables

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

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

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

Problem 1. (a) (8 points total, 2 points each. **You must provide brief explanations to justify your answers to get credit on all parts.**)

(i) In words, explain what it means for a distribution to be memoryless.

Solution: Answers may vary.

(ii) **True/False** Let G_1 and G_2 be independent Erdös-Renyi random graphs on n vertices with probabilities p_1 and p_2, respectively. Let $G = G_1 \cup G_2$, that is, G is generated by combining the edges from G_1 and G_2. Then, G is an Erdös-Renyi random graph on n vertices with probability $p_1 + p_2$.

Solution: False. An edge appears in G if it appears in G_1 or G_2, which occurs with probability $p_1 + p_2 - p_1 p_2$ by the Inclusion-Exclusion Principle.

(iii) **True/False** Consider independent events A and B. Then for any event C, $P(A, B|C) = P(A|C)P(B|C)$

Solution: False, suppose A and B are independent fair coin flips (associating 0 to heads and 1 to tails), and C is the event that the total number of heads is 1.

(iv) **True/False** Suppose that you are involved in a first price auction with one other bidder and that both you and the other draw valuations uniformly on the interval $(0, 1)$. If your valuation is x and the other bidder bids her valuation, then your optimal bid is $\frac{x}{2}$.

Solution: True.

(b) (10 points) Consider a random bipartite graph $\mathcal{G}$ in which there are K left nodes and M right nodes. The random graph is generated according to the following rule: for each left node i, choose a right node j uniformly at random and draw the edge (i, j) as well as the edge $(i, j+1)$ (if $j = M$, then the two edges will be (i, M) and $(i, 1)$). Define a singleton to be a right node with degree 1. Give an exact answer for both parts (no approximations).

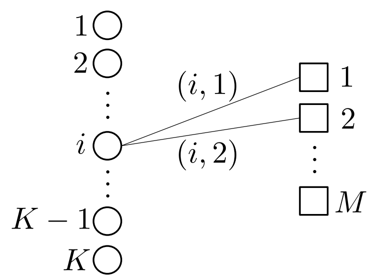

Figure 1: Bipartite graph with K left nodes and M right nodes.

(i.) (4 points) What is the expected number of singletons in $\mathcal{G}$?

Solution: The distribution of the degree of the right node is $\mathrm{Bin}(K, \frac{2}{M})$. Let $p = \frac{2}{M}$. By linearity of expectation, the expected number of singletons is:

$$M\binom{K}{1}p(1-p)^{K-1}$$

(ii.) (6 points) What is the expected number of left nodes connected to at least one singleton in $\mathcal{G}$?

Solution: WLOG, consider the first left node and suppose it was thrown into the first two right nodes. Let A_i, A_{i+1} be the events that the ith and $i+1$th right nodes are singletons given that they contain the same left node. We are interested in $P(A_1 \cup A_2)$:

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1, A_2) \\ &= 2(1 - \frac{2}{M})^{K-1} - (1 - \frac{3}{M})^{K-1} \end{aligned}$$

Thus, the expected number of left nodes connected to at least one singleton is:

$$K[2(1 - \frac{2}{M})^{K-1} - (1 - \frac{3}{M})^{K-1}]$$

(c) (8 points) Consider independent random variables X and Y, each of which is uniform on the interval $(0, 1)$.

(i.) (4 points) Are $X + Y$ and $X - Y$ uncorrelated?

Solution: Yes: $E[(X + Y)(X - Y)] = 0$ and $E[X - Y] = 0$.

(ii.) (4 points) Are $X + Y$ and $X - Y$ independent?

Solution: No. Consider $P(X + Y < \frac{1}{2}, X - Y < \frac{1}{2}) = 0$. However, $P(X + Y < \frac{1}{2}) > 0$ and $P(X - Y < \frac{1}{2}) > 0$. Thus they are not independent.

(d) (9 points) Consider a graph with n vertices. For each pair of nodes in the graph, draw an edge between them with probability $\frac{1}{2}$, independently of all other edges. Suppose that X is the number of isolated nodes in this graph. Find $\mathrm{Var}(X)$ (give an exact answer).

