# HOMEWORK 3

PSTAT 120A: Summer 2022

**Due: 11:59pm on Friday, July 1**
Instructor: Ethan P. Marzban

***Instructions:***

- Please submit your work to Gradescope by no later than the due date posted above.
- Be sure to show your work; correct answers with no supporting work will not be awarded full points.
- 2 randomly selected questions/parts will be graded, but you must still turn in your work for all problems in order to be eligible to earn full credit.

1. Let $X$ be a random variable with probability mass function (p.m.f.) given by

$$p_X(k) = \begin{cases} \dfrac{1}{k(k+1)} & \text{if } k = 1, 2, 3, \cdots \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that $p_X(k)$ is a valid p.m.f. Be sure to show **all** of your work; don't use WolframAlpha! **Hint:** Partial Fraction Decomposition.

**Solution:** First we check the nonnegativity condition. For every $k \in \{1, 2, \cdots\}$ we have that both $k$ and $k+1$ are positive, meaning their product is positive and hence $1/[k(k+1)] > 0 \geq 0$. Therefore, $p_X(k) \geq 0$ for every $k \in \mathbb{R}$.

Now, we must also check that $\sum_k p_X(k) = 1$; in other words, we wish to evaluate

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$$

This is not a sum we immediately recognize; as such, let us first find the $n^{\text{th}}$ partial sum

$$S_n := \sum_{k=1}^{n} \frac{1}{k(k+1)}$$

This is where the hint comes into play: let us perform Partial Fraction Decomposition (PFD) on the summand. We write

$$\frac{1}{k(k+1)} = \frac{A}{k} + \frac{B}{k+1} = \frac{A(k+1) + Bk}{k(k+1)} = \frac{k(A+B) + A}{k(k+1)}$$

which leads to the system of equations

$$\begin{cases} A + B & = 0 \\ A & = 1 \end{cases} \implies A = 1,\ B = -1 \implies \frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$$

Therefore,

$$\begin{aligned} S_n &:= \sum_{k=1}^{n} \frac{1}{k(k+1)} \\ &= \sum_{k=1}^{n} \left( \frac{1}{k} - \frac{1}{k+1} \right) \\ &= \left(1 - \cancel{\frac{1}{2}}\right) + \left(\cancel{\frac{1}{2}} - \cancel{\frac{1}{3}}\right) + \left(\cancel{\frac{1}{3}} - \cancel{\frac{1}{4}}\right) + \cdots + \left(\cancel{\frac{1}{n}} - \frac{1}{n+1}\right) \\ &= 1 - \frac{1}{n} \end{aligned}$$

Therefore,

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} := \lim_{n\to\infty} \left( \sum_{k=1}^{n} \frac{1}{k(k+1)} \right) = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right) = 1 \checkmark$$

Hence, since $p_X(k) \geq 0$ for all $k \in \mathbb{R}$ and $\sum_k p_X(k) = 1$ we can conclude that $p_X(k)$ is indeed a valid p.m.f.

(b) Compute $\mathbb{E}[X]$.

**Solution:** From the definition of expectation,

$$\mathbb{E}[X] = \sum_k k p_X(k) = \sum_{k=1}^{\infty} k \cdot \frac{1}{k(k+1)} = \sum_{k=1}^{\infty} \frac{1}{k+1}$$

which diverges to $\infty$. Hence, $\mathbb{E}[X] = \infty$.

(c) Compute $\mathbb{E}[X^2]$. Use this to compute $\text{Var}(X)$.

**Solution:** From the LOTUS,

$$\mathbb{E}[X^2] = \sum_k k^2 p_X(k) = \sum_{k=1}^{\infty} k \cdot \frac{1}{k(k+1)} = \sum_{k=1}^{\infty} \frac{k}{k+1}$$

Recall that a necessary (but not sufficient) condition for the sum $\sum_k a_k$ to converge is that $a_k \to 0$ as $k \to \infty$. Here, the summand goes to 1 meaning $\sum_k k^2 p_X(k)$ diverges to $\infty$, and thus $\mathbb{E}(X^2) = \infty$. Now, technically it is correct to say that the variance is undefined as it is unclear exactly which of the terms ($\mathbb{E}[X^2]$ or $\mathbb{E}[X]$) dominate. But I would have also accepted $\infty$ as an answer.

