CS 70 Spring 2018 | Discrete Mathematics and Probability Theory | Babak Ayazifar and Satish Rao | DIS 9B

1 Telebears

Lydia has just started her Telebears appointment. She needs to register for a marine science class and CS 70. There are no waitlists, and she can attempt to enroll once per day in either class or both. The Telebears system is strange and picky, so the probability of enrolling in the marine science class is μ and the probability of enrolling in CS 70 is κ. These events are independent. Let M be the number of days it takes to enroll in the marine science class, and C be the number of days it takes to enroll in CS 70.

(a) What distribution do M and C follow? Are M and C independent?

(b) For some integer $k \geq 1$, what is $\mathbb{P}[C \geq k]$?

(c) For some integer $k \geq 1$, what is the probability that she is enrolled in both classes before day k?

Solution:

(a) $M \sim \text{Geometric}(\mu)$, $C \sim \text{Geometric}(\kappa)$. Yes they are independent.

(b) We are looking for the probability that it takes at least k days to enroll in CS 70. Using the geometric distribution, this is $(1-\kappa)^{k-1}$.

(c) Use independence. Let X be the number of days before she is enrolled in both.

$$\begin{aligned}\mathbb{P}[X < k] &= \mathbb{P}[M < k]\mathbb{P}[C < k] = (1 - \mathbb{P}[M \geq k])(1 - \mathbb{P}[C \geq k]) \\ &= (1-(1-\mu)^{k-1})(1-(1-\kappa)^{k-1})\end{aligned}$$

2 Sum of Bernoulli and Geometric Distributions

We know that the sum of i.i.d. Bernoulli random variables follow a binomial distribution. Now, we will consider the sum of i.i.d. geometric random variables.

1. Show that the expectation and variance of the sum of Bernoulli random variables, X_i, matches those of a binomial random variable, B.

2. Find the expectation and variance of the sum of geometric random variables, $G_i \sim \text{Geometric}(p)$.

3. Say you flip a coin until you get k heads. What is the expected number of flips? Variance?

Solution:

1. **Expectation**

$$\begin{aligned} B &= X_1 + X_2 + \cdots + X_n \\ \mathbb{E}[B] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} \mathbb{E}[X_i] \end{aligned}$$

Note that $\mathbb{E}[X_i] = p$ because X_i is a Bernoulli r.v. Therefore,

$$\mathbb{E}[B] = np.$$

which is the expectation for a binomial distribution!

Variance

Since all X_is are independent, we can use the linearity of variance.

$$\begin{aligned} \operatorname{var}(B) &= \operatorname{var}\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} \operatorname{var}(X_i) \\ &= np(1-p) \end{aligned}$$

which is the variance for a binomial distribution!

2. **Expectation**

$$\begin{aligned} G &= G_1 + G_2 + \cdots + G_n \\ \mathbb{E}[G] &= \mathbb{E}\left[\sum_{i=1}^{n} G_i\right] \\ &= \sum_{i=1}^{n} \mathbb{E}[G_i] \end{aligned}$$

Note that $\mathbb{E}[G_i] = \frac{1}{p}$ because G_i is a geometric r.v. Therefore,

$$\mathbb{E}[B] = \frac{n}{p}.$$

Variance

Since all G_is are independent, we can use linearity of variance.

$$\begin{aligned}\mathrm{var}(G) &= \mathrm{var}\left(\sum_{i=1}^{n} G_i\right) \\ &= \sum_{i=1}^{n} \mathrm{var}(G_i) \\ &= \frac{n(1-p)}{p^2}\end{aligned}$$

Note: The sum of geometric distributions equal the Pascal distribution! This distribution is out of the scope of the class but is useful when analyzing Bernoulli processes.

3. Using the expectation and variance found in part (b):

$$\begin{aligned}\mathbb{E}[G] &= \frac{k}{p} \\ \mathrm{var}(G) &= \frac{k(1-p)}{p^2}\end{aligned}$$

3 Fishy Computations

Use the Poisson distribution to answer these questions:

(a) Suppose that on average, a fisherman catches 20 salmon per week. What is the probability that he will catch exactly 7 salmon this week?