Solution: Let X_i be the indicator random variable which takes the value 1 if node i is isolated. Then, we have:

$$\begin{aligned}\mathrm{Var}(X) &= n\mathrm{Var}(X_1) + n(n-1)\mathrm{cov}(X_1, X_2)\\ &= n(\frac{1}{2})^{n-1}(1-(\frac{1}{2})^{n-1}) + n(n-1)(\frac{1}{2})^{2n-2}\end{aligned}$$

(e) (10 points) Consider the joint PDF of X, Y shown in Figure .

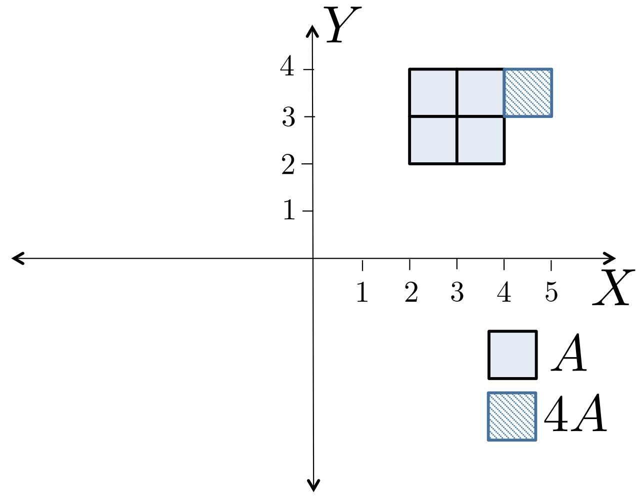

Figure 2: Joint PDF of X, Y.

(i.) (4 points) Find A.

Solution: $A = \frac{1}{8}$

(ii.) (6 points) Find $\text{cov}(X, Y)$

Solution: Note that the covariance is invariant to arbitrary shifts by a constant in either dimension. Thus, we shift the figure so that the 2×2 square is centered at the origin: Now, letting E_1 be the event that a random point is in the 2×2 square and E_2 that a

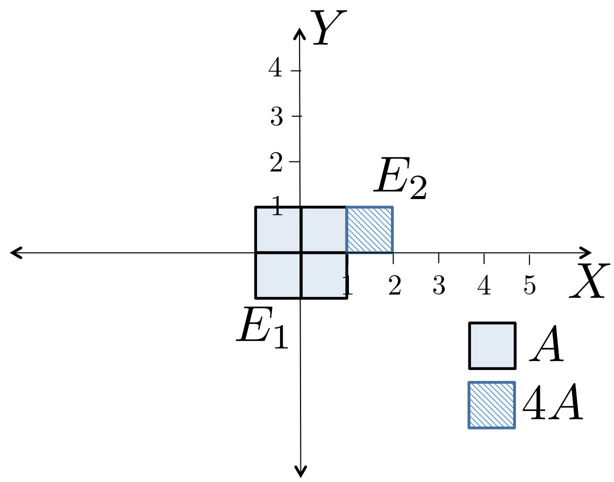

Figure 3: Joint PDF of X, Y.

random point is in the shaded square. We have by the law of iterated expectation that $E[XY] = E[XY|E_1]P(E_1) + E[XY|E_2]P(E_2)$. Thus, $E[XY] = \frac{3}{8}$. Similarly, we see that $E[X] = \frac{3}{4}$ and $E[Y] = \frac{1}{4}$. Thus, $\text{cov}(X, Y) = \frac{3}{16}$

Problem 2. (15 points) You are playing a card game with your friend: you take turns picking a card from a deck (you may assume that you never run out of cards, i.e. that the deck is infinite). If you draw one of the special "bullet" cards, then you lose the game. Unfortunately, you do not know the contents of the deck. Your friend claims that $1/3$ of the deck is filled with "bullet" cards. You don't trust your friend fully, however: he is lying with probability $1/4$. You assume that if your friend is lying, then the opposite is true: $2/3$ of the deck is filled with "bullet" cards.

(a.) (7 points) Suppose that you draw first. Let p denote the probability that you draw a bullet card. Find p.

