CS 70 Fall 2018 — Discrete Mathematics and Probability Theory — Alistair Sinclair and Yun Song — DIS 9B

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. Recall that $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

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.)

Solution:

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 $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ directly using linearity of expectation, but now how can we find $\mathbb{E}[X^2]$? Recall that

$$\mathbb{E}[X^2] = \mathbb{E}[(X_1 + \cdots + X_n)^2] = \mathbb{E}\left[\sum_{i,j} X_i X_j\right] = \sum_{i,j} \mathbb{E}[X_i X_j] = \sum_{i} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j].$$

The first term is simple to calculate:

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

meaning that

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

$X_i X_j = 1$ 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

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

Finally, we plug in to see that

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

2 Probabilistically Buying Probability Books

Chuck will go shopping for probability books for K hours. Here, K is a random variable and is equally likely to be 1, 2, or 3. The number of books N that he buys is random and depends on how long he shops. We are told that

$$\mathbb{P}[N = n | K = k] = \begin{cases} \frac{c}{k} & \text{for } n = 1, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

for some constant c.

(a) Compute c.

(b) Find the joint distribution of K and N.

(c) Find the marginal distribution of N.

Solution:

(a) For any k, we know that probabilities conditioned on $K = k$ must sum to 1, i.e

$$\sum_n \mathbb{P}[N = n | K = k] = 1 \quad ,$$

so it must be that

$$1 = \sum_{n=1}^{k} \mathbb{P}[N = n | K = k] = k \times \frac{c}{k} = c \quad .$$

Thus, $c = 1$.

(b) The joint distribution specifies $\mathbb{P}[N = n \cap K = k]$ for all n and k. Note that

$$\mathbb{P}[N = n \cap K = k] = \mathbb{P}[N = n | K = k]\mathbb{P}[K = k]$$

and we know $\mathbb{P}[N = n | K = k]$ and $\mathbb{P}[K = k]$ (it says all $k \in \{1, 2, 3\}$ are equally likely). We use this formula to calculate $\mathbb{P}[N = n \cap K = k]$ for each n, k and list the result in a table:

$n \setminus k$	1	2	3
1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{9}$
2	0	$\frac{1}{6}$	$\frac{1}{9}$
3	0	0	$\frac{1}{9}$

Alternatively, we can define the joint distribution as a formula with specified domain. $\mathbb{P}[N = n, K = k] = \mathbb{P}[N = n \mid K = k]\mathbb{P}[K = k] = \frac{1}{k}\frac{1}{3}$ whenever it is nonzero. So,

$$\mathbb{P}[N = n, K = k] = \begin{cases} \frac{1}{3k} & k \in \{1,2,3\}, n \in \{1,\dots,k\} \\ 0 & \text{otherwise} \end{cases}$$

(c) The marginal distribution of N is given by the value of $\mathbb{P}[N = n]$, for each possible value of n. By the total probability rule,

$$\mathbb{P}[N = n] = \mathbb{P}[N = n \cap K = 1] + \mathbb{P}[N = n \cap K = 2] + \mathbb{P}[N = n \cap K = 3] \; .$$

Thus, we get

$$\mathbb{P}[N = n] = \begin{cases} \frac{1}{3} + \frac{1}{6} + \frac{1}{9} & \text{if } n = 1 \\ \frac{1}{6} + \frac{1}{9} & \text{if } n = 2 \\ \frac{1}{9} & \text{if } n = 3 \end{cases} = \begin{cases} \frac{11}{18} & \text{if } n = 1 \\ \frac{5}{18} & \text{if } n = 2 \\ \frac{2}{18} & \text{if } n = 3 \end{cases}$$

3 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 such that $\mathbb{P}(X = x), \mathbb{P}(Y = y) > 0$.

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 $(1,0), (0,1), (-1,0), (0,-1)$. See Figure 1. Then

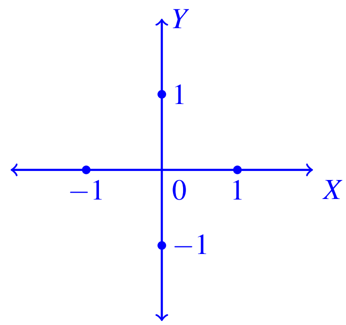

Figure 1: Choose one of the four points shown uniformly at random.

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

To see why, observe that $XY = 0$ always because at least one of X and Y is always 0, and furthermore $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ because both X and Y are symmetric around 0. So, there is no linear relationship, but X and Y are not independent (for example, $\mathbb{P}(Y = 0) = 1/2$ but $\mathbb{P}(Y = 0 \mid X = 1) = 1$).