(d) Find $F_X(x)$, the cumulative mass function (c.m.f.) of $X$. **Hint:** You can reuse some of the work you did in part (a).

**Solution:** Suppose that $x$ is an integer in the set $\{1, 2, 3, \cdots\}$; then

$$F_X(x) = \sum_{k=1}^{x} \frac{1}{k(k+1)} = 1 - \frac{1}{x+1}$$

where we have utilized our working from part (a) to evaluate the partial sum. If instead $x$ is a real number from the set $[1, \infty]$, we can see that

$$F_X(x) = \sum_{k=1}^{\lfloor x \rfloor} \frac{1}{k(k+1)} = 1 - \frac{1}{\lfloor x \rfloor + 1}$$

(if this is not clear, try examining a test case; say, $x = 1.5$. We know that $F_X(1.5) = p_X(1)$, meaning our sum should go up to the largest integer smaller than $x$) meaning we can write our c.m.f. as

$$F_X(x) = \begin{cases} 1 - \dfrac{1}{\lfloor x \rfloor + 1} & \text{if } x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Note that this is a step function, as is to be expected.

2. A recent survey has revealed that 30% of cars are Electric Vehicles (EV's), 30% are Hybrid, and the remaining 40% are gas. A surveyor goes from parking lot to parking lot, recording the status (i.e. EV, Hybrid, or Gas) cars one at a time. The surveyor, however, is a bit forgetful, and sometimes records the same car twice.

   (a) What is the probability that, among the first 10 cars surveyed, the surveyor observes exactly 3 EV's?

   **Solution:** Let $X$ denote the number of EV's observed in the first 10 cars; then $X \sim \mathrm{Bin}(10, 0.3)$ and
   $$\mathbb{P}(X = 3) = \binom{10}{3}(0.3)^3(0.7)^7$$

   (b) What is the probability that the $10^{\text{th}}$ car the surveyor examines is her first Hybrid vehicle?

   **Solution:** Let $Y$ denote the number cars needed to observe the $1^{\text{st}}$ Hyrbid; then $Y \sim \mathrm{Geom}(0.3)$ and so
   $$\mathbb{P}(Y = 10) = (0.7)^9(0.3)$$

   (c) What is the probability that the $24^{\text{th}}$ car the surveyor examines is her fourth Gas car?

   **Solution:** Let $Z$ denote the number of cars needed to observe the fourth Gas car; then $Z \sim \mathrm{NegBin}(4, 0.4)$ and
   $$\mathbb{P}(Z = 24) = \binom{24-1}{4-1}(0.4)^4(0.6)^{24-4} = \binom{23}{3}(0.4)^4(0.6)^{20}$$

   (d) On average, how many cars (including the final "successful" car) does the surveyor need to examine before observing her fourth Gas car?

   **Solution:** Letting $Z$ be as in part (c), we can see
   $$\mathbb{E}[Z] = \frac{4}{0.4} = 10$$

   (e) What is the variance of the number of Hybrid vehicles the surveyor observes among the first 30 cars she examines?

   **Solution:** Let $W$ denote the number of Hybrids observed among the first 30 cars; then $W \sim \mathrm{Bin}(30, 0.3)$ and
   $$\mathrm{Var}(W) = 30(0.3)(0.7)$$

3. Chen-7 is a highly contagious disease that is capable of killing a Time Lord in under a day. Thankfully, only 7% of the population of Gallifrey has been infected (so far...). Additionally, there exists a test for Chen-7 but it is imperfect: 80% of the time it correctly identifies a healthy Time Lord as disease-free, but 30% of the time it incorrectly identifies a diseased Time Lord as disease-free.