Solution: Let L be the event that your friend is lying, and let B be the event that the randomly selected card is a "bullet" card. Using the Law of Total Probability,

$$\Pr(B) = \Pr(B \mid L)\Pr(L) + \Pr(B \mid \bar{L})\Pr(\bar{L}) = \frac{2}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{3}{4} = \frac{5}{12}.$$

(b.) (8 points) You win if your opponent draws the bullet card. Supposing that p is the true fraction of cards which are bullet cards, compute the probability of winning if you draw first.

Solution: Since each card drawn has probability p of being a "bullet" card, and the game ends when the first "bullet" card is drawn, the number of turns before the game ends, X, is a geometric random variable with probability p. The probability that you win is the probability that X is even, so we have

$$\Pr(X \text{ is even}) = \sum_{\substack{k=1 \\ k \text{ is even}}}^{\infty} p(1-p)^{k-1} = p(1-p)\sum_{j=0}^{\infty}(1-p)^{2j} = \frac{p(1-p)}{1-(1-p)^2}$$
$$= \frac{p(1-p)}{2p-p^2} = \frac{1-p}{2-p}.$$

Plugging in $p = 5/12$, we see that the probability of winning is $7/19$.

Problem 3. (25 points, 5 points each) The Donald is holding a press conference in which he answers n questions. With probability p, he answers with an alternative fact and with probability $1-p$, he asks if he can phone a friend. Let K be the random variable denoting the number of questions he answers with alternative facts.

(a.) What is the expected number of questions asked before The Donald answers with an alternative fact? (You may assume $n \to \infty$ for this part)

Solution: Note that this is a shifted geometric random variable and has expectation $\frac{1}{p} - 1$.

(b.) Suppose that $K = m$, find the probability that the first m questions were answered with alternative facts.

Solution: By symmetry, $\frac{1}{\binom{n}{m}}$.

Let X_i be the ith question to which he answers with an alternative fact. For example, if he is asked 4 questions and answers the second and fourth questions with alternative facts, $X_1 = 2$ and $X_2 = 4$.

(c.) Find $E[X_1 + X_2 + \cdots + X_m | K = m]$.

Solution: Let $(T_1, T_2, \ldots, T_m)$ denote m numbers drawn from $\{1, 2, \ldots, n\}$ without replacement. Note that when $K = m$, $\sum_{i=1}^{m} X_i$ and $\sum_{i=1}^{m} T_i$ have the same probability mass functions. Therefore, $\mathbb{E}[\sum_{i=1}^{m} X_i | K = m] = \mathbb{E}[\sum_{i=1}^{m} T_i]$. Since

$$\mathbb{E}[T_1] = \frac{1}{n} \sum_{i=1}^{n} i = \frac{n+1}{2},$$

we have

$$\mathbb{E}[\sum_{i=1}^{m} T_i] = \sum_{i=1}^{m} \mathbb{E}[T_i] = m\frac{n+1}{2} = \mathbb{E}[\sum_{i=1}^{m} X_i | K = m].$$

Suppose that The Donald holds m press conferences, each of which has n questions, and answers exactly 1 question with an alternative fact in each press conference. Let Y_i be the question The Donald answers with an alternative fact during the ith press conference. Consider the array $[Y_1, Y_2, \ldots Y_m]$ and move all of the Y_i with value 1 to the back of the array. For example, the array

$$[6, 4, 1, 1, 5, 3, 1, 5, 2, 3]$$

becomes

$$[6, 4, 5, 3, 5, 2, 3, 1, 1, 1]$$

Suppose Y_i has value $\neq 1$. After the array has been altered, Y_i has been moved to index X.

(d.) Find $E[X]$.

Solution: First note that Y_i are independent samples drawn uniformly from the set $\{1, 2, \ldots, n\}$. Let X_j be the indicator that the jth entry of the original array is 1, for $j \in \{1, \ldots, i-1\}$. Then, the ith entry is moved backwards $\sum_{j=1}^{i-1} X_j$, positions, so

$$E[X] = i - \sum_{j=1}^{i-1} E[X_j] = i - \frac{i-1}{n} = \frac{(n-1)i+1}{n}.$$

(e.) Find $\mathrm{Var}(X)$.

