CS 70 Spring 2019 | Discrete Mathematics and Probability Theory Babak Ayazifar and Satish Rao | HW 10

1 Binomial Variance

Throw n balls into m bins uniformly at random. For a specific ball i, what is the variance of the number of roommates it has (i.e. the number of other balls that it shares its bin with)?

Solution:

When we concentrate on the bin that ball i is in, we care about how many of the other $n-1$ balls land in that same bin. Notice that when determining whether the balls land in this bin, each ball is independent of every other ball, which makes variance much easier to compute as we can take advantage of linearity of variance. This is a binomial distribution, with $n-1$ trials and probability $1/m$ of success for each trial, where success is defined as having the ball land in the same bin as ball i.

Therefore, the variance is

$$(n-1)\left(\frac{1}{m}\right)\left(1-\frac{1}{m}\right).$$

2 Working with Distributions

1. For each of the following scenarios, find the distribution of each of the following random variables. That means, find the possible values that it can take on and their associated probabilities.
   (a) Five fair coins are flipped and the random variable Y is defined as the number of tails observed.
   (b) Two dice are rolled and the random variable Z is defined as the product of the two numbers rolled.
2. Suppose a fair six sided dice is rolled until a number smaller than 3 is observed. Let N be the total number of times the dice is rolled. Find $\mathbb{P}(N=k)$ for $k=1,2,3,\ldots$.
3. Now suppose two six-sided dice are rolled and the two numbers observed are defined as X and Y.
   (a) Calculate $\mathbb{P}(X>3 \mid Y=1)$.
   (b) Let $Z=X+Y$. What is the range of Z?

(c) Calculate $\mathbb{P}(X = 1 \mid Z < 4)$.

Solution:

1. (a) The sample space of Y includes the values: $0, 1, 2, 3, 4, 5$.

$$\mathbb{P}(Y = 0) = \frac{1}{32}$$

$$\mathbb{P}(Y = 1) = \frac{5}{32}$$

$$\mathbb{P}(Y = 2) = \frac{10}{32}$$

$$\mathbb{P}(Y = 3) = \frac{10}{32}$$

$$\mathbb{P}(Y = 4) = \frac{5}{32}$$

$$\mathbb{P}(Y = 5) = \frac{1}{32}$$

(b) All possible values of Z and their probabilities are:

$$\mathbb{P}(Z = 1) = \frac{1}{36}$$

$$\mathbb{P}(Z = 2) = \frac{2}{36}$$

$$\mathbb{P}(Z = 3) = \frac{2}{36}$$

$$\mathbb{P}(Z = 4) = \frac{3}{36}$$

$$\mathbb{P}(Z = 5) = \frac{2}{36}$$

$$\mathbb{P}(Z = 6) = \frac{4}{36}$$

$$\mathbb{P}(Z = 8) = \frac{2}{36}$$

$$\mathbb{P}(Z = 9) = \frac{1}{36}$$

$$\mathbb{P}(Z = 10) = \frac{2}{36}$$

$$\mathbb{P}(Z = 12) = \frac{4}{36}$$

$$\mathbb{P}(Z = 15) = \frac{2}{36}$$

$$\mathbb{P}(Z=16)=\frac{1}{36}$$
$$\mathbb{P}(Z=18)=\frac{2}{36}$$
$$\mathbb{P}(Z=20)=\frac{2}{36}$$
$$\mathbb{P}(Z=24)=\frac{2}{36}$$
$$\mathbb{P}(Z=25)=\frac{1}{36}$$
$$\mathbb{P}(Z=30)=\frac{2}{36}$$
$$\mathbb{P}(Z=36)=\frac{1}{36}$$

2. In every trial, the probability that the value observed is less than 3 is $1/3$. Therefore, we can think of each trial as a Bernoulli experiment where the success probability, $p=1/3$, and we therefore get that N is a Geometric random variable.

   Thus the probabilities can be expressed as:
$$\mathbb{P}(N=k)=\frac{1}{3}\left(\frac{2}{3}\right)^{k-1}.$$

3. (a) Since X and Y are independent and therefore the condition $Y=1$ does not affect $\mathbb{P}(X>3)$. Therefore the solution is $1/2$.

   (b) Z ranges from 2 to 12.

   (c) Given $Z<4$, we therefore limit the possible values of X and Y. The only possible dice rolls would therefore be:
$$X=1, Y=1$$
$$X=2, Y=1$$
$$X=1, Y=2$$
   Based on this, we can clearly see that $\mathbb{P}(X=1 \mid Z<4)=2/3$.

