1 Variance

(a) Let X be a random variable representing the outcome of the roll of one fair 6-sided die. What is $\text{Var}(X)$?

(b) Let Z be a random variable representing the average of n rolls of a fair die 6-sided die. What is $\text{Var}(Z)$?

Solution:

(a) Recall that $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$. We can compute each of the individual terms using the definition of expectation:

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} \\ \mathbb{E}\left[X^2\right] &= \frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right) \\ &= \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}\end{aligned}$$

Now, we plug back into the variance expression:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}\end{aligned}$$

(b) Because each die roll is independent of the others, we can utilize the fact that for independent random variables X and Y, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$. Let X_i be a random variable representing the outcome of the ith dice roll. We now have:

$$\begin{aligned}
\text{Var}(Z) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \text{Var}(X_i) && \text{All } X_i\text{'s are independent.} \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \frac{35}{12} && \text{We computed the variance of one roll in part (a).} \\
&= \left(\frac{1}{n}\right)^2 \cdot n \cdot \frac{35}{12} = \frac{35}{12n}
\end{aligned}$$

2 Coupon Collector Variance

It's that time of the year again - Safeway is offering its Monopoly Card promotion. Each time you visit Safeway, you are given one of n different Monopoly Cards with equal probability. You need to collect them all to redeem the grand prize.

Let X be the number of visits you have to make before you can redeem the grand prize. Show that $\text{Var}(X) = n^2 \left(\sum_{i=1}^{n} i^{-2}\right) - \mathbb{E}(X)$. *[Hint: Try to break this problem down using indicators as with the coupon collector's problem. Are the indicators independent?]*

Solution:

Note that this is the coupon collector's problem, but now we have to find the variance. Let X_i be the number of visits we need to make before we have collected the ith unique Monopoly card actually obtained, given that we have already collected $i-1$ unique Monopoly cards. Then $X = \sum_{i=1}^{n} X_i$ and each X_i is geometrically distributed with $p = (n-i+1)/n$. Moreover, the indicators themselves are independent, since each time you collect a new card, you are starting from a clean slate.

$$\begin{aligned}
\text{Var}(X) &= \sum_{i=1}^{n} \text{Var}(X_i) && \text{(as the } X_i \text{ are independent)} \\
&= \sum_{i=1}^{n} \frac{1-(n-i+1)/n}{[(n-i+1)/n]^2} && \text{(variance of a geometric r.v. is } (1-p)/p^2) \\
&= \sum_{j=1}^{n} \frac{1-j/n}{(j/n)^2} && \text{(by noticing that } n-i+1 \text{ takes on all values from 1 to } n) \\
&= \sum_{j=1}^{n} \frac{n(n-j)}{j^2} \\
&= \sum_{j=1}^{n} \frac{n^2}{j^2} - \sum_{j=1}^{n} \frac{n}{j} \\
&= n^2 \left(\sum_{j=1}^{n} \frac{1}{j^2} \right) - \mathbb{E}(X) && \text{(using the coupon collector problem expected value).}
\end{aligned}$$

3 Shuttles and Taxis at Airport

In front of terminal 3 at San Francisco Airport is a pickup area where shuttles and taxis arrive according to a Poisson process. The shuttles arrive at a rate $\lambda_1 = 1/20$ (i.e. 1 shuttle per 20 minutes) and the taxis arrive at a rate $\lambda_2 = 1/10$ (i.e. 1 taxi per 10 minutes) starting at 00:00. The shuttles and the taxis arrive independently.

(a) What is the distribution of the following:

  (i) The number of taxis that arrive between times 00:00 and 00:20?

  (ii) The number of shuttles that arrive between times 00:00 and 00:20?

  (iii) The total number of pickup vehicles that arrive between times 00:00 and 00:20?

(b) What is the probability that exactly 1 shuttle and 3 taxis arrive between times 00:00 and 00:20?

(c) Given that exactly 1 pickup vehicle arrived between times 00:00 and 00:20, what is the conditional probability that this vehicle was a taxi?

(d) Suppose you reach the pickup area at 00:20. You learn that you missed 3 taxis and 1 shuttle in those 20 minutes. What is the probability that you need to wait for more than 10 mins until either a shuttle or a taxi arrives?

Solution:

(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 Poisson$(\lambda_2 \cdot 20) = $ Poisson(2), i.e.

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

(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 Poisson$(\lambda_1 \cdot 20) = $ Poisson(1), i.e.

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

(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 $N[(0,20)] \sim$ Poisson(3), i.e.

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

(b) We have

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

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

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

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

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

Thus, we get

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

(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 $N[(20,30)] \sim$ Poisson$((\lambda_1 + \lambda_2) \cdot 10) = $ Poisson$(3/2)$. Since Poisson arrivals in disjoint intervals are independent, we have

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