Solution: The variance is also easy to compute, since the X_j are independent. Then, $\mathrm{var}(X_j) = (1/n)((n-1)/n) = \frac{n-1}{n^2}$, so

$$\mathrm{var}(X) = \mathrm{var}\left(i - \sum_{j=1}^{i-1} X_j\right) = \sum_{j=1}^{i-1} \mathrm{var}(X_j) = \frac{(n-1)(i-1)}{n^2}.$$

Problem 4. (25 points, 5 points each) You would like to compute a job and have four machines at your disposal. Three of the machines complete the job according to an exponential distribution with rate $\lambda_s = 1$ and one machine completes the job according to an exponential distribution with rate $\lambda_f = 2$, but you do not know which machine has what rate. (Note: an exponential distribution with rate λ has PDF $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$)

(a.) Suppose you run your job on two machines, one of which happens to be the fast machine. What is the probability, p_a that the fast machine finishes first?

Solution: Let X_1 and X_2 denote the time it takes the machines to finish their service, where X_1 is the fast machine and X_2 is a slow machine.

$$\begin{aligned}\Pr(X_2 > X_1) &= \int_0^\infty P(X_2 > X_1 | X_1 = t) f_{X_1}(t) dt \\ &= \int_0^\infty e^{-t} 2e^{-2t} dt \\ &= \frac{2}{3}\end{aligned}$$

(b.) Suppose you are in the same setting as part (a); i.e. you send your job to two machines, one of which is fast. Let the service time of the fast server be T_f and that of the slow server be T_s. You observe that the first of the two servers finishes at time $Z = 1$. What is the expected total amount of time for the second server to finish? (you may leave your answer in terms of p_a and Z)

Solution: The probability that the faster server finishes first is $\frac{2}{3}$ and the probability that the slower server finishes first is $\frac{1}{3}$. Thus, the expected amount of time for the second server to finish is $1 + \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} = 1 + \frac{5}{6} = \frac{11}{6}$.

(c.) The administrator decides to let you choose two machines at random and send your job to both machines. Your job is finished when either of the machines is done computing. What is the expected amount of time to finish computing your job?

Solution: The probability that the two selected jobs were both slower machines is $\frac{1}{2}$ and the probability that one was fast and the other was slow is $\frac{1}{2}$. Thus, the expected amount of time is $\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{12}$.

(d.) Feeling generous, the administrator now allows you to reproduce your job on all four machines, but with a catch. The result of the first machine that finishes is kept by the administrator so that your job is complete when the second machine finishes. What is the expected amount of time until your job is complete?

Solution: We are looking for the second minimum amongst all machines. The expected amount of time for the first machine is $\frac{1}{5}$. Then, the probability that the fast machine finished first is $\frac{2}{5}$. Thus, the expected amount of time for the second machine to finish is $\frac{1}{5} + \frac{2}{5} \cdot \frac{1}{3} + \frac{3}{5} \cdot \frac{1}{4} = \frac{29}{60}$.

(e.) The administrator would like you to run the job on only one machine, but allows you to first send four test jobs to all four machines. On all four test jobs, machine 1 finishes first. You send your job to machine 1 and observe that your job has taken T seconds and is not complete. What is the expected amount of time left until your job is finished?

Solution: Let X denote the service time of the first machine and let p be the probability that the first machine is the fast machine. Additionally let F denote the event that the first machine is the fast machine and A denote the event that the first machine finished the four test jobs fastest. We have:

$$\begin{aligned} p &= P(F|X > T, A) \\ &= \frac{P(A, X > T|F)P(F)}{P(A, X > T|F^c)P(F^c)} \\ &= \frac{(\frac{2}{5})^4 e^{-2T} \cdot \frac{1}{4}}{(\frac{2}{5})^4 e^{-2T} \cdot \frac{1}{4} + (\frac{1}{5})^4 e^{-T} \cdot \frac{3}{4}} \end{aligned}$$

Thus, the expected amount of time left is:

$$E[X - T|X > T, A] = p\frac{1}{2} + (1 - p)$$

END OF THE EXAM.

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