CS 70 Spring 2018 — Discrete Mathematics and Probability Theory — Babak Ayazifar and Satish Rao — DIS 10A

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

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

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

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 Family Planning

Mr. and Mrs. Brown decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Browns have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

(b) Compute the joint distribution of G and C. Fill in the table below.

	$C=1$	$C=2$	$C=3$
$G=0$			
$G=1$			

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

$\mathbb{P}(G=0)$	
$\mathbb{P}(G=1)$	

$\mathbb{P}(C=1)$	$\mathbb{P}(C=2)$	$\mathbb{P}(C=3)$

(d) Are G and C independent?

(e) What is the expected number of girls the Browns will have? What is the expected number of children that the Browns will have?

Solution:

(a) The sample space is the set of all possible sequences of children that the Browns can have: $\Omega = \{g, bg, bbg, bbb\}$. The probabilities of these sample points are:

$$\mathbb{P}(g) = \frac{1}{2}$$
$$\mathbb{P}(bg) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$\mathbb{P}(bbg) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$
$$\mathbb{P}(bbb) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

(b)

	$C=1$	$C=2$	$C=3$
$G=0$	0	0	$\mathbb{P}(bbb)=1/8$
$G=1$	$\mathbb{P}(g)=1/2$	$\mathbb{P}(bg)=1/4$	$\mathbb{P}(bbg)=1/8$

(c) Marginal distribution for G:

$$\mathbb{P}(G=0) = 0+0+\frac{1}{8} = \frac{1}{8}$$
$$\mathbb{P}(G=1) = \frac{1}{2}+\frac{1}{4}+\frac{1}{8} = \frac{7}{8}$$

Marginal distribution for C:

$$\mathbb{P}(C=1) = 0+\frac{1}{2} = \frac{1}{2}$$
$$\mathbb{P}(C=2) = 0+\frac{1}{4} = \frac{1}{4}$$
$$\mathbb{P}(C=3) = \frac{1}{8}+\frac{1}{8} = \frac{1}{4}$$

(d) No, G and C are not independent. If two random variables are independent, then

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

To show this dependence, consider an entry in the joint distribution table, such as $\mathbb{P}(G=0,C=3)=1/8$. This is not equal to $\mathbb{P}(G=0)\mathbb{P}(C=3)=(1/8)\cdot(1/4)=1/32$, so the random variables are not independent.

(e) We can apply the definition of expectation directly for this problem, since we've computed the marginal distribution for both random variables.

$$\mathbb{E}(G)=0\cdot\mathbb{P}(G=0)+1\cdot\mathbb{P}(G=1)=1\cdot\frac{7}{8}=\frac{7}{8}$$
$$\mathbb{E}(C)=1\cdot\mathbb{P}(C=1)+2\cdot\mathbb{P}(C=2)+3\cdot\mathbb{P}(C=3)=1\cdot\frac{1}{2}+2\cdot\frac{1}{4}+3\cdot\frac{1}{4}=\frac{7}{4}$$

3 Coupon Collection

Suppose you take a deck of n cards and repeatedly perform the following step: take the current top card and put it back in the deck at a uniformly random position (the probability that the card is placed in any of the n possible positions in the deck — including back on top — is $1/n$).

Consider the card that starts off on the bottom of the deck. What is the expected number of steps until this card rises to the top of the deck? (For large n, you may use the approximation $\sum_{i=1}^{n}\frac{1}{i}\approx\ln n$)

[*Hint*: Let T be the number of steps until the card rises to the top. We have $T=T_n+T_{n-1}+\cdots+T_2$, where the random variable T_i is the number of steps until the bottom card rises from position i to position $i-1$. Thus, for example, T_n is the number of steps until the bottom card rises off the bottom of the deck, and T_2 is the number of steps until the bottom card rises from second position to top position. What is the distribution of T_i?]

