CS 70 Discrete Mathematics and Probability Theory
Fall 2017 Kannan Ramchandran and Satish Rao

DIS 10B

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

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

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

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

(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

$$\mathbb{E}[X^2] = \sum_{i}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j].$$

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

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

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

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

2 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 Binomial Conditioning

Let $n \in \mathbb{Z}_+$ and $p, q \in [0,1]$. Let $X \sim \text{Binomial}(n,p)$ and suppose that conditioned on $X = x$, $Y \sim \text{Binomial}(x,q)$. What is the unconditional distribution of Y?

Solution:

Y takes on values in $\{0,\dots,n\}$. So, let $y \in \{0,\dots,n\}$.

$$\begin{aligned}
\mathbb{P}(Y=y) &= \sum_{x=0}^{n} \mathbb{P}(X=x, Y=y) = \sum_{x=0}^{n} \mathbb{P}(X=x)\mathbb{P}(Y=y \mid X=x) \\
&= \sum_{x=y}^{n} \mathbb{P}(X=x)\mathbb{P}(Y=y \mid X=x) = \sum_{x=y}^{n} \binom{n}{x} p^x (1-p)^{n-x} \binom{x}{y} q^y (1-q)^{x-y} \\
&= \sum_{x=y}^{n} \frac{n!}{x!(n-x)!} \frac{x!}{y!(x-y)!} p^x (1-p)^{n-x} q^y (1-q)^{x-y} \\
&= \frac{n!}{y!(n-y)!} (pq)^y \sum_{x=y}^{n} \frac{(n-y)!}{(n-x)!(x-y)!} \big(p(1-q)\big)^{x-y} (1-p)^{n-x} \\
&= \binom{n}{y} (pq)^y \sum_{x=0}^{n-y} \frac{(n-y)!}{(n-y-x)!x!} \big(p(1-q)\big)^{x} (1-p)^{n-y-x} \\
&= \binom{n}{y} (pq)^y \sum_{x=0}^{n-y} \binom{n-y}{x} \big(p(1-q)\big)^{x} (1-p)^{n-y-x} \\
&= \binom{n}{y} (pq)^y \big(p(1-q) + (1-p)\big)^{n-y} = \binom{n}{y} (pq)^y (1-pq)^{n-y}.
\end{aligned}$$

We see that $Y \sim \text{Binomial}(n, pq)$.