2024Spring Probability & Statistics for EECS 2024/03/27

Homework 4

Professor: Ziyu Shao *Due: 2024/03/31 10:59pm*

1. Let X have PMF
$$P(X = k) = cp^k/k \text{ for } k = 1, 2, \ldots,$$
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. Let a random variable X satisfies Hypergeometric distribution with parameters w, b, n.
   (a) Find $E\left[\binom{X}{2}\right]$
   (b) Use the result of (a) to find the variance of X.

3. 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 $\text{Var}(N)$

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, independently of past types collected, it is type j with probability p_j, and $\sum_{j=1}^{n} p_j = 1$. Let N denote the number of different types of toys that appear among the first m collected toys. Find $E(N)$ and $\text{Var}(N)$.

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 $I_i I_j$ 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

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

6. (**Optional Challenging Problem**) 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, independently of past types collected, it is type j with probability p_j, and $\sum_{j=1}^{n} p_j = 1$. Let N denote the number of toys needed until you have a complete set. Find $E(N)$ and $\mathrm{Var}(N)$.