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

Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up. (*signature here*)

1 Fundamentals

True or false? For the following statements, provide either a proof or a simple counterexample. Let X,Y,Z be arbitrary random variables.

(a) If (X,Y) are independent and (Y,Z) are independent, then (X,Z) are independent.

(b) If (X,Y) are dependent and (Y,Z) are dependent, then (X,Z) are dependent.

(c) Assume X is discrete. If $\text{var}(X) = 0$, then X is a constant.

(d) $\mathbb{E}[X]^4 \leq \mathbb{E}[X^4]$

Solution:

(a) FALSE. Let X,Y be i.i.d Bernoulli$(1/2)$ random variables, and let $Z = X$. Then (X,Y) and (Y,Z) are independent by construction, but (X,Z) are not independent because $Z = X$.

(b) FALSE. Let X,Z be i.i.d Bernoulli$(1/2)$ random variables, and let $Y = XZ$. Then (X,Z) are independent by construction, but (X,Y) are not independent, since $\mathbb{P}[X = 0 \wedge Y = 1] = 0 \neq \mathbb{P}[X = 0]\mathbb{P}[Y = 1] = (1/2)(1/4)$. Also, (Y,Z) are not independent by the same logic.

(c) TRUE. Let $\mu = \mathbb{E}[X]$. By definition,

$$0 = \text{var}(X) = \mathbb{E}[(X-\mu)^2] = \sum_{\omega\in\Omega} \mathbb{P}[\omega](X(\omega)-\mu)^2$$

The RHS is the sum of non-negative numbers, so if the sum is 0, each term must be 0.

So $\mathbb{P}[\omega] > 0 \implies (X(\omega)-\mu)^2 = 0 \implies X(\omega) = \mu$. Therefore X is constant (equal to $\mu = \mathbb{E}[X]$).

(d) TRUE. First, for an arbitrary random variable Y, we have:

$$0 \le \mathbb{E}[(Y-\mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$$

So $\mathbb{E}[Y]^2 \le \mathbb{E}[Y^2]$. Now applying this twice, once for $Y = X$ and once for $Y = X^2$:

$$\mathbb{E}[X]^4 = (\mathbb{E}[X]^2)^2 \le (\mathbb{E}[X^2])^2 \le \mathbb{E}[(X^2)^2] = \mathbb{E}[X^4]$$

2 Balls and Bins

Throw n balls into m bins, where m and n are positive integers. Let X be the number of bins with exactly one ball. Compute $\operatorname{var} X$.

Solution:

Let X_i be the indicator that bin i has exactly one ball, for each $i = 1,\ldots,m$. Since $X = \sum_i X_i$, we can use the computational formula for variance:

$$\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{m} X_i\right)^2\right] - \left(\mathbb{E}[\sum_{i=1}^{m} X_i]\right)^2 \\
&= \mathbb{E}\left[\sum_{i\neq j} X_iX_j + \sum_{i=1}^{m} X_i^2\right] - \left(\sum_{i=1}^{m}\mathbb{E}[X_i]\right)^2 \\
&= \sum_{i\neq j}\mathbb{E}[X_iX_j] + \sum_{i=1}^{m}\mathbb{E}[X_i] - \left(\sum_{i=1}^{m}\mathbb{E}[X_i]\right)^2,
\end{aligned}$$

where the last line followed from linearity of expectation and recognizing that $X_i^2 = X_i$, since it can only take on the values 0 or 1.

One has

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

and for $j \in \{1,\ldots,n\}$, $j \neq i$,

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

Noting that $\sum_{i\neq j}$ has $m(m-1)$ terms, and the rest of the sums have m terms, we find

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

3 Proof with Indicators

Let $n \in \mathbb{Z}_+$. Let $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ and let $A_1, \dots, A_n$ be events. Prove that $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbb{P}(A_i \cap A_j) \geq 0$. Note that α_i can be less than 0.

Solution:

We write the summation with indicators. Let X_i be the indicator for event A_i, $i = 1, \dots, n$. Then,

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbb{P}(A_i \cap A_j) &= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbb{E}[X_i X_j] = \mathbb{E}\Big[\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j X_i X_j\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{i=1}^n \alpha_i X_i\Big)\Big(\sum_{j=1}^n \alpha_j X_j\Big)\Big] = \mathbb{E}\Big[\Big(\sum_{i=1}^n \alpha_i X_i\Big)^2\Big] \geq 0.
\end{aligned}$$

4 Boutique Store

(a) 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 simply 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 $X = Y + Z$). What is the probability that $Y = k$ for a given k? How about $\mathbb{P}[Z = k]$? Prove that Y and Z are Poisson random variables themselves.

