Probability & Statistics for EECS: Homework #01

Due on 2023-10-15 at 23:59

Name:
Student ID:

Problem 1

(Story Proof) Define $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ as the number of ways to partition $\{1, 2, \dots, n\}$ into k non-empty subsets, or the number of ways to have n students split up into k groups such that each group has at least one student. For example, $\left\{ \begin{matrix} 4 \\ 2 \end{matrix} \right\} = 7$ because we have the following possibilities:

•$\{1\}, \{2, 3, 4\}$	•$\{1, 2\}, \{3, 4\}$
•$\{2\}, \{1, 3, 4\}$	•$\{1, 3\}, \{2, 4\}$
•$\{3\}, \{1, 2, 4\}$	•$\{1, 4\}, \{2, 3\}$
•$\{4\}, \{1, 2, 3\}$	

Prove the following identities:
(a)

$$\left\{ \begin{matrix} n+1 \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\} + k \left\{ \begin{matrix} n \\ k \end{matrix} \right\}.$$

Hint: I'm either in a group by myself or I'm not.
(b)

$$\sum_{j=k}^{n} \binom{n}{j} \left\{ \begin{matrix} j \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n+1 \\ k+1 \end{matrix} \right\}.$$

Hint: First decide how many people are not going to be in my group.

Solution:
(a) When n turns to $n+1$, there may be two possible case

1, the new item becomes a new subset which is independent to the subset of $\left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\}$, then in this case we have $\left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\}$ because the new item is fixed in one subset.

2, the new item belongs to one subset of $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$. since item can join any any one of the existing k sunsets, in this case we have $k \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ possibilities. In conclusion, we can prove that

$$\left\{ \begin{matrix} n+1 \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\} + k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \tag{1}$$

(b) According to the identities in sub-problem(a), if we extend k to $k+1$, then we have:

$$\left\{ \begin{matrix} n+1 \\ k+1 \end{matrix} \right\} = \left\{ \begin{matrix} n \\ k \end{matrix} \right\} + (k+1) \left\{ \begin{matrix} n \\ k+1 \end{matrix} \right\} \tag{2}$$

then the identity to be proved in sub-problem(b) is equivalent to

$$\sum_{j=k}^{n-1} \binom{n}{j} \left\{ \begin{matrix} j \\ k \end{matrix} \right\} = (k+1) \left\{ \begin{matrix} n \\ k+1 \end{matrix} \right\} \tag{3}$$

Then we try to observe the generation process of $\left\{ \begin{matrix} n \\ k+1 \end{matrix} \right\}$. Firstly consider then situation that we have k subsets, obviously we need at least $j = k$ items to make sure that each subset contains at least one item. j should be at most $n-1$ because we require at least one item to fill the last($k+1$th) subset. For each j, obviously we have $\binom{n}{k-1}$ choices. Noting that in this process there may be repetitive situation because

Problem 1 continued on next page. . .

any one of the $k+1$ subsets can be seen as the last subset which is not considered in the potability discussion of $\left\{ \begin{array}{c} j \\ k \end{array} \right\}$, which means that $\sum_{j=k}^{n-1} \left(\begin{array}{c} n \\ j \end{array} \right) \left\{ \begin{array}{c} j \\ k \end{array} \right\}$ repeats $k+1$ times of $\left\{ \begin{array}{c} n \\ k+1 \end{array} \right\}$, therefore we can prove the identity in equ3, which is equivalent to the original identity in problem(b).

Problem 2

A *norepeatword* is a sequence of at least one (and possibly all) of the usual 26 letters a, b, c, .., z, with repetitions not allowed. For example, "course" is a norepeatword, but "statistics" is not. Order matters, e.g., "course" is not the same as "source". A norepeatword is chosen randomly, with all norepeatwords equally likely. Show that the probability that it uses all 26 letters is very close to $1/e$.

Solution:
Let n represents the length of the norepeatword, $N(n)$ repesents the number of norepeatword whose length is n. Obviously $N(n)$ is equal to the num of ordered sample with no replacement, which is equal to $\frac{K!}{(K-n)!}$, where $K = 26$. Thus the probability that the length of the selected norepeatword letter is 26 can be written as:

$$\frac{K!}{\sum_{j=1}^{K} \frac{K!}{(K-j)!}} = \frac{1}{\sum_{j=1}^{K} \frac{1}{(K-j)!}} = \frac{1}{\sum_{i=0}^{K-1} \frac{1}{i!}} \tag{4}$$

noting that the limitation of the numerator in above number is $1/e$ because the limitation of $\sum_{i=1}^{K} \frac{1}{i!}$ is $e-1$. Thus we can show that the probability that it uses all 26 letters is very close to $1/e$.

Problem 3

Given $n \geq 2$ numbers $(a_1, a_2, \dots, a_n)$ with no repetitions, a bootstrap sample is a sequence $(x_1, x_2, \dots, x_n)$ formed from the a_j 's by sampling with replacement with equal probabilities. Bootstrap samples arise in a widely used statistical method known as the bootstrap. For example, if $n = 2$ and $(a_1, a_2) = (3, 1)$, then the possible bootstrap samples are $(3, 3), (3, 1), (1, 3)$, and $(1, 1)$.
(a) How many possible bootstrap samples are there for $(a_1, \dots, a_n)$?
(b) How many possible bootstrap samples are there for $(a_1, \dots, a_n)$, if order does not matter (in the sense that it only matters how many times each a_j was chosen, not the order in which they were chosen)?
(c) One random bootstrap sample is chosen (by sampling from $a_1, \dots, a_n$ with replacement, as described above). Show that not all unordered bootstrap samples (in the sense of (b)) are equally likely. Find an unordered bootstrap sample $\mathbf{b}_1$ that is as likely as possible, and an unordered bootstrap sample $\mathbf{b}_2$ that is as unlikely as possible. Let p_1 be the probability of getting $\mathbf{b}_1$ and p_2 be the probability of getting $\mathbf{b}_2$ (so p_i is the probability of getting the specific unordered bootstrap sample $\mathbf{b}_i$). What is p_1/p_2 ? What is the ratio of the probability of getting an unordered bootstrap sample whose probability is p_1 to the probability of getting an unordered sample whose probability is p_2 ?

Solution:
(a) It is obvious that every time we can choose any one of the n numbers, with n times, so there are n^n possible samples in total.
(b) According to Bose-Einstein Counting, the result is given by $\binom{n+n-1}{n} = \binom{2n-1}{n}$
(c) Since the order does not matter, so a bootstrap with all the elements different has the highest probability to be chosen, since there are many ordered permutations can be merged to it. On the contrary, bootstraps with all the elements the same have the lowest probability to be chosen. According to the statement, p_1 and p_2 can be respectively given by $p_1 = n!/n^n$ and $p_2 = 1/n^n$. Thus, we have $p_1/p_2 = n!$. Because there are n bootstraps which contains the identical number in it, so the probability of getting a bootstrap whose probability is p_1 to the probability of getting a bootstrap whose probability is p_2 is given by $P = n!/n = (n-1)!$.

Problem 4

(Geometric Probability) You get a stick and break it randomly into three pieces. What is the probability that you can make a triangle using such three pieces?

Solution:
It is denoted that the length of the three pieces are x, y, and $1-x-y$, respectively. It is obvious that we have $0 \leq x \leq 1$, $0 \leq y \leq 1$ and $0 \leq (1-x-y) \leq 1$. In order to make a triangle successfully, it is necessary that

- $x+y > 1-x-y \Longrightarrow x+y > 1/2$;
- $x+1-x-y > y \Longrightarrow y < 1/2$;
- $y+1-x-y > x \Longrightarrow x < 1/2$.

With the help of Figure 1, the probability is 1/4.

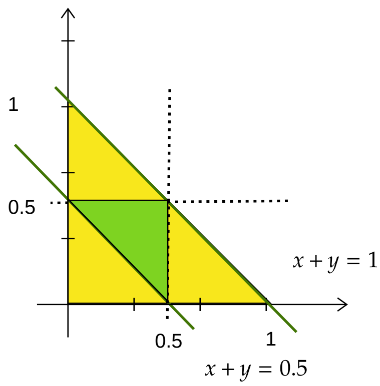


Figure 1: Problem 4.

Problem 5

In the birthday problem, we assumed that all 365 days of the year are equally likely (and excluded February 29). In reality, some days are slightly more likely as birthdays than others. For example, scientists have long struggled to understand why more babies are born 9 months after a holiday. Let $\mathbf{p} = (p_1, p_2, \ldots, p_{365})$ be the vector of birthday probabilities, with p_j the probability of being born on the j th day of the year (February 29 is still excluded, with no offense intended to Leap Dayers). The k th elementary symmetric polynomial in the variables $x_1, \ldots, x_n$ is defined by

$$e_k(x_1, \ldots, x_n) = \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} x_{j_1} \ldots x_{j_k}.$$

This just says to add up all of the $\binom{n}{k}$ terms we can get by choosing and multiplying k of the variables. For example, $e_1(x_1, x_2, x_3) = x_1 + x_2 + x_3$, $e_2(x_1, x_2, x_3) = x_1x_2 + x_1x_3 + x_2x_3$, and $e_3(x_1, x_2, x_3) = x_1x_2x_3$
Now let $k \geq 2$ be the number of people.
(a) Find a simple expression for the probability that there is at least one birthday match, in terms of $\mathbf{p}$ and

Problem 5 continued on next page...

an elementary symmetric polynomial.
(b) Explain intuitively why it makes sense that P (at least one birthday match) is minimized when $p_j = \frac{1}{365}$ for all j, by considering simple and extreme cases.
(c) The famous arithmetic mean-geometric mean inequality says that for $x, y \geq 0$

$$\frac{x+y}{2} \geq \sqrt{xy}.$$

This inequality follows from adding $4xy$ to both sides of $x^2 - 2xy + y^2 = (x-y)^2 \geq 0$ Define $\mathbf{r} = (r_1, \ldots, r_{365})$ by $r_1 = r_2 = (p_1 + p_2)/2, r_j = p_j$ for $3 \leq j \leq 365$. Using the arithmetic mean-geometric mean bound and the fact, which you should verify, that

$$e_k(x_1, \ldots, x_n) = x_1 x_2 e_{k-2}(x_3, \ldots, x_n) + (x_1 + x_2) e_{k-1}(x_3, \ldots, x_n) + e_k(x_3, \ldots, x_n)$$

show that $P($ at least one birthday match $\mid \mathbf{p}) \geq P($ at least one birthday match $\mid \mathbf{r})$ with strict inequality if $\mathbf{p} \neq \mathbf{r}$, where the given $\mathbf{r}$ notation means that the birthday probabilities are given by $\mathbf{r}$. Using this, show that the value of $\mathbf{p}$ that minimizes the probability of at least one birthday match is given by $p_j = \frac{1}{365}$ for all j.

Solution:
(a) It is easier to consider the problem inversely, i.e., what is the probability that no person shares the same birthday. Since $e_k(\mathbf{p})$ multiplies and sums k **different** days, and there are $k!$ ways to map each condition to everyone, so the probability for no matching can be given by $k!e_k(\mathbf{p})$, thereby the probability for at least one birthday matching can be given by $1 - k!e_k(\mathbf{p})$.
(b) By considering the simple case where $k = 2$, we have

$$\begin{aligned} & P(\text{at least one birthday matches}) \\ =& 1 - 2e_2(\mathbf{p}) \\ =& (\sum_{i=1}^{365} p_i)^2 - 2 \sum_{1 \leq i < j \leq n} p_i p_j \\ =& \sum_{i=1}^{365} p_i^2 \\ \geq& 365 \cdot \left(\frac{\sum_{i=1}^{365} p_i}{365} \right)^2 \\ =& \frac{1}{365}, \end{aligned} \tag{5}$$

where the equation holds if and only if $\forall i, p_i = 1/365$.
(c) For each terms in the expansion of $e_k(x_1, \cdots, x_n)$, it either contains x_1 or not. For those terms not containing x_1, their sum can be written as $e_k(x_2, \cdots, x_n)$. For those terms containing x_1, by extracting the common factor (i.e., x_1), they can be written as $x_1 \cdot e_{k-1}(x_2, \cdots, x_n)$ (this is correct, because there are $\binom{n}{k}$ terms in $e_k(x_1, \cdots, x_n)$, $\binom{n-1}{k}$ terms in $e_k(x_2, \cdots, x_n)$, $\binom{n-1}{k-1}$ terms in $e_{k-1}(x_2, \cdots, x_n)$, and $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$). By applying the same steps to x_2, this conclusion holds.

$$\begin{aligned} & P(\text{at least one birthday match}|\mathbf{p}) \\ =& 1 - k!e_k(\mathbf{p}) \\ =& 1 - k!(p_1 p_2 e_{k-2}(p_3, \cdots, p_n) + (p_1 + p_2) e_{k-1}(p_3, \cdots, p_n) + e_k(p_3, \cdots, p_n)) \\ \geq& 1 - k! \left(\frac{(p_1 + p_2)^2}{4} e_{k-2}(p_3, \cdots, p_n) + (p_1 + p_2) e_{k-1}(p_3, \cdots, p_n) + e_k(p_3, \cdots, p_n) \right) \\ =& 1 - k!(r_1 r_2 e_{k-2}(r_3, \cdots, r_n) + (r_1 + r_2) e_{k-1}(r_3, \cdots, r_n) + e_k(r_3, \cdots, r_n)) \\ =& 1 - k!e_k(\mathbf{r}) \\ =& P(\text{at least one birthday match}|\mathbf{r}) \end{aligned} \tag{6}$$

Problem 5 continued on next page...

The proposition can be then proved by contradiction. Assume that $\mathbf{p} = (365^{-1}, \cdots, 365^{-1})$, and $\mathbf{p}' = (p'_1, \cdots, p'_n) \neq \mathbf{p}$ satisfies that $\mathbf{p}'$ minimizes the probability. Since $\mathbf{p}' \neq \mathbf{p}$, there are at least two elements, saying p'_i and p'_j, satisfy that $p'_i \neq p'_j$. Then, a corresponding $\mathbf{r}' = (r'_1, \cdots, r'_n)$ can be further given by $r'_i = r'_j = (p'_i + p'_j)/2$, $r'_k = p'_k (k \neq i, j)$. It is obvious that we have $P(\text{at least one birthday match}|\mathbf{p}') \geq P(\text{at least one birthday match}|\mathbf{r}')$, which contradicts with the assumption. Therefore, the probability is minimized only when $\mathbf{p} = (365^{-1}, \cdots, 365^{-1})$.

Problem 6

(Coupon Collection) If each box of a brand of crispy instant noodle contains a coupon, and there are 108 different types of coupons. Given $n \geq 200$, what is the probability that buying n boxes can collect all 108 types of coupons? You also need to plot a figure to show how such probability changes with the increasing value of n. When such probability is no less than 95%, what is the minimum number of n ?

Solution:

To sample an arbitrary type of coupons from all 108 types for n times, there are 108^n possibilities in total. In order to collect all the 108 types with n boxes, it is equivalent to find a division of n, where it is divided into 108 non-empty subsets and each type of coupon is placed into the corresponding subsets, whose result is given by $\left\{ {n \atop 108} \right\}$. Since the order of these types does not matter, there are 108! permutations in total, so the final probability can be written as $P = 108! \left\{ {n \atop 108} \right\} / 108^n$.

With the help of the formula of Stirling number, i.e., $\left\{ {n \atop m} \right\} = \frac{1}{m!} \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n = \sum_{k=0}^{m} (-1)^k \frac{(m-k)^n}{k!(m-k)!}$, the probability can be expanded by $P = \sum_{k=0}^{108} (-1)^k \frac{108!}{k!(108-k)!} \cdot \left(\frac{108-k}{108} \right)^n$.

With numerical experiments, the minimal number of boxes for $P \geq 0.95$ is 823, as shown in Figure 2.

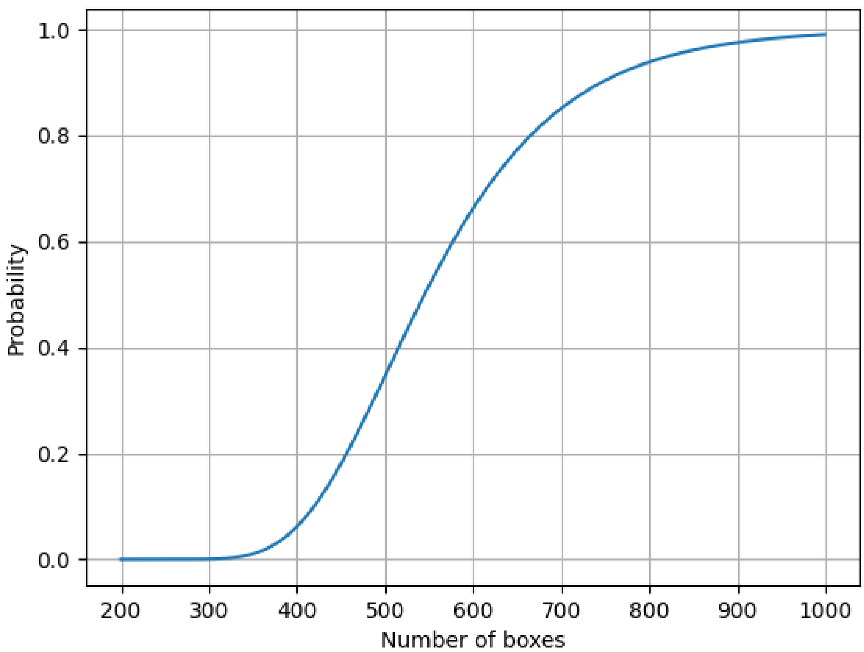


Figure 2: Problem 6.

Probability & Statistics for EECS: Homework #02

Due on Oct 22, 2023 at 23:59

Name:
Student ID:

Oct 17, 2023

Problem 1

Alice is trying to communicate with Bob, by sending a message (encoded in binary) across a channel.
(a) Suppose for this part that she sends only one bit (a 0 or 1), with equal probabilities. If she sends a 0, there is a 5% chance of an error occurring, resulting in Bob receiving a 1; if she sends a 1, there is a 10% chance of an error occurring, resulting in Bob receiving a 0. Given that Bob receives a 1, what is the probability that Alice actually sent a 1?
(b) To reduce the chance of miscommunication, Alice and Bob decide to use a repetition code. Again Alice wants to convey a 0 or a 1, but this time she repeats it two more times, so that she sends 000 to convey 0 and 111 to convey 1. Bob will decode the message by going with what the majority of the bits were. Assume that the error probabilities are as in (a), with error events for different bits independent of each other. Given that Bob receives 110, what is the probability that Alice intended to convey a 1?

Solution:

(a) Suppose event A represents: Alice sends 1, event A^c represents: Alice sends 0, event B represents: Bob receives 1 event B^c represents: Bob receives 0.

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{p(B)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\ &= \frac{0.5 \cdot 0.9}{0.5 \cdot 0.05 + 0.5 \cdot 0.9} \\ &\approx 0.9474. \end{aligned}$$

(b) Suppose event A_1 represents Alice sends 111, event A_0 represents Alice sends 000, and event B represents Bob receives 110.

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} \\ &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{0.9^2 \cdot 0.1 \cdot 0.5}{0.9^2 \cdot 0.1 \cdot 0.5 + 0.05^2 \cdot 0.95 \cdot 0.5} \\ &\approx 0.9715. \end{aligned}$$

Problem 2

Fred decides to take a series of n tests, to diagnose whether he has a certain disease (any individual test is not perfectly reliable, so he hopes to reduce his uncertainty by taking multiple tests). Let D be the event that he has the disease, $p = P(D)$ be the prior probability that he has the disease, and $q = 1 - p$. Let T_j be the event that he tests positive on the jth test.

(a) Assume for this part that the test results are conditionally independent given Fred's disease status. Let $a = P(T_j \mid D)$ and $b = P(T_j \mid D^c)$, where a and b don't depend on the jth test. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the n tests.

(b) Suppose that Fred tests positive on all n tests. However, some people have a certain gene that makes them always test positive. Let G be the event that Fred has the gene. Assume that $P(G) = 1/2$ and that D and G are independent. If Fred does not have the gene, then the test results are conditionally independent given his disease status. Let $a_0 = P(T_j \mid D, G^c)$ and $b_0 = P(T_j \mid D^c, G^c)$, where a_0 and b_0 don't depend on j. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the tests.

Solution

(a) We need to calculate $P\left(D \mid \cap_{j=1}^n T_j\right)$. Use Bayes' formula, LOTP and conditional independence of T_j (if D is given) to obtain following

$$\begin{aligned}
P\left(D \mid \cap_{j=1}^n T_j\right) &= \frac{P(D)P\left(\cap_{j=1}^n T_j \mid D\right)}{P\left(\cap_{j=1}^n T_j\right)} = \frac{P(D)\prod_{j=1}^n P(T_j \mid D)}{P\left(\cap_{j=1}^n T_j \mid D\right)P(D) + P\left(\cap_{j=1}^n T_j \mid D^c\right)P(D^c)} \\
&= \frac{P(D)\prod_{j=1}^n P(T_j \mid D)}{P(D)\prod_{j=1}^n P(T_j \mid D) + P(D^c)\prod_{j=1}^n P(T_j \mid D^c)} \\
&= \frac{p \cdot a^n}{p \cdot a^n + (1-p)b^n}.
\end{aligned}$$

(b) Again, we need to calculate $P\left(D \mid \cap_{j=1}^n T_j\right)$. Same as in (a), obtain that is

$$P\left(D \mid \cap_{j=1}^n T_j\right) = \frac{P(D)P\left(\cap_{j=1}^n T_j \mid D\right)}{P(D)P\left(\cap_{j=1}^n T_j \mid D\right) + P(D^c)P\left(\cap_{j=1}^n T_j \mid D^c\right)}.$$

Use LOTP and independence of G and D to calculate

$$\begin{aligned}
P\left(\cap_{j=1}^n T_j \mid D\right) &= P\left(\cap_{j=1}^n T_j \mid D, G\right)P(G \mid D) + P\left(\cap_{j=1}^n T_j \mid D, G^c\right)P(G^c \mid D) \\
&= P\left(\cap_{j=1}^n T_j \mid D, G\right)P(G) + P\left(\cap_{j=1}^n T_j \mid D, G^c\right)P(G^c) \\
&= \frac{1}{2}P\left(\cap_{j=1}^n T_j \mid D, G\right) + \frac{1}{2}P\left(\cap_{j=1}^n T_j \mid D, G^c\right) \\
&= \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \prod_{j=1}^n P(T_j \mid D, G^c) \\
&= \frac{1}{2} + \frac{1}{2}{a_0}^n.
\end{aligned}$$

Similarly we get that

$$P\left(\cap_{j=1}^n T_j \mid D^c\right) = \frac{1}{2} + \frac{1}{2}{b_0}^n.$$

Plug all these information in to obtain that

$$P\left(D \mid \cap_{j=1}^n T_j\right) = \frac{p\left(\frac{1}{2} + \frac{1}{2}{a_0}^n\right)}{p\left(\frac{1}{2} + \frac{1}{2}{a_0}^n\right) + (1-p)\left(\frac{1}{2} + \frac{1}{2}{b_0}^n\right)}.$$

Problem 3

We want to design a spam filter for email. A major strategy is to find phrases that are much more likely to appear in a spam email than in a no spam email. In that exercise,we only consider one such phrase:"free money". More realistically, suppose that we have created a list of 100 words or phrases that are much more likely to be used in spam than in non-spam. Let W_j be the event that an email contains the jth word or phrase on the list. Let

$$p = P(spam), p_j = P(W_j|spam), r_j = P(W_j|not \quad spam)$$

where "spam" is shorthand for the event that the email is spam.
Assume that $W_1, ..., W_{100}$ are conditionally independent given that the email is spam, and also conditionally independent given that it is not spam. A method for classifying emails (or other objects) based on this kind of assumption is called a naive Bayes classifier. (Here "naive" refers to the fact that the conditional independence is a strong assumption, not to Bayes being naive. The assumption may or may not be realistic, but naive Bayes classifiers sometimes work well in practice even if the assumption is not realistic.)
Under this assumption we know, for example, that

$$P(W_1, W_2, W_3^c, W_4^c, ..., W_{100}^c|spam) = p_1p_2(1-p_3)(1-p_4)...(1-p_{100}).$$

Without the naive Bayes assumption, there would be vastly more statistical and computational difficulties since we would need to consider 2100 ≈ 1.31030 events of the form$A1 \cap A2... \cap A100$ with each A_j equal to either W_j orW_j^c . A new email has just arrived, and it include the 23rd, 64th, and 65th words or phrases on the list(but not the other 97). So we want to compute

$$P(spam|W_1^c, ..., W_{22}^c, W_{23}, W_{24}^c, ..., W_{63}^c, W_{64}, W_{65}, W_{66}^c, ..., W_{100}^c).$$

Note that we need to condition on all the evidence, not just the fact that $W_{23} \cap W_{64} \cap W_{65}$ occurred. Find the condition probability that the new email is spam (in terms of p and the p_j and r_j).

