Probability and Statistics Chapter 2 Questions

Sanjit K. Kaul

I. Textbook Problems

End of Chapter Problems 2.2.6, 2.2.7, 2.3.7, 2.3.11, 2.3.13, 2.4.1, 2.5.1, 2.5.2, ..., 2.5.9, 2.6.1, 2.6.5, 2.6.6, 2.7.5, 2.7.6, 2.7.9, 2.8.9, 2.9.6, 2.9.7 from the textbook by Roy Yates and David Goodman (2nd Ed).

Question 1.

1) Give an example of a range space of a Bernoulli RV.
2) Write down the PMF of a Bernoulli RV with parameter p.
3) Write down the CDF of a Bernoulli RV with parameter p.

Question 2. You have a coin. However, you don't know its model. While the coin is two faced like any other coin, the probability P with which tossing it gives heads is unknown. You know that P must take a value in the set $\{0.1, 0.3, 0.5, 0.9\}$. You start with the belief that P takes the values with equal probability (the prior PMF of P).

You perform an experiment in which you toss the coin five times and get three heads and two tails. How does the outcome of the experiment change your belief (updated PMF) about P?

Question 3. There are N sensors that measure energy consumption in a building. They send their measurements to a sink node for further processing. Sensors take turns in a round robin manner (as per the sequence, sensors 1, 2,...,N, which repeats) to send their measurements, as messages, to the sink.

A sensor may take one or more turns to send a message. A turn has a max of k attempts. During a turn, a sensor attempts to send a message till the message is sent successfully or till the sensor exhausts the k attempts. An attempt is successful with probability p independently of other attempts.

Let L be the number of turns that a sensor takes to successfully send a message.

1) What is the PMF of L?
2) What is the PMF of the number of messages transmitted successfully by a sensor in m turns?
3) Consider the turn during which a sensor's message is sent successfully. What is the PMF of the number of attempts made by the sensor during such a turn?

Question 4. A lecture is 60 minutes long. You arrive at the beginning of the lecture. The process of recording attendance is completed by a certain time selected uniformly and randomly during the lecture. You leave soon after attendance is recorded. Calculate the expected value of the time you spend attending the lecture?

Question 5. You attend a 60 minute lecture. During every minute of the lecture your attention is demanded by the lecture slide, your WhatsApp application, and your Facebook application. You always pay attention to the lecture slide during the first and the last minute. For every other minute, you pay attention to the lecture slide with probability 0.4 and to the WhatsApp application with probability 0.5. Calculate the expected number of minutes during which you pay attention to the lecture slides?

Question 6. Number of assignments in a probability and statistics course offered over the duration of 1 semester are often well modeled by the Poisson random variable of rate 20 per semester. Suppose half the assignments are 40 minutes long and the rest are 20 minutes long. Calculate the expected value of the time students spend on assignments? [Hint: The PMF of the Poisson RV K is $P[K = k] = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$, where λ is the rate and T is the duration of observation.]

Question 7. A bus may arrive with equal probability at the start of one of the 60 minutes (indexed $\{1, 2, \ldots, 60\}$) in an hour. You catch the bus in case you arrive at the same minute as the bus or you arrive earlier than the bus.

You arrive at the bus stop at minute 4. Calculate the probability that you miss the bus (arrived after the bus). Calculate the average value of the minute at the start of which the bus arrives.

After you arrive at minute 4, you are told that you have missed the bus. What is the average number of minutes by which you missed the bus?

Question 8. You have 2 movie tickets to sell within 4 minutes. During the first 2 minutes, you decide to price your tickets at ₹ 2. In the last 2 minutes you plan to sell any remaining tickets at ₹ 1. The number of buyers that approach you to buy tickets is a Poisson distributed random variable with rate $1/2$ per minute. Any buyer buys exactly one ticket. Buyers arrive independently of each other. What is the expected value of the money you get as a result of your ticket sale? [Hint: The PMF of a Poisson RV K is $P[K = k] = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$, $k = 0, 1, \ldots$, and $P[K = k] = 0$ otherwise.]