The Doctor is worried that they may have contracted Chen-7, so they takes 20 independent tests. Of these 20 tests, 10 indicate that they have the disease and the remaining 10 indicate that they do not. With this information, what is the probability that the Doctor actually has Chen-7?

**Solution:** As always, let's establish some notation first. Let $D$ denote "The Doctor has Chen-7," $-$ denote "a randomly taken test returns a negative result," and $+$ denote "a randomly taken test returns a positive result." Additionally, let $X$ denote the number of tests in 20 randomly taken tests that return a positive result. With this information we have

$$\mathbb{P}(- \mid D^{\complement}) = 0.8; \quad \mathbb{P}(- \mid D) = 0.3$$

which enablees us to compute

$$\mathbb{P}(+ \mid D^{\complement}) = 0.2; \quad \mathbb{P}(+ \mid D) = 0.7$$

The quantity we seek is $\mathbb{P}(D \mid X = 10)$; by Bayes' Rule, this is computable as

$$\mathbb{P}(D \mid X = 10) = \frac{\mathbb{P}(X = 10 \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(X = 10)}$$

Let's focus on the numerator for a bit. Conditioned on $D$, the random variable $X$ follows the Binomial distribution with parameters $n = 20$ and $p = \mathbb{P}(+ \mid D) = 0.7$, meaning

$$\mathbb{P}(X = 10 \mid D) = \binom{20}{10} (0.7)^{10} (0.3)^{10}$$

For the denominator of Bayes' Rule, we utilize the Law of Total Probability, also noting that $(X \mid D^{\complement}) \sim \text{Bin}(20, 0.2)$:

$$\begin{aligned}
\mathbb{P}(X = 10) &= \mathbb{P}(X = 10 \mid D) \cdot \mathbb{P}(D) + \mathbb{P}(X = 10 \mid D^{\complement}) \cdot \mathbb{P}(D^{\complement}) \\
&= \binom{20}{10} (0.7)^{10} (0.3)^{10} \cdot (0.07) + \binom{20}{10} (0.2)^{10} (0.8)^{10} \cdot (0.93)
\end{aligned}$$

Therefore, putting everything together,

$$\begin{aligned}
\mathbb{P}(D \mid X = 10) &= \frac{\mathbb{P}(X = 10 \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(X = 10)} \\
&= \frac{\binom{20}{10} (0.7)^{10} (0.3)^{10} \cdot (0.07)}{\binom{20}{10} (0.7)^{10} (0.3)^{10} \cdot (0.07) + \binom{20}{10} (0.2)^{10} (0.8)^{10} \cdot (0.93)} \\
&= \boxed{\frac{(0.7)^{10} (0.3)^{10} \cdot (0.07)}{(0.7)^{10} (0.3)^{10} \cdot (0.07) + (0.2)^{10} (0.8)^{10} \cdot (0.93)} \approx 53.31\%}
\end{aligned}$$

4. Let $X$ be a discrete random variable with support contained entirely within the positive real line (such a random variable is said to be "nonnegative"). Prove the so-called **tail-sum formula**, which states

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$

Here is a rough guide to help you through your proof:

- Write $\mathbb{P}(X \geq k)$ as a sum, thereby obtaining a double-sum on the RHS of the Tail-Sum formula.
- Change the order of summation (a dot diagram may be helpful)
- Show that the resulting summation, with the order of summation reversed, is in fact equal to $\mathbb{E}[X]$.

