SI 140A-02 Probability & Statistics for EECS, Fall 2024
Homework 5

Name: **Wenye Xiong**
Student ID: 2023533141

Due on Nov. 5, 2024, 11:59 UTC+8

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Any form of plagiarism will lead to 0 point of this homework.

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 total number of stops, and we get $X = X_2 + X_3 + \cdots + X_n$, where X_i is the indicator random variable for the event that the elevator stops at floor i. Then we have:

$$E(X_i) = P(\text{at least one people stop at floor i}) = 1 - P(\text{no people stop at floor i}) = 1 - \left(1 - \frac{1}{n-1}\right)^k = 1 - \left(\frac{n-2}{n-1}\right)^k$$

Thus we have:

$$E(X) = E(X_2) + E(X_3) + \cdots + E(X_n) = (n-1)(1 - \left(\frac{n-2}{n-1}\right)^k)$$

(b):
As we have discussed in (a), we have:

$$E(X_i) = 1 - (1 - p_i)^k$$

Thus we have:

$$E(X) = E(X_2) + E(X_3) + \cdots + E(X_n) = \sum_{i=2}^{n}(1 - (1 - p_i)^k) = n - 1 - \sum_{i=2}^{n}(1 - p_i)^k$$

Problem 2

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 (as is sometimes the case, for example, with toys included in cereal boxes or included with kids' meals from a fast food restaurant). 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 Var(N).

Solution

Let N be the random variable that represents the number of toys needed to collect a complete set of n types of toys. We first want to know the Expectation of N. Let I_i denote the sequence of the type of toys that we collect at the i-th time($1 \leq C_i \leq n$).
We also let X_i be the time to collect the i-th new type of toys after we have collected $i-1$ types of toys. Then we have: $N = X_1 + X_2 + \cdots + X_n$.
Think of the probability of collecting a new type of toy is $p_i = \frac{n-i+1}{n}$, then we know that X_i follows a geometric distribution with parameter p_i. Thus we have: $E(X_i) = \frac{1}{p_i} = \frac{n}{n-i+1}$.
By the linearity of expectation, we have: $E(N) = E(X_1) + E(X_2) + \cdots + E(X_n) = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} = n(1 + \frac{1}{2} + \cdots + \frac{1}{n}) = nH_n$, where H_n is the n-th harmonic number. For large n, we have $H_n \approx \ln n + \gamma$, where γ is the Euler-Mascheroni constant.

Now we want to know the variance of N. We have:

$$Var(N) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = \sum_{i=1}^{n} \frac{1-p_i}{p_i^2} = \sum_{i=1}^{n} \frac{n\cdot(i-1)}{(n-i+1)^2} = \sum_{i=0}^{n-1} \frac{ni}{(n-i)^2}$$
$$= \sum_{i=1} n\frac{n(n-i)}{i^2} = n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n\sum_{i=1}^{n} \frac{1}{i} = n^2 \sum_{i=1}^{n} \frac{1}{i^2} - nH_n = \frac{\pi^2}{6}n^2 - n\ln n - n\gamma$$

Problem 3

Given a six-sided dice, let X denote the number obtained by rolling the dice one time. The PMF of X is: $P(X=1)=P(X=2)=\frac{1}{7}, P(X=3)=P(X=4)=\frac{1}{5}$, $P(X=5)=\frac{2}{35}, P(X=6)=\frac{9}{35}$. Now the dice is rolled five times independently. What is more likely: a sum of 24 or a sum of 25 ?

Solution

Consider the Probability Generating Function of X, we have:

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

Since the dice is rolled five times independently, X_i are i.i.d. random variables. Thus we have:

$$E(t^X)=E(t^{X_1+X_2+X_3+X_4+X_5})=E(t^{X_1})^5=(\frac{1}{7}t+\frac{1}{7}t^2+\frac{1}{5}t^3+\frac{1}{5}t^4+\frac{2}{35}t^5+\frac{9}{35}t^6)^5$$