Hint: You can use the identity

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

(b) 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,

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

Hence, Y follows the Poisson distribution with parameter $\lambda(1-p)$. The case for Z is completely analogous:

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

and Z follows the Poisson distribution with parameter λp.

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

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

5 Ordering Random Variables

Here we will investigate the properties of *ordered collections* of identically distributed (not identical!) random variables. The techniques in this problem can be applied to any kind of distribution, but here we will consider a specific case. Let $X_1, X_2, \ldots, X_n$ be iid geometric with parameter p.

(a) Find $\mathbb{P}[X_i > k]$ and $\mathbb{P}[X_i < k]$ for all i and k

(b) Let $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ be the random variables corresponding to the *ordered* collection from above. In other words, consider the set of random variables $S = \{X_i \mid 1 \leq i \leq n\}$, and let $X_{(j)}$ be the jth smallest element in that set.

What is $X_{(1)}$, as a function of $X_1 \ldots X_n$? (We're just looking to see that you understand the definition. No work necessary)

(c) Find $\mathbb{P}[X_{(1)} \leq k]$.

(d) Find $\mathbb{P}[X_{(i)} \le k]$.

Hint: There are no tricks to simplify it, like the last case. You will end up with a sum. First try to translate "$X_{(i)} \le k$" into a statement about all of $X_{(1)}$ through $X_{(n)}$ (what do we know about each if $X_{(i)} \le k$?). Then relate this statement to events involving $X_1, \ldots X_n$.

(e) Calculate $\mathbb{P}[X_{(1)} = k_1, X_{(2)} = k_2, \ldots, X_{(n)} = k_n]$ using symmetry arguments, assuming all the k_i are distinct. That is, assume $k_1 \ne k_2 \cdots \ne k_n$.

(f) The probabilities in the previous part are associated with the *joint distribution* of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. If we want to completely specify the joint distribution, we cannot limit ourselves to only cases where the k_i are distinct.

How would you modify your calculation to account for the possibility that not all the k_i are distinct? Either an explanation in words or an explicit calculation is fine.

Solution:

(a) Recall that the geometric distribution characterizes the length of a series of independent coin flips up to and including the first heads. So $\mathbb{P}[X_i > k]$ is the probability that the coin has at least k tails: $\boxed{\mathbb{P}[X_i > k] = (1-p)^k}$. We know $\mathbb{P}[X_i = k] = (1-p)^{k-1}p$, so

$$\boxed{\mathbb{P}[X_i < k] = 1 - (1-p)^k - (1-p)^{k-1}p}.$$

These probabilities hold regardless of i, because the X_i are identically distributed. The formula holds for all possible values of k, namely $\mathbb{Z}^+$ (we must flip the coin a positive number of times before getting a heads).

Notice two things: first, we found a simple interpretation for the event "$X_i > k$" and used complements to calculate the more difficult probability. Second, we used the fact that "$X_i = i$" and "$X_i = j$" are disjoint events for $i \ne j$ (this is true for any random variable!).

(b) $\min(X_1, X_2, \ldots, X_n)$

(c) $\mathbb{P}[X_{(1)} \le k] = 1 - \mathbb{P}[X_{(1)} > k] = 1 - \mathbb{P}[\min(X_1, \ldots, X_n) > k] = 1 - (\mathbb{P}[X_1 > k]\mathbb{P}[X_2 > k] \cdots \mathbb{P}[X_n > k])$, where the last equality follows from independence. We calculated each one of these terms in part (a), so we have $\mathbb{P}[X_{(1)} \le k] = 1 - \left((1-p)^k\right)^n$.

