CS 70   Discrete Mathematics and Probability Theory
Summer 2023   Huang, Suzani, and Tausik   **DIS 5C**

1   Dice Variance

(a) Let X be a random variable representing the outcome of the roll of one fair 6-sided die. What is $\text{Var}(X)$?

(b) Let Z be a random variable representing the average of n rolls of a fair 6-sided die. What is $\text{Var}(Z)$?

Solution:

(a) Recall that $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$. We can compute each of the individual terms using the definition of expectation:

$$\mathbb{E}[X] = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}$$
$$\mathbb{E}\left[X^2\right] = \frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right)$$
$$= \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}$$

Now, we plug back into the variance expression:

$$\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$
$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

(b) Because each die roll is independent of the others, we can utilize the fact that for independent random variables X and Y, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$. Let X_i be a random variable

representing the outcome of the ith dice roll. We now have:

$$\text{Var}(Z) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)$$

$$= \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^{n}X_i\right)$$

$$= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n}\text{Var}(X_i) \qquad\qquad (X_i\text{'s are independent})$$

$$= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n}\frac{35}{12} \qquad\qquad\qquad (\text{from (a)})$$

$$= \left(\frac{1}{n}\right)^2 \cdot n \cdot \frac{35}{12} = \frac{35}{12n}$$

2 Elevator Variance

A building has n upper floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each person gets off at one of the n upper floors uniformly at random and independently of everyone else. What is the *variance* of the number of floors the elevator *does not* stop at?

Solution: Let N be the number of floors the elevator does not stop at. We can represent N as the sum of the indicator variables $I_1, \ldots, I_n$, where $I_i = 1$ if no one gets off on floor i. Thus, we have

$$\mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

and from linearity of expectation,

$$\mathbb{E}[N] = \sum_{i=1}^{n}\mathbb{E}[I_i] = n\left(\frac{n-1}{n}\right)^m.$$

To find the variance, we cannot simply sum the variance of our indicator variables. However, since $\text{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2$ the only piece we don't already know is $\mathbb{E}[N^2]$. We can calculate this by again expanding N as a sum:

$$\mathbb{E}[N^2] = \mathbb{E}[(I_1 + \cdots + I_n)^2] = \mathbb{E}\left[\sum_{i,j}I_iI_j\right] = \sum_{i,j}\mathbb{E}[I_iI_j] = \sum_{i}\mathbb{E}[I_i^2] + \sum_{i\neq j}\mathbb{E}[I_iI_j].$$

The first term is simple to calculate: since I_i is an indicator, $I_i^2 = I_i$, so we have

$$\mathbb{E}[I_i^2] == \mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

meaning that

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

From the definition of the variables I_i, we see that $I_i I_j = 1$ when both I_i and I_j are 1, which means no one gets off the elevator on floor i and floor j. This happens with probability

$$\mathbb{P}[I_i = I_j = 1] = \mathbb{P}[I_i = 1 \cap I_j = 1] = \left(\frac{n-2}{n}\right)^m.$$

Thus we now know

$$\sum_{i \neq j} \mathbb{E}[I_i I_j] = n(n-1)\left(\frac{n-2}{n}\right)^m,$$

and we can assemble everything we've done so far to see that

$$\mathrm{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - n^2\left(\frac{n-1}{n}\right)^{2m}.$$

3 Covariance

Note 16

(a) 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 events of the first and second ball being red, respectively. What is $\mathrm{cov}(X_1, X_2)$? Recall that $\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

(b) Now, we have two bags A and B, with 5 red and 5 blue balls each. Draw a ball uniformly at random from A, record its color, and then place it in B. Then draw a ball uniformly at random from B and record its color. Let X_1 and X_2 be indicator random variables for the events of the first and second draws being red, respectively. What is $\mathrm{cov}(X_1, X_2)$?

Solution:

(a) We can use the formula $\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$.

$$\mathbb{E}[X_1] = \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2},$$
$$\mathbb{E}[X_2] = \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2},$$
$$\mathbb{E}[X_1 X_2] = \frac{5}{10} \cdot \frac{4}{9} \times 1 + \left(1 - \frac{5}{10} \cdot \frac{4}{9}\right) \times 0 = \frac{2}{9}.$$

Therefore,

$$\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{2}{9} - \frac{1}{2} \times \frac{1}{2} = -\frac{1}{36}.$$

(b) Again, we use the formula $\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$.

$$\mathbb{E}[X_1] = \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2}$$
$$\mathbb{E}[X_2] = \left(\frac{5}{10} \times \frac{6}{11} + \frac{5}{10} \times \frac{5}{11}\right) \times 1 + \left(\frac{5}{10} \times \frac{5}{11} + \frac{5}{10} \times \frac{6}{11}\right) \times 0 = \frac{1}{2}$$
$$\mathbb{E}[X_1 X_2] = \frac{5}{10} \times \frac{6}{11} \times 1 = \frac{30}{110}.$$

Therefore,

$$\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{30}{110} - \frac{1}{4} = \frac{1}{44}.$$

Note that in part (a), if one event happened, the other would be less likely to happen, and thus the covariance was negative. Similarly, in part (b), if one event happened, the other would be more likely to happen, and thus the covariance was positive.