SI 140A-02 Probability & Statistics for EECS, Fall 2024
Homework 4

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

Due on Oct. 29, 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

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

(a):
Define Z_i, where $Z_i = 1$ if $X_i = Y_i = 1$ and $Z_i = 0$ otherwise.
Then $p(Z = k) = (1 - p_1p_2)^{k-1}p_1p_2 \sim Geom(p_1p_2)$. Thus $E[Z] = \frac{1}{p_1p_2}$.
(b):
Similarly, define W_i, where $W_i = 0$ if $X_i = Y_i = 0$ and $W_i = 1$ otherwise.
Then $p(W = k) \sim Geom(1 - (1 - p_1)(1 - p_2))$. Thus $E[W] = \frac{1}{1-(1-p_1)(1-p_2)} = \frac{1}{p_1+p_2-p_1p_2}$.
(c):
Let $p = p_1 = p_2$, N denotes Nick's first success and P denotes Penny's first success.
Then $p(N = P) = \sum_{k=1}^{\infty} p^2((1 - p)^2)^{k-1} = p^2 \sum_{k=0}^{\infty}(1 - p)^{2k} = \frac{p^2}{1-(1-p)^2} = \frac{p^2}{2p-p^2} = \frac{p}{2-p}$.
According to the symmetry, $p(N < P) = p(P < N)$, thus $p(N > P) = \frac{1-\frac{p}{2-p}}{2} = \frac{1-p}{2-p}$.

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 $X \sim \text{Bern}(p)$ and $Y \sim \text{Bern}(1/2)$, 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

$$Y_J = \oplus_{j \in J} X_j.$$

Show that $Y_J \sim \text{Bern}(1/2)$ and that these $2^n - 1$ R.V.s are pairwise independent, but not independent.

Solution

(a):
The distribution of $X \oplus Y$ is $\frac{1}{2}$, no matter what the value of p is.

$$\begin{aligned} P(X \oplus Y = 1) &= P(X \oplus Y = 1|X = 1)P(X = 1) + P(X \oplus Y = 1|X = 0)P(X = 0) \\ &= P(Y = 0)P(X = 1) + P(Y = 1)P(X = 0) = \tfrac{p}{2} + \tfrac{1-p}{2} = \tfrac{1}{2} \end{aligned}$$

(b):
The conditional distribution of $X \oplus Y|(X = x)$ is $\frac{1}{2}$, as shown in (a), which is independent of X.
This result and the result from (a) make sense intuitively: adding Y destroys all information about X, resulting in a fair coin flip independent of X. Given $X = x$, $X \oplus Y$ is x with probability $\frac{1}{2}$, and 1-x with probability $\frac{1}{2}$, so $X \oplus Y|(X = x) \sim Bern(1/2)$.

If $p = 1/2$, the distribution formula indicates that $X \oplus Y$ is independent of X and Y.
But if $p \neq 1/2$, $X \oplus Y$ is not independent of Y, as the distribution of $X \oplus Y|(Y = y)$ is $Bern(p(1-y)+(1-p)y)$.

$$P(X \oplus Y = 1|Y = y) = P(X \neq y) = p(1 - y) + (1 - p)y$$

(c):
To show these $2^n - 1$ R.V.s are not independent, we can consider the case where we know, for example, $Y_{\{1\}}$ and $Y_{\{2\}}$. Then we get $Y_{\{1,2\}} = Y_{\{1\}} \oplus Y_{\{2\}}$.
But these $2^n - 1$ R.V.s are pairwise independent, let's say we have Y_J and Y_K, we can write them in their stated form by partioning the set $J \cup K$ into $J \cap K$ and $J \cap K^c$ and $J^c \cap K$
Assume that $J \cup K$ is a nonempty set, then by (a) $A \sim Bern(1/2)$, we have the following for $y \in \{0, 1\}, z \in \{0, 1\}$:

$$\begin{aligned} P(Y_J = y, Y_K = z) &= \tfrac{1}{2}P(A \oplus B = y, A \oplus C = z|A = 1) + \tfrac{1}{2}P(A \oplus B = y, A \oplus C = z|A = 0) \\ &= \tfrac{1}{2}P(B = y, C = z) + \tfrac{1}{2}P(B \neq y, C \neq z) = \tfrac{1}{2}\tfrac{1}{4} + \tfrac{1}{2}\tfrac{1}{4} = \tfrac{1}{4} \\ &= P(Y_J = y)P(Y_K = z) \end{aligned}$$

If $J \cup K$ is an empty set, then obviously $Y_J = B$ and $Y_K = C$ are independent.
Thus these $2^n - 1$ R.V.s are pairwise independent.