Solution:
Let W represents $W_1^c, ..., W_{22}^c, W_{23}, W_{24}^c, ..., W_{63}^c, W_{64}, W_{65}, W_{66}^c, ..., W_{100}^c$. Then using Baye's formula, we have:

$$p(spam|W) = \frac{p(spam) \cdot p(W|spam)}{p(spam) \cdot p(W|spam) + p(spam^c) \cdot p(W|spam^c)},$$

Because:
$p(W|spam) = p(W_1^c|spam) \cdot ... \cdot p(W_{22}^c|spam) \cdot p(W_{23}|spam) \cdot p(W_{24}^c|spam) \cdot ... \cdot p(W_{64}|spam) \cdot p(W_{65}|spam) \cdot p(W_{66}^c|spam) \cdot ... \cdot p(W_{100}^c|spam) = (1-p_1)...(1-p_{22})p_{23}(1-p_{24})...p_{64}p_{65}(1-p_{66})...(1-p_{100})$,

$p(W|spam^c) = p(W_1^c|spam^c) \cdot ... \cdot p(W_{22}^c|spam^c) \cdot p(W_{23}|spam^c) \cdot p(W_{24}^c|spam^c) \cdot ... \cdot p(W_{64}|spam^c) \cdot p(W_{65}|spam^c) \cdot p(W_{66}^c|spam^c) \cdot ... \cdot p(W_{100}^c|spam^c) = (1-r_1)...(1-r_{22})r_{23}(1-r_{24})...r_{64}r_{65}(1-r_{66})...(1-r_{100})$,
the equation above can be written as

$$\begin{aligned} &P(spam|W) \\ =&\frac{p(1-p_1)...(1-p_{22})p_{23}(1-p_{24})...p_{64}p_{65}(1-p_{66})(1-p_{100})}{p(1-p_1)...p_{23}(1-p_{24})...p_{65}(1-p_{66})(1-p_{100}) + (1-p)(1-r_1)...(1-r_{22})r_{23}(1-r_{24})...r_{64}r_{65}(1-r_{66})(1-r_{100})} \\ =&\frac{P}{P+Q}, \end{aligned}$$

where,
$P = p(1-p_1)...(1-p_{22})p_{23}(1-p_{24})...p_{64}p_{65}(1-p_{66})...(1-p_{100})$,
$Q = (1-p)(1-r_1)...(1-r_{22})r_{23}(1-r_{24})...r_{64}r_{65}(1-r_{66})...(1-r_{100})$.

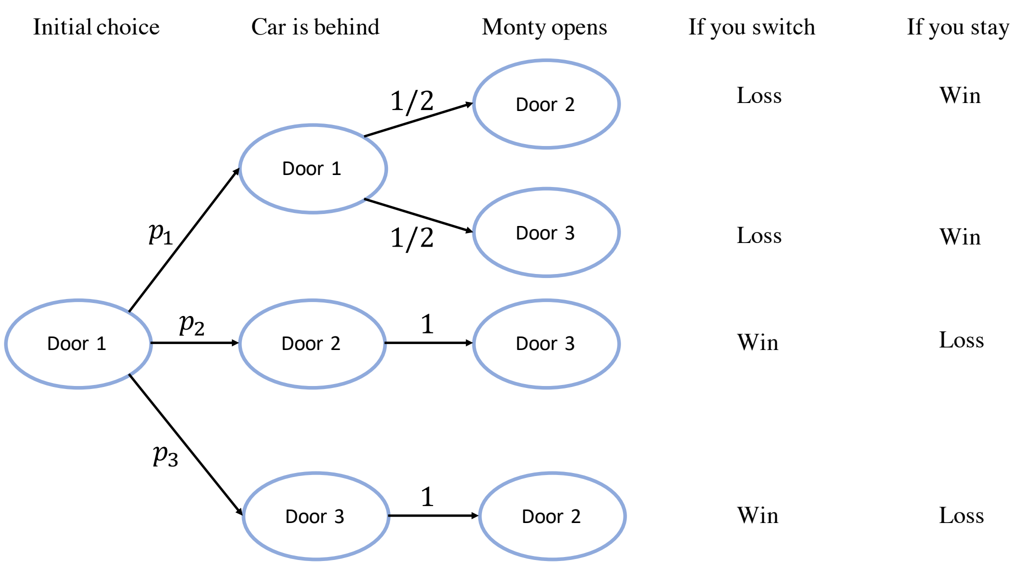
Problem 4

In Monty Hall problem, now suppose the car is not placed randomly with equal probability behind the three doors. Instead, the car is behind door one with probability p_1, behind door two with probability p_2, and behind door three with probability p_3. Here $p_1 + p_2 + p_3 = 1$ and $p_1 \geq p_2 \geq p_3 > 0$. You are to choose one of the three doors, after which Monty will open a door he knows to conceal a goat. Monty always chooses randomly with equal probability among his options in those cases where your initial choice is correct. What strategy should you follow?

Solution:
We define

1. $P_i^{\text{switch}}(\text{win})$ as the probability of winning if "choosing door i first and then switching".
2. $P_i^{\text{stay}}(\text{win})$ as the probability of winning if "choosing door i first and then sticking to initial choice".

When we choose door 1 first, outcomes are shown in the following diagram:



Therefore, we have

$$P_1^{\text{switch}}(\text{win}) = p_2 + p_3,$$
$$P_1^{\text{stay}}(\text{win}) = p_1.$$

Similarly, when we choose door 2 first,

$$P_2^{\text{switch}}(\text{win}) = p_1 + p_3,$$
$$P_2^{\text{stay}}(\text{win}) = p_2.$$

When we choose door 3 first,

$$P_3^{\text{switch}}(\text{win}) = p_1 + p_2,$$
$$P_3^{\text{stay}}(\text{win}) = p_3.$$

Remind that $p_1 \geq p_2 \geq p_3 > 0$. It's not difficult to find that $P_3^{\text{switch}}(\text{win})$ has the maximum winning probability. Hence, in this case, the optimal strategy should be: **choose door 3 first, then switch to the unopened door** after Monty opens some door. Intuitively, we are actually choosing the most unlikely door at the beginning, and then switch to the surviving door, which is the most likely one to be our target.

Problem 5

Consider the Monty Hall problem, except that Monty enjoys opening door 2 more than he enjoys opening door 3, and if he has a choice between opening these two doors, he opens door 2 with probability p, where $1/2 \leq p \leq 1$. To recap: there are three doors, behind one of which there is a car (which you want), and behind the other two of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door, which for concreteness we assume is door 1.
(a) Find the unconditional probability that the strategy of always switching succeeds (unconditional in the sense that we do not condition on which of doors 2 or 3 Monty opens).
(b) Find the probability that the strategy of always switching succeeds, given that Monty opens door 2.
(c) Find the probability that the strategy of always switching succeeds, given that Monty opens door 3.

Solution

(a) Let C_j be the event that the car is hidden behind door j and let W be the event that we win using the switching strategy. Using the law of total probability, we can find the unconditional probability of winning in the same way as in class:

$$\begin{aligned}P(W) &= P\left(W \mid C_1\right) P\left(C_1\right)+P\left(W \mid C_2\right) P\left(C_2\right)+P\left(W \mid C_3\right) P\left(C_3\right) \\ &= 0 \cdot 1/3+1 \cdot 1/3+1 \cdot 1/3=2/3.\end{aligned}$$

(b) A tree method works well here (delete the paths which are no longer relevant after the conditioning, and reweight the remaining values by dividing by their sum), or we can use Bayes' rule and the law of total probability (as below).
Let D_i be the event that Monty opens Door i. Note that we are looking for $P\left(W \mid D_2\right)$, which is the same as $P\left(C_3 \mid D_2\right)$ as we first choose Door 1 and then switch to Door 3. By Bayes' rule and the law of total probability,

$$\begin{aligned}P\left(C_3 \mid D_2\right) &= \frac{P\left(D_2 \mid C_3\right) P\left(C_3\right)}{P\left(D_2\right)} \\ &= \frac{P\left(D_2 \mid C_3\right) P\left(C_3\right)}{P\left(D_2 \mid C_1\right) P\left(C_1\right)+P\left(D_2 \mid C_2\right) P\left(C_2\right)+P\left(D_2 \mid C_3\right) P\left(C_3\right)} \\ &= \frac{1 \cdot 1/3}{p \cdot 1/3+0 \cdot 1/3+1 \cdot 1/3} \\ &= \frac{1}{1+p}.\end{aligned}$$

(c) The structure of the problem is the same as part (b) (except for the condition that $p \geq 1/2$, which was no needed above). Imagine repainting doors 2 and 3, reversing which is called which. By part (b) with $1-p$ in place of p, $P\left(C_2 \mid D_3\right)=\frac{1}{1+(1-p)}=\frac{1}{2-p}$.

Problem 6

A/B testing is a form of randomized experiment that is used by many companies to learn about how customers will react to different treatments. For example, a company may want to see how users will respond to a new feature on their website (compared with how users respond to the current version of the website) or compare two different advertisements. As the name suggests, two different treatments, Treatment A and Treatment B, are being studied. Users arrive one by one, and upon arrival are randomly assigned to one of the two treatments. The trial for each user is classified as "success" (e.g., the user made a purchase) or "failure". The probability that the n-th user receives Treatment A is allowed to depend on the outcomes for the previous users. This set-up is known as a two-armed bandit. Many algorithms for how to randomize the treatment assignments have been studied. Here is an especially simple (but fickle) algorithm, called a "stay-with-a-winner" procedure:
(i) Randomly assign the first user to Treatment A or Treatment B, with equal probabilities.
(ii) If the trial for the n-th user is a success, stay with the same treatment for the $(n+1)$-st user; otherwise, switch to the other treatment for the $(n+1)$-st user.
Let a be the probability of success for Treatment A, and b be the probability of success for Treatment B. Assume that $a \neq b$, but that a and b are unknown (which is why the test is needed). Let p_n be the probability of success on the n-th trial and a_n be the probability that Treatment A is assigned on the n-th trial (using the above algorithm).
(a) Show that

$$p_n = (a-b)a_n + b, \ a_{n+1} = (a+b-1)a_n + 1 - b.$$

(b) Use the results from (a) to show that p_{n+1} satisfies the following recursive equation:

$$p_{n+1} = (a+b-1)p_n + a + b - 2ab.$$

(c) Use the result from (b) to find the long-run probability of success for this algorithm, $\lim_{n\to+\infty} p_n$, assuming that this limit exists.
Solution:
(a)

$$\begin{aligned}
p_n =& P\{n\text{-th trial succeed}\} \\
=& P\{n\text{-th trial succeed}|\text{Treatment A is assigned on the } n\text{-th trial}\}P\{\text{Treatment A is assigned on the } n\text{-th trial}\} \\
&+ P\{n\text{-th trial succeed}|\text{Treatment B is assigned on the } n\text{-th trial}\}P\{\text{Treatment B is assigned on the } n\text{-th trial}\} \\
=& a \cdot a_n + b \cdot (1-a_n) \\
=& (a-b)a_n + b.
\end{aligned} \tag{1}$$

$$\begin{aligned}
a_{n+1} =& P\{\text{Treatment A is assigned on the } (n+1)\text{-th trial}\} \\
=& P\{\text{Treatment A is assigned on the } n\text{-th trial}\}P\{n\text{-th trial succeed}\} \\
&+ P\{\text{Treatment B is assigned on the } n\text{-th trial}\}P\{n\text{-th trial failed}\} \\
=& a_n a + (1-a_n)(1-b) \\
=& (a+b-1)a_n + 1 - b.
\end{aligned} \tag{2}$$

(b)

$$\begin{aligned}p_{n+1} =& P\{(n+1)\text{-th trial succeed}\}\\ =& P\{(n+1)\text{-th trial succeed}|\text{Treatment A is assigned on the } (n+1)\text{-th trial}\}\\ & \cdot P\{\text{Treatment A is assigned on the } (n+1)\text{-th trial}\}\\ & + P\{(n+1)\text{-th trial succeed}|\text{Treatment B is assigned on the } (n+1)\text{th trial}\}\\ & \cdot P\{\text{Treatment B is assigned on the } (n+1)\text{-th trial}\}\\ =& a a_{n+1} + b(1-a_{n+1})\\ =& (a-b)[(a+b-1)a_n + 1 - b] + b\\ =& (a+b-1)p_n + a + b - 2ab.\end{aligned} \tag{3}$$

(c) It is denoted that $\lim_{n\to+\infty} p_n = p$. Since the limitation exists, we have

$$p = (a+b-1)p + a + b - 2ab, \tag{4}$$

that is,

$$p = \frac{a+b-2ab}{2-a-b}. \tag{5}$$

Probability & Statistics for EECS: Homework #03

Due on Oct 10, 2023 at 23:59

Name:

Student ID:

Problem 1

A fair die is rolled repeatedly, and a running total is kept (which is, at each time, the total of all the rolls up until that time). Let p_n be the probability that the running total is ever exactly n (assume the die will always be rolled enough times so that the running total will eventually exceed n, but it may or may not ever equal n).
(a) Write down a recursive equation for p_n (relating p_n to earlier terms p_k in a simple way). Your equation should be true for all positive integers n, so give a definition of p_0 and p_k for $k < 0$ so that the recursive equation is true for small values of n.
(b) Find p_7.
(c) Give an intuitive explanation for the fact that $p_n \to 1/3.5 = 2/7$ as $n \to \infty$.

Solution

(a) For an arbitrary integer n, it can be rolled by rolling a $(n-1)$ and a 1, or a $(n-2)$ and a 2, ..., or a $(n-6)$ and a 6, where temporarily ignore the limitation of positive integers. Thus we have

$$p_n = \frac{1}{6}(p_{n-1} + p_{n-2} + p_{n-3} + p_{n-4} + p_{n-5} + p_{n-6}) \tag{1}$$

On one hand, it is easy to find that $p_1 = 1/6$. On the other hand, $p_1 = 1/6 \cdot (p_0 + p_{-1} + p_{-2} + p_{-3} + p_{-4} + p_{-5})$. Consider practical scenarios, where a 0 can always be rolled without rolling it, and a negative number can never be rolled. Thereby, we have

$$\begin{cases} p_0 = 1, \\ p_i = 0, \quad i = -1, \dots, -5. \end{cases} \tag{2}$$

(b) Since

$$p_7 = \frac{1}{6}(p_6 + p_5 + p_4 + p_3 + p_2 + p_1) \tag{3}$$

We have

$$\begin{aligned}
p_1 &= \frac{1}{6} * p_0 = \frac{1}{6} \\
p_2 &= \frac{1}{6} * (p_0 + p_1) = \frac{1}{6} * (1 + \frac{1}{6}) = \frac{7}{36} \\
p_3 &= \frac{1}{6} * (p_0 + p_1 + p_2) = \frac{1}{6} * (1 + \frac{1}{6} + \frac{7}{36}) = \frac{49}{216} \\
p_4 &= \frac{1}{6} * (p_0 + p_1 + p_2 + p_3) = \frac{1}{6} * (1 + \frac{1}{6} + \frac{7}{36} + \frac{49}{216}) = \frac{343}{1296} \\
p_5 &= \frac{1}{6} * (p_0 + p_1 + p_2 + p_3 + p_4) = \frac{1}{6} * (1 + \frac{1}{6} + \frac{7}{36} + \frac{49}{216} + \frac{343}{1296}) = \frac{2401}{7776} \\
p_6 &= \frac{1}{6} * (p_0 + p_1 + p_2 + p_3 + p_4 + p_5) = \frac{1}{6} * (1 + \frac{1}{6} + \frac{7}{36} + \frac{49}{216} + \frac{343}{1296} + \frac{2401}{7776}) = \frac{16807}{46656}
\end{aligned}$$

Then, we have

$$p_7 = \frac{70993}{279936}$$

(c) As $n \to +\infty$, the gap of rolling different numbers is negligible, where the expectation for each "increment" is given by $1/6 \cdot (1 + 2 + 3 + 4 + 5 + 6) = 7/2$, so the probability for rolling a certain number is given by $p_n \approx 1/(7/2) = 2/7$.

Problem 2

A message is sent over a noisy channel.The message is a sequence $x_1, x_2, ..., x_n$ of n bits ($x_i \in \{0, 1\}$). Since the channel is noisy, there is a chance that any bit might be corrupted, resulting in an error ($a_0$becomes a_1 or vice versa). Assume that the error events are independent. Let p be the probability that an individual bit has an error($0 < p < 1/2$).Let $y_1, y_2, ..., y_n$ be the received message (so $y_i = x_i$ if there is no error in that bit,but $y_i = 1 - x_i$ if there is an error there).
To help detect errors, the n th bit is reserved for a parity check:x_n is defined to be 0 if $x_1 + x_2 + ... + x_{n-1}$ is even, and 1 if $x_1 + x_2 + ... + x_{n-1}$ is odd. When the message is received, the recipient checks whether y_n has the same parity as $y_1 + y_2 + ... + y_{n_1}$. If the parity is wrong, the recipient knows that at least one error occurred; otherwise, the recipient assumes that there were no errors.
(a) For $n = 5, p = 0.1$, what is the probability that the received message has errors which go undetected?
(b) For general n and p, write down an expression (as a sum) for the probability that the received message has errors which go undetected.
(c) Give a simplified expression, not involving a sum of a large number of terms, for the probability that the received message has errors which go undetected.

solution:

(a) If the error message is undetected, the total number of error bits should be even: if the number of error bits in $x_1, x_2, \ldots, x_{n-1}$ is even(> 0), then the last bit x_n should be right; if the number of error bits in $x_1, x_2, \ldots, x_{n-1}$ is odd, then the last bit x_n should be wrong.

$$\begin{aligned} P(\text{undetected error message}) &= \binom{5}{4} p^4(1-p) + \binom{5}{2} p^2(1-p)^3 \\ &= 0.07335 \end{aligned}$$

(b) According to (a), we have

$$P(\text{undetected error message}) = \sum_{k \text{ is even}}^{n} \binom{n}{k} p^k(1-p)^{n-k}$$

(c)

$$\begin{aligned} P(\text{undetected error message}) &= \frac{(p+(1-p))^n + (-1)^n(p-(1-p))^n}{2} - (1-p)^n \\ &= \frac{1+(1-2p)^n}{2} - (1-p)^n \end{aligned}$$

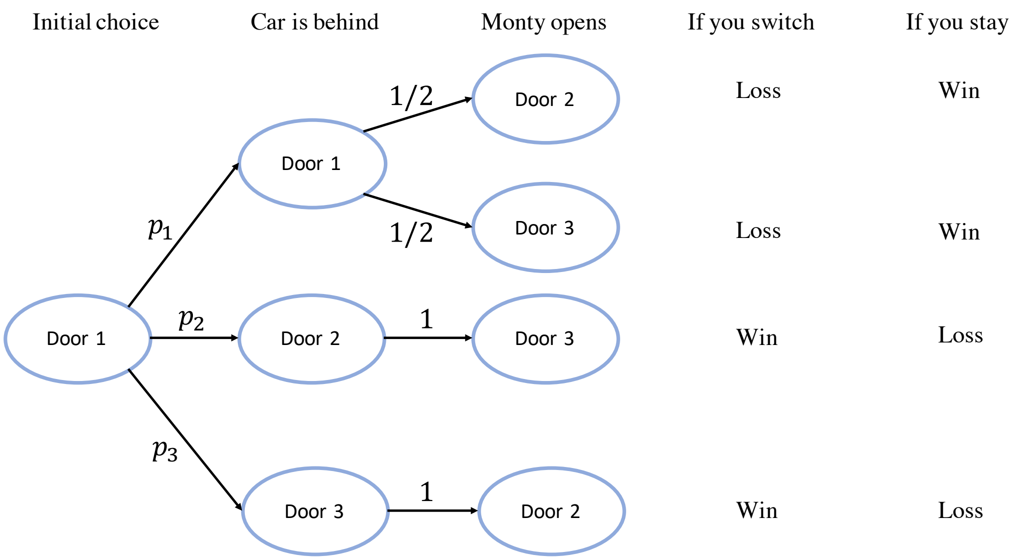
Problem 3

In Monty Hall problem, now suppose the car is not placed randomly with equal probability behind the three doors. Instead, the car is behind door one with probability p_1, behind door two with probability p_2, and behind door three with probability p_3. Here $p_1 + p_2 + p_3 = 1$ and $p_1 \geq p_2 \geq p_3 > 0$. You are to choose one of the three doors, after which Monty will open a door he knows to conceal a goat. Monty always chooses randomly with equal probability among his options in those cases where your initial choice is correct. What strategy should you follow?

Solution:
We define

1. $P_i^{\text{switch}}(\text{win})$ as the probability of winning if "choosing door i first and then switching".
2. $P_i^{\text{stay}}(\text{win})$ as the probability of winning if "choosing door i first and then sticking to initial choice".

When we choose door 1 first, outcomes are shown in the following diagram:



Therefore, we have

$$P_1^{\text{switch}}(\text{win}) = p_2 + p_3,$$
$$P_1^{\text{stay}}(\text{win}) = p_1.$$

Similarly, when we choose door 2 first,

$$P_2^{\text{switch}}(\text{win}) = p_1 + p_3,$$
$$P_2^{\text{stay}}(\text{win}) = p_2.$$

When we choose door 3 first,

$$P_3^{\text{switch}}(\text{win}) = p_1 + p_2,$$
$$P_3^{\text{stay}}(\text{win}) = p_3.$$

Remind that $p_1 \geq p_2 \geq p_3 > 0$. It's not difficult to find that $P_3^{\text{switch}}(\text{win})$ has the maximum winning probability. Hence, in this case, the optimal strategy should be: **choose door 3 first, then switch to the unopened door** after Monty opens some door. Intuitively, we are actually choosing the most unlikely door at the beginning, and then switch to the surviving door, which is the most likely one to be our target.

Consider the following 7-door version of the Monty Hall problem. There are 7 doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens 3 goat doors, and offers you the option of switching to any of the remaining 3 doors.

(a) (10 points) Assume that Monty Hall knows which door has the car, will always open 3 goat doors and offer the option of switching, and that Monty chooses with equal probabilities from all his choices of which goat doors to open. Should you switch? What is your probability of success if you switch to one of the remaining 3 doors?

(b) Generalize the above to a Monty Hall problem where there are $n \geq 3$ doors, of which Monty opens m goat doors, with $1 \leq m \leq n-2$.

Solution

(a) Assume the doors are labeled such that you choose Door 1 (to simplify notation), and suppose first that you follow the "stick to your original choice" strategy. Let S be the event of success in getting the car, and let C_j be the event that the car is behind Door j. Conditioning on which door has the car, we have

$$P(S) = \frac{1}{7}.$$

Let M_{ijk} be the event that Monty opens Doors i, j, k. Then

$$P(S) = \sum_{i,j,k} P(S|M_{i,j,k}), 2 \leq i < j < k \leq 7$$

By symmetry, this gives $P(S|M_{i,j,k}) = P(S) = \frac{1}{7}$. Thus the conditional probability that the car is behind 1 of the remaining 3 doors is 6/7, which gives 2/7 for each. So you should switch, thus making your probability of success 2/7 rather than 1/7

(b) The problem becomes the following: Consider the following n-door version of the Monty Hall problem. There are n doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens m goat doors, and offers you the option of switching to any of the remaining $n-m-1$ doors.

Without loss of generality, we can assume the contestant picked door 1 (if she didn't pick door 1, we could simply relabel the doors, or rewrite this solution with the door numbers permuted). Let S be the event of success in getting the car, and let C_j be the event that the car is behind door j. Conditioning on which door has the car, by LOTP we have

$$P(S) = P(S|C_1) \cdot \frac{1}{n} + \cdots + P(S|C_n) \cdot \frac{1}{n}.$$

Suppose you employs the non-switching strategy. The only possibility of success is that the car is indeed behind door 1, which implies

$$P_{\text{non-switching}}(S) = 1 \cdot \frac{1}{n} + 0 \cdot \frac{1}{n} + \cdots + 0 \cdot \frac{1}{n} = \frac{1}{n}.$$

Suppose you employs the switching strategy. If the car is behind door 1, then switching will fail, so $P(\text{get car}|C_1) = 0$. Otherwise, since Monty always reveals m goat, the probability of getting a car by switching to a remaining unopened door is $\frac{1}{n-m-1}$. Thus,

$$P_{\text{switching}}(S) = 0 \cdot \frac{1}{n} + \frac{1}{n-m-1} \cdot \frac{1}{n} + \cdots + \frac{1}{n-m-1} \cdot \frac{1}{n} = \frac{n-1}{(n-m-1)n}.$$

This value is greater than $\frac{1}{n}$, so you should switch.

Problem 5

A/B testing is a form of randomized experiment that is used by many companies to learn about how customers will react to different treatments. For example, a company may want to see how users will respond to a new feature on their website (compared with how users respond to the current version of the website) or compare two different advertisements. As the name suggests, two different treatments, Treatment A and Treatment B, are being studied. Users arrive one by one, and upon arrival are randomly assigned to one of the two treatments. The trial for each user is classified as "success" (e.g., the user made a purchase) or "failure". The probability that the n-th user receives Treatment A is allowed to depend on the outcomes for the previous users. This set-up is known as a two-armed bandit. Many algorithms for how to randomize the treatment assignments have been studied. Here is an especially simple (but fickle) algorithm, called a "stay-with-a-winner" procedure:

(i) Randomly assign the first user to Treatment A or Treatment B, with equal probabilities.

(ii) If the trial for the n-th user is a success, stay with the same treatment for the $(n+1)$-st user; otherwise, switch to the other treatment for the $(n+1)$-st user.

Let a be the probability of success for Treatment A, and b be the probability of success for Treatment B. Assume that $a \neq b$, but that a and b are unknown (which is why the test is needed). Let p_n be the probability of success on the n-th trial and a_n be the probability that Treatment A is assigned on the n-th trial (using the above algorithm).

(a) Show that

$$p_n = (a-b)a_n + b,\ a_{n+1} = (a+b-1)a_n + 1 - b.$$

(b) Use the results from (a) to show that p_{n+1} satisfies the following recursive equation:

$$p_{n+1} = (a+b-1)p_n + a + b - 2ab.$$

(c) Use the result from (b) to find the long-run probability of success for this algorithm, $\lim_{n\to+\infty} p_n$, assuming that this limit exists.