(d) In the general case, we aren't so lucky as to be able to solve for the complement easily. Instead we consider what it means for $X_{(i)}$ to be less than or equal to k. We would need all of $X_{(1)}, \ldots X_{(i)}$ to be less than or equal to k, and then $X_{(i+1)}, \ldots, X_{(n)}$ just need to be greater than or equal to $X_{(i)}$ – some of them could be less than k, some could be greater. So, we need to divide these events further. Notice that if $X_{(i+1)} \le k$, $X_{(i)}$ is necessarily $\le k$ as well. So, $X_{(i)} \le k$ is the event that *for some* $j \ge i$, $X_{(1)}, X_{(2)}, \ldots X_{(j)}$ are all $\le k$. We can use the Law of Total Probability to solve for that:

$$\mathbb{P}[X_{(i)} \le k] = \sum_{j=i}^{n} \mathbb{P}[X_{(1)} \le k, X_{(2)} \le k, \ldots, X_{(j)} \le k, X_{(j+1)} > k, \ldots X_{(n)} > k]$$

Notice that, by forcing the variables $j+1$ through n to be strictly larger than k, we made the events $X_{(1)},\ldots,X_{(j)} \leq k$ disjoint for different choices of j. That's why we can sum over all possible choices. It's not clear why this equation is useful yet – it's still in terms of the ordered variables $X_{(i)}$! To relate it back to the original geometric random variables, let's write this slightly differently:

$$\mathbb{P}[X_{(i)} \leq k] = \mathbb{P}[\text{exactly } i \text{ X's} \leq k] + \mathbb{P}[\text{exactly } i+1 \text{ X's} \leq k] + \cdots + \mathbb{P}[\text{exactly } n \text{ X's} \leq k]$$

Each of these events is well-defined in terms of the original variables X_i: there are exactly $\binom{n}{j}$ ways to choose exactly j of the original variables to be $\leq k$ and $n-j$ to be $> k$:

$$\begin{aligned}
\mathbb{P}[X_{(i)} \leq k] &= \sum_{j=i}^{n} \mathbb{P}[X_{(1)} \leq k, X_{(2)} \leq k, \ldots, X_{(j)} \leq k, X_{(j+1)} > k, \ldots X_{(n)} > k] \\
&= \sum_{j=i}^{n} \binom{n}{j} \mathbb{P}[X_1 \leq k, X_2 \leq k, \ldots, X_j \leq k, X_{j+1} > k, \ldots, X_n > k] \\
&= \sum_{j=i}^{n} \binom{n}{j} \mathbb{P}[X_1 \leq k]^j \mathbb{P}[X_1 > k]^{n-j} \\
&= \sum_{j=i}^{n} \binom{n}{j} (1-(1-p)^k)^j (1-p)^{k(n-j)}
\end{aligned}$$

(e) The trick here is to realize that the joint event $J(k_1,k_2,\ldots,k_n) = \{X_{(1)} = k_1, X_{(2)} = k_2, \ldots X_{(n)} = k_n\}$ corresponds to a specific set of events with the original random variables:

$$J(k_1,k_2,\ldots,k_n) = \{\text{some permutation of } X_1,X_2,\ldots,X_n \text{ is } (k_1,k_2\ldots,k_n)\}.$$

We know the joint distribution of the X_i's, so all that remains is to find *how many* permutations correspond to each event J. For each set of k_i's, there are n choices for which X_i is k_1, $n-1$ choices for which X_i is k_2, and so on. Therefore, there are $n!$ permutations of the X_i's that correspond to each $J(k_1,\ldots,k_n)$, all disjoint. So, we have in summary:

$$\begin{aligned}
\mathbb{P}[X_{(1)} = k_1,\ldots,X_{(n)} = k_n] &= n!\mathbb{P}[X_1 = k_1,\ldots,X_n = k_n] && k_1 < k_2 < \ldots < k_n \\
&= n! \prod_{i=1}^{n} (1-p)^{k_i-1} p && k_1 < k_2 < \ldots < k_n
\end{aligned}$$

(f) We perform exactly the same calculation (since the connection between $J(k_1,\ldots,k_n)$ and X_i remains). However, our multiplicative factor changes. Instead of $n!$, we have to account for the possibility that, for example $k_1 = k_2$, in which case we've overcounted by $2!$. So we divide out the number of ways to permute all the duplicate k_i's.

The above is sufficient, but if we wanted to explicitly calculate this, we could introduce the following notation. Let there be m distinct k_i's, and let d_i be the number of times each k_i occurs. Then, our multiplicative factor is precisely $\frac{n!}{\prod_{i=1}^{m}(d_i!)}$. So, we find in general

$$\mathbb{P}[X_{(1)} = k_1,\ldots,X_{(n)} = k_m] = \frac{n!}{\prod_{i=1}^{m}(d_i!)} \prod_{i=1}^{m} \left((1-p)^{k_i-1}p\right)^{d_i} \quad k_1 < k_2 < \ldots < k_m$$

6 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.

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

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