3 Geometric and Poisson

Let X be geometric with parameter p, Y be Poisson with parameter λ, and $Z=\max(X,Y)$. Assume X and Y are independent. For each of the following parts, your final answers should not have summations.

(a) Compute $P(X>Y)$.

(b) Compute $P(Z \geq X)$.

(c) Compute $P(Z \leq Y)$.

Solution:

(a) Condition on Y so you can use the nice property of geometric random variables that $P(X > k) = (1-p)^k$:

$$\begin{aligned}
P(Y < X) &= \sum_{y=0}^{\infty} P(X > Y | Y = y) P(Y = y) \\
&= \sum_{y=0}^{\infty} (1-p)^y \frac{e^{-\lambda}\lambda^y}{y!} \\
&= e^{-\lambda p} e^{\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda(1-p))^y}{y!} \\
&= e^{-\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda(1-p)}(\lambda(1-p))^y}{y!} \\
&= e^{-\lambda p}
\end{aligned}$$

To simplify the last summation we observed that the sum could be interpreted as the sum of the probabilities for a Poisson($\lambda(1-p)$) random variable, which is equal to 1.

Alternative solution: Since we know that Y is the limit of $Y_n \sim \text{Bin}(n, q_n = \lambda/n)$, we can write $\mathbb{P}(X > Y) = \lim_{n\to\infty} \mathbb{P}(X > Y_n)$, where

$$\begin{aligned}
\mathbb{P}(X > Y_n) &= \sum_{k=0}^{n} \mathbb{P}(Y_n = k)\mathbb{P}(X > k \mid Y_n = k) = \sum_{k=0}^{n} \binom{n}{k} q_n^k (1-q_n)^{n-k} \cdot (1-p)^k \\
&= \sum_{k=0}^{n} \binom{n}{k} (q_n(1-p))^k (1-q_n)^{n-k} = (q_n(1-p) + (1-q_n))^n \\
&= \left(1 - \frac{\lambda p}{n}\right)^n
\end{aligned}$$

which as $n \to \infty$ becomes $e^{-\lambda p}$ as above.

(b) 1, the max of X, Y is always at least X.

(c) $P(Z \leq Y) = P(\max(X,Y) \leq Y) = P(X \leq Y) = 1 - P(X > Y) = 1 - e^{-\lambda p}$

4 Exploring the Geometric Distribution

(a) Let $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(q)$ are independent. Find the distribution of $\min\{X,Y\}$ and justify your answer.

(b) 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)

(c) Prove that U and V are independent.

Solution:

(a) X is the number of coins we flip until we see a heads from flipping a coin with bias p, and Y is the same as flipping a coin with bias q. Imagine we flip the bias p coin and the bias q coin at the same time. The min of the two random variables represents how many simultaneous flips occur before at least one head is seen.

The probability of not seeing a head at all on any given simultaenous flip is $(1-p)(1-q)$, so the probability that there will be a success on any particular trial is $p+q-pq$. Therefore, $\min\{X,Y\} \sim \text{Geometric}(p+q-pq)$.

We can also solve it algebraically. The probability that $\min\{X,Y\} = k$ for some positive integer k is the probability that the first $k-1$ coin flips for both X and Y were tails, then times the probability that we get heads on the k-th toss. Specifically,

$$((1-p)(1-q))^{k-1} \cdot (p+q-pq)$$

We recognize this as the formula for a geometric random variable with parameter $p+q-pq$.

(b) 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:

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

(c) 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.

5 Boutique Store

Consider a boutique store in a busy shopping mall. Every hour, a large number of people visit the mall, and each independently enters the boutique store with some small probability. The store owner decides to model X, the number of customers that enter her store during a particular hour, as a Poisson random variable with mean λ.

Suppose that whenever a customer enters the boutique store, they leave the shop without buying anything with probability p. Assume that customers act independently, i.e. you can assume that they each flip a biased coin to decide whether to buy anything at all. Let us denote the number of customers that buy something as Y and the number of them that do not buy anything as Z (so $X = Y + Z$).

(a) What is the probability that $Y = k$ for a given k? How about $\mathbb{P}[Z = k]$? *Hint*: You can use the identity

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

(b) State the name and parameters of the distribution of Y and Z.

(c) Prove that Y and Z are independent.