Solution:

(a)

$$\begin{aligned}p_n =& P\{n\text{-th trial succeed}\}\\ =& P\{n\text{-th trial succeed}|\text{Treatment A is assigned on the } n\text{-th trial}\}P\{\text{Treatment A is assigned on the } n\text{-th trial}\}\\ &+ P\{n\text{-th trial succeed}|\text{Treatment B is assigned on the } n\text{-th trial}\}P\{\text{Treatment B is assigned on the } n\text{-th trial}\}\\ =& a\cdot a_n + b\cdot(1-a_n)\\ =& (a-b)a_n + b.\end{aligned} \tag{4}$$

$$\begin{aligned}a_{n+1} =& P\{\text{Treatment A is assigned on the } (n+1)\text{-th trial}\}\\ =& P\{\text{Treatment A is assigned on the } n\text{-th trial}\}P\{n\text{-th trial succeed}\}\\ &+ P\{\text{Treatment B is assigned on the } n\text{-th trial}\}P\{n\text{-th trial failed}\}\\ =& a_n a + (1-a_n)(1-b)\\ =& (a+b-1)a_n + 1 - b.\end{aligned} \tag{5}$$

(b)

$$\begin{aligned}
p_{n+1} =& P\{(n+1)\text{-th trial succeed}\} \\
=& P\{(n+1)\text{-th trial succeed}|\text{Treatment A is assigned on the } (n+1)\text{-th trial}\} \\
& \cdot P\{\text{Treatment A is assigned on the } (n+1)\text{-th trial}\} \\
& + P\{(n+1)\text{-th trial succeed}|\text{Treatment B is assigned on the } (n+1)\text{th trial}\} \\
& \cdot P\{\text{Treatment B is assigned on the } (n+1)\text{-th trial}\} \\
=& a a_{n+1} + b(1-a_{n+1}) \\
=& (a-b)[(a+b-1)a_n + 1 - b] + b \\
=& (a+b-1)p_n + a + b - 2ab.
\end{aligned} \tag{6}$$

(c) It is denoted that $\lim_{n\to+\infty} p_n = p$. Since the limitation exists, we have

$$p = (a+b-1)p + a + b - 2ab, \tag{7}$$

that is,

$$p = \frac{a+b-2ab}{2-a-b}. \tag{8}$$

Probability & Statistics for EECS: Homework #04

Due on Oct 10, 2023 at 23:59

Name:
Student ID:

Problem 1

Nick and Penny are independently performing independent Bernoulli trials. For concreteness, assume that Nick is flipping a nickel with probability p_1 of Heads and Penny is flipping a penny with probability p_2 of Heads. Let $X_1, X_2, \ldots$ be Nick's results and $Y_1, Y_2, \ldots$ be Penny's results, with $X_i \sim Bern(p_1)$ and $Y_j \sim Bern(p_2)$.
(a) Find the distribution and expected value of the first time at which they are simultaneously successful, i.e., the smallest n such that $X_n = Y_n = 1$.
Hint: Define a new sequence of Bernoulli trials and use the story of the Geometric.
(b) Find the expected time until at least one has a success (including the success).
Hint:Define a new sequence of Bernoulli trials and use the story of the Geometric.
(c)For $p_1 = p_2$, find the probability that their first successes are simultaneous, and use this to find the probability that Nick's first success precedes Penny's.
Solution:

1. Let $Z_i = 1$, if $X_i = Y_i = 1$, otherwise, $Z_i = 0$.

$$p(Z = k) = (1 - P_1P_2)^{K-1}P_1P_2 \sim Geom(p_1p_2).$$

   Thus, $E(Z) = \frac{1}{p_1p_2}$.

2. Let $Z_i = 0$, if $X_i = Y_i = 0$, otherwise, $Z_i = 1$.

$$p(Z = k) \sim Geom\left(1 - (1 - p_1)(1 - p_2)\right).$$

   Thus, $E(Z) = \frac{1}{p_1+p_2-p_1p_2}$.

3. Let X denotes Nick first success and Y denotes Penny first success, and $p_1 = p_2 = p$,$q = 1 - p$.

$$\begin{aligned} p(X = Y) &= \sum_{k=1}^{\infty} p^2(q^2)^{k-1} \\ &= p^2 \sum_{k=0}^{\infty}(q^2)^k \\ &= \frac{p^2}{1 - q^2} \\ &= \frac{p}{2 - p}. \end{aligned}$$

   Based on above,

$$\begin{aligned} p(X > Y) &= \frac{1 - \frac{p}{2-p}}{2} \\ &= \frac{1 - p}{2 - p}. \end{aligned}$$

Problem 2

For x and y binary digits (0 or 1), let $x \oplus y$ be 0 if $x = y$ and 1 if $x \neq y$ (this operation is called exclusive or (often abbreviated to XOR), or addition mod 2).
(a) Let $X \sim Bern(p)$ and $Y \sim Bern(1/2)$, independently. What is the distribution of $X \oplus Y$
(b) With notation as in sub-problem(a), is $X \oplus Y$ independent of X? Is $X \oplus Y$ independent of Y? Be sure to consider both the case $p = 1/2$ and the case $p \neq 1/2$.
(c) Let $X_1, ..., X_n$ be i.i.d. (i.e., independent and identically distributed) Bern(1/2) R.V.s. For each nonempty subset J of $\{1, 2, ..., n\}$, let

$$Y_J = \oplus_{Y \in J} X_J.$$

Show that Y_J $Bern(1/2)$ and that these $2^n - 1$ R.V.s are pairwise independent, but not independent.
Solution:

1. Let $Z = X \oplus Y$. When $Z = 1$ is the same as $X = 1, Y = 0$ or $X = 0, Y = 1$, also $X \sim Bern(p)$ and $Y \sim Bern(1/2)$, thus we can get:

$$\begin{aligned} p(Z = 1) &= p(X = 1, Y = 0) + p(X = 0, Y = 1) \\ &= p(X = 1)p(Y = 0) + +p(X = 0)p(Y = 1) \\ &= p * 1/2 + (1 - p) * 1/2 \\ &= 1/2. \end{aligned}$$

   For the same reason, we can get $p(Z = 0) = 1/2$. Therefor, $Z \sim Bern(1/2)$.

2. To show whether Z is independent of X, it is the same to verify whether:

$$p(Z = z, X = x) = p(Z = z)p(X = x)$$

   Let's consider the left side respectively:

$$p(Z = 0, X = 1) = p(Y = 1, X = 1) = \frac{1}{2}p,$$

$$p(Z = 0, X = 0) = p(Y = 0, X = 0) = \frac{1}{2}(1 - p),$$

$$p(Z = 1, X = 1) = p(Y = 0, X = 1) = \frac{1}{2}p,$$

$$p(Z = 1, X = 0) = p(Y = 1, X = 0) = \frac{1}{2}(1 - p).$$

   Then consider the right side:

$$p(Z = 0)p(X = 1) = \frac{1}{2}p,$$

$$p(Z = 0)p(X = 0) = \frac{1}{2}(1 - p),$$

$$p(Z = 1)p(X = 1) = \frac{1}{2}p,$$

$$p(Z = 1)p(X = 0) = \frac{1}{2}(1 - p),$$

   No matter the value of p, the equation above is always true. Thus, Z is independent of X for all the case.

   To show whether Z is independent of Y, it is very similar to the above. For the left side:

$$p(Z = 0, Y = 1) = \frac{1}{2}p,$$

Problem 2 continued on next page...

$$p(Z=0,Y=0)=\frac{1}{2}(1-p),$$

$$p(Z=1,Y=1)=\frac{1}{2}(1-p),$$

$$p(Z=1,Y=0)=\frac{1}{2}p.$$

For the right side:

$$p(Z=z)p(Y=Y)=\frac{1}{2}*\frac{1}{2}=\frac{1}{4}.$$

To make sure the equation is always true, we should guarantee $\frac{1}{2}p=\frac{1}{2}(1-p)=\frac{1}{4}$ for all time. This is true only when $p=\frac{1}{2}$.Thus, only when $p=\frac{1}{2}$, Z is independent of Y.

3. Let l denotes the length of subset J. Then use Mathematical induction to prove the equation. As we know:
   When $l=1$, $p(Y_J=1)=p(X_J=1)=\frac{1}{2}$, $Y_J\sim Bern(1/2)$.
   Suppose $l=k$, $Y_J\sim Bern(1/2)$.
   Then when $l=k+1$, let $\hat{J}=J\cup\{\hat{j}\}$, where the length of J is k, Thus:

$$\begin{aligned}p(Y_{\hat{J}}=1)&=p(Y_J=1,X_{\hat{j}}=0)+p(Y_J=0,X_{\hat{j}}=1)\\&=\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\\&=1/2.\end{aligned}$$

   Therefore, $Y_J\sim Bern(1/2)$.
   To show that they are pairwise independent,but not independent, is equal to verify:

$$p(Y_mY_n)=p(Y_m)p(Y_n),\forall m,n,$$

$$p(Y_1Y_2...Y_{2^n-1})\neq p(Y_1)p(Y_2)...p(Y_{2^n-1}).$$

   For the first equation, there are two occasions.
   $Y_m\cap Y_n=\emptyset$:
   as $X_1,X_2...,X_n$ are IID distribution, Y_m and Y_n are obviouly independent.
   $Y_m\cap Y_n\neq\emptyset$:
   let $p=Y_m\cap Y_n$, $s=Y_m-Y_m\cap Y_n$, $q=Y_n-Y_m\cap Y_n$,

$$\begin{aligned}P(Y_m=1,Y_n=1)=&=P(p=1)p(s=0)p(q=0)+P(p=0)p(s=1)p(q=1)\\&=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\\&=1/4\\&=P(Y_m=1)P(Y_n=1).\end{aligned}$$

   Thus, they are pairwise independent.

   For the second equation, we can list a counterexample:

$$P(Y_1,Y_2,Y_3)=\frac{1}{4},$$

$$P(Y_1)P(Y_2)P(Y_3)=\frac{1}{8},$$

   where $Y_1=x_1\oplus x_2$, $Y_2=x_2\oplus x_3$, $Y_3=x_3\oplus x_1$. Therefore, they are not independent.

Problem 3

Let a random variable X satisfies Hypergeometric distribution with parameters w, b, n.

(a) Find $\mathbb{E}\left(\binom{X}{2}\right)$

(b) Use the result of (a) to find the variance of X.

Solution

(a)

Consider an urn with w white balls and b black balls. We draw n balls out of the urn at random without replacement. Let X be the number of white balls in the sample. Then $X \sim \text{HGeom}(w, b, n)$.
Let A_i be the event that the ith chosen ball is white, $i = 1, 2, \ldots, n$.
Let I_i be the indicator of A_i, $i = 1, 2, \ldots, n$.
Since $\binom{X}{2} = \sum_{i<j} I_i I_j$, we have

$$\mathbb{E}\left(\binom{X}{2}\right) = \sum_{i<j} P(A_i \cap A_j)$$

We consider the order of balls, then there are $\binom{w+b}{n} = \frac{(w+b)!}{(w+b-n)!n!}$ ways to choose n balls from $w+b$ balls, $\binom{w}{2} = \frac{w(w-1)}{2}$ ways to choose 2 from the w white balls, and $\binom{w+b-2}{n-2}$ ways to choose the remaining $n-2$ balls from the rest $w+b-2$ balls. Thus we have

$$P(A_i \cap A_j) = \frac{\binom{w}{2}\binom{w+b-2}{n-2}}{\binom{w+b}{n}} = \frac{\frac{w(w-1)}{2} \cdot \frac{(w+b-2)!}{(n-2)!(w+b-n)!}}{\frac{(w+b)!}{n!(w+b-n)!}} = \frac{n!w(w-1)(w+b-2)!}{2(n-2)!(w+b)!} = \frac{n(n-1)w(w-1)}{2(w+b)(w+b-1)}$$

Or by symmetry, $P(A_i \cap A_j) = P(A_1 \cap A_2) = P(A_1)P(A_2|A_1) = \frac{w}{w+b}\frac{w-1}{w+b-1}$ for all $i \neq j$.
It follows that

$$\mathbb{E}\left(\binom{X}{2}\right) = \binom{n}{2} \cdot \frac{w}{w+b} \cdot \frac{w-1}{w+b-1} = \frac{n(n-1)}{2} \cdot \frac{w}{w+b} \cdot \frac{w-1}{w+b-1}$$

(b)

We have

$$\mathbb{E}\left(\binom{X}{2}\right) = \mathbb{E}\left(\frac{X(X-1)}{2}\right) = \frac{1}{2}\mathbb{E}[X(X-1)] = \frac{1}{2}\mathbb{E}[X^2 - X], \quad \mathbb{E}(X) = n \cdot \frac{w}{w+b}$$

By combining the above two equations we have

$$E(X^2) = 2E\left[\binom{X}{2}\right] + E(X) = \frac{n(n-1)w(w-1)}{(w+b)(w+b-1)} + \frac{wn}{w+b} = \frac{wn(wn-n+b)}{(w+b)(w+b-1)}$$

Then we have

$$Var(X) = E(X^2) - (EX)^2 = \frac{w(w-1)n(n-1)}{(w+b)(w+b-1)} - \frac{w^2n^2}{(w+b)^2} + \frac{wn}{w+b} = \frac{wnb(w+b-n)}{(w+b)^2(w+b-1)}$$

Problem 4

Suppose there are n types of toys, which you are collecting one by one, with the goal of getting a complete set. When collecting toys, the toy types are random. Assume that each time you collect a toy, it is equally likely to be any of the n types. Let N denote the number of toys needed until you have a complete set. Find $\mathrm{Var}(N)$.

Solution

Consider a state where the collector has already collected m coupons. Let N_m denote the number of coupons he needs to collect to get to $m+1$ type. Then, if the total coupons needed is N, we have:

$$N = \sum_{m=1}^{n} N_m$$

Every coupon collected from here is like a coin toss where with probability $\frac{m}{n}$, the collector hits a coupon he already has and makes no progress. With probability $\frac{n-m}{n}$, he collects a new coupon. So, this becomes a geometric random variable with $p = \frac{n-m}{n}$. We know that a geometric random variable has a mean $\frac{1}{p}$ and variance $\frac{1-p}{p^2}$. Hence,

$$E(N_m) = \frac{n}{n-m}$$

Further, we have:

$$E(N) = E\left(\sum_{m=1}^{n} N_m\right) = \sum_{m=1}^{n} E(N_m) = \sum_{m=1}^{n} \frac{n}{n-m} = n \sum_{m=1}^{n} \frac{1}{n-m}$$

Substituting $m = n - m$ we get:

$$E(N) = n \sum_{m=1}^{n} \frac{1}{m}$$

Since the random variables N_m are independent, the variance of their sum is equal to the sum of their variances. Therefore, similar to the reasoning for expectation, we have the variance as follows:

$$\mathrm{Var}(N) = n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n \sum_{k=1}^{n} \frac{1}{k}$$

Problem 5

People are arriving at a party one at a time. While waiting for more people to arrive they entertain themselves by comparing their birthdays. Let X be the number of people needed to obtain a birthday match, i.e., before person X arrives there are no two people with the same birthday, but when person X arrives there is a match. Assume for this problem that there are 365 days in a year, all equally likely. By the result of the birthday problem form Chapter 1, for 23 people there is a 50.7% chance of a birthday match (and for 22 people there is a less than 50% chance). But this has to do with the median of X; we also want to know the mean of X, and in this problem we will find it, and see how it compares with 23.
(a) A median of an r.v. Y is a value m for which $P(Y \leq m) \geq 1/2$ and $P(Y \geq m) \geq 1/2$. Every distribution has a median, but for some distributions it is not unique. Show that 23 is the unique median of X.
(b) Show that $X = I_1 + I_2 + \cdots + I_{366}$, where I_j is the indicator r.v. for the event $X \geq j$. Then find $E(X)$ in terms of p_j's defined by $p_1 = p_2 = 1$ and for $3 \leq j \leq 366$,

$$p_j = (1 - \frac{1}{365})(1 - \frac{2}{365}) \cdots (1 - \frac{j-2}{365})$$

(c) Compute $E(X)$ numerically.
(d) Find the variance of X, both in terms of the p_j's and numerically.
Hint:What is I_i^2, and what is I_iI_j for $i < j$? Use this to simplify the expansion

$$X^2 = I_1^2 + \cdots + I_{366}^2 + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_iI_j.$$

Note: In addition to being an entertaining game for parties, the birthday problem has many applications in computer science, such as in a method called the birthday attack in cryptography. It can be shown that if there are n days in a year and n is large, then $E(X) \approx \sqrt{\pi n/2}$. In Volume 1 of his masterpiece The Art of Computer Programming, Don Knuth shows that an even better approximation is

$$E(X) \approx \sqrt{\frac{\pi n}{2}} + \frac{2}{3} + \sqrt{\frac{\pi}{288n}}.$$

Solution:
(a) For an arbitrary pair of people, the probability of having the same birthday is $1/365$. It is denoted that the number of birthday match is Z. Since in the corresponding number of samples is relatively large and the probability is small, we have

$$P(\text{At least one birthday match}) = P(Z \geq 1) = 1 - P(Z = 0) \approx 1 - \mathrm{e}^{\lambda}, \tag{1}$$

where $\lambda = \binom{m}{2}p$, m is the number of people, and p is the probability. Therefore, we have

$$P(X \leq 23) \approx 1 - \mathrm{e}^{\lambda} \approx 0.5002 \geq 0.5. \tag{2}$$

On the other hand, we have

$$\begin{aligned} P(X \geq 23) =& P(\text{No match before 23}) \approx \mathrm{e}^{\lambda} \\ =& \mathrm{e}^{\binom{22}{2}\cdot\frac{1}{365}} \approx 0.531 \geq 0. \end{aligned} \tag{3}$$

Thus, 23 is the unique median of X.
(b) For X, it can always be expressed with the sum of binary indicators since it is not decreasing. Then we have

$$\begin{aligned} E(X) =& E(I_1 + I_2 + \cdots + I_{366}) \\ =& E(I_1) + E(I_2) + \cdots + E(I_{366}) \\ =& \sum_{i=1}^{366} p_j. \end{aligned} \tag{4}$$

(c)

$$\begin{aligned} E(X) =&\sum_{i=1}^{366} p_j \\ =&1+1+\left(1-\frac{1}{365}\right)+\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)+\cdots+\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{364}{365}\right) \\ \approx&24.62 \end{aligned} \tag{5}$$

(d)

$$\begin{aligned} E(X^2) =&E\left(I_1^2+I_2^2+I_3^2+...+I_{366}^2+2\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_i I_j\right) \\ =&E(I_1^2)+E(I_2^2)+E(I_3^2)+...+E(I_{366}^2)+2E\left(\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_j\right) \\ =&\sum_{j=1}^{366} p_j + 2\sum_{j=1}^{366}(j-1)E(I_j) \\ =&\sum_{j=1}^{366}(2j-1)p_j. \end{aligned} \tag{6}$$

Thus, we have

$$D(X) = E(X^2) - [E(X)]^2 \approx 148.64. \tag{7}$$

Probability & Statistics for EECS: Homework #5 Solutions

Professor Ziyu Shao

Problem 1

A building has n floors, labeled $1, 2, \ldots, n$. At the first floor, k people enter the elevator, which is going up and is empty before they enter. Independently, each decides which of floors $2, 3, \ldots, n$ to go to and presses that button (unless someone has already pressed it).
(a) Assume for this part only that the probabilities for floors $2, 3, \ldots, n$ are equal. Find the expected number of stops the elevator makes on floors $2, 3, \ldots, n$.
(b) Generalize (a) to the case that floors $2, 3, \ldots, n$ have probabilities $p_2, \ldots, p_n$ (respectively); you can leave your answer as a finite sum.
Solution:

(a) Let X be the number of stops. $X = X_2 + \ldots + X_n$, where $X_i = 1$ if someone stops at ith floor, otherwise $X_i = 0$.

$$\begin{aligned} E(X_i) &= 1 \cdot \mathrm{P}(\text{at least one person stop at } i\text{th floor}) \\ &= 1 - (\frac{n-2}{n-1})^k. \end{aligned}$$

Thus,

$$\begin{aligned} E(X) &= E(X_2 + \cdots + X_n) \\ &= E(X_2) + \cdots + E(X_n) \\ &= (n-1)\left[1 - (\frac{n-2}{n-1})^k\right] \end{aligned}$$

(b) According to (a)

$$\begin{aligned} E(X_i) &= 1 \cdot \mathrm{P}(\text{at least one person stop at } i\text{th floor}) \\ &= 1 - (1-p_i)^k. \end{aligned}$$

Thus, we have

$$E(X) = \sum_{i=2}^{n} 1 - (1-p_i)^k = n - 1 - (\sum_{i=2}^{n} (1-p_i)^k)$$

Problem 2

Given a six-sided dice, let X denote the number obtained by rolling the dice one time. The PMF of X is:

$$P(X=1)=P(X=2)=\frac{1}{7},\quad P(X=3)=P(X=4)=\frac{1}{5},\quad P(X=5)=\frac{2}{35},\quad P(X=6)=\frac{9}{35}.$$

Now the dice is rolled five times independently. What is more likely: a sum of 24 or a sum of 25?
Method1:PGF The PGF is:

$$E(t^{X_1})=\sum_{k=0}^{6}P(X_1=k)t^k=\frac{1}{7}t+\frac{1}{7}t^2+\frac{1}{5}t^3+\frac{1}{5}t^4+\frac{2}{35}t^5+\frac{9}{35}t^6.$$

$$E(t^{X_2}),\cdots,E(t^{X_5}) \text{ are the same as } E(t^{X_1}). \text{ Thus,}$$

$$E(t^X)=\prod_{i=1}^{5}E(t^{X_i})=\left(\frac{t+t^2}{7}+\frac{t^3+t^4}{5}+\frac{2}{35}t^5+\frac{9}{35}t^6\right)^5$$

Then the probability of get 24 and 25 can be obtained via calculating the coefficient of t^{24} and t^{25}. Finally their numerical values are:

$$P(X=24)=\frac{375916}{10504375}\approx 0.0358$$
$$P(X=25)=\frac{1340537}{52521875}\approx 0.0255$$

Thus, we know that a sum of 24 is more likely.
Method2:Enumerate and Comparison
Considering the fact that 24 and 25 are relative big numbers for 5 dice rolling, we can enumerate all their possible cases.
For 24, there are:

$$(4,5,5,5,5),(4,4,5,5,6),(4,4,4,6,6),(3,5,5,5,6),(3,4,5,6,6),(3,3,6,6,6),(2,5,5,6,6),(2,4,6,6,6),(1,5,6,6,6)$$

For 25, there are:

$$(5,5,5,5,5),(4,5,5,5,6),(4,4,5,6,6),(3,5,5,6,6),(3,4,6,6,6),(2,5,6,6,6),(1,6,6,6,6)$$

Then we compare the sub-cases of them by pairs.

$$\begin{aligned}
(4,5,5,5,5) &\quad > (5,5,5,5,5)\\
(4,4,5,5,6) &\quad > (4,5,5,5,6)\\
(4,4,4,6,6) &\quad > (4,4,5,6,6)\\
(3,5,5,5,6) &\\
(3,4,6,6,6) &\quad > (3,5,5,6,6)\\
(3,3,6,6,6) &\quad = (3,4,6,6,6)\\
(2,5,6,6,6) &\\
(2,4,6,6,6) &\quad > (2,5,6,6,6)\\
(1,5,6,6,6) &\\
&\quad\ \ (1,6,6,6,6)
\end{aligned}$$

Finally it's not hard to obtain that $P(\text{Pair}((3,5,5,5,6)))+P(\text{Pair}((2,5,5,6,6)))+P(\text{Pair}((1,5,6,6,6)))>P(\text{Pair}((1,6,6,6,6)))$. So, the sum of 24 is more likely.

Problem 3

Given a random variable $X \sim \text{Pois}(\lambda)$ where $\lambda > 0$, show that for any non-negative integer k, we have the following identity:

$$E\left[\binom{X}{k}\right] = \frac{\lambda^k}{k!}.$$

solution:
via LOTUS, we have

$$\begin{aligned}
E\left(X_k\right) &= \sum_{i=0}^{\infty} \binom{i}{k} \cdot \frac{\lambda^i}{i!} e^{-\lambda} \\
&= \sum_{i=0}^{\infty} \frac{i!}{k!(i-k)!} \cdot \frac{\lambda^i}{i!} e^{-\lambda} \\
&= \sum_{i=k}^{\infty} \frac{\lambda^i}{k!(i-k)!} e^{-\lambda} \\
&= \sum_{i=k}^{\infty} \frac{i!}{k!(i-k)!} \cdot \frac{\lambda^i}{i!} e^{-\lambda} \\
&= \frac{e^{-\lambda}}{k!} \sum_{i=k}^{\infty} \frac{\lambda^i}{(i-k)!} \\
&= \frac{e^{-\lambda}}{k!} \lambda^k \sum_{t=0}^{\infty} \frac{\lambda^t}{t!} \\
&= \frac{e^{-\lambda}}{k!} \lambda^k \cdot e^{\lambda} \\
&= \frac{\lambda^k}{k!}
\end{aligned} \tag{1}$$

Problem 4

(a) Use LOTUS to show that for $X \sim \text{Pois}(\lambda)$ and any function g, $E(Xg(X)) = \lambda E(g(X+1))$. This is called the *Stein-Chen identity* for the Poisson.
(b) Find the moment $E(X^4)$ for $X \sim \text{Pois}(\lambda)$ by using the identity from (a) and a bit of algebra to reduce the calculation with the fact that X has mean λ and variance λ.
solution:

(a) From $X \sim \text{Poisson}(\lambda)$ we have $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, $k \in \mathbb{N}$. Denote $f(x) = Xg(X)$, we have

$$\begin{aligned}
E[Xg(X)] &= \sum_{x=0}^{+\infty} f(x)P(X = x) \\
&= \sum_{x=0}^{+\infty} xg(x)\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \lambda \sum_{x=0}^{+\infty} g(x)\frac{\lambda^{(x-1)} e^{-\lambda}}{(x-1)!}
\end{aligned}$$

Denote $Y = X - 1$, we have

$$\begin{aligned}
E[Xg(X)] &= \lambda \sum_{x=0}^{+\infty} g(x)\frac{\lambda^{(x-1)} e^{-\lambda}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{+\infty} g(y+1)\frac{\lambda^{(y)} e^{-\lambda}}{(y)!} \\
&= \lambda E(g(Y)) \\
&= \lambda E(g(X+1))
\end{aligned}$$

(b) Let $g(X) = X^4$

$$\begin{aligned}
E(X^4) &= \lambda E[(X+1)^3] \\
&= \lambda[E(X^3) + 3E(X^2) + 3E(X) + 1] \\
&= \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda
\end{aligned}$$

Problem 5

Suppose a fair coin is tossed repeatedly, and we obtain a sequence of H and T (H denotes Head and T denotes Tail). Let N denote the number of tosses to observe the first occurrence of the pattern "HH". Find the PMF of N.

