CS 70 Discrete Mathematics and Probability Theory
Summer 2021 **Dis 05C**

1 Ball in Bins

You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(a) What is $\mathbb{E}[X_i]$?

let $A_i = \begin{cases} 1 & \text{if } i\text{th ball lands in bin } i \\ 0 & \text{o.w.} \end{cases}$

$X_i = A_1 + \ldots + A_K \quad E[X_i] = E\left[\sum_{i=1}^{K} A_i\right] = \sum_{i=1}^{K} E[A_i] = \frac{K}{n}$

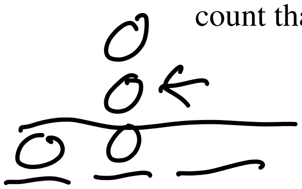
- by Linearity of Expectation

(b) What is the expected number of empty bins?

let $B_i = \begin{cases} 1 & \text{if } i\text{th bin is empty} \\ 0 & \text{o.w.} \end{cases}$

$E\left[\sum_{i=1}^{n} B_i\right] = \sum_{i=1}^{n} E[B_i] = \sum_{i=1}^{n} \left(\frac{n-1}{n}\right)^K = n\left(\frac{n-1}{n}\right)^K$

(c) Define a collision to occur when a ball lands in a non-empty bin (if there are n balls in a bin, count that as $n-1$ collisions). What is the expected number of collisions?



$E[\text{collisions}] = E[\underbrace{\#\text{ of balls}}^{K} - \#\text{ of occupied bins}]$

$= K - E[\#\text{ of occupied bins}]$

$= K - E[\#\text{ of bins} - \#\text{ of empty bins}]$

$= K - n + n\left(\frac{n-1}{n}\right)^K$

2 Variance X, Y independent r.v.'s , $Var(X+Y) = Var(X) + Var(Y)$

If the random variables are independent, we could just sum up the variances individually. If not, we generally use this technique that we will show in this problem. This problem will give you practice to compute the variance of a sum of random variables that are not pairwise independent. Recall that $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

(a) A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the expected number of floors the elevator stops at (not counting the ground floor)?

let $A_i = \begin{cases} 1 & \text{if elevator stopped at floor } i \\ 0 & \text{o.w.} \end{cases}$

let $A = A_1 + \ldots + A_n$ be # of floors stopped at

$E[A_i] = P(A_i = 1) = 1 - P(\text{no one got off at floor } i)$

$= 1 - \left(\frac{n-1}{n}\right)^m$

$E[A] = E\left[\sum_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} E[A_i] = \sum_{i=1}^{n}\left[1 - \left(\frac{n-1}{n}\right)^m\right]$

$= n\left[1 - \left(\frac{n-1}{n}\right)^m\right]$

(b) What is the *variance* of the number of floors the elevator *does not* stop at? (In fact, the variance of the number of floors the elevator *does* stop at must be the same (make sure you understand why), but the former is a little easier to compute.)

let X be # of floors elevator does not stop at

let $X_i = \begin{cases} 1 & \text{if no one gets off at floor } i \\ 0 & \text{o.w.} \end{cases}$

$X = X_1 + \dots + X_n$

$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(X_i = 1)$

$= \sum_{i=1}^{n} \left(\frac{n-1}{n}\right)^m$

$= n\left(\frac{n-1}{n}\right)^m$

ex. $(X_1 + X_2)(X_1 + X_2)$

$X_1^2 + X_1X_2 + X_2X_1 + X_2^2$

want $Var(X) = E[X^2] - E[X]^2$

$E[X^2] = E[(X_1 + \dots + X_n)^2]$

$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right]$ $\quad \sum_{i,j} := \sum_{i=1}^{n}\sum_{j=1}^{n}$

$= \sum_{i,j} E[X_i X_j]$

$$= \sum_{i=1}^{n} E[X_i^2] + 2\sum_{i<j} E[X_iX_j]$$

$$E[X_i^2] = 0^2 \cdot \cancel{P(X_i = 0)} + 1^2 \cdot P(X_i = 1)$$

$$= \left(\frac{n-1}{n}\right)^m$$

$$\Rightarrow \sum_{i=1}^{n} E[X_i^2] = n\left(\frac{n-1}{n}\right)^m$$

$$E[X_iX_j] = \cancel{0 \cdot P(X_iX_j = 0)}^{\,0} + 1 \cdot P(X_iX_j = 1)$$

$$= P(X_i = 1, X_j = 1)$$ * no one got off at floor i & floor j

$$= \left(\frac{n-2}{n}\right)^m$$

$$2\sum_{i<j} E[X_iX_j] = 2\binom{n}{2}\left(\frac{n-2}{n}\right)^m = n(n-1)\left(\frac{n-2}{n}\right)^m$$

$$\Rightarrow E[X^2] = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m$$

$$\Rightarrow Var(X) = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - n^2\left(\frac{n-1}{n}\right)^{2m}$$

3 Covariance

We have a bag of 5 red and 5 blue balls. We take two balls uniformly at random from the bag without replacement. Let X_1 and X_2 be indicator random variables for the first and second ball being red. What is $\text{cov}(X_1, X_2)$? Recall that $\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.

$$\text{cov}(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2]$$

$$E[X_1] = 0 \cdot P(\text{1st ball blue}) + 1 \cdot P(\text{1st ball red})$$

$$= \frac{1}{2}$$

$$E[X_2] = 0 \cdot P(\text{2nd ball blue}) + 1 \cdot P(\text{2nd ball red})$$

$$= \frac{1}{2}$$

$$E[X_1 X_2] = 1 \cdot P(\text{both balls red}) + 0 \cdot P(\text{at } \cancel{\text{least}} \overset{0}{} \text{ 1 ball blue})$$

$$= \frac{5}{10} \cdot \frac{4}{9}$$

$$= \frac{2}{9}$$

$$\Rightarrow \text{cov}(X_1, X_2) = \frac{2}{9} - \frac{1}{2}\left(\frac{1}{2}\right) = -\frac{1}{36}$$