Distributions and Continuous Probability

Geometric Distribution

1. One of the most important properties of the geometric distribution $X \approx Geom(p)$ is the memoryless property: $\Pr[X = j | X > i] = \Pr[X = j - i]$ for all $j > i$. Intuitively, what does this property say? Another way to write it is
$$\Pr[X - i = j - i | X > i] = \Pr[X = j - i].$$
Imagine that you are flipping a coin repeatedly and you have just observed that the first i flips were all tails. Verify that this property holds.

Solution: By using the definition of conditional probability we obtain:
$$\Pr[X = j | X > i] = \frac{\Pr[X = j \cap X > i]}{\Pr[X \geq i+1]} = \frac{\Pr[X = j]}{\Pr[X \geq i+1]}.$$
Now since we know that X has the geometric distribution, we know that $\Pr[X \geq i+1] = (1-p)^i$ (from the notes) and $\Pr[X = j] = p(1-p)^{j-1}$. Plugging this in we obtain :
$$\frac{p(1-p)^{j-1}}{(1-p)^i} = p(1-p)^{j-i-1} = \Pr[X = j - i].$$

2. Recall the elevator question from homework 10. Now assume that n people are getting on the elevator, but there are still 10 floors. Each person still gets off at a randomly selected floor, and each person's destination is independent of everyone else's. What is the expected number of people who need to get on the elevator until the elevator is required to stop at every floor?

Solution: This is very similar to the coupon collector's problem in the notes. Let X_i be the number of people who need to get on the elevator before the elevator stops at i distinct floors (starting immediately after the elevator stops at $i-1$ distinct floors). Then we're looking for $\mathbb{E}(X)$ where $X = X_1 + X_2 + \ldots + X_{10}$.

X_1 is trivial; the elevator stops at 1 distinct floor as soon as 1 person gets on. Therefore, $\Pr[X_1 = 1] = 1$ so $\mathbb{E}(X_1) = 1$. For X_2, note that each time a new person gets on the elevator, the elevator stops at the same floor as the first person with probability $\frac{1}{10}$ and a distinct floor with probability $\frac{9}{10}$. Therefore, $\Pr[X_2 = i] = (\frac{1}{10})^{i-1}(\frac{9}{10})$; X_2 has the geometric distribution with parameter $p = \frac{9}{10}$ and $\mathbb{E}(X) = \frac{10}{9}$.

Next consider X_3. Each time a new person gets on the elevator, the elevator stops at one of the earlier two distinct floors with probability $\frac{2}{10}$ and a new floor with probability $\frac{8}{10}$. So X_3 has the geometric distribution with parameter $p = \frac{8}{10}$ and $\mathbb{E}(X_3) = \frac{10}{8}$. We can make the same argument for all X_i to obtain that X_i has the geometric distribution with parameter $p = \frac{10-(i-1)}{10}$ and $\mathbb{E}(X_i) = \frac{10}{11-i}$.

To find $\mathbb{E}(X)$, we simply use linearity of expectation to obtain:

$$\mathbb{E}(X) = \sum_{i=1}^{10} \mathbb{E}(X_i) = \sum_{i=1}^{10} \frac{10}{11-i} = 10 \sum_{i=1}^{10} \frac{1}{i}.$$

Poisson Distribution

On average, .7 cars drive by a gas station per hour. Each car stops independently at the gas station with probability p.

1. Let X be the number of cars that drive by the gas station. For $i \geq 0$, what is $\Pr[X = i]$?

   Solution: We model X as a Poisson random variable with parameter $\lambda = .7$. Then $\Pr[X = i] = \frac{\lambda^i}{i!} e^{-\lambda}$.

2. Let Y be the number of cars that stop at the gas station in the next hour. For $i \geq 0$, what is $\Pr[Y = i]$? You can use the following identity: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

   Solution: Assume i cars drive by the gas station in the next hour. Then the probability that i cars stop at the gas station is p^i. If n cars drive by the gas station, the probability that i cars stop at the gas station is $\binom{n}{i} p^i (1-p)^{n-i}$. This is because we are choosing i of the n cars to stop at the gas station, and they stop with probability p while the rest do not stop with probability $1-p$.