Solution:

Denote $p_k = P(N = k)$, $p_0 = p_1 = 0$, $p_2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$, $p_3 = \left(1 - \frac{1}{2}\right)\left(\frac{1}{2}\right)^2 = \frac{1}{8}$.

Via First-Step analysis, we can know that for $k \geq 3$, there is

$$p_k = P(N = k) = P(N = k, S_1 = H) + P(N = k, S_1 = T)$$

thus we have

$$p_k = \frac{1}{4}p_{k-2} + \frac{1}{2}p_{k-1}$$

Via solving the characteristic function of the above recursion, $x^2 = \frac{1}{2}x + \frac{1}{4}$, we can obtain the two roots:

$$r_1 = \frac{1+\sqrt{5}}{4}, r_2 = \frac{1-\sqrt{5}}{4} \tag{2}$$

thus we can obtain the solution for p_k as follows:

$$p_k = m\left(\frac{1+\sqrt{5}}{4}\right)^k + n\left(\frac{1-\sqrt{5}}{4}\right)^k$$

substitute $p_3 = \frac{1}{8}$ and $p_4 = \frac{1}{8}$, $m = \frac{5-\sqrt{5}}{10}$ and $n = \frac{5+\sqrt{5}}{10}$

Finally, the PMF is

$$P(N = k) = p_k = \frac{5-\sqrt{5}}{10}\left(\frac{1+\sqrt{5}}{4}\right)^k + \frac{5+\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{4}\right)^k$$

Problem 6

Suppose a fair coin is tossed repeatedly, and we obtain a sequence of H and T (H denotes Head and T denotes Tail). Let N denote the number of tosses to observe the first occurrence of the pattern "HTHT". Find $E(N)$ and $\mathrm{Var}(N)$.

Solution:

Denote S_k as The kth time we toss the coin, and let $p_k = P(N = k)$ for simplicity.

It's easy to know that:

$$p_0 = 0, \quad p_1 = 0, \quad p_2 = 0, \quad p_3 = 0, \quad p_4 = \frac{1}{16}$$

For $k \geq 5$, via LOTP, we have

$$p_k = P(N = k) = P(N = k, S_1 = H) + P(N = k, S_1 = T)$$

where $P(N = k, S_1 = H)$ can be further decompsed as follows

$$\begin{aligned}
P(N = k, S_1 = H) &= P(N = k, S_1 = H, S_2 = H) + P(N = k, S_1 = H, S_2 = T) \\
&= \frac{1}{2}P(N = k-1, S_1 = H) + P(N = k, S_1 = H, S_2 = T, S_3 = H) + P(N = k, S_1 = H, S_2 = T, S_3 = T) \\
&= \frac{1}{2}P(N = k-1, S_1 = H) + P(N = k, S_1 = H, S_2 = T, S_3 = H) + \frac{1}{8}P(N = k-3) \\
&= \frac{1}{2}P(N = k-1, S_1 = H) + \frac{1}{8}P(N = k-3, S_1 = H) + \frac{1}{8}P(N = k-3) \\
&= \frac{1}{2}(p_{k-1} - \frac{1}{2}p_{k-2}) + \frac{1}{8}(p_{k-3} - \frac{1}{2}p_{k-4}) + \frac{1}{8}p_{k-3} \\
&= \frac{1}{2}p_{k-1} - \frac{1}{4}p_{k-2} + \frac{1}{4}p_{k-3} - \frac{1}{16}p_{k-4}
\end{aligned}$$

Therefore we have:

$$p_k = p_{k-1} - \frac{1}{4}p_{k-2} + \frac{1}{4}p_{k-3} - \frac{1}{16}p_{k-4} \tag{3}$$

Denote $g(t)$ as the PGF of N, then we have:

$$g(t) = E[t^N] = \sum_{k=0}^{\infty} p_k t^k = \sum_{k=4}^{\infty} p_k t^k = \frac{1}{16}t^4 + \sum_{k=5}^{\infty} p_k t^k$$

Using the result from equ.3, we have

$$\begin{aligned}
\sum_{k=5}^{\infty} p_k t^k &= \sum_{k=5}^{\infty} \left(p_{k-1} - \frac{1}{4}p_{k-2} + \frac{1}{4}p_{k-3} - \frac{1}{16}p_{k-4}\right) t^k \\
&= \sum_{k=5}^{\infty} p_{k-1}t^k - \frac{1}{4}\sum_{k=5}^{\infty} p_{k-2}t^k + \frac{1}{4}\sum_{k=5}^{\infty} p_{k-3}t^k - \frac{1}{16}\sum_{k=5}^{\infty} p_{k-4}t^k \\
&= t\sum_{k=5}^{\infty} p_{k-1}t^{k-1} - \frac{1}{4}t^2\sum_{k=5}^{\infty} p_{k-2}t^{k-2} + \frac{1}{4}t^3\sum_{k=5}^{\infty} p_{k-3}t^{k-3} - \frac{1}{16}t^4\sum_{k=5}^{\infty} p_{k-4}t^{k-4} \\
&= t \cdot g(t) - \frac{1}{4}t^2 \cdot g(t) + \frac{1}{4}t^3 \cdot g(t) - \frac{1}{16}t^4 \cdot g(t)
\end{aligned} \tag{4}$$

Thus we have

$$\Rightarrow g(t) = \frac{\frac{1}{16}t^4}{1 - t + \frac{1}{4}t^2 - \frac{1}{4}t^3 + \frac{1}{16}t^4} = \frac{t^4}{16 - 16t + 4t^2 - 4t^3 + t^4}$$

$$g'(t) = \frac{64t^3 - 48t^4 + 8t^5 - 4t^6}{(16 - 16t + 4t^2 - 4t^3 + t^4)^2}$$

$$g''(t) = \frac{8t^2(t^4 - 4t^3 + 4(t-2)^2)(t^7 - 3t^6 + 24t^5 - 52t^4 + 240t^2 - 512t + 384)}{(t^8 - 8t^7 + 24t^6 - 64t^5 + 176t^4 - 256t^3 + 384t^2 - 512t + 256)^2}$$

The numerical value of them are $g'(1) = 20, \quad g''(1) = 656$. Finally, we can obtain that

$$E(N) = g'(1) = 20$$

$$\mathrm{Var}(N) = E(N^2) - E(N)^2 = g''(1) + g'(1) - (g'(1))^2 = 276$$

Probability & Statistics for EECS: Homework #06

Due on Oct 29, 2023 at 23:59

Name:
Student ID:

Oct 22, 2023

Problem 1

The Cauchy distribution has PDF

$$f(x) = \frac{1}{\pi(1+x^2)} \tag{1}$$

for all x. Find the CDF of a random variable with the Cauchy PDF.
Hint: Recall that the derivative of the inverse tangent function $\tan^{-1}(x)$ is $\frac{1}{1+x^2}$.
Solution: Given that the PDF of the Cauchy distribution is

$$f(x) = \frac{1}{\pi(1+x^2)}, \tag{2}$$

and the hint that the derivative of the inverse tangent function $\tan^{-1}(x)$ is $\frac{1}{1+x^2}$, we can calculate the CDF of the Cauchy distribution by definition, i.e., integrating the PDF over range $(-\infty, x]$.
Therefore, the CDF of the Cauchy distribution $F(x)$ is as follows:

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\,dt = \frac{1}{\pi}\tan^{-1}(t)\Big|_{-\infty}^{x} && (3)\\ &= \frac{1}{\pi}\tan^{-1}(x) - \frac{1}{\pi}\left(-\frac{\pi}{2}\right) && (4)\\ &= \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x), && (5) \end{aligned}$$

where $x \in (-\infty, \infty)$.

Problem 2

The Pareto distribution with parameter $a > 0$ has PDF

$$f(x) = \frac{a}{x^{a+1}} \tag{6}$$

for $x \geq 1$ (and 0 otherwise). This distribution is often used in statistical modeling. Find the CDF of a Pareto random variable with parameter a; check that it is a valid CDF.

Solution: Given that the PDF of the Pareto distribution with parameter $a > 0$ is

$$f(x) = \begin{cases} \frac{a}{x^{a+1}}, & x \geq 1, \\ 0, & \text{Otherwise} \end{cases} \tag{7}$$

we can calculate the CDF of the Pareto distribution by definition, i.e., integrating the PDF over range $(-\infty, x]$.

Therefore, the CDF of the Pareto distribution $F(x)$ is as follows:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{1}^{x} \frac{a}{t^{a+1}}\, dt = -\frac{1}{t^a}\Big|_1^x = 1 - \frac{1}{x^a}, \tag{8}$$

where $x \in [1, \infty)$. When $x \in (-\infty, 1)$, by definition, $F(x) = 0$.

We then check if $F(x)$ is a valid CDF as follows:

- **Increasing**: Due to the fact that $\frac{1}{x^a}, a > 0$ is decreasing over $[1, \infty)$, CDF $F(x) = 1 - \frac{1}{x^a}$ is increasing over the corresponding support $[1, \infty)$.

- **Right-continuous**: Due to the fact that $1 - \frac{1}{x^a}, a > 0$ is continuous over $[1, \infty)$, CDF $F(x)$ is right-continuous over the corresponding support $[1, \infty)$.

- **Convergence to 0 and 1 in the limits**: Due to the fact that $F(x) = 0, x < 1$ and $\lim\limits_{x\to\infty} \frac{1}{x^a} = 0$ when $a > 0$, CDF $F(x)$ have its limits as follows:

$$\lim_{x\to-\infty} F(x) = 0, \tag{9}$$

$$\lim_{x\to\infty} F(x) = 1 - 0 = 1. \tag{10}$$

In summary, the CDF $F(x)$ is valid.

Problem 3

The Beta distribution with parameters $a = 3$, $b = 2$ has PDF

$$f(x) = 12x^2(1-x), \quad \text{for } 0 < x < 1. \tag{11}$$

Let X have this distribution.

(a) Find the CDF of X.

(b) Find $P(0 < X < 1/2)$.

(c) Find the mean and variance of X (without quoting results about the Beta distribution).

Solution:
The Beta distribution with parameters $a = 3$, $b = 2$ has PDF

$$f(x) = 12x^2(1-x), \quad \text{for } 0 < x < 1. \tag{12}$$

Let X have this distribution.

(a) The CDF of X is

$$\begin{aligned} F(X) &= \int_0^x f(t)\,dt = \int_0^x 12t^2(1-t)\,dt \\ &= \int_0^x 12t^2\,dt - \int_0^x 12t^3\,dt \\ &= 4t^3\Big|_0^x - 3t^4\Big|_0^x \\ &= x^3(4-3x), \quad \text{for } 0 < x < 1. \end{aligned}$$

(b) According to CDF $F(x)$, $P(0 < X < 1/2) = F(1/2) = \frac{5}{16}$.

(c) According to PDF, the mean of X is

$$\begin{aligned} E(X) &= \int_0^1 xf(x)\,dx = \int_0^1 12x^3(1-x)\,dx \\ &= \int_0^1 12x^3\,dx - \int_0^1 12x^4\,dx \\ &= \frac{3}{5}. \end{aligned}$$

We have

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 f(x)\,dx = \int_0^1 12x^4(1-x)\,dx \\ &= \int_0^1 12x^4\,dx - \int_0^1 12x^5\,dx \\ &= \frac{2}{5}. \end{aligned}$$

Thus, we have

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{1}{25}.$$

Problem 4

The Exponential is the analog of the Geometric in continuous time. This problem explores the connection between Exponential and Geometric in more detail, asking what happens to a Geometric in a limit where the Bernoulli trials are performed faster and faster but with smaller and smaller success probabilities.
Suppose that Bernoulli trials are being performed in continuous time; rather than only thinking about first trial, second trial, etc., imagine that the trials take place at points on a timeline. Assume that the trials are at regularly spaced times $0, \Delta t, 2\Delta t, \ldots$, where Δt is a small positive number. Let the probability of success of each trial be $\lambda\Delta t$, where λ is a positive constant. Let G be the number of failures before the first success (in discrete time), and T be the time of the first success (in continuous time).

(a) Find a simple equation relating G to T. Hint: Draw a timeline and try out a simple example.

(b) Find the CDF of T. Hint: First find $P(T > t)$.

(c) Show that as $\Delta t \to 0$, the CDF of T converges to the Expo(λ) CDF, evaluating all the CDFs at a fixed $t \geq 0$.

Solution:

1. The Bernoulli trials occur at times $0, \Delta t, 2\Delta t, \ldots$. G is the number of failures before the first success, so the first success occurs at the $(G+1)$-th trial. Since each trial is spaced by Δt units of time, the time T of the first success is:

$$T = G\Delta t$$

2. The trials occur at times $0, \Delta t, 2\Delta t, \ldots$. The number of trials conducted up to time t is $n = \left\lceil \dfrac{t}{\Delta t} \right\rceil$. The probability of success in each trial is $p = \lambda\Delta t$. The probability of failure in each trial is $1-p = 1-\lambda\Delta t$. Therefore, the probability of n consecutive failures is:

$$P(\text{No success in } n \text{ trials}) = (1-\lambda\Delta t)^n$$

Thus we have:

$$P(T > t) = (1-\lambda\Delta t)^n$$

So the CDF of T is

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - (1-\lambda\Delta t)^n$$

3. The CDF of T can be rewrite as follows:

$$\begin{aligned} P(T > t) &= (1-\lambda\Delta t)^{\lceil \frac{t}{\Delta t} \rceil} \\ &= (1-\lambda\Delta t)^{\frac{t}{\Delta t} + (\lceil \frac{t}{\Delta t} \rceil - \frac{t}{\Delta t})} \\ &= (1-\lambda\Delta t)^{\frac{t}{\Delta t} + d} \end{aligned} \tag{13}$$

where $d = \lceil \frac{t}{\Delta t} \rceil - \frac{t}{\Delta t}$.
Note that

$$\lim_{\Delta t \to 0} P(T > t) = e^{\lambda t}(1-\lambda\Delta t)^d \tag{14}$$

We know that $\lim_{\Delta t \to 0} d = 0$, therefore we have $\lim_{\Delta t \to 0} F_T(t) = 1 - \lim_{\Delta t \to 0} P(T > t) = 1 - e^{-\lambda t}$, thus $P(T \leq t)$ converges to $1 - e^{-\lambda t}$, which is the CDF of Exponential Distribution.

Problem 5

The *Gumbel distribution* is the distribution of $-logX$ with $X \sim \text{Expo}(1)$.

(a) Find the CDF of the Gumbel distribution.

(b) Let $X_1, X_2, \ldots$ be i.i.d. $Expo(1)$ and let $M_n = max(X_1, \ldots, X_n)$. Show that $M_n - \log n$ converges in distribution to the Gumbel distribution, i.e., as $n \to \infty$ the CDF of $Mn - \log n$ converges to the Gumbel CDF.

Solution:

(a) Let G be Gumbel and $X \sim \text{Expo}(1)$. The CDF of G is

$$\begin{aligned} P(G \leq t) &= P(-\log X \leq t) \\ &= P(X \geq e^{-t}) \\ &= e^{-e^{-t}}. \end{aligned}$$

(b) CDF of $M_n - \log n$ can be written as

$$P(M_n - \log n \leq t) = P(X_1 \leq t + \log n, X_2 \leq t + \log n, ..., X_n \leq t + \log n) = P(X_1 \leq t + \log n)^n.$$

According to (a) and $\lim_{n\to\infty}(1 + \frac{x}{n})^n = e^x$. Thus,

$$\begin{aligned} \lim_{n\to\infty} P(M_n - \log n \leq t) &= \lim_{n\to\infty} (1 - e^{-(t+\log n)})^n \\ &= \lim_{n\to\infty} (1 - \frac{e^{-t}}{n})^n \\ &= e^{-e^{-t}}. \end{aligned}$$

Probability & Statistics for EECS: Homework #7 Solutions

Professor Ziyu Shao

Problem 1

	Y **discrete**	Y **continuous**
X **discrete**	$P(Y=y\|X=x)=\frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x)=\frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X **continuous**	$P(Y=y\|X=x)=\frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x)=\frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

- X discrete, Y discrete: we can directly write it via Bayes Rule
- X continuous, Y continuous: According to Bayes Rule, we have

  $$P(Y\in(y-\varepsilon,y+\varepsilon)|X\in(x-\varepsilon,x+\varepsilon))=\frac{P(X\in(x-\varepsilon,x+\varepsilon)|Y\in(y-\varepsilon,y+\varepsilon))P(Y\in(y-\varepsilon,y+\varepsilon))}{P(X\in(x-\varepsilon,x+\varepsilon))}$$

  By letting $\varepsilon\to 0$, we have

  $$\begin{aligned}&\lim_{\varepsilon\to 0}P(Y\in(y-\varepsilon,y+\varepsilon)|X\in(x-\varepsilon,x+\varepsilon))=f_{Y|X}(y|x)2\varepsilon\\&=\lim_{\varepsilon\to 0}\frac{P(X\in(x-\varepsilon,x+\varepsilon)|Y\in(y-\varepsilon,y+\varepsilon))P(Y\in(y-\varepsilon,y+\varepsilon))}{P(X\in(x-\varepsilon,x+\varepsilon))}=\frac{f_{X|Y}(x|y)f_Y(y)2\varepsilon}{f_X(x)}\\&\Rightarrow\ f_{Y|X}(y=x)=\frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}\end{aligned}$$

- X discrete, Y continuous:
  According to the continuous Bayes' rule, we have

  $$P(Y\in(y-\varepsilon,y+\varepsilon)|X=x)=\frac{P(X=x|Y\in(y-\varepsilon,y+\varepsilon)P(Y\in(y-\varepsilon,y+\varepsilon))}{P(X=x)}.$$

  By letting $\varepsilon\to 0$, we have

  $$\lim_{\varepsilon\to 0}\ P(Y\in(y-\varepsilon,y+\varepsilon)|X=x)=\lim_{\varepsilon\to 0}\ f_Y(y|X=x)\cdot 2\varepsilon,$$

  and

  $$\lim_{\varepsilon\to 0}\ \frac{P(X=x|Y\in(y-\varepsilon,y+\varepsilon)P(Y\in(y-\varepsilon,y+\varepsilon))}{P(X=x)}=\lim_{\varepsilon\to 0}\ \frac{P(X=x|Y=y)f_Y(y)\cdot 2\varepsilon}{P(X=x)}.$$

  Therefore, we can finish the proof by canceling the term 2ε in the following equation:

  $$\begin{aligned}\lim_{\varepsilon\to 0}\ f_Y(y|X=x)\cdot 2\varepsilon&=\lim_{\varepsilon\to 0}\ \frac{P(X=x|Y=y)f_Y(y)\cdot 2\varepsilon}{P(X=x)}\\\Rightarrow f_Y(y|X=x)&=\frac{P(X=x|Y=y)f_Y(y)}{P(X=x)}.\end{aligned}$$

- X continuous, Y discrete:

  $$\begin{aligned}P(Y=y|X=x)&=\lim_{\varepsilon\to 0}\ P(Y=y|X\in(x-\varepsilon,x+\varepsilon))\\&=\lim_{\varepsilon\to 0}\ \frac{P(X\in(x-\varepsilon,x+\varepsilon)|Y=y)P(Y=y)}{P(X\in(x-\varepsilon,x+\varepsilon))}\\&=\lim_{\varepsilon\to 0}\ \frac{2\varepsilon\cdot f_X(x|Y=y)P(Y=y)}{2\varepsilon\cdot f_X(x)}\\&=\frac{f_X(x|Y=y)P(Y=y)}{f_X(x)}\end{aligned}$$

	Y **discrete**	Y **continuous**
X **discrete**	$P(X=x)=\sum_y P(X=x \vert Y=y)P(Y=y)$	$P(X=x)=\int_{-\infty}^{\infty} P(X=x \vert Y=y)f_Y(y)dy$
X **continuous**	$f_X(x)=\sum_y f_X(x \vert Y=y)P(Y=y)$	$f_X(x)=\int_{-\infty}^{\infty} f_{X \vert Y}(x \vert y)f_Y(y)dy$

- X discrete, Y continuous:

$$P(X=x|Y\in(y-\varepsilon,y+\varepsilon))=\frac{P(Y\in(y-\varepsilon,y+\varepsilon)|X=x)P(X=x)}{P(Y\in(y-\varepsilon,y+\varepsilon))}.$$

By letting $\varepsilon\to 0$, we have

$$\lim_{\varepsilon\to 0}P(X=x|Y\in(y-\varepsilon,y+\varepsilon))=P(X=x|Y=y),$$

and

$$\begin{aligned}&\lim_{\varepsilon\to 0}\frac{P(Y\in(y-\varepsilon,y+\varepsilon)|X=x)P(X=x)}{P(Y\in(y-\varepsilon,y+\varepsilon))}\\ &=\lim_{\varepsilon\to 0}\frac{f_Y(y|X=x)\cdot 2\varepsilon\cdot P(X=x)}{f_Y(y)\cdot 2\varepsilon}\\ &=\frac{f_Y(y|X=x)P(X=x)}{f_Y(y)}.\end{aligned}$$

By combining the two equations, we can get

$$\begin{aligned}P(X=x|Y=y)&=\frac{f_Y(y|X=x)P(X=x)}{f_Y(y)}\\ \Rightarrow P(X=x|Y=y)f_Y(y)&=f_Y(y|X=x)P(X=x).\end{aligned}$$

By integrating on both sides of the equation with respect y, we can get

$$\begin{aligned}\int_{-\infty}^{\infty}P(X=x|Y=y)f_Y(y)dy&=\int_{-\infty}^{\infty}f_Y(y|X=x)P(X=x)dy\\ &=P(X=x)\int_{-\infty}^{\infty}f_Y(y|X=x)dy\\ &=P(X=x).\end{aligned}$$

- X continuous, Y discrete:

$$P(X\in(x-\varepsilon,x+\varepsilon))=\sum_y P(X\in(x-\varepsilon,x+\varepsilon)|Y=y)P(Y=y).$$

By letting $\varepsilon\to 0$, we have

$$\lim_{\varepsilon\to 0}P(X\in(x-\varepsilon,x+\varepsilon))=\lim_{\varepsilon\to 0}f_X(x)\cdot 2\varepsilon,$$

and

$$\lim_{\varepsilon\to 0}\sum_y P(X\in(x-\varepsilon,x+\varepsilon)|Y=y)P(Y=y)=\lim_{\varepsilon\to 0}\sum_y f_X(x|Y=y)\cdot 2\varepsilon\cdot P(Y=y).$$

By combining the two equations, we can get

$$\begin{aligned}\lim_{\varepsilon\to 0}f_X(x)\cdot 2\varepsilon&=\lim_{\varepsilon\to 0}\sum_y f_X(x|Y=y)\cdot 2\varepsilon\cdot P(Y=y)\\ f_X(x)&=\sum_y f_X(x|Y=y)P(Y=y).\end{aligned}$$

- X continuous, Y continuous:

  According to the continuous version Bayes rule in problem(a) we have

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy = f_X(x) \int_{-\infty}^{\infty} f_{Y|X}(y|x) dy = f_X(x)$$

Problem 2

A chicken lays a Pois(λ) number N of eggs. Each egg hatches a chick with probability p, independently. Let X be the number which hatch, and Y be the number which do NOT hatch.

(a) Find the joint PMF of N, X, Y. Are they independent?

(b) Find the joint PMF of N, X. Are they independent?

(c) Find the joint PMF of X, Y. Are they independent?

(d) Find the correlation between N (the number of eggs) and X (the number of eggs which hatch). Simplify; your final answer should work out to a simple function of p (the λ should cancel out).

Solution

Using the chicken-egg story, we can obtain that X is distributed Pois($p\lambda$) and Y similarly Pois($q\lambda$) with $q = 1 - p$ and that these random variables are independent!

(a) For non-negative integer i, j, n, if $i + j \neq n$, $P(X = i, Y = j, N = n) = 0$.
If $i + j = n$, then

$$P(X = i, Y = j, N = n) = P(X = i, Y = j \mid N = n)P(N = n) = \binom{n}{i} p^i q^{n-i} \cdot \frac{\lambda^n}{n!} e^{-\lambda}.$$

X, Y, and N are not independent because we have that for $i, j, n > 0$ such that $i + j \neq n$

$$P(X = i, Y = j, N = n) = 0.$$

But obviously we have that

$$P(X = i)P(Y = j)P(N = n) > 0.$$

(b) For $n \geq i \geq 0$,

$$P(X = i, N = n) = P(X = i \mid N = n)P(N = n) = \binom{n}{i} p^i q^{n-i} \cdot \frac{\lambda^n}{n!} e^{-\lambda}, n \geq i \geq 0.$$

Otherwise, $P(X = i, N = n) = 0$.

X and N are not independent since from the story, $X \sim \text{Pois}(p\lambda)$, then we have

$$P(X = i)P(N = n) = \frac{(\lambda p)^i}{i!} e^{-\lambda p} \cdot \frac{\lambda^n}{n!} e^{-\lambda}$$

which is obviously not equal to the joint PMF. (We could also see this by observing that for $i > n$ we have that $P(X = i, N = n) = 0$.)