Problem 3

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

(a) Find $E\left[\begin{pmatrix} X \\ 2 \end{pmatrix}\right]$

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

Solution

(a):

In Hypergeometric, $\begin{pmatrix} X \\ 2 \end{pmatrix}$ is the number of ways to choose 2 white balls.

Thus $E\left[\begin{pmatrix} X \\ 2 \end{pmatrix}\right] = \begin{pmatrix} n \\ 2 \end{pmatrix} \frac{w}{w+b} \frac{w-1}{w+b-1} = \frac{n(n-1)w(w-1)}{2(w+b)(w+b-1)}$.

(b):

By (a) we have $EX^2 - EX = E(X(X-1)) = \frac{n(n-1)p(w-1)}{N-1}$, where $p = \frac{w}{w+b}$ and $N = w + b$.

So $Var(X) = E(X^2) - (EX)^2 = \frac{n(n-1)p(w-1)}{N-1} + np - (np)^2 = np(\frac{(N-n)(N-w)}{N(N-1)}) = \frac{N-n}{N-1}np(1-p)$.

Problem 4

Let X have PMF

$$P(X = k) = cp^k/k \ for \ k = 1, 2, \ldots,$$

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

$$-log(1-p) = p + \frac{p^2}{2} + \frac{p^3}{3} + \ldots$$

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.

Solution

We first solve the mean of X:

$$E[X] = \textstyle\sum_{k=1}^{\infty} k \cdot p(X = k) = c \cdot \sum_{k=1}^{\infty} k \cdot \frac{p^k}{k} = c \cdot \frac{p}{1-p} = -\frac{p}{(1-p)log(1-p)}.$$

Then we can find E[X^2] by the same method, and Var(X) = E[X^2] - (E[X])^2.

$$E[X^2] = \textstyle\sum_{k=1}^{\infty} k^2 \cdot p(X = k) = c \cdot \sum_{k=1}^{\infty} k^2 \cdot \frac{p^k}{k} = cp \sum_{k=1}^{\infty} k \cdot p^{k-1} = cp \cdot \frac{1}{(1-p)^2}$$
$$Var(X) = E[X^2] - (E[X])^2 = \frac{cp(1-cp)}{(1-p)^2}$$

Problem 5

Let X be a discrete R.V. whose distinct possible values are $a_1, a_2, ..., a_n$, with probabilities $p_1, p_2..., p_n$, respectively (so $p_1 + p_2 + \cdots + p_n = 1$). The entropy of X is defined to be the average surprise of learning the value of X:

$$H(X) = \sum_{j=1}^{n} p_j \log_2(1/p_j).$$

Show that the maximum entropy for X is when its distribution is uniform over $a_1, a_2, ..., a_n$, i.e., $p_j = 1/n, \forall j$.

Solution

To show that the maximum entropy for X is when its distribution is uniform over $a_1, a_2, ..., a_n$, we can use the Jensen's inequality.
Consider a r.v. Y whose possible values are the probabilities $p_1, p_2, ..., p_n$. Then $H(x) = E(log_2(1/Y))$ and $E(\frac{1}{Y}) = \sum_{j=1}^{n} p_j \cdot \frac{1}{p_j} = n$.
By Jensen's inequality, $E(log_2(1/Y)) \leq log_2(E(1/Y)) = log_2(n)$.
The equality holds if and only if Y is a constant, i.e., $p_j = 1/n, \forall j$, which means the distribution of X is uniform over $a_1, a_2, ..., a_n$.

Problem 6

(**Optional Challenging Problem I**) Show the following theorems:

1. 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 $(1 < m < n)$.

2. 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
$$P[M(r_1, r_2, \ldots, r_n) = 0] \leq \frac{d}{|S|}.$$