5.3 Unit 5 Practice Problems:

Problem 1: Randomly, k distinguishable balls are placed into n distinguisable boxes, with all possibilities equally likely. Find the expected number of empty boxes as a function of k and n

- Let I_j be indicator r.v that j^{th} box is empty, So the expected number of empty boxes:

$$E(X) = E(I_1 + I_2 + \dots + I_n)$$
$$= E(I_1) + E(I_2) + \dots + E(I_n)$$
$$= \sum_{j=1}^{n} E(I_j)$$

- To find $E(I_j)$, we need to calculate the probability that a given box j is empty after placing all k balls.

A box is empty if none of the k balls are placed in it. Since each ball is placed in box j with probability $\frac{1}{n}$, the probability that a ball is not placed in j^{th} box is $\left(1 - \frac{1}{n}\right)$.

Since the balls are placed independently, probability that all k balls not placed in box j is:

$$P(\text{box } j \text{ empty}) = \left(1 - \frac{1}{n}\right)^k$$

- By the fundamental bridge:

$$P(\text{box } j \text{ emty}) = E(I_j) = \left(1 - \frac{1}{n}\right)^k$$

- So expected number of empty boxes is:

$$E(X) = \sum_{j=1}^{n} E(I_j) = n\left(1 - \frac{1}{n}\right)^k$$

Problem 2: Alice and Bob each has 50 friends out of 1000 people in town. They think that they are unlikely to have friends in common since "each of them are friends with only 5% of the people in town, so their 5% are unlikely to overlap"

a) Compute the expected number of mutual friends Alice and Bob have.

- Let I_j be the indicator r.v for the j^{th} person being a mutual friend. Then

$$E\left(\sum_{j=1}^{1000} I_j\right) = 1000 \cdot E(I_j)$$
$$= 1000 \cdot P(I_j) \qquad \text{<fundamental bridge>}$$
$$= 1000 \cdot \left(\frac{50}{1000}\right)^2$$
$$= 2.5$$

b) Let X be the number of mutual friends they have. Find $P(X=2)$

- Condition on who Alice friends are, and then count the number of ways that Bob can be friends with exactly k of them. This gives:

$$P(X=k \mid \text{Alice's friends}) = \frac{\binom{50}{k}\binom{950}{50-k}}{\binom{1000}{50}}$$

choose the remaining $(50-k)$ friends from the friend that are not Alice's friend

choose k mutual friends from Alice's friends

- So $P(X=2) \approx 0.266$

c) What is the distribution of X ?

Its Hypergeometric since we can think of the problem as "tagging" Alice friends and then see how many tagged people are also Bob's friends.

Problem 3: A group of n people play "Secret Santa" where each person puts their name on a slip of paper in a hat. Then each person picks a name randomly from the hat and has to buy gift for that person. Unfortunately, they overlook the possibility of drawing one's own name, so some may have to buy gift for themselves.

a) Find the expected value of the number X of people who pick their own names

$$E(X) = E\left(\sum_{j=1}^{n} I_j\right) \quad , \text{ where } I_j \text{ is indicator r.v if a person pick his own name.}$$

$$= n\, E(I_j) \quad \langle \text{linearity} \rangle$$

$$= n\, P(I_j = 1) \quad \langle \text{fundamental bridge} \rangle$$

$$= n \frac{1}{n} = 1$$

b) Let X be the number of people who pick their own names. What is the approximate distribution of X if n is large?

- Since n is large, $p_j = P(I_j)$ are small and I_j are independent or weakly dependent:

$$X \sim \text{Pois}(1) \quad , \text{ where } \lambda = \sum_{j=1}^{n} p_j = n \cdot \frac{1}{n} = 1$$

Problem 4: Let $Z \sim N(0,1)$. A measuring device is used to observe Z, but the device can only handle positive values, and gives a reading of 0 if $Z \leq 0$; this is an example of censored data. So assume that $X = Z I_{Z>0}$ is observed rather than Z, where $I_{Z>0}$ is the indicator of $Z > 0$

a) Find E(X)

- $$E(X) = \int_0^\infty x \cdot f(x)\, dx = \int_0^\infty x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_0^\infty x e^{-x^2/2}\, dx$$