Question 9. You are given the task of testing a wireless link between Delhi and Hyderabad. You do so by sending many bits (each bit is either 0 or 1) over the link from Delhi to Hyderabad. A receiver at Hyderabad records, for every sent bit, its estimate (0 or 1) of what was sent. At the end of the experiment you come up with the following data. You find that 90% of the bits sent are received correctly. You also find that amongst the correctly received bits, 70% of the bits were sent as 1. Also, 30% of the wrongly received bits were sent as 1.

The cost of sending a bit that is correctly received is -100 and that of sending a bit that is wrongly received is 50. What is the expected cost of sending a bit over the link?

Question 10. A student gives three exams (one after the other) to pass a certification course on Intelligently Dumb Systems. In each exam, the student scores either a 0 or a 1. The probability that the student scores a 1 in the first exam is 0.6. The performance of a student in an exam is impacted by the student's performance in the previous exam. Specifically, if the student scores a 1 in the first exam, the student scores a 1 in the second exam with probability 0.8. Similarly, if a student scores a 1 in the second exam, the student scores a 1 in the third exam with probability 0.8. If the student scores a 0 in the first exam, the student scores a 1 in the second exam with probability 0.3. Similarly, if the student scores a 0 in the second exam, the student scores a 1 in the third exam with probability 0.3. Answer the following questions. [Hint: You may want to draw a tree diagram.]

(A) Calculate the probability that the total score at the end of the three exams is 3.
(B) Calculate the expected value of the score in the first exam.
(C) Calculate the expected value of the score in the second exam.
(D) Calculate the expected value of the total score at the end of the three exams.
(E) Calculate the expected value of the score in the first exam, given that score in the second exam is 1.
(F) Calculate the expected value of the total score, given that score in the second exam is 1.

Question 11. An autonomous shuttle waits for passengers in five minute intervals. In case, at least 10 passengers arrive, it sets out on its route at the end of the corresponding five minute interval. To exemplify, once the shuttle starts waiting it must wait for at least 5 minutes. If less than 10 passengers arrive in the 5 minutes, it waits for another 5 minutes, and so on, it waits till the end of that five minute interval during which the total number of passengers equals or exceeds 10. Note that the shuttle leaves after waiting for a certain random X minutes, where the range space $S_X = \{5, 10, 15, \ldots\}$. Assume that the passengers that arrive don't leave the shuttle.

The number of passengers that arrive in an interval of T minutes is a Poisson random variable with rate $\lambda = 0.5$ passengers per minute. The number of passengers that arrive in an interval is independent of the number of passengers that arrive in any other disjoint interval of time. To exemplify, the intervals of time $[2, 3]$ and $[3, 4]$ are disjoint. However, $[2, 5]$ and $[3, 6]$ are not. Answer the following questions. [Hint: The PMF of a Poisson RV K with parameter $\alpha = \lambda T$ is $P[K = k] = e^{-\alpha}\alpha^k/k!$, $k \in \{0, 1, 2, \ldots\}$.]

(A) Suppose that 5 customers arrive in the first 10 minutes. Calculate the probability that 1 customer arrived in the first 5 minutes? [Hint: Make judicious use of the property of independence stated in the question.]

(B) Calculate the probability that the shuttle leaves after waiting for 20 minutes.

(C) Calculate the expected value of the customers in the shuttle given that the shuttle departs after 20 minutes.

Question 12. A game of TT has two players that occupy opposite ends of a table. A TT game may be thought of as multiple rallies between the players. A rally is a sequence of alternate hits by the players that ends when a player's hit ends in error. We will call an erroneous hit a *miss*. As an example, a rally may consist of a *hit* by player 1, followed by a hit by player 2, and so on, and may end with a miss by player 2. Suppose that player 1 hits with probability 0.6 and misses with probability 0.4. Player 2 on the other hand hits with probability 0.4. Answer the following questions. Assume a rally always starts with player 1.

1) Calculate the probability that a rally has less than three hits?
2) Calculate the probability that a rally ends with a miss by player 1. If you will, leave your answer as a fraction.
3) Calculate the expectation of the length of a rally, where the length includes all hits and the final miss.
4) Calculate the expectation of the number of rallies, amongst 10 rallies, that end due to a miss by player 2?

