Quick Review

We can measure how "unpredictable" an RV is using variance:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

- "How far away are we from the mean on average?"
- By Linearity + algebra, we can write

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

(Much easier for calculations)

- Standard Deviation: $\sigma_X = \sqrt{\mathrm{Var}(X)}$

We can also measure how two RV's are related via covariance:

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(XY - \mathbb{E}[XY])^2\right]$$

- Sign matters, but magnitude is trickier.
- Again by Linearity + algebra,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Can augment definition to account for magnitude to get correlation:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

<u>Identities to know:</u>

Let X, Y be RVs and let $c \in \mathbb{R}$.

$$\text{Var}(cX) = c^2 \,\text{Var}(X)$$

If X, Y are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

$$\text{Var}(X) = \sigma_X^{\,2} = \text{Cov}(X, X)$$

X, Y independent $\Rightarrow$ $\text{Cov}(X, Y) = 0$

Converse is <u>not</u> true!

$\Rightarrow E[XY] = E[X]\,E[Y]$

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Positive Correlation

Negative Correlation

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 $\mathbb{E}[X_i]$?

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

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

2 Variance

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)?

(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.)

(c) A group of three friends has n books they would all like to read. Each friend (independently of the other two) picks a random permutation of the books and reads them in that order, one book per week (for n consecutive weeks). Let X be the number of weeks in which all three friends are reading the same book. Compute $\mathrm{Var}(X)$.

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 $\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.