- Let $u = z^2/2$, we have:

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-u}\, du = \frac{1}{\sqrt{2\pi}} \approx 0.3989.$$

b) Find Var(X)

- $$Var(X) = E(X^2) - (EX)^2 = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^2 e^{-x^2/2} - \frac{1}{\sqrt{2\pi}}$$

$$= \frac{1}{2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}}_{1} - \frac{1}{\sqrt{2\pi}}$$

$$= \frac{1}{2} \cdot 1 - \frac{1}{\sqrt{2\pi}} \approx 0.341 \quad \langle N(0,1) \text{ has variance } 1 \rangle$$

5.4 Unit 5 Homework Problems

Problem 1: Bobo, the amoeba, currently lives in a pond. After one minute, Bobo will either die, split into 2 amoebas, or stay the same, with equal probability.

Find the expectation and variance of the number of amoebas in the pond after one minute.

- Let X be the number of amoebas in the pond after one minute:
- $$E(X) = P(X=0).x_0 + P(X=1).x_1 + P(X=2).x_2$$
  $$= \frac{1}{3}.0 + \frac{1}{3}.1 + \frac{1}{3}.2$$
  $$= 1$$
- $$Var(X) = E(X^2) - (EX)^2$$
  $$= \left(\frac{1}{3}.0^2 + \frac{1}{3}.1^2 + \frac{1}{3}.2^2\right) - 1$$
  $$= \frac{2}{3}$$

Problem 2: Two researchers independently select simple random samples from a population size N, with sample sizes m and n.

Find the expected size of the overlap of the 2 samples as a function of N, n and m. For $m=20$, $n=30$, $N=100$, find the expected size of the overlap of the two samples.

- Let I_j be indicator r.v that the j^{th} sample in population N that belongs to both m and n samples. Then:

$$E\left(\sum_{j=1}^{N} I_j\right) = N.E(I_j) \quad \langle \text{linearity} \rangle$$
$$= N.P(I_j=1) \quad \langle \text{fundamental bridge} \rangle$$
$$= N.\left(\frac{m}{N}\right)\left(\frac{n}{N}\right) \quad \langle P(I_j=1) = P(\text{sample is in both } m \text{ and } n) \rangle$$
$$= \frac{m.n}{N} = \frac{20.30}{100} = 6$$

Problem 3: For $X \sim Pois(3)$, find $E(2^X)$ if it is finite

- $E(X) = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!}$
- By LOTUS: $E(2^X) = \sum_{k=0}^{\infty} 2^k \cdot \frac{e^{-\lambda}\lambda^k}{k!}$

$$= e^{-\lambda} \sum_{k=0}^{\infty} 2^k \cdot \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!}$$

$$= e^{-\lambda} e^{2\lambda} = e^{\lambda} = e^3 \approx 20.086$$

Problem 4: Three students are working independently on their probability hw. They all start at the same time. Each takes Exponential time with mean 6 hours to complete the homework. How many hours will it take until all 3 students have completed the homework, on average?

- Let $\begin{cases} X_j \text{ be how long it takes student } j \text{ to complete the homework.} \\ T \text{ be the time it takes for all 3 students to complete the homework.} \end{cases}$
- So: $T = T_1 + T_2 + T_3$

where $\begin{cases} T_1 = \min(T_1, T_2, T_3) \text{ is how long it takes for one student to complete the hw.} \\ T_2 \text{ is the additional time for the second student to complete the hw.} \\ T_3 \text{ is the additional time for the third student to complete the hw.} \end{cases}$

also
- $T_1 \sim Expo\left(\frac{3}{6}\right)$ since $\min(T_1, T_2, T_3) \sim Expo(\lambda_1 + \lambda_2 + \lambda_3)$

  <Example 4.6.3>
- $T_2 \sim Expo\left(\frac{2}{6}\right)$ since by memoryless property, when the first student completes, the other 2 start fresh so $\min(T_2, T_3) \sim Expo(\lambda_2 + \lambda_3)$
- $T_3 \sim Expo\left(\frac{1}{6}\right)$ again similar reasoning as above using memoryless property