1 Variance

This problem will give you practice using the "standard method" to compute the variance of a sum of random variables that are not pairwise independent (so you cannot use "linearity" of variance).

(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 person gets off at one of the n floors uniformly at random (independently of everybody else). 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, but the former is a little easier to compute.)

(b) 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 $\text{var}(X)$.

Solution:

(a) Let X be the number of floors the elevator does not stop at. We can represent X as the sum of the indicator variables $X_1,\ldots,X_n$, where $X_i = 1$ if no one gets off on floor i. Thus, we have

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

and from linearity of expectation,

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\left(\frac{n-1}{n}\right)^m.$$

To find the variance, we cannot simply sum the variance of our indicator variables. However, we can still compute $\text{var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2$ directly using linearity of expectation, but now how can we find $\mathbb{E}\left(X^2\right)$? Recall that

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \mathbb{E}\left((X_1 + \ldots + X_n)^2\right) \\
&= \mathbb{E}\left(\sum_{i,j} X_i X_j\right) \\
&= \sum_{i,j} \mathbb{E}\left(X_i X_j\right) \\
&= \sum_{i}^{n} \mathbb{E}\left(X_i^2\right) + \sum_{i \neq j} \mathbb{E}\left(X_i X_j\right).
\end{aligned}$$

The first term is simple to calculate:

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

meaning that

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

$X_iX_j = 1$ when both X_i and X_j are 1, which means no one gets off the elevator on floor i and floor j. This happens with probability

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

Thus, we can now compute

$$\sum_{i \neq j} \mathbb{E}\left(X_iX_j\right) = n(n-1)\left(\frac{n-2}{n}\right)^m.$$

Finally, we plug in to see that

$$\operatorname{var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - \left(n\left(\frac{n-1}{n}\right)^m\right)^2.$$

(b) Let $X_1, \ldots, X_n$ be indicator variables such that $X_i = 1$ if all three friends are reading the same book on week i. Thus, we have

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

and from linearity of expectation,

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}.$$

As before, we know that

$$\mathbb{E}\left(X^2\right) = \sum_{i}^{n} \mathbb{E}\left(X_i^2\right) + \sum_{i \neq j} \mathbb{E}\left(X_iX_j\right).$$

Furthermore, because X_i is an indicator variable, $\mathbb{E}\left(X_i^2\right) = 1^2\mathbb{P}[X_i = 1] = 1/n^2$, and

$$\sum_{i}^{n} \mathbb{E}\left(X_i^2\right) = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}.$$

Again, because X_i and X_j are indicator variables, we are interested in

$$\mathbb{P}[X_i = X_j = 1] = \mathbb{P}[X_i = 1 \cap X_j = 1] = \frac{1}{(n(n-1))^2},$$

the probability that all three friends pick the same book on week i and week j. Thus,

$$\sum_{i \neq j} \mathbb{E}(X_i X_j) = n(n-1)\left(\frac{1}{(n(n-1))^2}\right) = \frac{1}{n(n-1)}.$$

Finally, we compute

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{n} + \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2.$$

2 Correlation and Independence

(a) What does it mean for two random variables to be uncorrelated?

(b) What does it mean for two random variables to be independent?

(c) Are all uncorrelated variables independent? Are all independent variables uncorrelated? If your answer is yes, justify your answer; if your answer is no, give a counterexample.

Solution:

(a) Recall that for two random variables X and Y,

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

Two random variables are uncorrelated iff their covariance is equal to zero. If X and Y are uncorrelated, then there is no linear relationship between them.

(b) Recall that two random variables X and Y are independent if and only if the following criteria are met (the three criteria are equivalent and connected by Bayes rule):

$$\mathbb{P}(X = x \mid Y = y) = \mathbb{P}(X = x)$$
$$\mathbb{P}(Y = y \mid X = x) = \mathbb{P}(Y = y)$$
$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

for all x, y.

If X and Y are independent, any information about one variable offers no information whatsoever about the other variable.

(c) Note that if two random variables are independent, they must have no relationship whatsoever, including linear relationships; therefore they must be uncorrelated. The converse, however, is not true: two uncorrelated variables may not be independent. Consider two variables X and Y that follow a uniform joint distribution over the points $(1,0),(0,1),(-1,0),(0,-1)$. Then

$$\text{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - (0)(0) = 0$$

so there is no linear relationship, but X and Y are not independent (for example, $\mathbb{P}(Y = 0) = 1/2$ but $\mathbb{P}(Y = 0 \mid X = 1) = 1$).

3 Covariance

We have a bag of 5 red and 5 blue balls. We take two balls 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 $\mathrm{cov}(X_1, X_2)$?

Solution:

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,

$$\mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)(X_2) = \frac{2}{9} - \frac{1}{2} \times \frac{1}{2} = -\frac{1}{36}.$$