(c) As we know from the chicken-egg story, we have that X and Y are independent, so the joint distribution is

$$P(X = i, Y = j) = P(X = i)P(Y = j) = \frac{(\lambda p)^i}{i!} e^{-\lambda p} \frac{(\lambda q)^j}{j!} e^{-\lambda q}, i, j \geq 0.$$

(d) By the property of covariance,

$$\text{Cov}(N, X) = \text{Cov}(X + Y, X) = \text{Cov}(X, X) + \text{Cov}(X, Y) = \text{Var}(X) = \lambda p$$

Since $N \sim \text{Pois}(\lambda), \text{Var}(N) = \lambda$, we have

$$\text{Corr}(N, X) = \frac{\text{Cov}(N, X)}{\sqrt{\text{Var}(N)\text{Var}(X)}} = \frac{\lambda p}{\sqrt{\lambda \cdot \lambda p}} = \sqrt{p}.$$

Problem 3

Let X and Y be i.i.d. Expo(λ), and $T = X + Y$.

(a) Find the conditional CDF of T given $X = x$. Be sure to specify where it is zero.

(b) Find the conditional PDF $f_{T|X}(t \mid x)$, and verify that it is a valid PDF.

(c) Find the conditional PDF $f_{X|T}(x \mid t)$, and verify that it is a valid PDF.

(d) In Example 8.2.4, we will show that the marginal PDF of T is $f_T(t) = \lambda^2 t e^{-\lambda t}$, for $t > 0$. Give a short alternative proof of this fact, based on the previous parts and Bayes' rule.

Solution

(a)

$$F_{T|X}(t|x) = P(T \le t|X = x) = P(X + Y \le t|X = x) = P(Y \le t - x) = \left(1 - e^{-\lambda(t-x)}\right) \cdot \chi_{t\ge x}.$$

P.S., view $\chi_{\{\cdot\}}$ as the indicator function $\mathbb{I}\{\cdot\}$.

(b) Take derivative from $F_{T|X}$ respect to t.

$$f_{T|X}(t|x) = \frac{\partial}{\partial t} F_{T|X}(t|x) = \frac{\partial}{\partial t}\left[\left(1 - e^{-\lambda(t-x)}\right) \cdot \chi_{t\ge x}\right] = \lambda e^{-\lambda(t-x)} \cdot \chi_{t\ge x}.$$

- Non-negativity:

$$f_{T|X}(t|x) = \lambda e^{-\lambda(t-x)} \cdot \chi_{t\ge x} = \begin{cases} 0 \ge 0, & t < x \\ \lambda e^{-\lambda(t-x)} \ge 0, & t \ge x \end{cases}$$

- Integrates to 1:

$$\int_{-\infty}^{\infty} f_{T|X}(t|x)dt = \int_x^{\infty} \lambda e^{-\lambda(t-x)}dt = -\, e^{-\lambda(t-x)}\Big|_{t=x}^{t=\infty} = 1$$

Therefore, $f_{T|X}(t|x)$ is valid PDF.

(c)

$$\begin{aligned} f_{X|T}(x|t) =& \frac{f_{T|X}(t|x) f_X(x)}{f_T(t)} \\ =& \frac{1}{f_T(t)} \lambda e^{-\lambda(t-x)} \cdot \lambda e^{-\lambda x} \cdot \chi_{t\ge x} \\ =& \frac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \cdot \chi_{t\ge x} \end{aligned}$$

- Non-negativity:

$$f_{X|T}(x|t) = \frac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \cdot \chi_{t\ge x} = \begin{cases} 0 \ge 0, & t < x \\ \frac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \ge 0, & t \ge x \end{cases}$$

- Integrates to 1: Note that $f_{X|T}(x|t) = \frac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \cdot \chi_{t\ge x}$ is constant with respect to x. In particular, $f_{X|T}(x|t)$ is a non-zero constant respect to x over support $(0, t)$ and zero otherwise. Via integrate $f_{T|X}(t|x) f_X(x)$ on x, we have $f_T(t) = \int_0^t \lambda^2 e^{-\lambda t} dx = t\lambda^2 e^{-\lambda t}$, thus we have $f_{X|T}(x|t) = \frac{1}{t}, x \in (0, t)$, hence a valid PDF $f_{X|T}(x|t)$ over support $(0, t)$.

(d) Recall in part (c) that $X|T = t \sim \text{Unif}(0, t)$, hence $f_{X|T}(x|t) = \frac{1}{t} \cdot \chi_{t\ge x}$. Therefore, we have

$$f_T(t) = \frac{f_{T|X}(t|x) f_X(x)}{f_{X|T}(x|t)} = \frac{\lambda e^{-\lambda(t-x)} \cdot \chi_{t\ge x} \cdot \lambda e^{-\lambda x}}{\frac{1}{t} \cdot \chi_{t\ge x}} = \lambda^2 t e^{-\lambda t}, t > 0.$$

Problem 4

Let U_1, U_2, U_3 be i.i.d. Unif$(0,1)$, and let $L = \min(U_1, U_2, U_3)$, $M = \max(U_1, U_2, U_3)$.

(a) Find the marginal CDF and marginal PDF of M, and the joint CDF and joint PDF of L, M.
Hint: For the latter, start by considering $P(L \geq l, M \leq m)$.

(b) Find the conditional PDF of M given L.

Solution

(a) The event $M \leq m$ is the same as the event that all 3 of the U_j are at most m, so the CDF of M is $F_M(m) = m^3$ and the PDF is $f_M(m) = 3m^2$, for $0 \leq m \leq 1$. The event $L \geq l$, $M \leq m$ is the same as the event that all 3 of the U_j are between l and m (inclusive), so

$$P(L \geq l, M \leq m) = (m-l)^3$$

for $m \geq l$ with $m, l \in [0,1]$. By the axioms of probability, we have

$$P(M \leq m) = P(L \leq l, M \leq m) + P(L > l, M \leq m)$$

So the joint CDF is

$$P(L \leq l, M \leq m) = m^3 - (m-l)^3,$$

for $m \geq l$ with $m, l \in [0,1]$. The joint PDF is obtained by differentiating this with respect to l and then with respect to m (or vice versa):

$$f(l,m) = 6(m-l),$$

for $m \geq l$ with $m, l \in [0,1]$. As a check, note that getting the marginal PDF of M by finding $\int_0^m f(l,m)dl$ does recover the PDF of M (the limits of integration are from 0 to m since the min can't be more than the max).

(b) The marginal PDF of L is $f_L(l) = 3(1-l)^2$ for $0 \leq l \leq 1$ since $P(L > l) = P(U_1 > l, U_2 > l, U_3 > l) = (1-l)^3$ (alternatively, use the PDF of M together with the symmetry that $1 - U_j$ has the same distribution as U_j, or integrate out m in the joint PDF of L, M). So the conditional PDF of M given L is

$$f_{m|L}(m|l) = \frac{f(l,m)}{f_L(l)} = \frac{2(m-1)}{(1-l)^2},$$

for all $m, l \in [0,1]$ with $m \geq l$.

Problem 5

This problem explores a visual interpretation of covariance. Data are collected for $n \geq 2$ individuals, where for each individual two variables are measured (e.g., height and weight). Assume independence across individuals (e.g., person l's variables gives no information about the other people), but not within individuals (e.g., a person's height and weight may be correlated).

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the n data points. The data are considered here as fixed, known numbers-they are the observed values after performing an experiment. Imagine plotting all the points (x_i, y_i) in the plane, and drawing the rectangle determined by each pair of points. For example, the points $(1,3)$ and $(4,6)$ determine the rectangle with vertices $(1,3), (1,6), (4,6), (4,3)$.

The signed area contributed by (x_i, y_i) and (x_j, y_j) is the area of the rectangle they determine if the slope of the line between them is positive, and is the negative of the area of the rectangle they determine if the slope of the line between them is negative. (Define the signed area to be 0 if $x_i = x_j$ or $y_i = y_j$, since then the rectangle is degenerate.) So the signed area is positive if a higher x value goes with a higher y value for the pair of points, and negative otherwise. Assume that the x_i are all distinct and the y_i are all distinct.

(a) The sample covariance of the data is defined to be

$$r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

are the sample means. (There are differing conventions about whether to divide by $n-1$ or n in the definition of sample covariance, but that need not concern us for this problem.)

Let (X, Y) be one of the (x_i, y_i) pairs, chosen uniformly at random. Determine precisely how $\operatorname{Cov}(X, Y)$ is related to the sample covariance.

(b) Let (X, Y) be as in (a), and $(\tilde{X}, \tilde{Y})$ be an independent draw from the same distribution. That is, (X, Y) and $(\bar{X}, \tilde{Y})$ are randomly chosen from the n points, independently (so it is possible for the same point to be chosen twice).

Express the total signed area of the rectangles as a constant times $E((X - \bar{X})(Y - \tilde{Y}))$. Then show that the sample covariance of the data is a constant times the total signed area of the rectangles.

Hint: Consider $E((X - \tilde{X})(Y - \tilde{Y}))$ in two ways: as the average signed area of the random rectangle formed by (X, Y) and $(\bar{X}, \bar{Y})$, and using properties of expectation to relate it to $\operatorname{Cov}(X, Y)$. For the former, consider the n^2 possibilities for which point (X, Y) is and which point $(\tilde{X}, \bar{Y})$; note that n such choices result in degenerate rectangles.

(c) Based on the interpretation from (b), give intuitive explanations of why for any r.v.s W_1, W_2, W_3 and constants a_1, a_2, covariance has the following properties:

(i) $\operatorname{Cov}(W_1, W_2) = \operatorname{Cov}(W_2, W_1)$;

(ii) $\operatorname{Cov}(a_1 W_1, a_2 W_2) = a_1 a_2 \operatorname{Cov}(W_1, W_2)$;

(iii) $\operatorname{Cov}(W_1 + a_1, W_2 + a_2) = \operatorname{Cov}(W_1, W_2)$;

(iv) $\operatorname{Cov}(W_1, W_2 + W_3) = \operatorname{Cov}(W_1, W_2) + \operatorname{Cov}(W_1, W_3)$.

Solution

(a) Since (X, Y) is chosen uniformly at random, we have

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}; \quad E(Y) = \sum_{i=1}^{n} y_i P(Y = y_i) = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}.$$

By definition, we know

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= \sum_{i=1}^{n} (x_i - E(X))(y_i - E(Y))P(X = x_i, Y = y_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = r.\end{aligned}$$

Thus we prove that $\mathrm{Cov}(X, Y)$ equals the sample covariance r.

(b)

- Denote the total signed area of the rectangles as S, then

$$S = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)(y_i - y_j).$$

  Since (X, Y) and $(\tilde{X}, \tilde{Y})$ are independent, we have

$$\begin{aligned}E((X - \tilde{X})(Y - \tilde{Y})) &= \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)(y_i - y_j)P(X = x_i, Y = y_i)P(\tilde{X} = x_j, \tilde{Y} = y_j) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)(y_i - y_j) = \frac{S}{n^2}.\end{aligned}$$

  Thusly we have $S = n^2 E((X - \tilde{X})(Y - \tilde{Y}))$.

- By the properties of expectation and considering that (X, Y) and $(\tilde{X}, \tilde{Y})$ are identically and independently sampled, we have

$$\begin{aligned}E((X - \tilde{X})(Y - \tilde{Y})) &= E(XY) - E(\tilde{X}Y) - E(X\tilde{Y}) + E(\tilde{X}\tilde{Y}) \\ &= E(XY) - E(\tilde{X})E(Y) - E(X)E(\tilde{Y}) + E(\tilde{X}\tilde{Y}) \\ &= E(XY) - E(X)E(Y) - E(\tilde{X})E(\tilde{Y}) + E(\tilde{X}\tilde{Y}) \\ &= 2[E(XY) - E(X)E(Y)] \\ &= 2\,\mathrm{Cov}(X, Y) = 2r.\end{aligned}$$

  Thusly we have $r = \frac{S}{2n^2}$.

(c)

- The claim (i) is true because it doesn't matter what is the base and what is the height of the rectangle, we can switch them.
- The claim (ii) is true because rescaling the one coordinate by the factor c yields that the total area of the rectangle rescales for c.
- The claim (iii) is true since the area of the rectangle is invariant on linear translation.
- The claim (iv) is true because the distributive property of the area: it doesn't matter if we calculate two areas with the same base and then sum them or first we add heights and then calculate the total area.

Probability & Statistics for EECS: Homework #08

Due on Dec 2, 2023 at 23:59

Name:
Student ID:

Problem 1

Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} cx^2y, & \text{if } 0 \le y \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of constant c.

(b) Find the conditional probability $P(Y \le X/4 \mid Y \le X/2)$.

Solution:

(a) According to the statement, we have

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^1 \int_0^x cx^2y\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^1 \frac{c}{2}x^4\,\mathrm{d}x \\ &= \frac{c}{10} \end{aligned} \tag{1}$$

So that, $c = 10$.

(b)

$$\begin{aligned} P\left(Y \le \frac{X}{4} \mid Y \le \frac{X}{2}\right) &= \frac{P(Y \le \frac{X}{4}, Y \le \frac{X}{2})}{P(Y \le \frac{X}{2})} \\ &= \frac{P(Y \le \frac{X}{4})}{P(Y \le \frac{X}{2})} \\ &= \frac{\int_0^1 \int_0^{\frac{x}{4}} 10x^2y\,\mathrm{d}y\,\mathrm{d}x}{\int_0^1 \int_0^{\frac{x}{2}} 10x^2y\,\mathrm{d}y\,\mathrm{d}x} \\ &= \frac{\int_0^1 \frac{x^4}{32}\,\mathrm{d}y\,\mathrm{d}x}{\int_0^1 \frac{x^4}{8}\,\mathrm{d}y\,\mathrm{d}x} \\ &= \frac{1}{4}. \end{aligned} \tag{2}$$

Problem 2

Let X and Y be two integer random variables with joint PMF

$$P_{X,Y}(x,y) = \begin{cases} \frac{1}{6 \cdot 2^{\min(x,y)}}, & \text{if } x, y \geq 0, |x-y| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the marginal distributions of X and Y.

(b) Are X and Y independent?

(c) Find $P(X = Y)$.

Solution:

(a) The marginal distributions of X is

$$P_X(x) = \sum_{y=0}^{\infty} P_{X,Y}(x,y).$$

When $x = 0$, we have

$$P_X(0) = P_{X,Y}(0,0) + P_{X,Y}(0,1) = \frac{1}{3}.$$

When $x \neq 0$, we have

$$P_X(x) = P_{X,Y}(x,x-1) + P_{X,Y}(x,x) + P_{X,Y}(x,x+1) = \frac{1}{6 \cdot 2^{x-2}}.$$

Thus, the marginal distribution of X is

$$P_X(x) = \begin{cases} \frac{1}{3}, & x = 0 \\ \frac{1}{6 \cdot 2^{x-2}}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

According to the symmetric, the marginal distribution of Y is

$$P_Y(y) = \begin{cases} \frac{1}{3}, & y = 0 \\ \frac{1}{6 \cdot 2^{y-2}}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(b) Since that

$$P_{X,Y}(0,0) = \frac{1}{6}, \tag{3}$$

and

$$P_X(0)P_Y(0) = \frac{1}{9}, \tag{4}$$

X and Y are not independent.

(c) According to symmetric, we have $P(X = Y) = P(X = Y - 1) = P(X = Y + 1)$ and $P(X = Y) + P(X = Y - 1) + P(X = Y + 1) = 1$. Thus, we have

$$P(X = Y) = \frac{1}{3}.$$

Problem 3

Let X and Y be i.i.d. $\mathcal{N}(0,1)$, and let S be a random sign (1 or -1, with equal probabilities) independent of (X,Y).

(a) Determine whether or not $(X, Y, X+Y)$ is Multivariate Normal.

(b) Determine whether or not $(X, Y, SX+SY)$ is Multivariate Normal.

(c) Determine whether or not (SX, SY) is Multivariate Normal.

Solution:

(a) Yes, $(X, Y, X+Y)$ is Multivariate Normal, because for any $a, b, c \in R$,

$$aX + bY + c(X+Y) = (a+c)X + (b+c)Y,$$

and any linear combination of independent normally distributed variables are Normal.

(b) Denote $Z = X + Y + SX + SY = (1+S)X + (1+S)Y$.
$Z = 0$ is in fact $S = -1$, hence, we have that

$$P(Z=0) = P(S=-1) = \frac{1}{2}.$$

Hence, Z is not normally distributed.

(c) Observe that random vector (X,Y) is identically distributed as (-X,-Y). So,

$$\begin{aligned}
P(SX+SY \leq k) &= P(SX+SY \leq k, S=1) + P(SX+SY \leq k, S=-1) \\
&= P(SX+SY \leq k|S=1)P(S=1) + P(SX+SY \leq k|S=-1)P(S=-1) \\
&= \frac{1}{2}P(X+Y \leq k) + \frac{1}{2}P(X+Y \geq -k) \\
&= \frac{1}{2}P(X+Y \leq k) + \frac{1}{2}P(X+Y \leq k) \\
&= P(X+Y \leq k).
\end{aligned}$$

So, (SX, SY) is equally distributed as (X, Y), and (X, Y) is Bivariate normal. Hence, (SX, SY) is Multivariate Normal.

Problem 4

Let Z_1, Z_2 be two i.i.d. random variables satisfying standard normal distributions, i.e., $Z_1, Z_2 \sim \mathcal{N}(0.1)$. Define

$$X = \sigma_X Z_1 + \mu_X;$$
$$Y = \sigma_Y(\rho Z_1 + \sqrt{1-\rho^2} Z_2) + \mu_Y,$$

where $\sigma_X > 0$, $\sigma_y > 0$, $-1 < \rho < 1$.

(a) Show that X and Y are bivariate normal.

(b) Find the correlation coefficient between X and Y, i.e., $\text{Corr}(X, Y)$.

(c) Find the joint PDF of X and Y.

Solution:

(a) For $a, b \in \mathbb{R}$, we have

$$aX + bY = (a + b\sigma_Y \rho)Z_1 + b\sqrt{1-\rho^2}\sigma_Y Z_2 + a\mu_X + b\mu_Y.$$

Since the linear combination of two Normal distribution follows Normal distribution, X and Y are bivariate normal.

(b) Since $Z_1, Z_2 \sim \mathcal{N}(0,1)$. We have $\rho Z_1 + \sqrt{1-\rho^2} Z_2 \sim \mathcal{N}(0,1)$. So $X \sim \mathcal{N}(\mu_X,)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$. Thus, we have

$$\begin{aligned}
Cov(X,Y) &= Cov(\sigma_X Z_1 + \mu_X, \sigma_Y(\rho Z_1 + \sqrt{1-\rho^2} Z_2) + \mu_Y) \\
&= \sigma_X \sigma_Y Cov(Z_1, \rho Z_1 + \sqrt{1-\rho^2} Z_2) \\
&= \sigma_X \sigma_Y \left(\rho Var(Z_1) + \sqrt{1-\rho^2} Cov(Z_1, Z_2)\right) \\
&= \sigma_X \sigma_Y \rho.
\end{aligned}$$

Then correlation coefficient between X and y is

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_X \sigma_Y \rho}{\sigma_X \sigma_Y} = \rho.$$

(c) Since Z_1 and Z_2 are i.i.d., we have

$$f_{Z_1,Z_2}(z_1, z_2) = f_{Z_1}(z_1) f_{Z_2}(z_2) = \frac{1}{2\pi} e^{-\frac{z_1^2 + z_2^2}{2}}.$$

Since $X = \sigma_X Z_1 + \mu_X$, $Y = \sigma_Y(\rho Z_1 + \sqrt{1-\rho^2} Z_2) + \mu_Y$, we have

$$Z_1 = \frac{X - \mu_X}{\sigma_X}$$

and

$$Z_2 = \frac{Y - \mu_Y}{\sqrt{1-\rho^2}\sigma_Y} - \rho \frac{X - \mu_X}{\sqrt{1-\rho^2}\sigma_X}.$$

Thus,

$$\begin{aligned}
f_{X,Y}(x,y) &= \left|\frac{\partial(z_1,z_2)}{\partial(x,y)}\right| f_{Z_1,Z_2}(z_1,z_2) \\
&= \begin{vmatrix} \frac{\partial z_1}{\partial x} & \frac{\partial z_1}{\partial y} \\ \frac{\partial z_2}{\partial x} & \frac{\partial z_2}{\partial y} \end{vmatrix} f_{Z_1,Z_2}(z_1,z_2) \\
&= \begin{vmatrix} \frac{1}{\sigma_X} & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}\sigma_X} & \frac{1}{\sqrt{1-\rho^2}\sigma_Y} \end{vmatrix} f_{Z_1,Z_2}(z_1,z_2) \\
&= \frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}} f_{Z_1,Z_2}(z_1,z_2) \\
&= \frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}} f_{Z_1,Z_2}\left(\frac{x-\mu_X}{\sigma_X}, \frac{y-\mu_Y}{\sqrt{1-\rho^2}\sigma_Y} - \rho\frac{x-\mu_X}{\sqrt{1-\rho^2}\sigma_X}\right) \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{(\frac{x-\mu_X}{\sigma_X})^2+(\frac{y-\mu_Y}{\sqrt{1-\rho^2}\sigma_Y}-\rho\frac{x-\mu_X}{\sqrt{1-\rho^2}\sigma_X})^2}{2}} \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{(\frac{x-\mu_X}{\sigma_X})^2-\frac{2\rho(x-\sigma_X)(Y-\sigma_Y)}{\sigma_X\sigma_Y}+(\frac{y-\mu_Y}{\sigma_Y})^2}{2(1-\rho^2)}}.
\end{aligned}$$

Problem 5

(a) Let X and Y be i.i.d. $\mathcal{N}(0,1)$, and $Z = \frac{X}{Y}$. Find the PDF of Z.

(b) Let X and Y be i.i.d. Unif(0, 1), $W = X \cdot Y$, and $Z = \frac{X}{Y}$. Find the joint PDF of (W, Z).

(c) A point (X, Y) is picked at random uniformly in the unit circle. Find the joint PDF of R and X, where $R = \sqrt{X^2 + Y^2}$.

(d) A point (X, Y, Z) is picked uniformly at random inside the unit ball of $\mathbb{R}^3$. Find the joint PDF of Z and R, where $R = \sqrt{X^2 + Y^2 + Z^2}$.

Solution:

1. Let $X = R\cos(\Theta)$ and $Y = R\sin(\Theta)$, where R is the radial distance from the origin to the point (X, Y), and Θ is the angle formed with the positive x-axis. The variables R and Θ are given by:

$$R = \sqrt{X^2 + Y^2}, \quad \Theta = \tan^{-1}\left(\frac{Y}{X}\right).$$

The joint PDF of X and Y, given that both are standard normal, is:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

To convert this joint PDF from Cartesian coordinates (x, y) to polar coordinates (r, θ), use the Jacobian of the transformation:

$$x = r\cos(\theta), \quad y = r\sin(\theta), \quad J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} = r.$$

Substituting the polar expressions into the original joint PDF and adjusting for the Jacobian, the new joint PDF becomes:

$$f_{R,\Theta}(r, \theta) = f_{X,Y}(r\cos(\theta), r\sin(\theta)) \cdot r = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \cdot r.$$

This expression confirms that R and Θ are independent, with R following a Rayleigh distribution with scale parameter 1 and Θ being uniformly distributed from $-\pi$ to π. Consider that the transformation $Z = \tan(\Theta)$ maps Θ to Z. To calculate the PDF of Z, we use the transformation of variables formula. The derivative of $\tan^{-1}(z)$ with respect to z is:

$$\frac{d}{dz} \tan^{-1}(z) = \frac{1}{1+z^2}.$$

This derivative represents how a small change in Z corresponds to a change in Θ, factoring into the new PDF. Combining the above derivation, the PDF of Z is given by:

$$f_Z(z) = f_\Theta(\tan^{-1}(z)) \left| \frac{d}{dz} \tan^{-1}(z) \right| = \frac{1}{\pi(1+z^2)}.$$

2. Since X and Y are i.i.d. Uniform(0, 1), the PDF of each variable, $f_X(x)$ and $f_Y(y)$, is:

$$f_X(x) = f_Y(y) = 1 \text{ for } x, y \in [0, 1].$$

Define the transformations:

$$W = X \cdot Y, \quad Z = \frac{X}{Y} \Rightarrow X = \sqrt{WZ}, \quad Y = \sqrt{\frac{W}{Z}}.$$

Problem 5 continued on next page...

Computing the partial derivatives, we have:

$$\frac{\partial x}{\partial w} = \frac{1}{2} z^{-\frac{1}{2}} w^{-\frac{1}{2}}, \quad \frac{\partial x}{\partial z} = \frac{1}{2} w^{\frac{1}{2}} z^{-\frac{3}{2}},$$

$$\frac{\partial y}{\partial w} = \frac{1}{2} z^{-\frac{1}{2}} w^{-\frac{1}{2}}, \quad \frac{\partial y}{\partial z} = -\frac{1}{2} w^{\frac{1}{2}} z^{-\frac{3}{2}}.$$

Thus, the Jacobian determinant is:

$$J = \begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial u}{\partial z} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} z^{-\frac{1}{2}} w^{-\frac{1}{2}} & \frac{1}{2} w^{\frac{1}{2}} z^{-\frac{3}{2}} \\ \frac{1}{2} z^{-\frac{1}{2}} w^{-\frac{1}{2}} & -\frac{1}{2} w^{\frac{1}{2}} z^{-\frac{3}{2}} \end{vmatrix} = -\frac{1}{2z}.$$

Therefore, the joint PDF $f_{W,Z}(w,z)$ is given by:

$$f_{W,Z}(w,z) = f_{X,Y}(x,y)|J| = 1 \cdot \left| -\frac{1}{2z} \right| = \frac{1}{2z},$$

for $x, y \in [0,1]$ (or w, z such that $0 < w \leq 1$, $z \geq w$, and $z \leq \frac{1}{w}$). The joint PDF of (W,Z), ($x, y \geq 0$ and $x, y \leq 1$), is:

$$f_{W,Z}(w,z) = \frac{1}{2z}$$

for $w \in (0,1)$ and $z \in \left(w, \frac{1}{w}\right)$.