Now we want to know the probability of the sum of 24 and 25, which is the coefficient of t^{24} and t^{25} in $E(t^X)$. We can expand $E(t^X)$ with the help of Wolfram Alpha, and we get:

$$\begin{gathered}E(t^X)=(59049t^{30})/52521875+(13122t^{29})/10504375+(51759t^{28})/10504375+(88047t^{27})/10504375+\\(22689t^{26})/1500625+(1340537t^{25})/52521875+(375916t^{24})/10504375+(544406t^{23})/10504375+\\(687591t^{22})/10504375+(836718t^{21})/10504375+(4808007t^{20})/52521875+(1020519t^{19})/10504375+\\(1047666t^{18})/10504375+(994101t^{17})/10504375+(896977t^{16})/10504375+(3815657t^{15})/52521875+\\(24134t^{14})/420175+(90441t^{13})/2100875+(2487t^{12})/84035+(7941t^{11})/420175+(186t^{10})/16807+\\(69t^9)/12005+(45t^8)/16807+(17t^7)/16807+(5t^6)/16807+t^5/16807\end{gathered}$$

We can see that the coefficient of t^{24} is $\frac{375916}{10504375}$, and the coefficient of t^{25} is $\frac{1340537}{52521875}$. Thus we know that the probability of the sum of 24 is more likely than the sum of 25.

Problem 4

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

Let $X \sim \text{Pois}(\lambda)$, then we have:

$$E\left[\binom{X}{k}\right] = \sum_{i=0}^{\infty} \binom{i}{k} \frac{e^{-\lambda}\lambda^i}{i!} = \sum_{i=k}^{\infty} \frac{i!}{k!(i-k)!} \frac{e^{-\lambda}\lambda^i}{i!} = \sum_{i=k}^{\infty} \frac{e^{-\lambda}\lambda^i}{k!(i-k)!} = \frac{e^{-\lambda}}{k!} \sum_{i=k}^{\infty} \frac{\lambda^i}{(i-k)!}$$

Let $j = i - k$, then we have:

$$E\left[\binom{X}{k}\right] = \frac{e^{-\lambda}}{k!} \sum_{j=0}^{\infty} \frac{\lambda^{j+k}}{j!} = \frac{e^{-\lambda}\lambda^k}{k!} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \frac{e^{-\lambda}\lambda^k}{k!} e^{\lambda} = \frac{\lambda^k}{k!}$$

Thus we have proved that $E\left[\binom{X}{k}\right] = \frac{\lambda^k}{k!}$.

Problem 5

(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 $E\left(X^4\right)$ for $X \sim \text{Pois}(\lambda)$ by using the identity from (a) with the fact that X has mean λ and variance λ.

Solution

(a):
Because $X \sim \text{Pois}(\lambda)$, we have $P(X=i) = \frac{e^{-\lambda}\lambda^i}{i!}$. Then we have:

$$E(Xg(X)) = \sum_{i=0}^{\infty} ig(i)\frac{e^{-\lambda}\lambda^i}{i!} = \lambda \sum_{i=0}^{\infty} g(i)\frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!}$$

Let $j = i-1$, then we have:

$$E(Xg(X)) = \lambda \sum_{j=0}^{\infty} g(j+1)\frac{e^{-\lambda}\lambda^j}{j!} = \lambda E(g(X+1))$$

Thus we have proved that $E(Xg(X)) = \lambda E(g(X+1))$.

(b):
Let $g(X) = X^4$, then we have:

$$\begin{gathered} E(X^4) = E(Xg(X)) = \lambda E(g(X+1)) = \lambda E((X+1)^3) = \lambda(E(X^3) + E(3X^2) + E(3X) + 1) \\ = \lambda(E(X^3) + 3(\lambda^2+\lambda) + 3\lambda + 1) = \lambda(\lambda E((X+1)^2) + 3\lambda^2 + 6\lambda + 1) \\ = \lambda(\lambda(E(X^2) + E(2X) + 1) + 3\lambda^2 + 6\lambda + 1) = \lambda(\lambda(\lambda^2 + \lambda + 2\lambda + 1) + 3\lambda^2 + 6\lambda + 1) \\ = \lambda(\lambda^3 + 6\lambda^2 + 7\lambda + 1) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda \end{gathered}$$