Question 13. A node in a wireless network experiences one among three possibilities in any wireless timeslot. In any slot, the node either idles (outcome I) or transmits successfully (T) or transmits unsuccessfully U, independently of outcomes in any other slot. Probability that a node idles in a slot is p, that it transmits successfully is q, and that it transmits unsuccessfully is r. Answer the following questions.

(a) What is the probability that the node transmits successfully in the first slot, has an unsuccessful transmission in the second slot, and a successful transmission in the third?

(b) You are interested in characterizing the random number of slots up to and including the slot in which the node transmits successfully for the first time. Derive the distribution (PMF) of this number.

(c) Derive the probability that the first successful transmission slot occurs earlier than the first unsuccessful transmission slot. [Hint: Think of a suitable event space.]

(d) Suppose that the first successful transmission slot is observed by the node before the first unsuccessful transmission slot. Derive the PMF of the random number of slots including and up to the first successful transmission. [Hint: The result you derived in (c) may be useful].

(e) As in (d), suppose that the first successful transmission slot is observed by the node before the first unsuccessful transmission slot. Derive the expected value of the number of slots up to and including the first *unsuccessful* transmission slot? [Hint: Think sum of expected values.]

Question 14. A professor visits his lab on any day with probability 0.5. A student working with the professor visits the lab on any day with probability 0.8. Let N_A be the number of days the professor visits the lab over a week (7 days). Let N_S be the number of days the student visits the lab over a week. Further let N be the number of days in a week that they both meet. Assume that they both meet on days both visit the lab.

(a) Derive the joint probability mass function of N_A and N.
(b) Using the definition of independent random variables show whether N_A and N are independent or not.
(c) Calculate the expectation $E[N_A|N = n]$ and use this to further calculate $E[N_A]$.
(d) Suppose we are interested in those weeks in which the professor visited the lab more number of days than the student. Derive the conditional expectation of N_A. [Hint: First write down the given event in terms of the described random variables. Calculate the appropriate conditional marginal PMF.]

Incorrect Solutions A professor who teaches probability and statistics is known to incorrectly solve questions he creates. The questions he creates can be broadly categorized into easy and difficult. He *incorrectly* solves an easy question with probability $1/10$ and a difficult question with probability $1/5$. His solving a question incorrectly is independent of how he solves other questions. The professor has created a question bank consisting of 10 difficult and 20 easy questions, together with their solutions.

Question 15. Suppose you pick a question randomly from the question bank. Derive the probability that your chosen question will be incorrectly solved in the question bank. Show your steps. **In the Google form, mention the events you considered to calculate the probability.**

Exam Using the Question Bank You are supposed to create an exam consisting of 5 questions. To reduce your effort, you decide to randomly choose 5 questions from the bank created by the professor we got introduced to in an earlier paragraph. Questions are chosen randomly from the bank independently of the other questions chosen from the bank. Two exams are considered to be the same if they have the same set of 5 questions.

Question 16. Answer the following questions.

(a) Your are picking questions independently of each other. Using the definition of independence, show that therefore you can not pick questions from the bank without replacement.
(b) Derive the probability that in such an exam the first three questions will be easy and the remaining two will be difficult.
(c) Derive from first principles the PMF of the random number D that counts the number of difficult questions that such an exam will have.
(d) Suppose a student solves questions in the order provided in the exam. Derive the PMF of the random variable that counts the number of easy questions that the student solves before attempting (if at all) the first difficult question.
(e) Derive from first principles the PMF of the random variable that counts the number of questions in the exam that have incorrect solutions in the question bank.
(f) Consider exams that have two questions with incorrect solutions. Calculate the probability that such exams have K difficult questions, for $K = 0, 1, 2, 3, 4, 5$.
(g) In continuation of the above, for exams that have two questions with incorrect solutions, calculate the expected value of K.

Question 17. Suppose over time, with contributions of many others, the professor's question bank becomes countably infinite. However, the fraction of easy and difficult questions remains as before. The fraction of solutions that are incorrect also remains the same as before. You have chosen to volunteer your time to correct three incorrect solutions and you are given access to the question bank. Not knowing which questions have incorrect solutions, you begin to solve questions one after another. No question is repeated. Answer the following.

