2023Fall Probability & Mathematical Statistics 2023/10/29

Homework 4

Professor: Ziyu Shao & Dingzhu Wen *Due: 2023/11/05 10:59pm*

1. Let X have PMF
$$P(X = k) = cp^k/k \text{ for } k = 1, 2, \dots,$$
where p is a parameter with $0 < p < 1$ and c is a normalizing constant. We have $c = -1/\log(1-p)$, as seen from the Taylor series
$$-\log(1-p) = p + \frac{p^2}{2} + \frac{p^3}{3} + \cdots.$$
This distribution is called the *Logarithmic* distribution (because of the log in the above Taylor series), and has often been used in ecology. Find the mean and variance of X.

2. 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 $X_1, X_2, \cdots$ be Nick's results and $Y_1, Y_2, \cdots$ be Penny's results, with $X_i \sim \text{Bern}(p_1)$ and $Y_j \sim \text{Bern}(p_2)$.

   (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.

3. A building has n floors, labeled $1, 2, ..., n$. 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 $2, 3, ..., n$ to go to and presses that button (unless someone has already pressed it).

   (a) Assume for this part only that the probabilities for floors $2, 3, ..., n$ are equal. Find the expected number of stops the elevator makes on floors $2, 3, ..., n$.

   (b) Generalize (a) to the case that floors $2, 3, ..., n$ have probabilities $p_2, ..., p_n$ (respectively); you can leave your answer as a finite sum.

$$\sum_k k g(k) \cdot e^{-\lambda} \frac{\lambda^k}{k!}$$

$\sum_{k} \lambda g(k+1) \frac{\lambda^k}{k!}$

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 third moment $E(X^3)$ 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.

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.

$X = \sum_j [X \geq j]$

(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 $E(X) \approx \sqrt{\pi n/2}$. In Volume 1 of his masterpiece *The Art of Computer Programming*, Don Knuth shows that an even better approximation is

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

$P(X \geq 1) = 1$

$1 \cdot \frac{364}{365}$

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 "HHH". Find $E(N)$ and $Var(N)$.

7. (**Optional Challenging Problem**) Show the following theorems:

   (a) Given a complete graph K_n ($n \geq 3$), if $\binom{n}{m} 2^{-\binom{m}{2}+1} < 1$, then it is possible to color the edges of K_n with two colors so that it has no monochromatic K_m subgraph ($1 < m < n$).

   (b) Let $M \in F(x_1, x_2, \ldots, x_n)$ be a non-zero polynomial of total degree $d \geq 0$ over a field F. Let S be a finite subset of F and let $r_1, r_2, \ldots, r_n$ be selected at random independently and uniformly from S. Then

$$P[M(r_1, r_2, \ldots, r_n) = 0] \leq \frac{d}{|S|}.$$