3. Given that $R = \sqrt{X^2 + Y^2}$ and X and Y are uniformly distributed on a unit disk i.e., $x^2 + y^2 \leq 1$, we have $0 \leq r \leq 1$. The joint PDF $f_{X,Y}(x,y)$ is $\frac{1}{\pi}$ over the unit circle. Now, we have $S = X, R = \sqrt{X^2 + Y^2} \Rightarrow X = S, Y = \pm\sqrt{R^2 - S^2}$. This implies that $|s| < r < 1$. Thus, the Jacobian determinant is:

$$J = \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial r} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial r} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \pm\frac{-2s}{2\sqrt{r^2 - s^2}} & \pm\frac{2r}{2\sqrt{r^2 - s^2}} \end{vmatrix} = \pm\frac{r}{\sqrt{r^2 - s^2}}.$$

Therefore, we have the joint distribution:

$$\begin{aligned} f_{R,S}(r,s) &= f_{X,Y}(x,y)|J| \\ &= f_{X,Y}(s, \sqrt{r^2 - s^2}) \frac{r}{\sqrt{r^2 - s^2}} + f_{X,Y}(s, -\sqrt{r^2 - s^2}) \frac{r}{\sqrt{r^2 - s^2}} \\ &= \frac{2r}{\pi\sqrt{r^2 - s^2}}, \quad |s| < r < 1, \quad -1 < s < 1. \end{aligned}$$

4. The point (X, Y, Z) is chosen uniformly within the unit ball, which implies that the probability density function (PDF) for (X, Y, Z) is constant inside the ball and zero outside. The volume of the unit ball in $\mathbb{R}^3$ is $\frac{4}{3}\pi$, so the uniform density inside the ball is $\frac{3}{4\pi}$. We convert the Cartesian coordinates (X, Y, Z) into spherical coordinates (R, θ, ϕ), where $R = \sqrt{X^2 + Y^2 + Z^2}$ ranges from 0 to 1 (radius of the unit ball), ϕ ranges from 0 to π (polar angle), θ ranges from 0 to 2π (azimuthal angle). The relationships between Cartesian and spherical coordinates are:

$$X = R\sin\phi\cos\theta, \quad Y = R\sin\phi\sin\theta, \quad Z = R\cos\phi.$$

We calculate the Jacobian of the transformation from spherical coordinates to Cartesian coordinates. The determinant of the Jacobian matrix helps in finding the transformed joint PDF:

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \sin\phi\cos\theta & -r\sin\phi\sin\theta & r\cos\phi\cos\theta \\ \sin\phi\sin\theta & r\sin\phi\cos\theta & r\cos\phi\sin\theta \\ \cos\phi & 0 & -r\sin\phi \end{vmatrix} = r^2\sin\phi.$$

Given the uniform distribution in the unit ball, the joint PDF in spherical coordinates $f_{R,\phi,\theta}(R,\phi,\theta)$ is proportional to the volume element:

$$f_{R,\phi,\theta}(r,\phi,\theta) = \frac{3}{4\pi} r^2 \sin\phi,$$

here we abuse the notation ϕ, θ as random variables and scalars at the same time.

Now, we need to find the joint PDF of R and Z. In spherical coordinates, $Z = R\cos\phi$, so Z is directly related to R and ϕ, but not to θ. This allows us to integrate out θ since the distribution is symmetric around the origin and does not depend on the azimuthal angle θ. Therefore, performing the integration:

$$f_{R,\phi}(r,\phi) = \int_0^{2\pi} \frac{3}{4\pi} r^2 \sin\phi \, d\theta = \frac{3}{4\pi} r^2 \sin\phi \times 2\pi = \frac{3}{2} r^2 \sin\phi.$$

Therefore, we finally have:

$$f_{R,Z}(r,z) = f_{R,\phi}(r,\phi) \left| \det \begin{bmatrix} \frac{\partial r}{\partial r} & \frac{\partial r}{\partial z} \\ \frac{\partial \phi}{\partial r} & \frac{\partial \phi}{\partial z} \end{bmatrix} \right| = \frac{3}{2} r^2 \sin\varphi \left| \frac{\partial \phi}{\partial z} \right| = \frac{3}{2} r^2 \sin\phi \left| \frac{-1}{r\sin\phi} \right| = \frac{3}{2} r, \quad |z| \le r, \quad r \le 1.$$

Probability and Statistics for EECS: Homework #10 Solutions

Professor Ziyu Shao

Problem 1

Let $U_i \sim \text{Unif}(0,1)$, $i \geq 1$ be i.i.d. random variables. Define N as follows:

$$N = \max\left\{n : \prod_{i=1}^{n} U_i \geq e^{-1}\right\}.$$

(a) Estimate $\mathbb{E}(N)$ by generating 5000 samples of N and then use the sample mean.

(b) Estimate $\text{Var}(N)$.

(c) Estimate $P(N = i)$, for $i = 0, 1, 2, 3$.

(d) Can you find the exact distribution of N?

Solution

Code:

```python
import numpy as np
from collections import Counter

num_samples = 5000
e_inverse = np.exp(-1)

def compute_N():
    product = 1.0
    n = 0
    while product >= e_inverse:
        U = np.random.uniform(0, 1)
        product *= U
        n += 1
    return n - 1
N_samples = [compute_N() for _ in range(num_samples)]

E_N = np.mean(N_samples)
Var_N = np.var(N_samples)
N_counts = Counter(N_samples)
prob_N = {i: N_counts.get(i, 0) / num_samples for i in range(4)}

print(f"E(N):{E_N}")
print(f"Var(N):{Var_N}")
print(f"P(N=i):{prob_N},")
```

1. $\mathbb{E}(N) = 1.02$
2. $\text{Var}(N) = 1.02$
3. $P(N = 0) = 0.359, P(N = 1) = 0.365, P(N = 2) = 0.193, P(N = 3) = 0.006$
4. Denote $X_i = -\log(U_i) \sim \text{Exp}(1)$, then we have

$$\begin{aligned}
P(N = n) &= P\left(\sum_{i=1}^{n} X_i \leq 1, \sum_{i=1}^{n+1} X_i > 1\right) \\
&= \int_0^1 P\left(\sum_{i=1}^{n} X_i = x\right) P\left(X_{n+1} > 1 - x\right) dx
\end{aligned}$$

Note that $P(\sum_{i=1}^{n} X_i = x)$ follows Gamma distribution Gamma$(n, 1)$ since $\sum_{i=1}^{n} X_i$ is the summation of n exponential distribution random variables, therefore we have

$$\begin{aligned}
P\left(N = n\right) &= \int_0^1 e^{-(1-x)} \frac{1}{\Gamma(n)} x^{n-1} e^{-x} \, dx \\
&= \int_0^1 e^{-(1-x)} \frac{1}{(n-1)!} x^{n-1} e^{-x} \, dx \\
&= \frac{e^{-1}}{(n-1)!} \int_0^1 x^{n-1} \, dx \\
&= \frac{e^{-1} 1^n}{n!}.
\end{aligned}$$

Thus, we know that $N \sim Poisson(1)$

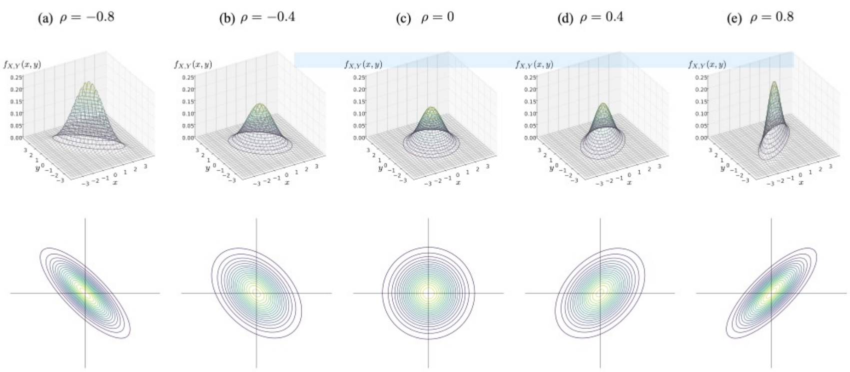
Problem 2

(a) Use the following transformation to generate samples from bivariate Normal distribution with correlation coefficient ρ:

$$X = Z$$
$$Y = \rho Z + \sqrt{1-\rho^2}W,$$

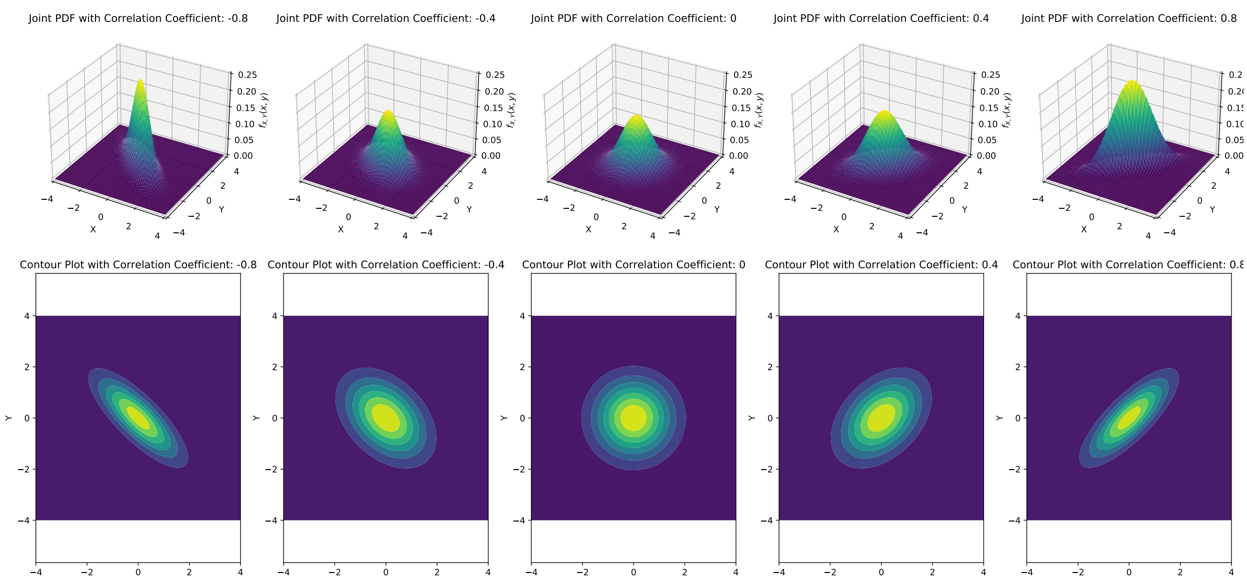
where $-1 < \rho < 1, Z$ and W are i.i.d. random variables following $\mathcal{N}(0,1)$.

(b) Plot the joint pdf function and the corresponding contour(or isocontour) as the following figure:



Solution

The following is the simulation result of sampling BVN:



Problem 3

Let $X_1 \sim \text{Expo}(\lambda_1), X_2 \sim \text{Expo}(\lambda_2)$ and $X_3 \sim \text{Expo}(\lambda_3)$ be independent.

(a) Find $E(X_1 \mid X_1 > 2024)$

(b) Find $E(X_1 \mid X_1 < 1997)$

(c) Find $E(X_1 + X_2 + X_3 \mid X_1 > 1997, X_2 > 2014, X_3 > 2025)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.

Solution

(a) According to the memoryless property of exponential distribution, we have $E(X-2024|X>2024) = E(X)$. Thus we can obtain the conditional expectation as follows:

$$E(X_1|X_1 > 2024) = 2024 + E(X_1 - 2024|X_1 > 2024) = 2024 + E(X_1) = 2024 + \frac{1}{\lambda_1}.$$

(b) From LOTUS, we have

$$E(X_1) = E(X_1|X_1 > 1997)P(X_1 > 1997) + E(X_1|X_1 < 1997)P(X_1 < 1997)$$

Thus we have

$$\begin{aligned}
E(X_1|X_1 < 1997) &= \frac{E(X_1) - E(X_1|X_1 > 1997)P(X_1 > 1997)}{P(X_1 < 1997)} \\
&= \frac{\frac{1}{\lambda_1} - (1997 + \frac{1}{\lambda_1})(1 - 1 + e^{-1997\lambda_1})}{1 - e^{-1997\lambda_1}} \\
&= \frac{\frac{1}{\lambda_1} - (1997 + \frac{1}{\lambda_1})e^{-1997\lambda_1}}{1 - e^{-1997\lambda_1}} \\
&= \frac{1 - (1997\lambda_1 + 1)e^{-1997\lambda_1}}{\lambda_1(1 - e^{-1997\lambda_1})}
\end{aligned}$$

(c) Since X_1, X_2, X_3 are independent, we have

$$\begin{aligned}
&E(X_1 + X_2 + X_3|X1 > 1997, X_2 > 2014, X_3 > 2025) \\
=\ &E(X_1|X_1 > 1997, X_2 > 2014, X_3 > 2025) \\
&\ + E(X_2|X1 > 1997, X_2 > 2014, X_3 > 2025) \\
&\ + E(X_3|X_1 > 1997, X_2 > 2014, X_3 > 2025) \\
=\ &E(X_1|X_1 > 1997) + E(X_2|X_2 > 2014) + E(X_3|X_3 > 2025) \\
=\ &E(X_1 - 1997|X_1 > 1997) + E(X_2 - 2014|X_2 > 2014) + E(X_3 - 2025|X_3 > 2025) + 6032 \\
=\ &E(X_1) + E(X_2) + E(X_3) + 6032 \\
=\ &\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + 6032
\end{aligned}$$

Problem 4

Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 6xy & \text{if } 0 \le x \le 1, 0 \le y \le \sqrt{x} \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal distributions of X and Y. Are X and Y independent?

(b) Find $E[X \mid Y = y]$ and $\text{Var}[X \mid Y = y]$ for $0 \le y \le 1$.

(c) Find $E[X \mid Y]$ and $\text{Var}[X \mid Y]$.

Solution

(a) The supports of X and Y are both $[0, 1]$. In this way, we have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= \int_0^{\sqrt{x}} 6xydy \\ &= 3xy^2\Big|_{y=0}^{y=\sqrt{x}} \\ &= 3x^2, \end{aligned}$$

and

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \int_{y^2}^{1} 6xydx \\ &= 3yx^2\Big|_{x=y^2}^{x=1} \\ &= 3y - 3y^5. \end{aligned}$$

Therefore,

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \begin{cases} 3y - 3y^5 & \text{if } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$, X and Y are not independent.

(b) Since

$$E[X|Y = y] = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx,$$

to calculate $E[X|Y = y]$, we need to first calculate $f_{X|Y}(x|y)$.

If $y^2 \le x \le 1$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2x}{1-y^4}.$$

In this way,

$$f_{X|Y}(x|y) = \begin{cases} \frac{2x}{1-y^4} & \text{if } y^2 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned}
E[X|Y=y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx \\
&= \int_{y^2}^{1} x \frac{2x}{1-y^4} dx \\
&= \frac{2}{3(1-y^4)} x^3 \Big|_{x=y^2}^{x=1} \\
&= \frac{2(1-y^6)}{3(1-y^4)} \\
&= \frac{2}{3} \cdot \frac{1+y^2+y^4}{1+y^2}
\end{aligned}$$

Since

$$\mathrm{Var}[X|Y=y] = E[X^2|Y=y] - (E[X|Y=y])^2,$$

to calculate $\mathrm{Var}[X|Y=y]$, we need to first calculate $E[X^2|Y=y]$.

Since

$$\begin{aligned}
E[X^2|Y=y] &= \int_{-\infty}^{\infty} x^2 f_{X|Y}(x|y)dx \\
&= \int_{y^2}^{1} x^2 \frac{2x}{1-y^4} dx \\
&= \frac{1}{2(1-y^4)} x^4 \Big|_{x=y^2}^{x=1} \\
&= \frac{1-y^8}{2(1-y^4)} \\
&= \frac{1+y^4}{2},
\end{aligned}$$

we have,

$$\begin{aligned}
\mathrm{Var}[X|Y=y] &= E[X^2|Y=y] - (E[X|Y=y])^2 \\
&= \frac{1+y^4}{2} - \left(\frac{2(1-y^6)}{3(1-y^4)}\right)^2 \\
&= \frac{1+y^4}{2} - \frac{4}{9} \cdot \frac{(1+y^2+y^4)^2}{(1+y^2)^2}
\end{aligned}$$

(c) According to the result in question(b), we have

$$\begin{aligned}
E[X|Y] &= \frac{2}{3} \cdot \frac{1+Y^2+Y^4}{1+Y^2}, \\
\mathrm{Var}[X|Y] &= \frac{1+Y^4}{2} - \frac{4}{9} \cdot \frac{(1+Y^2+Y^4)^2}{(1+Y^2)^2}.
\end{aligned}$$

Problem 5

Let X be a discrete r.v. whose distinct possible values are $x_0, x_1, \ldots$, and let $p_k = P(X = x_k)$. The entropy of X is $H(X) = \sum_{k=0}^{\infty} p_k \log_2(1/p_k)$.

(a) Find $H(X)$ for $X \sim \text{Geom}(p)$.

(b) Let X and Y be i.i.d. discrete r.v.s. Show that $P(X = Y) \geq 2^{-H(X)}$. Hint: Jensen's Inequality.

Solution

(a) The PMF of X is $P(X = k) = p(1-p)^k$ since there is $X \sim \text{Geom}(p)$. Thus we have

$$\begin{aligned} H(X) &= -\sum_{k=0}^{\infty} p(1-p)^k \log_2 \left(p(1-p)^k\right) \\ &= -p\sum_{k=0}^{\infty} k(1-p)^k \log_2(1-p) - p\log_2 p \sum_{k=0}^{\infty}(1-p)^k \\ &= -\log_2 p - \frac{1-p}{p}\log_2(1-p) \end{aligned}$$

(b) Since X and Y are i.i.d random variables, via LOTP, we have

$$\begin{aligned} P(X = Y) &= \sum_{k=0}^{\infty} P(X = Y|Y = k)\cdot P(Y = k) \\ &= \sum_{k=0}^{\infty} P(X = k)\cdot P(Y = k) = \sum_{k=0}^{\infty} p_k^2 \end{aligned}$$

Denote Z as a new discrete random variable such that $P(Z = p_k) = p_k$, then we have:

$$E(Z) = \sum_{k=0}^{\infty} p_k \times p_k = P(X = Y)$$

Since $\log(\cdot)$ is a convex function, according to Jensen's inequality, we have $E(\log(Z)) \leq \log(E(Z))$, thus there is

$$\begin{aligned} &\sum p_k \log_2 p_k \leq \log_2 \sum p_k^2 \\ \Leftrightarrow &-H(X) \leq \log_2 P(X = Y) \\ \Leftrightarrow &P(X = Y) \geq 2^{-H(X)}. \end{aligned}$$

Problem 6

Instead of predicting a single value for the parameter, we give an interval that is likely to contain the parameter: A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that $P(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]) \geq 1-\delta$. Now we toss a coin with probability p landing heads and probability $1-p$ landing tails. The parameter p is unknown and we need to estimate its value from experiment results. We toss such coin N times. Let $X_i = 1$ if the ith result is head, otherwise 0. We estimate p by using

$$\hat{p} = \frac{X_1 + \ldots + X_N}{N}.$$

Find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N.

(a) Method 1: Adopt Chebyshev inequality to find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N.

(b) Method 2: Adopt Hoeffding bound to find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N.

(c) Discuss the pros and cons of the above two methods.

Solution

Since $X_i \sim \text{Bern}(p), X_i \in \{0,1\}$, we have $\mathbb{E}[X_i] = p$ and $\mathbb{V}[X_i] = p(1-p)$. Therefore, we have

$$\mathbb{E}[\hat{p}] = p, \mathbb{V}[\hat{p}] = \frac{p(1-p)}{N}.$$

Besides, we know that

$$P(p \in [\hat{p}-\varepsilon, \hat{p}+\varepsilon]) \geq 1-\delta \Leftrightarrow P(|\hat{p}-p| \geq \varepsilon) \leq \delta.$$

(a) Applying Chebyshev's inequality on random variable $\hat{p}$, we have

$$P(|\hat{p}-p| \geq \epsilon) \leq \frac{p(1-p)}{N\epsilon^2} \Rightarrow \delta = \frac{p(1-p)}{N\epsilon^2}, \epsilon = \sqrt{\frac{p(1-p)}{N\delta}}$$

Therefore, we know that δ negatively correlates with ϵ, i.e., given a fixed number of samples N, there is natural trade-off between accuracy and confidence. Besides, 1) Fix the confidence interval parametrized by δ, reducing the estimation error ϵ requires increasing the number of samples N. 2) Fix the estimation error ϵ, narrowing the confidence interval requires increasing the number of samples N. That is, the impacts of N is on both the "estimation accuracy" and "estimation confidence".

(b) Applying Hoeffding's inequality on random variable $\hat{p}$, we have

$$P(|\hat{p}-p| \geq \epsilon) \leq 2e^{-2N\epsilon^2} \Rightarrow \delta = 2e^{-2N\epsilon^2}, \epsilon = \sqrt{\frac{\ln(2/\delta)}{2N}}$$

The effects of δ and N are similarly discussed as in (a).

(c) Chebyshev's inequality (see Cantelli's inequality for the one-side improvement):

- Pros: 1) sharp bound and cannot be improved in general (given no extra assumption). 2) can be improved with extra distributional information on polynomial moments.
- Cons: 1) requires the existence of moments until the second order. 2) quadratic convergence rate.

Hoeffding's inequality (see Theorem 2.8 and 2.9 of paper "old and new concentration inequalities" for the one-side improvement):

- Pros: 1) exponential convergence rate. 2) does not require assumption on moments.
- Cons: 1) works only for sub-Gaussian (e.g., bounded random variables). 2) in general not sharp when the variance is small (e.g., see popoviciu's inequality on variances and Bernstein's inequality).

Probability and Statistics for EECS: Homework #9 Solutions

Professor Ziyu Shao

Problem 1

Use the methods of inverse transform sampling to obtain samples from each of the following continuous distributions:

(a) Logistic distribution with CDF $F(x) = e^x/(1+e^x), x \in \mathbb{R}$.

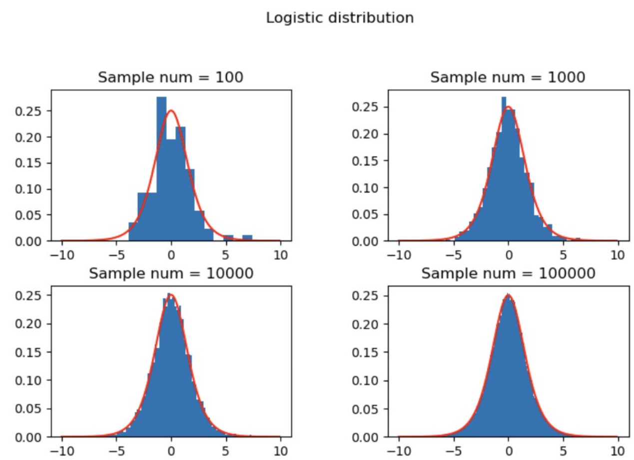
(b) Rayleigh distribution with CDF $F(x) = 1 - e^{-x^2/2}, x > 0$.

(c) Exponential distribution with CDF $F(x) = 1 - e^{-x}, x > 0$.

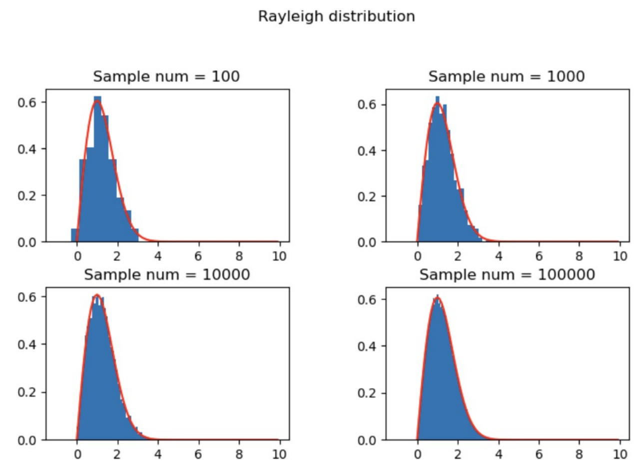
After obtaining enough samples, please plot the corresponding histogram and corresponding theoretical PDF.

Solution

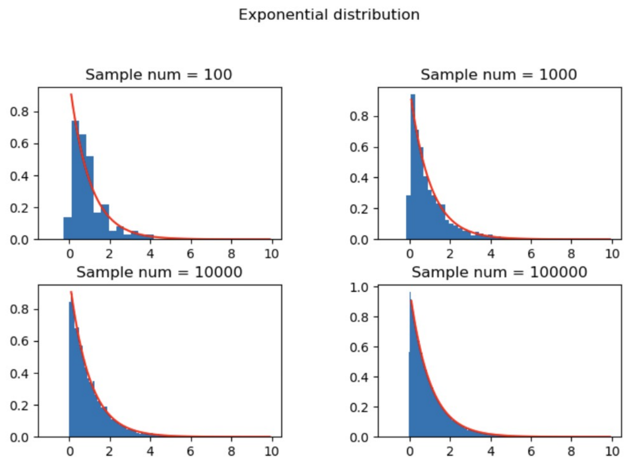
(a) The CDF of logistic distribution can be represented by $F(x) = \frac{1}{1+e^{-x}}, \forall x \in \mathbb{R}$, which is continuous and strictly increasing on the support of the distribution. So we can obtain its inverse function $F^{-1}(x) = -\log(1/x - 1)$ and the PDF is $f(x) = \frac{e^x}{(1+e^x)^2}$.