(b) Suppose that on average, you go to Fisherman's Wharf twice a year. What is the probability that you will go at most once in 2018?

(c) Suppose that in March, on average, there are 5.7 boats that sail in Laguna Beach per day. What is the probability there will be *at least* 3 boats sailing throughout the *next two days* in Laguna?

Solution:

(a) $X \sim \mathrm{Poiss}(20)$.

$$\mathbb{P}[X=7] = \frac{20^7}{7!}e^{-20} \approx 5.23 \cdot 10^{-4}.$$

(b) $X \sim \mathrm{Poiss}(2)$.

$$\mathbb{P}[X \le 1] = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} \approx 0.41.$$

(c) Let Y be the number of boats that sail in the next two days. We can approximate Y as a Poisson distribution $Y \sim \text{Poiss}(\lambda = 11.4)$, where λ is the average number of boats that sail over two days. Now, we compute

$$\begin{aligned}
\mathbb{P}[Y \geq 3] &= 1 - \mathbb{P}[Y < 3] \\
&= 1 - \mathbb{P}[Y = 0 \cup Y = 1 \cup Y = 2] \\
&= 1 - (\mathbb{P}[Y = 0] + \mathbb{P}[Y = 1] + \mathbb{P}[Y = 2]) \\
&= 1 - \left(\frac{11.4^0}{0!} e^{-11.4} + \frac{11.4^1}{1!} e^{-11.4} + \frac{11.4^2}{2!} e^{-11.4} \right) \\
&\approx 0.999.
\end{aligned}$$

We can show what we did above formally with the following claim: the sum of two independent Poisson random variables is Poisson. We won't prove this, but from the above, you should intuitively know why this is true. Now, we can model sailing boats on day i as a Poisson distribution $X_i \sim \text{Poiss}(\lambda = 5.7)$. Now, let X_1 be the number of sailing boats on the next day, and X_2 be the number of sailing boats on the day after next. We are interested in $Y = X_1 + X_2$. Thus, we know $Y \sim \text{Poiss}(\lambda = 5.7 + 5.7 = 11.4)$.

4 Sum of Poisson Variables

Assume that you were given two independent Poisson random variables X_1, X_2. Assume that the first has mean λ_1 and the second has mean λ_2. Prove that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

Hint: Recall the binomial theorem.

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Solution:

To show that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$, we have show that

$$\mathbb{P}[(X_1 + X_2) = i] = \frac{(\lambda_1 + \lambda_2)^i}{i!} \mathrm{e}^{-(\lambda_1 + \lambda_2)}.$$

We proceed as follows:

$$\begin{aligned}
\mathbb{P}[(X_1 + X_2) = i] &= \sum_{k=0}^{i} \mathbb{P}[X_1 = k, X_2 = (i-k)] = \sum_{k=0}^{i} \frac{\lambda_1^k}{k!} \mathrm{e}^{-\lambda_1} \cdot \frac{\lambda_2^{i-k}}{(i-k)!} \mathrm{e}^{-\lambda_2} \\
&= \mathrm{e}^{-\lambda_1} \mathrm{e}^{-\lambda_2} \sum_{k=0}^{i} \frac{1}{k!(i-k)!} \lambda_1^k \lambda_2^{i-k} = \frac{\mathrm{e}^{-\lambda_1} \mathrm{e}^{-\lambda_2}}{i!} \sum_{k=0}^{i} \frac{i!}{k!(i-k)!} \lambda_1^k \lambda_2^{i-k} \\
&= \frac{\mathrm{e}^{-(\lambda_1 + \lambda_2)}}{i!} \sum_{k=0}^{i} \binom{i}{k} \lambda_1^k \lambda_2^{i-k} = \frac{\mathrm{e}^{-(\lambda_1 + \lambda_2)}}{i!} (\lambda_1 + \lambda_2)^i
\end{aligned}$$

In the last line, we use the binomial expansion.