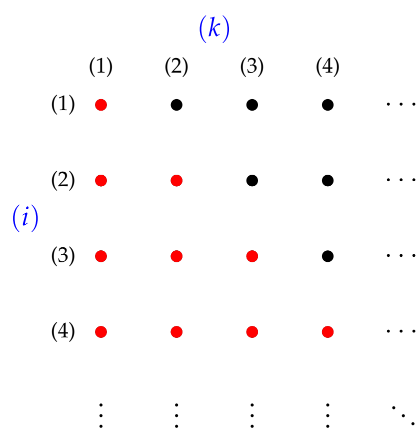
**Solution:** Let's follow the hints closely. We begin by writing

$$\mathbb{P}(X \geq k) = \sum_{i=k}^{\infty} p_X(i)$$

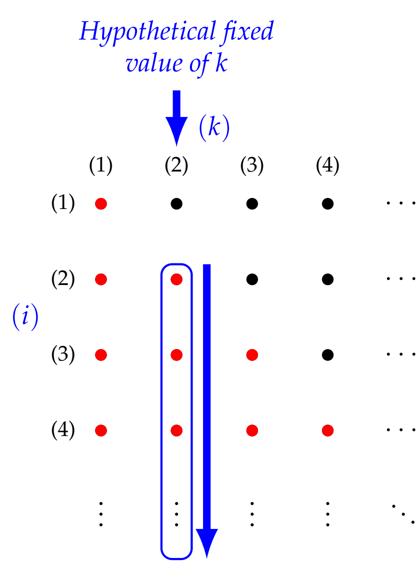
and so

$$\sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} p_X(i)$$

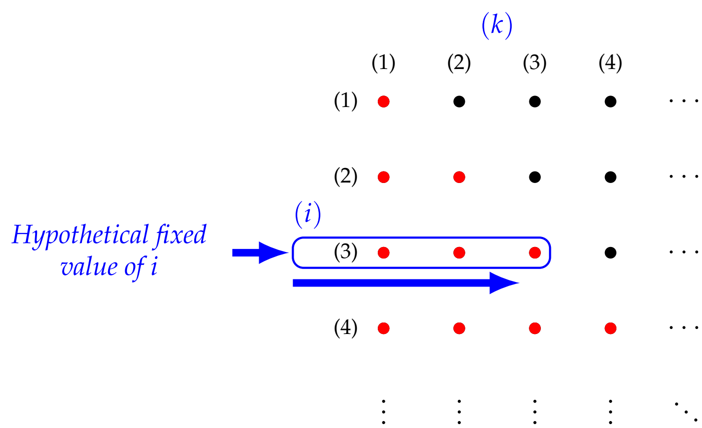
Note that a double sum is nothing but a sum over certain pairs of indices $(i, k)$. Let us utilize a dot diagram to visualize all pairs $(i, k)$ that are included in this double sum:




Here's the trick: in the double sum written above, we can imagine first fixing a value of $k$ between 1 and $\infty$, then letting $i$ vary. For a fixed value of $k$, $i$ varies from $k$ to $\infty$:



When we reverse the order of the summation, the pairs of indices included in the sum should still remain the same; that is, we still have the same red dots highlighted. Now, however, we consider fixing a value of $i$ (between 1 and $\infty$), and letting $k$ vary.




What we can see is that, for a fixed value of $i$, $k$ ranges from 1 to $i$. Therefore,

$$\sum_{k=1}^{\infty} \sum_{i=k}^{\infty} p_X(i) = \sum_{i=1}^{\infty} \sum_{k=1}^{i} p_X(i)$$

Now, note the summand on the RHS above does not depend on $k$, and is therefore treated as a constant by the inner sum (which ranges over $k$); that is to say,

$$\begin{aligned} \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} p_X(i) &= \sum_{i=1}^{\infty} \sum_{k=1}^{i} p_X(i) \\ &= \sum_{i=1}^{\infty} i p_X(i) \end{aligned}$$

which is precisely our formula for $\mathbb{E}[X]$ (note that we are told that $X$ has support on the *positive* real line, meaning we can assume $S_X = \{1, 2, \cdots\}$. If instead $X$ admitted 0 into its state space, we could have also re-indexed the sum to begin at 0.) ■