Solution:

Since a card at location i moves to location $i-1$ when the current top card is placed in any of the locations $i, i+1, \ldots, n$, it will rise with probability $p = (n-i+1)/n$. Thus, $T_i \sim \text{Geometric}(p)$, and $\mathbb{E}[T_i] = 1/p = n/(n-i+1)$. We now can see how this is exactly the coupon collector's problem, but with one fewer term (namely, without T_1). Finally, we can apply linearity of expectation to compute

$$\mathbb{E}[T] = \sum_{i=2}^{n} \mathbb{E}[T_i] = \sum_{i=2}^{n} \frac{n}{n-i+1} = n \sum_{i=2}^{n} \frac{1}{n-i+1} = n \sum_{i=1}^{n-1} \frac{1}{i} \approx n \ln(n-1).$$

4 Exploring the Geometric Distribution

In this question, we will further investigate the geometric distribution. Let X, Y be i.i.d. geometric random variables with parameter p. Let $U = \min\{X,Y\}$ and $V = \max\{X,Y\} - \min\{X,Y\}$. Compute the joint distribution of (U,V) and prove that U and V are independent. [*Hint*: If $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(q)$ are independent, then $\min\{X,Y\} \sim \text{Geometric}(p+q-pq)$.]

Solution:

One has, for $u,v \in \mathbb{N}$, $u,v \geq 1$:

$$\begin{aligned}
\mathbb{P}(U = u, V = v) &= \mathbb{P}(\min\{X,Y\} = u, \max\{X,Y\} = u+v) \\
&= \mathbb{P}(X = u, Y = u+v) + \mathbb{P}(X = u+v, Y = u) \\
&= \mathbb{P}(X = u)\mathbb{P}(Y = u+v) + \mathbb{P}(X = u+v)\mathbb{P}(Y = u) \\
&= p(1-p)^{u-1}p(1-p)^{u+v-1} + p(1-p)^{u+v-1}p(1-p)^{u-1} = 2p^2(1-p)^{2u+v-2}.
\end{aligned}$$

Also, for $u \in \mathbb{N}$, $u \geq 1$:

$$\begin{aligned}
\mathbb{P}(U = u, V = 0) &= \mathbb{P}(X = Y = u) = \mathbb{P}(X = u)\mathbb{P}(Y = u) = p(1-p)^{u-1}p(1-p)^{u-1} \\
&= p^2(1-p)^{2u-2}.
\end{aligned}$$

Putting it together, we have:

$$\mathbb{P}(U = u, V = v) = \begin{cases} 2p^2(1-p)^{2u+v-2} & u,v \in \mathbb{N}, u \geq 1, v \geq 1 \\ p^2(1-p)^{2u-2} & u \in \mathbb{N}, u \geq 1, v = 0 \\ 0 & \text{otherwise} \end{cases}$$

Now, to show that U and V are independent, we must compute their marginal distributions. Note that $U = \min\{X,Y\} \sim \text{Geometric}(2p-p^2)$, so

$$\mathbb{P}(U = u) = p(2-p)(1-p)^{2u-2}, \qquad u \in \mathbb{N},\ u \geq 1.$$

(We are using the fact that the minimum of two independent geometric random variables is also geometric.) On the other hand, how do we compute the distribution of V? If $v \in \mathbb{N}$, $v \geq 1$:

$$\begin{aligned}
\mathbb{P}(V = v) &= \sum_{k=1}^{\infty} (\mathbb{P}(X = k, Y = k+v) + \mathbb{P}(X = k+v, Y = k)) \\
&= \sum_{k=1}^{\infty} (\mathbb{P}(X = k)\mathbb{P}(Y = k+v) + \mathbb{P}(X = k+v)\mathbb{P}(Y = k)) \\
&= \sum_{k=1}^{\infty} (p(1-p)^{k-1} p(1-p)^{k+v-1} + p(1-p)^{k+v-1} p(1-p)^{k-1}) \\
&= 2p^2(1-p)^{v-2} \sum_{k=1}^{\infty} ((1-p)^2)^k = 2p^2(1-p)^{v-2} \cdot \frac{(1-p)^2}{1-(1-p)^2} = \frac{2p(1-p)^v}{2-p}.
\end{aligned}$$

Otherwise, if $v = 0$:

$$\begin{aligned}
\mathbb{P}(V = 0) &= \sum_{k=1}^{\infty} \mathbb{P}(X = Y = k) = \sum_{k=1}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y = k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} p(1-p)^{k-1} \\
&= p^2(1-p)^{-2} \sum_{k=1}^{\infty} ((1-p)^2)^k = p^2(1-p)^{-2} \cdot \frac{(1-p)^2}{1-(1-p)^2} = \frac{p}{2-p}.
\end{aligned}$$

It is easily verified that

$$\mathbb{P}(U = u, V = v) = \mathbb{P}(U = u)\mathbb{P}(V = v) \qquad \forall u, v \in \mathbb{R},$$

so U and V are independent.