(a) Derive, by correctly identifying the relevant events, the PMF of the number of questions that you must solve to satisfy your goal.

(b) Suppose the questions are sequenced in a manner such that the first two questions with incorrect solutions have 10 questions with correct solutions in between. Derive the expected value of the number of questions that you must solve. [Hint: The number is a sum of two random variables and a constant. Identify these and use the properties of the expectation operator.]

II. Other Questions

Question 1. There are a total of l lectures that are to be attended by a class of k students. The instructor is working on an attendance strategy. He decides that, given the length of a lecture, he will only roll call exactly $j < k$ students. Every lecture the instructor will choose j students randomly to call from the list of k students. Answer the following questions. Show all your steps.

(a) Assume that a certain student attends any lecture, independently of the other, with probability q. What is the probability p that the student's name is called during a lecture?

(b) Let X_i be the random variable that governs whether the student's name is called during lecture i, where $i \in \{1, 2, \ldots, l\}$. $X_i = 1$ in case the student's name is called during lecture i. Else $X_i = 0$. Let $M = \sum_{i=1}^{l} X_i$ be the random number of lectures during which the student's name is called. Derive the PMF of X_i and that of M.

(c) Let Y_i be the random variable that governs whether the student attends lecture i. $Y_i = 1$ in case the student attends lecture i. Else $Y_i = 0$. Let $N = \sum_{i=1}^{l} Y_i$. Derive the distribution of the random variable N.

(d) Consider the random variable $Z = \sum_{i=1}^{l} X_i Y_i$. Derive its distribution.

(e) Derive the conditional distribution of the random variable N, given the event $Z = z$.

(f) Use the above derived conditional distribution to derive the corresponding conditional expectation.

Next we will derive the conditional expectation $E[N|Z = z]$ of N, given the event $Z = z$, in an alternate manner. To do so, answer the following questions.

(aa) Consider the claim $E[N|Z = z] \geq z$. Is the claim correct? Explain your answer.

(bb) The z lectures are those that were attended by the student and during which the student's name was also called. Consider the other $l - z$ lectures, given that $Z = z$. For any of the other $l - z$ lectures, derive the probability $P[Y_i = 1|Z_i = 0]$.

(cc) Use the expressions derived in the above two parts to derive $E[N|Z = z]$. [Hint: What random variable models each of the $l - z$ lectures, given $Z = z$?]

(dd) Derive $E[N]$ starting with your expression for $E[N|Z = z]$.

Question 2. The state government has come up with a scheme to encourage use of buses for travel. It creates a large bin consisting of n identities, where n is the total number of people in the state. Every day, the bin containing n identities is mixed thoroughly and m identities are drawn from it without replacement. The chosen m people are given a free bus pass for the day.

(a) Derive the probability that any person in the city is chosen for a free pass for the day.

(b) Derive the distribution (PMF) of the random number of days a person must wait to get his first free pass.

(c) Derive the distribution of the random number of days a person must wait to get his k^{th}, $k > 1$, free pass.

(d) Derive the distribution of the number of free passes a person obtains over a month of 30 days.

Suppose we monitor allocations of free passes to two randomly chosen people in the state. Let X_i and Y_i, respectively, count the number of free passes received by the two individuals on day i of a 30 day month. We have $S_{X_i} = S_{Y_i} = \{0, 1\}$. Also $i \in \{1, 2, \ldots, 30\}$. Let $X = \sum_{i=1}^{30} X_i$ and $Y = \sum_{i=1}^{30} Y_i$. Define $Z = \sum_{i=1}^{30} X_i Y_i$ and let $Z_i = X_i Y_i$, for any i.

(aa) Derive $P[Z_i = 1]$, for any i. Keep calm and work out the probability from first principles.

(bb) Derive the conditional PMF of X given $Z = 5$.

(cc) Derive the conditional PMF of X given $Z = 5$ and $Y = 10$.

(dd) Derive the conditional PMF of X_{15} given $Z = 5$.

(ee) Derive $E[X_{15}|Z_{15} = 1]$.

(ff) Derive $E[X_{15}|Z_{15} = 0]$.