(b) The CDF $F(x) = 1 - e^{-x^2/2}, \forall x > 0$, which is continuous and strictly increasing on the support of the distribution. So we can obtain its inverse function $F^{-1}(x) = \sqrt{-2\log(1-y)}$ and the PDF is $f(x) = xe^{-x^2/2}$.



(c) The CDF $F(x) = 1 - e^{-x}, \forall x > 0$, which is continuous and strictly decreasing on the support of the distribution. So we can obtain its inverse function $F^{-1}(x) = \sqrt{-2\log(1-y)}$ and the PDF is $f(x) = -\log(1-x)$.



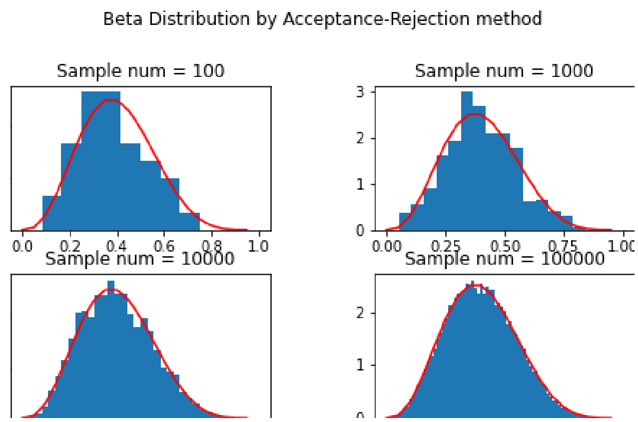
Problem 2

Acceptance-Rejection method

(a) Use the Acceptance-Rejection Method to obtain samples from Beta distribution Beta(4, 6)

(b) Use both the Box-Muller method and the Acceptance-Rejection Method to obtain samples from the standard Normal distribution N (0, 1), then discuss the pros and cons of each method

Solution



(a) We use the Acceptance-Rejection method to generate the Beta distribution.

- Generate two independent random variables, U_1 and U_2, from Unif(0,1). Set $V_1 = U_1^{1/a}$ and $V_2 = U_2^{1/b}$.
- If $V_1 + V_2 \le 1$, set $X = V_1/(V_1 + V_2)$; otherwise go back to step 1.

Proof. Let $W = V_1 + V_2$. Then the CDF of X is

$$F(x) = P(X \le x) = P\left(\frac{V_1}{W} \le x | W \le 1\right) = \frac{P(V_1 \le xW, W \le 1)}{P(W \le 1)}.$$

By changing of variables, we have

$$\begin{aligned} f_{V_1}(v_1) &= f_{U_1}(v_1^a)|av_1^{a-1}| = av_1^{a-1}, \quad 0 < v_1 < 1, \\ f_{V_2}(v_2) &= f_{U_2}(v_2^b)|bv_2^{b-1}| = bv_2^{b-1}, \quad 0 < v_2 < 1. \end{aligned}$$

The Jacobian matrix of transmission for $V_1, V_2 = W - V_1$ is

$$J = \begin{pmatrix} \frac{\partial v_1}{\partial v_1} & \frac{\partial v_1}{\partial w} \\ \frac{\partial v_2}{\partial v_1} & \frac{\partial v_2}{\partial w} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Then the joint pdf of V_1 and W is

$$f_{V_1,W}(v_1, w) = f_{V_1,V_2}(v_1, v_2)|J| = abv_1^{a-1}(w - v_1)^{b-1}, \quad 0 < v_1 < 1, v_1 < w < v_1 + 1.$$

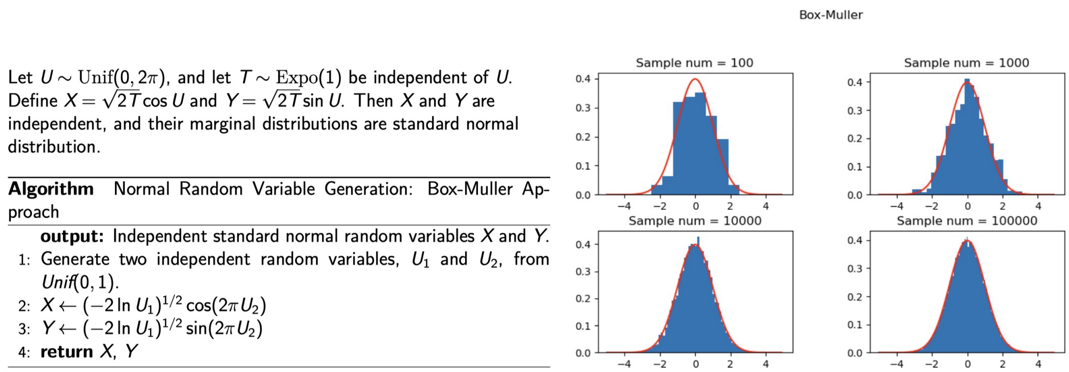
Therefore, we have

$$\begin{aligned} F(x) &= \frac{P(V_1 \le xW, W \le 1)}{P(W \le 1)} \\ &= \frac{\int_0^1 \int_0^{wx} abv_1^{a-1}(w - v_1)^{b-1} dv_1 dw}{P(W \le 1)} \\ &= \frac{\int_0^1 \int_0^{wx} v_1^{a-1}(w - v_1)^{b-1} dv_1 dw}{c}, \end{aligned}$$

where c is a constant. Taking the derivate and applying the fundamental theorem of calculus, we have

$$f(x) = \frac{1}{c}\int_0^1 w(wx)^{a-1}(w - wx)^{b-1}dw = \frac{1}{c(a+b)}x^{a-1}(1-x)^{b-1}.$$

(b) The following is the pseudocode and simulation results of Box-Muller method.



- Generate Exponential distribution with parameter $\lambda = 1$ from Unif(0,1).
  The CDF of Expo(1) is $F(x) = 1 - e^{-x}, x \geq 0$. The inverse of the CDF is $F^{-1}(u) = -\ln(1-u)$. If $u \sim \text{Unif}(0,1)$, then $y = -\ln(1-U) \sim \text{Expo}(1)$.
- Generate $U \sim \text{Unif}(0,1)$.
- Let $Z \sim N(0,1)$, we will generate $X \sim |Z|$ firstly.
  The pdf of X is $p(x) = \frac{2}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, x \geq 0$. The pdf of Expo(1) is $q(x) = e^{-x}, x \geq 0$.
  Let $c = \sup_x \frac{p(x)}{q(x)} = \sup_x \sqrt{\frac{2}{\pi}}e^{x-\frac{1}{2}x^2} = \sqrt{\frac{2e}{\pi}}$.
  If $u < \frac{p(y)}{cq(y)} = e^{-\frac{1}{2}(y-1)^2}$, set $x = y$; otherwise go back to step 1.
- $Z = \begin{cases} x, & \text{w.p. } 0.5 \\ -x, & \text{w.p. } 0.5 \end{cases}$. Then $Z \sim N(0,1)$.

(c) In terms of sampling Normal distribution, their variance are similar, while the sample efficiency of BoxMuller is higher with also higher running speed.

Box-Muller:

- Pros: It is easy to implement, and the method only uses Unif$(0,1)$ as the basis data sample, which is simple to sample.
- Cons: Only the standard normal distribution can be sampled by this method.

Acceptance-Rejection:

- Pros: It can sample many kinds of probability distribution including many distributions that is difficult to sample directly.
- Cons: The domain of function $g(x)$ must cover the domain of function $f(x)$. If c is closed to 1, the basis distribution g is still difficult to sample; while if c is closed to 0, the probability of acceptance success will be small, which will cause low efficiency.

Problem 3

Monte Carlo Integration

(a) Evaluate the integration

$$\int_0^1 \frac{4}{1+x^2} dx.$$

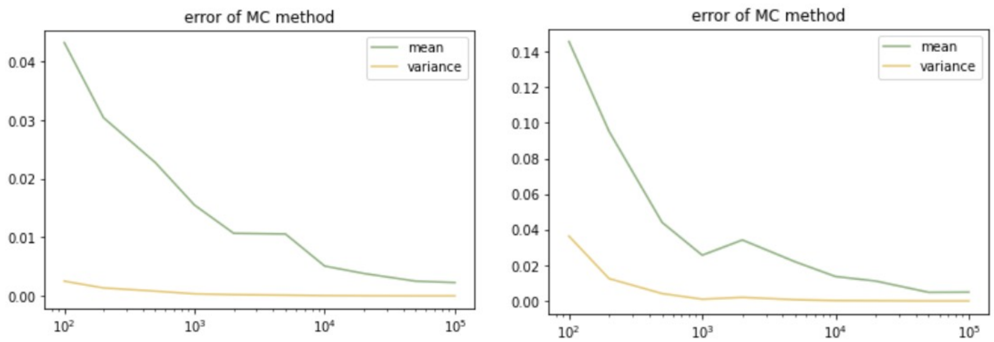
(b) Evaluate the integration

$$\int_0^4 \sqrt{x+\sqrt{x+\sqrt{x+\sqrt{x}}}}dx.$$

(c) Evaluate the probability of rare event $c = P(Y > 8)$, where $Y \sim \mathcal{N}(0, 1)$.

Solution

The following figures cover the results from (a)-(b)



- Ground truth 3.1415926535897936, Estimation with 2×10^6 samples: 3.14160782958
- Ground truth: 7.6766100019, Estimation with 2×10^6 samples: 7.6765401290

(c)
- Without importance sampling

$$c = P(Y > 8) = \mathbb{E}[I(Y > 8)] = \int_{-\infty}^{\infty} I(Y > 8) f(y) dy, f \sim \mathcal{N}(0, 1),$$
$$c \approx \frac{1}{n}\sum_{j=1}^{n} I(Y_j > 8), Y_j \sim \mathcal{N}(0, 1).$$

- With importance sampling

$$c \approx \frac{1}{n}\sum_{j=1}^{n} \frac{h(Y_j)f(Y_j)}{g(Y_j)} = \frac{1}{n}\sum_{j=1}^{n} I(Y_j > 8) \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Y_j^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Y_j-8)^2}} = \frac{1}{n}\sum_{j=1}^{n} I(Y_j > 8) e^{-8Y_j+32}, Y_j \sim g = \mathcal{N}(8, 1)$$

The following is the simulation result with 10^7 samples:

- With importance sampling: 6.228×10^{-16}
- Without importance sampling: 0.0

Problem 4

Use your own words to describe the geometric perspective of Jacobian Matrix and Jacobian Determinant.

Solution

Geometrically, the Jacobian Matrix at a point encapsulates the best linear approximation of a nonlinear transformation near that point. Each entry in the Jacobian represents the partial derivative of one transformed coordinate with respect to an original coordinate, effectively describing how each direction stretches, compresses, or rotates under the transformation.

The Jacobian Determinant provides a scalar measure of this transformation's local scaling effect on volume. Specifically, when a region in the original space is transformed, the absolute value of the Jacobian Determinant indicates how the volume of that region changes. For example:

- If the determinant is greater than 1, the transformation locally expands the volume.
- If it is between 0 and 1, the transformation contracts the volume.
- A negative determinant also indicates a reflection in addition to scaling.

Problem 5

The PDF of the Gamma distribution Gamma(a, $\lambda) is$:

$$f(x) = \frac{1}{\Gamma(a)}(\lambda x)^a e^{-\lambda x}\frac{1}{x}, \quad x > 0,$$

where $a > 0, \lambda > 0$, and $\Gamma(a)$ is the Gamma Function:

$$\Gamma(a) = \int_0^\infty z^{a-1}e^{-z}\,dz.$$

(a) If $X \sim \text{Gamma}(a, \lambda)$, find $E(X)$ and $\text{Var}(X)$.
(b) If $Y \sim \mathcal{N}(0, 1)$, show that $Y^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$.
(c) If $V = Y_1^2 + \ldots + Y_n^2$, where $Y_i, i = 1, \ldots, n$ are i.i.d. random variables and $Y_i \sim \mathcal{N}(0, 1)$, then V satisfies the chi-square distribution, i.e., $V \sim \chi_n^2$. Show that $V \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$ and find the PDF of V.
(d) If Y and V are independent, define the random variable Z as follows

$$Z = \frac{Y}{\sqrt{\frac{V}{n}}}.$$

Then Z satisfies Student's t-distribution, i.e., $Z \sim t_n$. Please adopt the change of variable method to find the PDF of Z.
(e) Given two independent random variables V_1 and V_2, where $V_1 \sim \chi_m^2$ and $V_2 \sim \chi_n^2$. Define random variable W as follows

$$W = \frac{\frac{V_1}{m}}{\frac{V_2}{n}}.$$

then W satisfies F-distribution, i.e. $W \sim F(m, n)$. Please adopt the change of variable method to find the PDF of W.

Solution

1. According to the definition of Gamma Distribution, we have:

$$
\begin{aligned}
E(X) &= \int_0^{\infty} x \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \frac{1}{x}\, dx \\
&= \frac{\lambda^a}{\Gamma(a)\lambda^{a+1}} \int_0^{\infty} (\lambda x)^a e^{-\lambda x}\, d(\lambda x) \\
&= \frac{\Gamma(a+1)}{\Gamma(a)\lambda} \\
&= \frac{a}{\lambda}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - E^2(X) \\
&= \int_0^{\infty} x^2 \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \frac{1}{x}\, dx - \frac{a^2}{\lambda^2} \\
&= \frac{\lambda^a}{\Gamma(a)\lambda^{a+2}} \int_0^{\infty} (\lambda x)^{a+1} e^{-\lambda x}\, d(\lambda x) - \frac{a^2}{\lambda^2} \\
&= \frac{\Gamma(a+2)}{\Gamma(a)\lambda^2} - \frac{a^2}{\lambda^2} \\
&= \frac{a(a+1)}{\lambda^2} - \frac{a^2}{\lambda^2} \\
&= \frac{a}{\lambda^2}
\end{aligned}
$$

2. For $Y \sim \mathcal{N}(0,1)$, its PDF is $f(Y=y) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$. Denote $Z = Y^2$, then we have $Y = \pm\sqrt{Z}$ and $\frac{dY}{dZ} = \pm\frac{1}{2\sqrt{Z}}$. Thus the PDF of Z can be obtained via:

$$
f_Z(z) = f_Y(\sqrt{z}) \left| \frac{dY}{dZ} \right| + f_Y(-\sqrt{z}) \left| \frac{dY}{dZ} \right| = \frac{-e^{-\frac{z}{2}}}{\sqrt{2\pi z}}
$$

which is exactly the PDF function of $\mathrm{Gamma}(\frac{1}{2}, \frac{1}{2})$.

3. Here we just need to prove the following property distribution: For $X_1 \sim \mathrm{Gamma}(\alpha_1, \beta)$ and $X_2 \sim \mathrm{Gamma}(\alpha_2, \beta)$, there is $X = X_1 + X_2 \sim \mathrm{Gamma}(\alpha_1 + \alpha_2, \beta)$

$$
\begin{aligned}
f_X(x) &= \int_0^{x} f_{X_1}(t) f_{X_2}(x-t)\, dt \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} e^{-x/\beta} \int_0^{x} t^{\alpha_1-1}(x-t)^{\alpha_2-1}\, dt \\
&= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} e^{-x/\beta} x^{\alpha_1+\alpha_2-1} \mathrm{Beta}(\alpha_1, \alpha_2) \\
&= \frac{x^{\alpha_1+\alpha_2-1} e^{-x/\beta}}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} \\
X &\sim \mathrm{Gamma}(\alpha_1, \alpha_2),
\end{aligned}
$$

where the third equation is obtained from the following integral transformation

$$
\int_0^{x} t^{\alpha_1-1}(x-t)^{\alpha_2-1}\, dt = x^{\alpha_1+\alpha_2-1} \mathrm{Beta}(\alpha_1, \alpha_2),
$$

where $\mathrm{Beta}(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$. Then we can directly obtain that $V \sim \mathrm{Gamma}(\frac{n}{2}, \frac{1}{2})$.

4. For the two independent random variable, the PDF of their joint distribution is

$$
\begin{aligned}
f_{Y,V}(y,v) &= f_Y(y)\cdot f_V(v)\\
&= \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\cdot\frac{v^{\frac{n}{2}-1}e^{-\frac{v}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}
\end{aligned}
$$

Then let $Z=\frac{Y}{\sqrt{\frac{V}{n}}}$, we have

$$\left|\frac{\partial(Y,V)}{\partial(Z,V)}\right|=\frac{\partial Y}{\partial Z}=\sqrt{\frac{V}{n}}$$

So,

$$f_{Z,V}(z,v)=f_{Y,V}\left(z\sqrt{\frac{v}{n}},v\right)\sqrt{\frac{v}{n}}$$

$$=\frac{1}{\sqrt{2\pi n}}\cdot\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}v^{\frac{n-1}{2}}e^{-\frac{v}{2}\left(1+\frac{z^2}{n}\right)}$$

The integral is of the form:

$$\int_0^\infty v^{\alpha-1}e^{-\beta v}\,dv=\frac{\Gamma(\alpha)}{\beta^\alpha},$$

with $\alpha=\frac{n}{2}-\frac{1}{2}+1=\frac{n}{2}+\frac{1}{2}$ $\quad$ (since $\frac{n}{2}-\frac{1}{2}=\frac{n-1}{2}$ and adding 1 gives $\frac{n+1}{2}$) $\quad$ and $\beta=\frac{n+Z^2}{2n}$.Thus $\alpha=\frac{n+1}{2}$.

Therefore:

$$\int_0^\infty v^{\frac{n}{2}-\frac{1}{2}}e^{-\frac{n+Z^2}{2n}v}\,dv=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\left(\frac{n+Z^2}{2n}\right)^{\frac{n+1}{2}}}.$$

We obtain the marginal PDF of Z:

$$f_Z(z)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}},\ z>0$$

5.

$$
\begin{aligned}
f_{V_1,V_2}(v_1,v_2) =&\frac{v_1^{\frac{m}{2}-1}e^{-\frac{v_1}{2}}}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}\cdot\frac{v_2^{\frac{n}{2}-1}e^{-\frac{v_2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\\
=&\frac{v_1^{\frac{m}{2}-1}v_2^{\frac{n}{2}-1}e^{-\frac{v_1+v_2}{2}}}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}
\end{aligned}
$$

Since $V_1=\frac{WV_2m}{n}$,The determinant Jacobian Matrix is $\left|\frac{\partial(V_1,V_2)}{\partial(W,V_2)}\right|=\frac{mW}{n}$

So the PDF is

$$
\begin{aligned}
f_{W,V_2}(w,v_2)&=f_{V_1,V_2}(\frac{mv_2w}{n},v_2)\frac{mw}{n}\\
&=\frac{\left(\frac{1}{2}\right)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{mw}{n}\right)^{\frac{m}{2}-1}v_2^{\frac{m+n}{2}-1}e^{-\frac{v_2}{2}\left(1+\frac{mw}{n}\right)}
\end{aligned}
$$

Via take integration on v_2 and simplification, we have

$$
\begin{aligned}
f_W(w) &= \frac{\left(\frac{1}{2}\right)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{mw}{n}\right)^{\frac{m}{2}-1}\int_0^\infty v_2^{\frac{m+n}{2}-1}e^{-\frac{v_2}{2}\left(1+\frac{mw}{n}\right)}\,dv_2 \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{w^{m/2-1}\left(\frac{m}{n}\right)^{m/2}}{\left(1+\frac{mw}{n}\right)^{\frac{n+m}{2}}}, w>0
\end{aligned}
$$

Probability & Statistics for EECS: Homework #11

Due on Dec 2, 2023 at 23:59

Problem 1

A coin with probability p of Heads is flipped repeatedly. For (a) and (b), suppose that p is a known constant, with $0 < p < 1$.

(a) What is the expected number of flips until the pattern HT is observed?

(b) What is the expected number of flips until the pattern HH is observed?

Solution

(a) This can be thought of as "Wait for the first Head, then wait for the first Tail afterwards," so the expected value is $\frac{1}{p} + \frac{1}{q}$ by the story of first success distributions, with $q = 1 - p$.

(b) Let X be the waiting time for HH and condition on the first toss, writing H for the occurrence of Head and T for the occurrence of Tail:

$$E[X] = E[X|\mathrm{H}]p + E[X|\mathrm{T}]q = E[X|\mathrm{H}]p + (1 + E[X])q.$$

To find $E[X|H]$, condition on the second toss:

$$E[X|\mathrm{H}] = E[X|\mathrm{HH}]p + E[X|\mathrm{HT}]q = 2p + (2 + E[X])q.$$

Solving for $E[X]$, we have

$$E[X] = \frac{1}{p} + \frac{1}{p^2}.$$

Problem 2

Given two random variables X and Y, the corresponding joint PDF is

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{Otherwise} \end{cases}$$

Find $\mathbb{E}[Y|X]$ and $L[Y|X]$.

Solution

1. **Finding** $\mathbb{E}[Y \mid X = x]$

   First, we compute the marginal PDF of X:

   $$f_X(x) = \int_0^1 (x+y)\,dy = x + \frac{1}{2}, \quad 0 \le x \le 1.$$

   The conditional PDF of Y given $X = x$ is:

   $$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}}, \quad 0 \le y \le 1.$$

   Then,

   $$\mathbb{E}[Y|X=x] = \int_0^1 y\frac{x+y}{x+\frac{1}{2}}\,dy = \frac{\frac{x}{2}+\frac{1}{3}}{x+\frac{1}{2}} = \frac{3x+2}{6x+3}.$$

   Thus, $\mathbb{E}[Y|X] = \frac{3X+2}{6X+3}$

2. **Finding** $L(Y \mid X)$,

   The LLSE is given by,

   $$L(Y|X) = \mathbb{E}(Y) + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}(X - \mathbb{E}(X)).$$

   We need to calculate $\mathbb{E}(X), \mathbb{E}(Y), \operatorname{Var}(X),$ and $\operatorname{Cov}(X,Y)$.

   $$\begin{aligned}\mathbb{E}(X) &= \int_0^1\int_0^1 x(x+y)\,dy\,dx = \int_0^1 x\left(x+\frac{1}{2}\right)dx. \\ &= \frac{1}{3}+\frac{1}{4} = \frac{7}{12} = \mathbb{E}(Y).\end{aligned}$$

   Thus, $\mathbb{E}(X) = \mathbb{E}(Y) = \frac{7}{12}$. Simlarly, we can get variance and covaraince.

   $$\mathbb{E}(X^2) = \int_0^1\int_0^1 x^2(x+y)\,dy\,dx = \int_0^1 x^2\left(x+\frac{1}{2}\right)dx = \frac{5}{12}.$$

   $$\operatorname{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{5}{12} - \frac{49}{144} = \frac{11}{144}.$$

   $$\mathbb{E}(XY) = \int_0^1\int_0^1 xy(x+y)\,dy\,dx = \frac{1}{3}$$

   Then

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{3} - \frac{49}{144} = -\frac{1}{144}.$$

Thus the LLSE is:

$L(Y|X) = \frac{7}{11} - \frac{1}{11}X.$

Problem 3

Let X be the height of a randomly chosen adult man, and Y be his father's height, where X and Y have been standardized to have mean 0 and standard deviation 1. Suppose that (X, Y) is Bivariate Normal, with $X, Y \sim \mathcal{N}(0,1)$ and $\operatorname{Corr}(X, Y)=\rho$.

(a) Let $y=ax+b$ be the equation of the best line for predicting Y from X (in the sense of minimizing the mean squared error), e.g., if we were to observe $X=1.3$ then we would predict that Y is $1.3a+b$. Now suppose that we want to use Y to predict X, rather than using X to predict Y. Give and explain an intuitive guess for what the slope is of the best line for predicting X from Y.

(b) Find a constant c (in terms of ρ) and an r.v. V such that $Y=cX+V$, with V independent of X. Hint: Start by finding c such that $\operatorname{Cov}(X, Y-cX)=0$.

(c) Find a constant d (in terms of ρ) and an r.v. W such that $X=dY+W$, with W independent of Y.

(d) Find $E(Y \mid X)$ and $E(X \mid Y)$.

(e) Reconcile (a) and (d), giving a clear and correct intuitive explanation.

Solution

(a) Since the parameter ρ tells us what is the rate of change of second variable respective to the first one, we can assume that ρ is the slope of the line, i.e. $a=\rho$. Now, in order to predict X from Y, we just have to consider the line that is inverse to the original line. From the basic algebra, we know that inverse has slope one over the original slope. Thus, the required slope is $\frac{1}{\rho}$.

(b) Since we have to find $V=Y-cX$ such that is independent from X, using the given hint, we have that

$$0=\operatorname{Cov}(X, Y-cX)=\operatorname{Cov}(X, Y)-c \operatorname{Var}(X)=\rho-c.$$

Hence, let's define $c=\rho$ and it is the only candidate for the constant c.

Let's check that X and V are independent. Observe that $Y-\rho X$ is also Normal (as the linear combination of two Bivariate Normals). So, the fact that two Normals that construct Bivariate Normal are independent is equivalent to the fact that they are uncorrelated. Since we have the last information, we have found the required.

(c) With the same calculation and discussion as in part (b), we have that the answer is also $d=\rho$.

(d) Using the definition of conditional density function, we have that

$$\begin{aligned}
f_{Y \mid X}(y \mid x) =& \frac{f(x, y)}{f(x)}=\frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)}\left(x^2+y^2-2xy\rho\right)\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)} \\
=& \frac{1}{\sqrt{2\pi\left(1-\rho^2\right)}} \exp\left(-\frac{x^2+y^2-2xy\rho}{2\left(1-\rho^2\right)}+\frac{x^2}{2}\right) \\
=& \frac{1}{\sqrt{2\pi\left(1-\rho^2\right)}} \exp\left(-\frac{(y-\rho x)^2}{2\left(1-\rho^2\right)}\right).
\end{aligned}$$

Now, we see that

$$Y \mid X=x \sim \mathcal{N}\left(\rho x, 1-\rho^2\right).$$

Hence, $E(Y \mid X)=\rho X$. Because of the symmetry, we also have that $E(X \mid Y)=\rho Y$.