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

Suppose $P(H) = p$ and $P(T) = 1 - p = q$. Let $P_k = P(N = k)$, then we have $P_0 = P_1 = P_2 = P_3 = 0$, $P_4 = p^2q^2$, $P_5 = p^2q^2$, $P_6 = (1 - pq)p^2q^2$.
Denote S_i as the result of the $i - th$ toss, then we have the following equations according to the law of total probability:

$$P_k = P(N = k) = P(N = k|S_1 = H)P(S_1 = H) + P(N = k|S_1 = T)P(S_1 = T) = P(N = k, S_1 = H) + P(N = k, S_1 = T)$$
$$P(N = k, S_1 = H) = P(N = k, S_1 = H|S_2 = H)P(S_2 = H) + P(N = k, S_1 = H|S_2 = T)P(S_2 = T) = P(N = k, S_1 = H, S_2 = H) + P(N = k, S_1 = H, S_2 = T)$$
$$P(N = k, S_1 = H) = \hat{P_k}$$
$$P(N = k, S_1 = H, S_2 = T) = pqP(N = k - 2) = pqP_{k-2}$$
$$P(N = k, S_1 = T) = qP(N = k - 1) = qP_{k-1}$$
$$P(N = k, S_1 = H, S_2 = H) = p^2P(N = k - 1, S_1 = H) = p^2\hat{P_{k-1}}$$

Take $p = \frac{1}{2}$, then we have:

$$\hat{P_k} = \frac{\hat{P_{k-1}}}{2} + \frac{\hat{P_{k-3}}}{8} + \frac{P_{k-3}}{4}$$

So $P_k = P_{k-1} - \frac{P_{k-2}}{2} + \frac{P_{k-3}}{2} - \frac{P_{k-4}}{8}$.
Then we have the Probability Generating Function:

$$G(t) = \frac{t^4}{16} + \sum_{i=5}^{\infty} P_i t^i = \frac{t^4}{16} + G(t)(t - \frac{t^2}{2} + \frac{t^3}{4} - \frac{t^4}{8})$$

Then we have:

$$E(N) = G'(1) = 20$$
$$Var(N) = G''(1) + G'(1) - (G'(1))^2 = 276$$

Problem 7

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 $I_i I_j$ for $i < j$? Use this to simplify the expansion

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

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 $E(X) \approx \sqrt{\frac{\pi n}{w}}$. 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

$$P(Z = 0) = (1 - \tfrac{1}{365})^n \approx e^{\lambda}$$
$$P(Z \geq 1) = 1 - P(Z = 0) \approx 1 - e^{\lambda}$$

where $\lambda = \binom{m}{2} p$, m is the number of people and p is the probability of having the same birthday.
Thus we have:

$$P(X \leq 23) \approx 1 - e^{\lambda} \approx 0.5002 > 0.5$$

On the other hand, we have:

$$P(X \geq 23) = e^{\lambda} \approx 0.531 \geq 0$$

Thus we have proved that 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

$$E(X) = E(I_1 + I_2 + \cdots + I_{366}) = E(I_1) + E(I_2) + \cdots + E(I_{366}) = p_1 + p_2 + \cdots + p_{366} = \sum_{j=1}^{366} p_j$$

(c):
We can compute $E(X)$ numerically by using the formula in (b). Using Python, we get that $E(X) \approx 24.62$.

(d):
$E(X^2) = E(I_1^2 + I_2^2 + \cdots + I_{366}^2 + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_i I_j) = E(I_1^2) + E(I_2^2) + \cdots + E(I_{366}^2) + 2E(\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_j)$
$= \sum_{j=1}^{366} p_j + 2\sum_{j=1}^{366}(j-1)E(I_j) = \sum_{j=1}^{366}(2j-1)p_j$. Thus we have: $D(X) = E(X^2) - (E(X))^2 \approx 148.64$

Problem 8

(**Optional Challenging Problem**)

(a) Let W be a bounded non-negative integer-valued random variable. If for all integer $k \geq 0$,

$$E\left[\binom{W}{k}\right] \approx \frac{\lambda^k}{k!}$$

Show that

$$P(W = k) \approx e^{-\lambda}\frac{\lambda^k}{k!}$$

(b) There are n people in a room. Assume each person's birthday is equally likely to be any of the 365 days of the year (we exclude February 29), and that peoples birthdays are independent (we assume there are no twins in the room). When the probability that three or more people in the group have the same birthday is 1/2, find n.