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 $X_1, X_2, \cdots$ 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.

Solution

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

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

(c) 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

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. $X = X_1 + X_2 + \ldots + X_n$, where $X_i = 1$ if someone stops at ith floor, otherwise $X_i = 0$.

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

Thus,

$$\begin{aligned} E(X) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= (n-1)\left[1 - (\frac{n-2}{n-1})^n\right] \end{aligned}$$

(b) According to (a)

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

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 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:

$$E\left[\left(\begin{array}{c} X \\ k \end{array}\right)\right] = \frac{\lambda^k}{k!}.$$

Solution

By definition of Poisson random variable and Taylor expansion of e^λ, we have

$$\begin{aligned}
&E\left[\left(\begin{array}{c} X \\ k \end{array}\right)\right] \\
=&\sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} \frac{e^{-\lambda}\lambda^n}{n!} \\
=&\frac{e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{\lambda^n}{(n-k)!} \\
=&\frac{e^{-\lambda}\lambda^k}{k!} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \\
=&\frac{e^{-\lambda}\lambda^k}{k!} e^{\lambda} \\
=&\frac{\lambda^k}{k!}.
\end{aligned}$$

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 third 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 $P(X=k) = \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}$, $k \in \mathbb{N}$. Denote $f(x) = Xg(X)$, we have

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

Denote $Y = X - 1$, we have

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

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

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

Since

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

we have

$$E(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$$

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 "HTHT". Find $E(N)$ and $\mathrm{Var}(N)$.

Solution

Define $P(H) = p, P(T) = 1 - P(H) = 1 - p = q$ and $p_k = P(N = k), k = 0, 1, 2, \cdots$ It is obvious that $p_0 = p_1 = p_2 = p_3 = 0, p_4 = p_5 = p^2q^2$. Define S_i as the result of the i^{th} toss, then by LOTP, we have

$$\begin{aligned}
&p_k = P(N = k) = P(N = k, S_1 = H) + P(N = k, S_1 = T)\\
&P(N = k, S_1 = T) = P(S_1 = T) \cdot P(N = k - 1) = qp_{k-1}\\
&P(N = k, S_1 = H) = P(N = k, S_1 = H, S_2 = H) + P(N = k, S_1 = H, S_2 = T).
\end{aligned}$$

Further, since we have

$$\begin{aligned}
&P(N = k, S_1 = H, S_2 = H)\\
=&P(N = k - 1, S_1 = H) \cdot P(S_1 = H)\\
=&p \cdot (P(N = k - 1) - P(N = k - 1, S_1 = T))\\
=&p \cdot (p_{k-1} - q \cdot p_{k-2})\\
=&pp_{k-1} - pqp_{k-2}
\end{aligned}$$

and

$$\begin{aligned}
&P(N = k, S_1 = H, S_2 = T)\\
&= P(N = k, S_1 = H, S_2 = T, S_3 = H) + P(N = k, S_1 = H, S_2 = T, S_3 = T)\\
&= P(S_1 = H) \cdot P(S_2 = T) \cdot P(S_3 = H) \cdot P(N = k - 3, S_1 = H)\\
&+ P(S_1 = H) \cdot P(S_2 = T) \cdot P(S_3 = T) \cdot P(N = k - 3)\\
&= p^2q \cdot (p_{k-3} - qp_{k-4}) - pq^2p_{k-3}\\
&= p_{k-3} - p^2q^2p_{k-4}.
\end{aligned}$$

Therefore, we have

$$p_k = (p + q)p_{k-1} - pqp_{k-2} + \left(p^2q + pq^2\right)p_{k-3} - p^2q^2p_{k-4}.$$

Now, we consider the PGF of N,

$$\begin{aligned}
g(t) = E\left[t^N\right] &= p_4t^4 + \sum_{k=5}^{\infty} p_kt^k\\
&= p_4t^4 + \sum_{k=5}^{\infty}\left((p + q)p_{k-1} - pqp_{k-2} + \left(p^2q + pq^2\right)p_{k-3} - p^2q^2p_{k-4}\right)t^k\\
&= p_4t^4 + (p + q)t\sum_{k=4}^{\infty} p_kt^k - pqt^2\sum_{k=3}^{\infty} p_kt^k + \left(p^2q + pq^2\right)t^3\sum_{k=2}^{\infty} p_kt^k - p^2q^2t^4\sum_{k=1}^{\infty} p_kt^k\\
&= p_4t^4 + \left((p + q)t - pqt^2 + \left(p^2q + pq^2\right)t^3 - p^2q^2t^4\right) \cdot g(t)
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&g(t) = p^2q^2t^4 + \left((p + q)t - pqt^2 + \left(p^2q + pq^2\right)t^3 - p^2q^2t^4\right) \cdot g(t)\\
\Rightarrow &g(t) = \frac{p^2q^2t^4}{1 - (p + q)t + pqt^2 - (p^2q + pq^2)t^3 + p^2q^2t^4} = \frac{t^4}{16 - 16t + 4t^2 - 4t^3 + t^4}
\end{aligned}$$