Solution:

(a) We consider all possible ways that the event $Y = k$ might happen: namely, $k + j$ people enter the store ($X = k + j$) and then exactly k of them choose to buy something. That is,

$$\begin{aligned}
\mathbb{P}[Y = k] &= \sum_{j=0}^{\infty} \mathbb{P}[X = k + j] \cdot \mathbb{P}[Y = k \mid X = k + j] \\
&= \sum_{j=0}^{\infty} \left(\frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \right) \cdot \left(\binom{k+j}{k} p^j (1-p)^k \right) \\
&= \sum_{j=0}^{\infty} \frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \cdot \frac{(k+j)!}{k!j!} p^j (1-p)^k = \frac{(\lambda(1-p))^k e^{-\lambda}}{k!} \cdot \sum_{j=0}^{\infty} \frac{(\lambda p)^j}{j!} \\
&= \frac{(\lambda(1-p))^k e^{-\lambda}}{k!} \cdot e^{\lambda p} = \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!}.
\end{aligned}$$

The case for Z is completely analogous:

$$\mathbb{P}[Z = k] = \frac{(\lambda p)^k e^{-\lambda p}}{k!}$$

(b) Y follows the Poisson distribution with parameter $\lambda(1-p)$ and Z follows the Poisson distribution with parameter λp.

(c) If Y and Z are independent, then $\mathbb{P}(Y = y, Z = z) = \mathbb{P}(Y = y) \cdot \mathbb{P}(Z = z)$:

$$\begin{aligned}
\mathbb{P}(Y = y, Z = z) &= \sum_{x=0}^{\infty} \mathbb{P}(X = x, Y = y, Z = z) = \sum_{x=0}^{\infty} \mathbb{P}(Y = y, Z = z \mid X = x) \mathbb{P}(X = x) \\
&= \mathbb{P}(Y = y, Z = z \mid X = y + z) \mathbb{P}(X = y + z) = \frac{\cancel{(y+z)!}}{y!z!} p^z (1-p)^y \frac{e^{-\lambda} \lambda^{y+z}}{\cancel{(y+z)!}} \\
&= \frac{e^{-\lambda(1-p)} (\lambda(1-p))^y}{y!} \cdot \frac{e^{-\lambda p} (\lambda p)^z}{z!} = \mathbb{P}(Y = y) \cdot \mathbb{P}(Z = z).
\end{aligned}$$

6 Student Life

In an attempt to avoid having to do laundry often, Marcus comes up with a system. Every night, he designates one of his shirts as his dirtiest shirt. In the morning, he randomly picks one of his shirts to wear. If he picked the dirtiest one, he puts it in a dirty pile at the end of the day (a shirt in the dirty pile is not used again until it is cleaned). When Marcus puts his last shirt into the dirty pile, he finally does his laundry, and again designates one of his shirts as his dirtiest shirt (laundry isn't perfect) before going to bed. This process then repeats.

(a) If Marcus has n shirts, what is the expected number of days that transpire between laundry events? Your answer should be a function of n involving no summations.

(b) Say he gets even lazier, and instead of organizing his shirts in his dresser every night, he throws his shirts randomly onto one of n different locations in his room (one shirt per location), designates one of his shirts as his dirtiest shirt, and one location as the dirtiest location. In the morning, if he happens to pick the dirtiest shirt, *and* the dirtiest shirt was in the dirtiest location, then he puts the shirt into the dirty pile at the end of the day and does not use that location anymore (it is too dirty now). What is the expected number of days that transpire between laundry events now? Again, your answer should be a function of n involving no summations.

Solution:

(a) The number of days that it takes for him to throw a shirt into the dirty pile can be represented as a geometric RV. For the first shirt, this is the geometric RV with $p = 1/n$. We can see this by noticing that every day the probability of getting the dirtiest shirt remains $1/n$.

We'll call X_i the number of days that goes until he throws the ith shirt into the dirty pile. Since on the ith shirt, there are $n-i+1$ shirts left, we can get the parameter of X_i is $1/(n-i+1)$. The number of days until he does his laundry is a sum of these variables. Therefore, we can get the following result:

$$\mathbb{E}[X] = \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (n-i+1) = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

(b) For this part we can use a similar approach but the probability for X_i becomes $1/(n-i+1)^2$. This is because the dirtiest shirt falls into the dirtiest spot with probability $1/(n-i+1)$ and we pick it after that with probability $1/(n-i+1)$, so the probability of picking the dirtiest shirt from the dirtiest spot for the ith shirt is $1/(n-i+1)^2$. Using the same approach, we get the following sum:

$$\mathbb{E}[X] = \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (n-i+1)^2 = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$