Discussion 5D

CS 70, Summer 2024

1 Rolling Dice

(a) Since rolling a die until obtaining a 6 involves independent rolls with a constant probability of success per roll, the number of times we roll follows a geometric distribution.

This question seeks to review basic formulas for the geometric distribution. The probability of rolling a 6 is $1/6$. Recall that the expectation is the inverse of the probability. Thus, the expectation is $1/(1/6) = 6$ rolls.

(b) If we roll the two dice, three outcomes are possible: the two dice show the same number, Die 1 is greater than Die 2, or Die 2 is greater than Die 1. The last two events occur with the same likelihood and the first event occurs with chance $n/n^2 = 1/n$, since there are n^2 possible rolls and n different numbers for which there could be duplicates. Thus the number of ways that Die 1 is smaller than Die 2 on a given roll is $(n^2 - n)/2$, so the probability that this occurs on a given roll is $(n^2 - n)/(2n^2) = (n-1)/(2n)$.

The number of times we roll is therefore geometrically distributed with

$$p = \frac{n-1}{2n}.$$

Recall that $\mathbb{E}(X) = \frac{1}{p}$ where $X \sim \text{Geo}(p)$. In our case, we have a geometric r.v. X with $p = \frac{(n-1)}{2n}$, so plugging into the formulas yields

$$\mathbb{E}(X) = \frac{1}{p} = \frac{1}{\frac{n-1}{2n}} = \frac{2n}{n-1}.$$

2 Shuttles and Taxis at Airport

(a) **(i)** Let $T([0,20])$ denote the number of taxis that arrive between times 00:00 and 00:20. This interval has length 20 minutes, so the number of taxis $T([0,20])$ arriving in this interval is distributed according to $\text{Poisson}(\lambda_2 \cdot 20) = \text{Poisson}(2)$, i.e.

$$\Pr[T([0,20]) = t] = \frac{2^t e^{-2}}{t!}, \text{ for } t = 0, 1, 2, \dots.$$

(ii) Let $S([0,20])$ denote the number of shuttles that arrive between times 00:00 and 00:20. This interval has length 20 minutes, so the number of shuttles $S([0,20])$ arriving in this interval is distributed according to $\text{Poisson}(\lambda_1 \cdot 20) = \text{Poisson}(1)$, i.e.

$$\Pr[S([0,20]) = s] = \frac{1^s e^{-1}}{s!}, \text{ for } s = 0, 1, 2, \dots.$$

(iii) Let $N([0,20]) = S([0,20]) + T([0,20])$ denote the total number of pickup vehicles (taxis and shuttles) arriving between times 00:00 and 00:20. Since the sum of independent Poisson random variables is Poisson distributed with parameter given by the sum of the individual parameters, we have $N[(0,20)] \sim \text{Poisson}(3)$, i.e.

$$\Pr[N([0,20]) = n] = \frac{3^n e^{-3}}{n!}, \text{ for } n = 0, 1, 2, \dots.$$

(b) We have

$$\Pr[T([0,20]) = 3] = \frac{2^3 e^{-2}}{3!} \text{ and } \Pr[S([0,20]) = 1] = \frac{1^1 e^{-1}}{1!}.$$

Since the taxis and the shuttles arrive independently, the probability that exactly 3 taxis and 1 shuttle arrive in this interval is given by the product of their individual probabilities, i.e.

$$\frac{2^3 e^{-2}}{3!} \frac{1^1 e^{-1}}{1!} = \frac{4}{3} e^{-3} \approx 0.0664.$$

(c) Let A be the event that exactly 1 taxi arrives between times 00:00 and 00:20. Let B be the event that exactly 1 vehicle arrives between times 00:00 and 00:20. We have

$$\Pr[B] = \frac{3^1 e^{-3}}{1!}.$$

Event $A \cap B$ is the event that exactly 1 taxi and 0 shuttles arrive between times 00:00 and 00:20. Hence

$$\Pr[A \cap B] = \frac{2^1 e^{-2}}{1!} \frac{1^0 e^{-1}}{0!}.$$

Thus, we get

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = 2/3.$$

(d) The event that you need to wait for more than 10 minutes starting 00:20 is equivalent to the event that no vehicle arrives between times 00:20 and 00:30. Let $N[20, 30]$ denote the number of vehicles that arrive between times 00:20 and 00:30. This interval has length 10 minutes, so $N[(20, 30)] \sim \text{Poisson}((\lambda_1 + \lambda_2) \cdot 10) = \text{Poisson}(3/2)$. Since Poisson arrivals in disjoint intervals are independent, we have

$$\Pr\left[N([20, 30]) = 0 \mid T([0, 20]) = 3, S([0, 20]) = 1\right] = \Pr[N([20, 30]) = 0] \sim \frac{1.5^0 e^{-1.5}}{0!} = e^{-1.5} \approx 0.2231.$$

3 Soccer Practice

Let X_1 denote the number of rounds until Messi's first miss, and X_2 denote the number of rounds until Ronaldo's first miss. We have that $X_1 \sim \text{Geometric}(p_1)$ and $X_2 \sim \text{Geometric}(p_2)$, where X_1, X_2 are independent r.v.'s.