   Let's generalize this reasoning by using the total probability rule:

   $$\Pr[Y = i] = \sum_{k=0}^{\infty} \Pr[Y = i | X = k+i] \Pr[X = k+i].$$

   We calculated $\Pr[Y = i | X = k+i]$ above: it is the probability that i cars stop at the gas station if $k+i$ drive by, so it is $\binom{k+i}{i} p^i (1-p)^k$. We calculated $\Pr[X = k+i]$ in the previous part of this question. We can plug in these values in the above expression:

   $$\Pr[Y = i] = \sum_{k=0}^{\infty} \frac{\lambda^{k+i}}{(k+i)!} e^{-\lambda} \binom{k+i}{i} p^i (1-p)^k.$$

   We can simplify the expression:

   $$\Pr[Y = i] = \frac{(\lambda p)^i e^{-\lambda}}{i!} \sum_{k=0}^{\infty} \frac{(\lambda(1-p))^k}{k!}.$$

   Now we'll use the following identity to simplify further: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Plugging in this identity we obtain:

   $$\Pr[Y = i] = \frac{(\lambda p)^i e^{-\lambda} e^{\lambda(1-p)}}{i!} = \frac{(\lambda p)^i e^{-\lambda p}}{i!}.$$

   Note that this is a Poisson distribution with parameter λp.

Continuous Probability

Let's revisit the bus problem from the previous discussion and from homework 10. Let's say there are two bus lines, 1 and 2, and each stops at your house once per hour. A line 1 bus stops at your house at the hour and a line 2 bus stops at your house half past the hour. However, line 1 takes you to work in 15 minutes and line 2 takes you to work in 45 minutes. You just get on whichever bus stops at your house first. What is the expected time it takes you to get to work?

1. What is the probability density function?

   Solution: It can take you anywhere between 15 minutes and 75 minutes to get to work, depending on when you wake up. If you wake up exactly at the hour, it takes you 15 minutes, and if you wake up right after the hour, it will take you 75 minutes (you have to wait about 30 minutes for a line 2 bus which takes 45 minutes to get to work). This is a continuous uniform distribution on the interval $[15,75]$. Therefore, the density function is :

$$f(x) = \begin{cases} 0 & \text{for } x < 0; \\ 1/60 & \text{for } 15 \leq x \leq 75; \\ 0 & \text{for } x > 75. \end{cases}$$

2. What is the expectation?

   Solution: Let X be the time it takes you to get to work. Then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{15}^{75} x \frac{1}{60} dx = \left[\frac{x^2}{120} \right]_{15}^{75} = 45.$$

3. What is the variance?

   Solution:

$$\text{Var}(X) = \int_{15}^{75} x^2 \frac{1}{60} dx - \mathbb{E}(X)^2 = \left[\frac{x^3}{180} \right]_{15}^{75} - (45)^2 = 300$$

4. Now let's say line 1 takes 5 minutes and line 2 still takes 45 minutes. What is the density function? The expectation?

   Solution: Now it can take you between 5 and 35 minutes to get to work, or between 45 and 75 minutes. It takes you 5 minutes if you wake up exactly at the hour and it takes you 35 minutes if you wake up just past the half hour. It takes you 45 minutes to get to work if you wake up exactly at the half hour, and 75 minutes if you wake up just past the hour. This is still a continuous uniform distribution, but it is on two disjoint intervals: $[5,35]$ and $[45,75]$. The density function is now:

$$f(x) = \begin{cases} 0 & \text{for } x < 0; \\ 1/60 & \text{for } 5 \leq x \leq 35; \\ 0 & \text{for } 35 < x < 45; \\ 1/60 & \text{for } 45 \leq x \leq 75; \\ 0 & \text{for } x > 75. \end{cases}$$

   We can compute the expectation as above:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{5}^{35} x \frac{1}{60} dx + \int_{45}^{75} x \frac{1}{60} dx = \left[\frac{x^2}{120} \right]_{5}^{35} + \left[\frac{x^2}{120} \right]_{45}^{75} = 40.$$