Then, we have $E(N)$ and $\mathrm{Var}(N)$ given $p = q = \frac{1}{2}$,

$$E(N) = g'(t)|_{t=1} = \left.\frac{64t^4 - 48t^3 + 8t^5 - 4t^6}{(16 - 16t + 4t^2 - 4t^3 + t^4)^2}\right|_{t=1} = 20$$

$$\mathrm{Var}(N) = g''(1) + g'(1) - \left(g'(1)\right)^2 = 656 + 20 - 400 = 276$$

Problem 6

(Optional Challenging Problem) An Erdos-Renyi random graph is formed on n vertices. Each unordered pair (edge) (i, j) of vertices is connected with probability p, independently of all the other pairs.

(a) A wedge (or path of length 2) is a tuple $(i, \{j, k\})$ where i, j, k are distinct and each of the edges (i, j) and (i, k) is connected. Let W denote the number of wedges contained in the random graph. Find appropriate condition under which W is approximately Poisson distributed.

(b) A triangle is a set of three vertices $\{i, j, k\}$ such that each of the three edges (i, j), (j, k) and (i, k) is connected. Let T denote the number of triangles contained in the random graph. Find appropriate condition under which T is approximately Poisson distributed.

Solution

Check book "Random Graphs and Complex Networks".

(a) A wedge in a graph is a path with two edges that share a common node. In the context of an Erdos-Renyi graph $G(n, p)$, the probability that a specific wedge exists is p^2 because the existence of each edge is independent and has probability p. We let

$$W = \frac{1}{2} \sum_{1 \le i,j,k \le n} \mathbf{1}_{\{ij, jk \in E\}}$$

denote the number of wedges in the graph G. The factor of two comes from the fact that the wedge ij, jk is the same as the wedge kj, ji, but it is counted twice.

Specifically, since there are n nodes, the number of possible wedges centered at a particular node i is $\binom{n-1}{2}$ because we must choose 2 other distinct nodes from the remaining $n - 1$ nodes to form the edges (i, j) and (i, k). The expected number of wedges centered at i is then $\binom{n-1}{2} p^2$.

To find the total expected number of wedges W in the graph, we multiply by n since any of the n nodes can be the center of the wedge:

$$\mathbb{E}[W] = n \binom{n-1}{2} p^2 = \frac{n(n-1)(n-2)}{6} p^2$$

The variable W is the sum of a large number of independent Bernoulli trials (one for each potential wedge), each with the same probability p^2. For W to be approximately Poisson distributed, we need n to be large and p to be small such that np^2 is a constant λ, where λ is the rate of the Poisson distribution.

(b) A triangle in the graph consists of three vertices where each pair is connected by an edge. The probability that a specific triangle exists is p^3, due to the independence of edge formations. We let

$$T = \frac{1}{6} \sum_{1 \le i,j,k \le n} \mathbf{1}_{\{ij, jk, ik \in E\}}$$

denote the number of triangles in G.

Specifically, the number of potential triangles is $\binom{n}{3}$ since we choose any 3 distinct vertices to form a triangle. The expected number of triangles T is thus:

$$\mathbb{E}[T] = \binom{n}{3} p^3 = \frac{n(n-1)(n-2)}{6} p^3$$

Similar to part (a), for T to be approximately Poisson distributed, we again need n to be large and p to be small, but this time such that np^3 is a constant λ, which will be the rate of the Poisson distribution for triangles.

In both cases, the Poisson approximation will be more accurate when the number of subgraphs (wedges or triangles) is large, and each has a small probability of existing. This situation typically occurs in sparse random graphs where p is much less than 1 but np^2 (for wedges) or np^3 (for triangles) remains constant as n goes to infinity.

This is a simplification and assumes that the graph is sparse enough for the subgraphs to be nearly independent. For more rigorous treatment, one would have to consider the dependencies between subgraphs and apply more sophisticated probabilistic techniques.