(e) From the results of (d), the slope of predicting X from Y is same as the slope of predicting Y from X. Based on the symmetry property of normal distribution, and the correlation of X and Y will determine the slope in both direction.

Problem 4

Show the following orthogonality properties of MMSE:

1. For any function $\phi(\cdot)$, one has
$$\mathbb{E}[(Y - \mathbb{E}[Y|X])\phi(X)] = 0$$
2. If the function $g(X)$ satisfied
$$\mathbb{E}[(Y - g(X))\phi(X)] = 0$$
then $g(X) = \mathbb{E}[Y|X]$

Solution

(a) We have

$$\begin{aligned}
\mathbb{E}[(Y - \mathbb{E}[Y|X])\phi(X)] &= \mathbb{E}[Y\phi(X) - \mathbb{E}[Y|X]\phi(X)] \\
&= \mathbb{E}[Y\phi(X)] - \mathbb{E}[\mathbb{E}[Y|X]\phi(X)] \\
&= \mathbb{E}[Y\phi(X)] - \mathbb{E}[\mathbb{E}[Y\phi(X)|X]] \\
&= \mathbb{E}[Y\phi(X)] - \mathbb{E}[Y\phi(X)] = 0
\end{aligned}$$

(b) We have

$$\begin{aligned}
\mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2] &= \mathbb{E}[(\mathbb{E}[Y|X] - g(X))(\mathbb{E}[Y|X] - Y + Y - g(X))] \\
&= \mathbb{E}[(\mathbb{E}[Y|X] - g(X))(\mathbb{E}[Y|X] - Y)] + \mathbb{E}[(\mathbb{E}[Y|X] - g(X))(Y - g(X))]
\end{aligned}$$

We know that $\mathbb{E}[Y|X] - g(X)$ is a function of X. For any function $\phi(\cdot)$, we have $\mathbb{E}[(Y - \mathbb{E}[Y|X])\phi(X)] = 0$. Then,

$$\mathbb{E}[(\mathbb{E}[Y|X] - g(X))(\mathbb{E}[Y|X] - Y)] = 0$$

And for any function $\phi(\cdot)$, we have $\mathbb{E}[(Y - g(X))\phi(X)] = 0$ which means

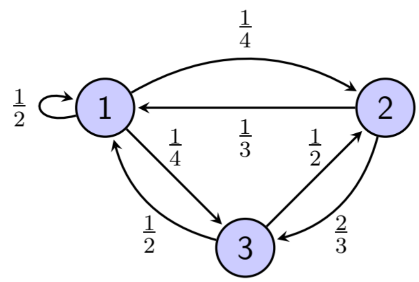
$$\mathbb{E}[(\mathbb{E}[Y|X] - g(X))(Y - g(X))] = 0$$

Eventually, $\mathbb{E}[(\mathbb{E}[Y|X] - g(X))^2] = 0$ which means $\mathbb{E}[Y|X] - g(X) = 0$. We have proved

$$g(X) = \mathbb{E}[Y|X]$$

Problem 5

Given a Markov chain with state-transition diagram shown as follows:



(a) Find $P(X_3 = 3 \mid X_2 = 2)$ and $P(X_4 = 1 \mid X_3 = 2)$.

(b) If $P(X_0 = 2) = \frac{2}{5}$, find $P(X_0 = 2, X_1 = 3, X_2 = 1)$.

(c) Find $P(X_2 = 1 \mid X_0 = 2), P(X_2 = 2 \mid X_0 = 2)$, and $P(X_2 = 3 \mid X_0 = 2)$.

(d) Find $E(X_2 \mid X_0 = 2)$.

Solution

The transition matrix of the Markov chain is

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

(a) From the state-transition diagram, we have

$$\begin{aligned} P(X_3 = 3|X_2 = 2) &= \frac{2}{3} \\ P(X_4 = 1|X_3 = 2) &= \frac{1}{3} \end{aligned} \tag{1}$$

(b)

$$\begin{aligned} P(X_0 = 2, X_1 = 3, X_2 = 1) &= P(X_1 = 3, X_2 = 1|X_0 = 2)P(X_0 = 2) \\ &= P(X_2 = 1|X_1 = 3, X_0 = 2)P(X_1 = 3|X_0 = 2)P(X_0 = 2) \\ &= P(X_2 = 1|X_1 = 3)P(X_1 = 3|X_0 = 2)P(X_0 = 2) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{5} = \frac{2}{15} \end{aligned} \tag{2}$$

(c)

$$
\begin{aligned}
&P(X_2 = 1|X_0 = 2) = \sum_{i=1}^{3} P(X_2 = 1|X_1 = i, X_0 = 2)P(X_1 = i|X_0 = 2) \\
&= \sum_{i=1}^{3} P(X_2 = 1|X_1 = i)P(X_1 = i|X_0 = 2) \\
&= \frac{1}{6} + 0 + \frac{2}{6} = \frac{1}{2} \\
&P(X_2 = 2|X_0 = 2) = \sum_{i=1}^{3} P\big(X_2 = 2\big|X_1 = i, X_0 = 2\big)P\big(X_1 = i\big|X_0 = 2\big) \\
&= \sum_{i=1}^{3} P(X_2 = 2|X_1 = i)P(X_1 = i|X_0 = 2) \\
&= \frac{1}{12} + 0 + \frac{2}{6} = \frac{5}{12} \\
&P(X_2 = 3|X_0 = 2) = \sum_{i=1}^{3} P(X_2 = 3|X_1 = i, X_0 = 2)P(X_1 = i|X_0 = 2) \\
&= \sum_{i=1}^{3} P(X_2 = 3|X_1 = i)P(X_1 = i|X_0 = 2) \\
&= \frac{1}{12} + 0 + 0 = \frac{1}{12}.
\end{aligned}
\tag{3}
$$

(d) The expectation is

$$
\mathrm{E}(X_2|X_0 = 2) = \sum_{i=1}^{3} iP(X_2 = i|X_0 = 2) = \frac{1}{2} + \frac{10}{12} + \frac{3}{12} = \frac{19}{12}.
$$

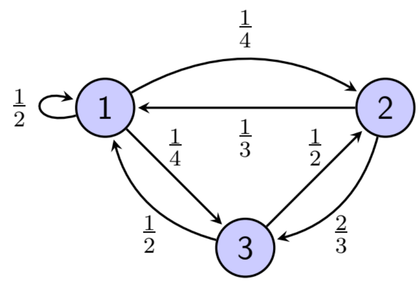
Probability & Statistics for EECS: Homework #12

Due on Dec 2, 2023 at 23:59

Name:
Student ID:

Problem 1

Given a Markov chain with state-transition diagram shown as follows:



(a) Is this chain irreducible?

(b) Is this chain aperiodic?

(c) Find the stationary distribution of this chain.

(d) Is this chain reversible?

Solution

The transition matrix of the Markov chain is

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

(a) Yes, because the elements in the matrix are $Q_{1,2}, Q_{2,1}, Q_{1,3}, Q_{3,1}, Q_{2,3}, Q_{3,2}$ are all non-zero.

(b) Yes, the diagram we know that 1 is a possible return time for state 1, thus $d(1) = 1$ since both $2, 3$ are possible for state 2 and 3, $d(2) = d(3) = 1$ because the chain is irreducible and $d(1) = d(2) = d(3) = 1$. Therefore the chain is aperiodic.

(c) Denote π as the stationary distribution for the chain, then there is $\pi Q = \pi$. Then we solve the problem $\pi(Q - I) = 0$, as follows:

$$\begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & -1 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & -1 \end{bmatrix} = 0$$

where $\sum_{i=1}^{3} \pi_i$. The solution is $\pi = (\frac{16}{35}, \frac{9}{35}, \frac{2}{7})$.

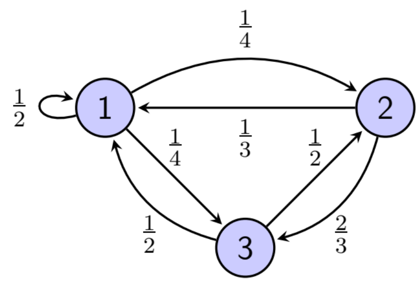
(d) No. If the chain is reversible, there exists a distribution π which satisfy:

$$\begin{cases} \pi_1 \cdot \frac{1}{4} = \pi_2 \cdot \frac{1}{3} \\ \pi_1 \cdot \frac{1}{4} = \pi_3 \cdot \frac{1}{2} \\ \pi_2 \cdot \frac{2}{3} = \pi_3 \cdot \frac{1}{2} \end{cases}$$

The solution of the above problem is $\pi_1 = \pi_2 = \pi_3 = 0$, which cannot satisfy the constraint $\sum \pi = 1$.

Problem 2

Given a Markov chain with state-transition diagram shown as follows:



(a) Find $P(X_9 = 3 \mid X_8 = 1)$ and $P(X_8 = 2 \mid X_7 = 3)$.

(b) If $P(X_0 = 3) = \frac{1}{2}$, find $P(X_0 = 3, X_1 = 1, X_2 = 2, X_4 = 3)$.

(c) Find $E(X_8 \mid X_6 = 2)$.

(d) Find $Var(X_7|X_5 = 3)$.

Solution

The transition matrix of the Markov chain is

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

(a) From the state-transition diagram, we have

$$\begin{aligned} P(X_9 = 3|X_8 = 1) &= \frac{1}{4} \\ P(X_8 = 2|X_7 = 3) &= \frac{1}{2} \end{aligned} \tag{1}$$

(b)

$$\begin{aligned} P(X_0 = 3, X_1 = 1, X_2 = 2, X_4 = 3) &= P(X_1 = 1, X_2 = 2, X_4 = 3|X_0 = 3)P(X_0 = 3) \\ &= P(X_2 = 2, X_4 = 3|X_1 = 1)P(X_1 = 1|X_0 = 3)P(X_0 = 3) \\ &= P(X_4 = 3|X_2 = 2)P(X_2 = 2|X_1 = 1)P(X_1 = 1|X_0 = 3)P(X_0 = 3) \end{aligned} \tag{2}$$

We need to calculate $P(X_4 = 3|X_2 = 2)$,

$$\begin{aligned} P(X_4 = 3|X_2 = 2) &= P(X_4 = 3, X_3 = 1|X_2 = 2) + P(X_4 = 3, X_3 = 2|X_2 = 2) + P(X_4 = 3, X_3 = 3|X_2 = 2) \\ &= P(X_4 = 3|X_3 = 1) * P(X_3 = 1|X_2 = 2) + P(X_4 = 3|X_3 = 2) * P(X_3 = 2|X_2 = 2) \\ &+ P(X_4 = 3|X_3 = 3) * P(X_3 = 3|X_2 = 2) \\ &= \frac{1}{3} * \frac{1}{4} + 0 + 0 \\ &= \frac{1}{12} \end{aligned} \tag{3}$$

Problem 2 continued on next page...

Thus we have

$$\begin{aligned} P(X_0 = 3, X_1 = 1, X_2 = 2, X_4 = 3) &= P(X_4 = 3|X_2 = 2)P(X_2 = 2|X_1 = 1)P(X_1 = 1|X_0 = 3)P(X_0 = 3) \\ &= \frac{1}{12} * \frac{1}{4} * \frac{1}{2} * \frac{1}{2} \\ &= \frac{1}{192} \end{aligned} \tag{4}$$

(c) We need to calculate Q^2,

$$Q^2 = \begin{bmatrix} \frac{11}{24} & \frac{1}{4} & \frac{7}{24} \\ \frac{1}{2} & \frac{5}{12} & \frac{1}{12} \\ \frac{5}{12} & \frac{1}{8} & \frac{11}{24} \end{bmatrix}$$

The expectation is

$$\mathrm{E}(X_8|X_6 = 2) = \sum_{i=1}^{3} iP(X_8 = i|X_6 = 2) = \frac{1}{2} + \frac{10}{12} + \frac{3}{12} = \frac{19}{12}.$$

(d) First we find $E(X_7|X_5 = 3)$,

$$\mathrm{E}(X_7|X_5 = 3) = \sum_{i=1}^{3} iP(X_7 = i|X_5 = 3) = \frac{5}{12} + \frac{2}{8} + \frac{33}{24} = \frac{49}{24}.$$

$$\mathrm{E}(X_7^2|X_5 = 3) = \sum_{i=1}^{3} i^2 P(X_7 = i|X_5 = 3) = \frac{5}{12} + \frac{4}{8} + \frac{99}{24} = \frac{121}{24}.$$

$$\mathrm{Var}(X_7|X_5 = 3) = \frac{121}{24} - (\frac{49}{24})^2 = \frac{503}{576}$$

Problem 3

There are two urns with a total of $2N$ distinguishable balls. Initially, the first urn has N white balls and the second urn has N black balls. At each stage, we pick a ball at random from each urn and interchange them. Let X_n be the number of black balls in the first urn at time n. This is a Markov chain on the state space $\{0, 1, \ldots, N\}$.

(a) (5 points) Find the transition probabilities of the chain.

(b) (5 points) Find the stationary distribution of the chain.

Solution

(a) We first note that

$$P(X_{n+1} = 1 \mid X_n = 0) = 1, \quad P(X_{n+1} = N-1 \mid X_n = N) = 1.$$

If $X_n = i$, $i \in \{1, \ldots, N-1\}$, we have $X_{n+1} \in \{i-1, i, i+1\}$.

- Observe that $X_{n+1} = i - 1$ if and only if we have chosen a black ball from the first urn and a white ball from the second:
$$P(X_{n+1} = i-1 \mid X_n = i) = \frac{i}{N} \cdot \frac{i}{N} = \frac{i^2}{N^2}.$$
- Similarly, we have that $X_{n+1} = i + 1$ if and only if we have chosen a white ball from the first urn and a black ball from the second:
$$P(X_{n+1} = i+1 \mid X_n = i) = \frac{N-i}{N} \cdot \frac{N-i}{N} = \frac{(N-i)^2}{N^2}.$$
- The number of black balls in the first urn remains the same if and only if we have picked different colors:
$$P(X_{n+1} = i \mid X_n = i) = \frac{i}{N} \cdot \frac{N-i}{N} + \frac{N-i}{N} \cdot \frac{i}{N} = \frac{2i(N-i)}{N^2}.$$

(b) Note two important observations:

- The Markov chain is irreducible.
- The Markov chain is a step-by-step analogy to the story of the Hypergeometric distribution.

These two observations lead to the guess of the stationary distribution as $\mathbf{s} = [s_0, \ldots, s_i, \ldots, s_N]$ with the PMF of the Hypergeometric distribution, i.e.,

$$s_i = \frac{\binom{N}{i}\binom{N}{N-i}}{\binom{2N}{N}}.$$

Due to irreducibility, we justify the proposed distribution by checking the detailed balance equation:

$$s_i q_{ij} = s_j q_{ji}, \quad \forall i, j \in \{0, 1, \ldots, N\}.$$

For state $i = 0$, the only non-trivial case we need to check is state $j = 1$ since there is no direct transition to other states. Therefore, we have that

$$s_0 q_{01} = s_1 q_{10},$$

which simplifies as:

$$\frac{\binom{N}{0}\binom{N}{N}}{\binom{2N}{N}} \cdot 1 = \frac{\binom{N}{1}\binom{N}{N-1}}{\binom{2N}{N}} \cdot \frac{1^2}{N^2}.$$

Problem 3 continued on next page. . .

We use the fact that $\binom{N}{1} = N$ and $\binom{N}{N-1} = N$ to simplify:

$$\frac{1}{\binom{2N}{N}} = \frac{N \cdot N}{\binom{2N}{N} \cdot N^2}.$$

Similarly, for $i = N$, the only non-trivial case is for $j = N - 1$, which is true using the same calculations. For $i = 1, \ldots, N-1$, non-trivial cases happen for $j = i - 1$ and $j = i + 1$. We are going to show that the equation holds for $1 < i \leq N - 1$ and for $j = i - 1$ (all other calculations are similar or we have already shown). We have that

$$s_i q_{i,i-1} = s_{i-1} q_{i-1,i}.$$

Substituting the values:

$$\frac{\binom{N}{i}\binom{N}{N-i}}{\binom{2N}{N}} \cdot \frac{i^2}{N^2} = \frac{\binom{N}{i-1}\binom{N}{N-i+1}}{\binom{2N}{N}} \cdot \frac{(N-i+1)^2}{N^2}.$$

Expanding the binomial coefficients:

$$\frac{\binom{N}{i} \cdot \binom{N}{N-i} \cdot i^2}{N!} = \frac{\binom{N}{i-1} \cdot \binom{N}{N-i+1} \cdot (N-i+1)^2}{N!}.$$

Rewriting in factorial form:

$$\frac{\frac{N!}{i!(N-i)!} \cdot \frac{N!}{(N-i)!i!} \cdot i^2}{N!} = \frac{\frac{N!}{(i-1)!(N-i+1)!} \cdot \frac{N!}{(N-i+1)!(i-1)!} \cdot (N-i+1)^2}{N!}.$$

Simplifying:

$$\frac{N!}{(i-1)!(N-i)!} \cdot \frac{1}{(N-i)!} \cdot \frac{i^2}{i} = \frac{N!}{(i-1)!(N-i)!} \cdot \frac{1}{(i-1)!} \cdot (N-i+1)^2.$$

Therefore:

$$s_i q_{i,i-1} = s_{i-1} q_{i-1,i}.$$

Hence, we have shown that the chain is reversible, and s is the stationary distribution.

Problem 4

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from the distribution $\mathcal{N}(\mu, \sigma^2)$, where both μ and σ^2 are unknown constants. Suppose the observed data is $\mathbf{x} = (x_1, \ldots, x_n)$, find both $\hat{\mu}$ (estimate of μ) and $\hat{\sigma}^2$ (estimate of σ^2) through the MLE (Maximum Likelihood Estimation) rule.

Solution

Since $X_1, X_2, \ldots, X_n$ are i.i.d. from $\mathcal{N}(\mu, \sigma^2)$, the joint pdf (likelihood) is:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\tfrac{(x_i-\mu)^2}{2\sigma^2}\Big).$$

To simplify the maximization, we take the natural logarithm of the likelihood:

$$\ell(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = \sum_{i=1}^{n} \ln\Big[\tfrac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\tfrac{(x_i-\mu)^2}{2\sigma^2}\big)\Big].$$

We can rewrite this as:

$$\ell(\mu, \sigma^2) = \sum_{i=1}^{n} \Big[-\tfrac{1}{2}\ln(2\pi\sigma^2) - \tfrac{(x_i-\mu)^2}{2\sigma^2}\Big].$$

Collecting terms, we get:

$$\ell(\mu, \sigma^2) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Differentiate $\ell(\mu, \sigma^2)$ with respect to μ:

$$\frac{\partial \ell}{\partial \mu} = \frac{\partial}{\partial \mu}\Big[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\Big] = -\frac{1}{2\sigma^2}\cdot(-2)\sum_{i=1}^{n}(x_i - \mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu).$$

Set this equal to zero to find the critical point:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \quad \Longrightarrow \quad \sum_{i=1}^{n}(x_i - \mu) = 0 \quad \Longrightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Hence, the MLE for μ is the sample mean.

Substituting $\mu = \hat{\mu}$ into the log-likelihood, we get:

$$\ell(\hat{\mu}, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

Differentiate this with respect to σ^2:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

Set this equal to zero:

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = 0.$$

Solving for σ^2, we obtain:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

The Maximum Likelihood Estimators for the normal distribution parameters are therefore:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

Problem 5

Given a coin with the probability p of landing heads. p is unknown and we need to estimate its value through data. In our data collection model, we have n independent tosses, result of each toss is either Head or Tail. Let X denote the number of heads in the total n tosses. Now we conduct experiments to collect data and find $X = k$. Then we need to find $\hat{p}$, the estimation of p.

(a) Assume p is an unknown constant. Find $\hat{p}$ through the MLE (Maximum Likelihood Estimation) rule.

(b) Assume p is a random variable with a prior distribution $p \sim \text{Beta}(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MAP (Maximum a Posterior Probability) rule.

(c) Assume p is a random variable with a prior distribution $p \sim \text{Beta}(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MMSE (Minimal Mean Squared Error) rule.

Solution

(a) Let X_i be the outcome of ith toss. Then $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Bern}(p)$, where p is an unknown constant. The PMF of X_i can be formulated as

$$P_{X_i}(x_i; p) = p^{x_i}(1-p)^{1-x_i}$$

since

$$p^{x_i}(1-p)^{1-x_i} = \begin{cases} p, & \text{if } x_i = 1, \\ 1-p, & \text{if } x_i = 0. \end{cases}$$

The likelihood function is

$$P_X(x; p) = \prod_{i=1}^{n} P_{X_i}(x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^k(1-p)^{n-k}$$

So the corresponding log-likelihood function is

$$g(p) = \log P_X(x; p) = \log p^{S_n}(1-p)^{n-S_n} = S_n \log p + (n - S_n)\log(1-p)$$

Now we try to find $\widehat{p}_{\text{MLE}}$ such that $g(\widehat{p}_{\text{MLE}})$ is the maximum of $g(p)$. We have

$$\begin{aligned} g'(p) &= \frac{k}{p} - \frac{n-k}{1-p}, \\ g''(p) &= -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} \leq 0 \end{aligned}$$

Let $g'(p) = 0$, we can get $p = \frac{k}{n}$. Since $g''(p) \leq 0$, then we know that

$$\widehat{p}_{\text{MLE}} = \frac{k}{n}$$

is the MLE of p.

(b) We know the posterior distribution

$$f_{p|X=k} \propto p^{a+k-1}(1-p)^{b+n-k-1},\ p \in (0,1)$$

by Beta-Binomial conjugacy. Then the MAP estimator

$$\widehat{p}_{\text{MAP}} = \arg\max_p f_{\theta|X=k} = \arg\max_p \log(f_{p|X=k})$$

since logarithmic function is monotonically increasing. Let

$$g(p) = \log(f_{p|X=k}) = (a+k-1)\log p + (b+n-k-1)\log(1-p),$$

where we don't consider the proportional constant. Our goal is to find p^* such that $g(p^*)$ is maximum of $g(p)$. We have

$$\begin{aligned} g'(p) &= \frac{a+k-1}{p} - \frac{b+n-k-1}{1-p}, \\ g''(p) &= -\frac{a+k+1}{p^2} - \frac{b+n-k-1}{(1-p)^2} < 0. \end{aligned}$$

Let $g'(p^*) = 0$. We have $p^* = \frac{a+k-1}{a+b+n-2}$, and $g(p^*)$ is maximum of $g(p)$ since $g''(p) < 0$.

Then we can get the MAP estimate

$$\widehat{p}_{\text{MAP}} = \arg\max_p f_{p|X=k} = \arg\max_p \log(f_{\theta|X=k}) = p^* = \frac{a+k-1}{a+b+n-2}.$$

(c) Since the prior distribution is $p \sim \text{Beta}(a,b)$ and the conditional distribution of X given p is $X|p \sim \text{Bin}(n,p)$, we can get the posterior distribution

$$\Theta|X = k \sim \text{Beta}(a+k, b+n-k)$$

by Beta-Binomial conjugacy. It follows that

$$E(p|X=k) = \frac{a+k}{a+b+n},$$

so the MMSE estimation of Θ is

$$\widehat{p}_{\text{MMSE}} = E(p|X=k) = \frac{a+k}{a+b+n}.$$

Problem 6

Two chess players, Vishy and Magnus, play a series of games. Given p, the game results are i.i.d. with probability p of Vishy winning, and probability $q = 1-p$ of Magnus winning (assume that each game ends in a win for one of the two players). But p is unknown, so we will treat it as an r.v. To reflect our uncertainty about p, we use the prior $p \sim \text{Beta}(a, b)$, where a and b are known positive integers and $a \geq 2$.

(a) Find the expected number of games needed in order for Vishy to win a game (including the win). Simplify fully; your final answer should not use factorials or Γ.

(b) Explain in terms of independence vs. conditional independence the direction of the inequality between the answer to (a) and $1 + E(G)$ for $G \sim \text{Geom}\left(\frac{a}{a+b}\right)$.

(c) Find the conditional distribution of p given that Vishy wins exactly 7 out of the first 10 games.

Solution

(a) Denote N as the number of games needed for Vishy to win the game for one time, then there is $N|p \sim \text{FS}(p)$. Via Adam's law, we have:

$$
\begin{aligned}
\mathrm{E}(N) &= \mathrm{E}(\mathrm{E}(N|p)) \\
&= \mathrm{E}\left(\frac{1}{p}\right) \\
&= \int_0^1 \frac{1}{\beta(a,b)} \frac{1}{p} p^{a-1} (1-p)^{b-1}\, dp \\
&= \frac{\beta(a-1,b)}{\beta(a,b)} \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} \\
&= \frac{a+b-1}{a-1}
\end{aligned}
$$

(b) Since $1 + E(G) = \frac{a+b}{a}$ can be seen as the expectation of the number of trials for Vishy to win the first game given $p = \frac{a}{a+b}$. The winning rate is constant.
$\frac{a+b-1}{a-1} = E(N)$ means that the games are conditionally independent given p while not independent between each other. When p is a r.v., and each time of Vishy's loss will decrease the belief of wining the game. Thus the expectation estimated by conditional probability is larger than the expectation given by prior distribution. Therefore $1 + E(G) < E(N)$.

(c) Via Beta-Binomial conjugacy, the conditional distribution of p give that Vish wins exactly 7 out out of the first 10 games is $\text{Beta}(a+7, b+3)$.