Probability: Discrete Distributions Problem Set

Cameron Stewart

June 2023

Hint: For the following questions, there are theorems from previous calculus and probability lectures regarding sums and series which may be very helpful.

1 Random Variables, PMFs, and CDFs

1. You are playing a board game in which you must roll two 4-sided dice and sum the outcomes together to give a score. What is the PMF and CDF for the distribution of scores? What is the probability of getting a score of 6 or more? Use this question to practice using correct notation.

2. You have decided to run an experiment to see how likely it is for two people in a group to share the same birthday. In your setup, you start with two randomly selected volunteers in an empty room. They tell you their birthday. If they do not share a birthday, you add another randomly selected volunteer to the room. If their birthday isn't shared by either of the other two volunteers, another volunteer is added. The process continues until two people share a birthday.

   Let the random variable X denote the number of people added until two people share a birthday. What is the support and PMF of X? Plot the PMF and the CDF (you can do this quickly in Wolfram Alpha). You may assume:

   - February 29th doesn't exist.
   - No day of the year is more likely to be a birthday than any other day.
   - Your supply of volunteers does not run out and the room never runs out of space.

   Remember, your PMF should sum to 1 over the support. Use this to help check that you haven't made a mistake.

3. You are a bus enthusiast, dismayed at the lack of regular scheduling for buses in your town. Over countless hours you have modelled the probability distribution of bus arrivals at you local bus stop on weekdays between 7am and 8am. For random variable X representing the number of arrivals, you have determined the PMF to be $f_X(x) = \frac{3^x e^{-3}}{x!}$ on support $\mathcal{X} = \mathbb{N}_0$.

   At 6:30am one morning you make a prediction that 3 buses will arrive in the hour. What is the probability that you are correct?

   At 6:45am, you receive information from the transport app on your phone that at least 2 buses will arrive during the hour. Given this new information, what is the probability now that your prediction is correct?

4. Determine the value of a for PMF $f(x) = -\frac{1}{10}(x+1)(x-a)$ with support $\{0, 1, 2\}$.

5. Determine the value of a for PMF $f(x) = a\frac{\lambda^x}{x!}$ with support $\mathbb{N}_0$. *Hint*: If $\frac{\mathrm{d}}{\mathrm{d}\lambda} g(\lambda) = g(\lambda)$, what can be said about g?

2 Expectation and Variance

1. Prove $\mathbb{E}[ag(X) + bh(X) + c] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)] + c$ for discrete random variable X with support $\mathcal{X}$ and PMF f_X, and some functions g and h.

2. Prove $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ for discrete random variable X with support $\mathcal{X}$ and PMF f_X.

3. Calculate the expectation, variance, and standard deviation of X with support $\mathcal{X} = \left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}\right\}$ and PMF $f_X(x) = \frac{1}{5}\left(\sin^2(x) + 1\right)$. Calculate these same statistics for $\sqrt{5}X + \pi$.

4. Prove the following:

   (a) If X has support $\mathcal{X} = \{0, 1\}$ and PMF $f_X(x) = p^x(1-p)^{1-x}$, then $\mathbb{E}[X] = p$.

   (b) If X has support $\mathcal{X} = \mathbb{N}_0$ and PMF $f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, then $\mathbb{E}[X] = \lambda$.

   (c) If X has support $\mathcal{X} = \{0, \ldots, n\}$ and PMF $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$, then $\mathbb{E}[X] = np$.

   (d) If X has support $\mathcal{X} = \mathbb{N}$ and PMF $f_X(x) = (1-p)^{x-1} p$, then $\mathbb{E}[X] = \frac{1}{p}$. *Hint*: Rewriting an expression as a derivative of some other expression might help.

3 Important Discrete Distributions

1. Prove that the binomial PMF sums to 1 over the support.

2. Prove that the geometric PMF sums to 1 over the support.

3. Prove that the limit of the binomial distribution is the Poisson distribution, as mentioned in the lecture. I.e. what is the PMF of $X \sim \text{Bin}\left(n, \frac{\lambda}{n}\right)$ when $n \to \infty$.

4. Prove that the sum of independent Poisson random variables is also a Poisson random variable. I.e. if $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$, what is the distribution of $Z = X + Y$?

5. Prove the memoryless property of the geometric distribution, showing that the distribution over the remaining trials until success is independent of the trials that have already happened. I.e. prove that $\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t)$ for $X \sim \text{Geo}(p)$.

6. Prove that the minimum of a set of independent geometric random variables is also geometric. *Hint*: Start with $\mathbb{P}(\min\{X_1, \ldots, X_n\} > x)$ for independent random variables $X_i \sim \text{Geo}(p_i)$, $i \in \{1, \ldots, n\}$. What could be an equivalent way of writing this expression?

4 Introduction to Stochastic Processes

We'll end the day with a coding exercise. In this exercise you will approximate a Poisson process and use it to generate samples from an approximate Poisson distribution. If you're comfortable with Python, I'd recommend using NumPy, SciPy (for the `scipy.special.factorial` function), and Matplotlib. Complete the following tasks in order:

1. Plot the Poisson PMF with parameter $\lambda = 5$ for integers 0 to 20. Matplotlib's stem plots are a good option for this.

2. Approximately simulate a Poisson process for t_{total} simulated seconds with rate $\lambda = 5$ points/second. To do this, create a Bernoulli process with n Bernoulli random variables per second, each representing $\frac{1}{n}$th of a second. Think about the most efficient way you can simulate this process, especially for large n.

3. Count the number of points/successes in each second. These t_{total} counts are t_{total} samples from some distribution. Plot the observed proportions of each count value. How does this compare to your plot in part 1? What happens when you increase t_{total} and/or n?