Now, for all $k \in \{1, 2, \dots\}$, $\min(X_1, X_2) = k$ is equivalent to $(X_1 = k) \cap (X_2 \geq k)$ or $(X_2 = k) \cap (X_1 > k)$. Hence,

$$\begin{aligned} \Pr[\min(X_1, X_2) = k] &= \Pr[(X_1 = k) \cap (X_2 \geq k)] + \Pr[(X_2 = k) \cap (X_1 > k)] \\ &= \Pr[X_1 = k] \cdot \Pr[X_2 \geq k] + \Pr[X_2 = k] \cdot \Pr[X_1 > k] \end{aligned}$$

(since X_1 and X_2 are independent)

$$= [(1 - p_1)^{k-1} p_1](1 - p_2)^{k-1} + [(1 - p_2)^{k-1} p_2](1 - p_1)^k$$

(since X_1 and X_2 are geometric)

$$\begin{aligned} &= \left((1 - p_1)(1 - p_2)\right)^{k-1} (p_1 + p_2(1 - p_1)) \\ &= (1 - p_1 - p_2 + p_1 p_2)^{k-1} (p_1 + p_2 - p_1 p_2). \end{aligned}$$

But this final expression is precisely the probability that a geometric RV with parameter $p_1 + p_2 - p_1 p_2$ takes the value k. Hence $\min(X_1, X_2) \sim \text{Geometric}(p_1 + p_2 - p_1 p_2)$, and $\text{E}[\min(X_1, X_2)] = (p_1 + p_2 - p_1 p_2)^{-1}$.

Then for $\text{E}[\max(X_1, X_2)]$, we can use inclusion-exclusion. Note that we always have $\max(X_1, X_2) = X_1 + X_2 - \min(X_1, X_2)$ for any value of X_1, X_2, then by linearity of expectation, we get $\text{E}[\max(X_1, X_2)] = \text{E}[X_1] + \text{E}[X_2] - \text{E}[\min(X_1, X_2)] = p_1^{-1} + p_2^{-1} - (p_1 + p_2 - p_1 p_2)^{-1}$.

An alternative, slightly cleaner approach is to work with the *tail probabilities* of the geometric distribution, rather than with the usual point probabilities as above. In other words, we can work with $\Pr[X \geq k]$ rather than with $\Pr[X = k]$; clearly the values $\Pr[X \geq k]$ specify the values $\Pr[X = k]$ since $\Pr[X = k] = \Pr[X \geq k] - \Pr[X \geq (k + 1)]$, so it suffices to calculate them instead. We then get the following argument:

$$\begin{aligned} \Pr[X \geq k] &= \Pr[\min(X_1, X_2) \geq k] = \Pr[(X_1 \geq k) \cap (X_2 \geq k)] \\ &= \Pr[X_1 \geq k] \cdot \Pr[X_2 \geq k] && \text{since } X_1, X_2 \text{ are independent} \\ &= (1 - p_1)^{k-1} (1 - p_2)^{k-1} && \text{since } X_1, X_2 \text{ are geometric} \\ &= \left((1 - p_1)(1 - p_2)\right)^{k-1} = (1 - p_1 - p_2 + p_1 p_2)^{k-1}. \end{aligned}$$

This is the tail probability of a geometric distribution with parameter $p_1 + p_2 - p_1 p_2$, so $\min(X_1, X_2) \sim \text{Geometric}(p_1 + p_2 - p_1 p_2)$. The rest is the same as the above approach.

4 Cookie Jars

Assume that you found jar 1 empty; the probability that $X = k$ and you found jar 1 empty is computed as follows.

In order for there to be k cookies remaining, you must have eaten a cookie for $2n - k$ days, and then you must have chosen jar 1 (to discover that it is empty). Within those $2n - k$ days, exactly n of those days you chose jar 1. The probability of this is $\binom{2n-k}{n} 2^{-(2n-k)}$.

Furthermore, the probability that you then discover jar 1 is empty the day after is $1/2$. So, the probability that $X = k$ and you discover jar 1 empty is $\binom{2n-k}{n} 2^{-(2n-k+1)}$. However, we assumed that we discovered jar 1 to be empty; the probability that $X = k$ and jar 2 is empty is the same by symmetry, so the overall probability that $X = k$ is:

$$\Pr[X = k] = \binom{2n-k}{n} \frac{1}{2^{2n-k}}, \qquad k \in \{0, \ldots, n\}.$$