1 Shuttles and Taxis at Airport

Note 19

In front of terminal 3 at San Francisco Airport is a pickup area where shuttles and taxis arrive according to a Poisson distribution. 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 $\text{Poisson}(\lambda_2 \cdot 20) = \text{Poisson}(2)$, 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 $\text{Poisson}(\lambda_1 \cdot 20) = \text{Poisson}(1)$, 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 $N[(0,20)] \sim \text{Poisson}(3)$, 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 $N[(20,30)] \sim \text{Poisson}((\lambda_1 + \lambda_2) \cdot 10) = \text{Poisson}(3/2)$. Since Poisson arrivals in disjoint intervals are independent, we have

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

2 Family Planning

Note 15

Mr. and Mrs. Johnson decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Johnsons have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

(b) Compute the joint distribution of G and C. Fill in the table below.

	$C=1$	$C=2$	$C=3$
$G=0$			
$G=1$			

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

$\mathbb{P}[G=0]$	
$\mathbb{P}[G=1]$	

$\mathbb{P}[C=1]$	$\mathbb{P}[C=2]$	$\mathbb{P}[C=3]$

(d) Are G and C independent?

(e) What is the expected number of girls the Johnsons will have? What is the expected number of children that the Johnsons will have?

Solution:

(a) The sample space is the set of all possible sequences of children that the Johnsons can have: $\Omega = \{g, bg, bbg, bbb\}$. The probabilities of these sample points are:

$$\mathbb{P}[g] = \frac{1}{2}$$
$$\mathbb{P}[bg] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$\mathbb{P}[bbg] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$
$$\mathbb{P}[bbb] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

(b)

	$C=1$	$C=2$	$C=3$
$G=0$	0	0	$\mathbb{P}[bbb] = 1/8$
$G=1$	$\mathbb{P}[g] = 1/2$	$\mathbb{P}[bg] = 1/4$	$\mathbb{P}[bbg] = 1/8$

(c) Marginal distribution for G:

$$\mathbb{P}[G=0] = 0 + 0 + \frac{1}{8} = \frac{1}{8}$$
$$\mathbb{P}[G=1] = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$$

Marginal distribution for C:

$$\mathbb{P}[C=1] = 0 + \frac{1}{2} = \frac{1}{2}$$
$$\mathbb{P}[C=2] = 0 + \frac{1}{4} = \frac{1}{4}$$
$$\mathbb{P}[C=3] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

(d) No, G and C are not independent. If two random variables are independent, then

$$\mathbb{P}[X=x, Y=y] = \mathbb{P}[X=x]\mathbb{P}[Y=y].$$

To show this dependence, consider an entry in the joint distribution table, such as $\mathbb{P}[G=0, C=3] = 1/8$. This is not equal to $\mathbb{P}[G=0]\mathbb{P}[C=3] = (1/8)\cdot(1/4) = 1/32$, so the random variables are not independent.

(e) We can apply the definition of expectation directly for this problem, since we've computed the marginal distribution for both random variables.

$$\mathbb{E}[G] = 0 \cdot \mathbb{P}[G=0] + 1 \cdot \mathbb{P}[G=1] = 1 \cdot \frac{7}{8} = \frac{7}{8}$$
$$\mathbb{E}[C] = 1 \cdot \mathbb{P}[C=1] + 2 \cdot \mathbb{P}[C=2] + 3 \cdot \mathbb{P}[C=3] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{7}{4}$$

3 Linearity

Note 15

Solve each of the following problems using linearity of expectation. Explain your methods clearly.

(a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability $1/3$ (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability $1/5$, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

(b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears? (*Hint*: Consider where the sequence "book" can appear in the string.)

Solution:

(a) Let A_i be the indicator you win the ith time you play game A and B_i be the same for game B. The expected value of A_i and B_i are

$$\mathbb{E}[A_i] = 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \frac{1}{3},$$
$$\mathbb{E}[B_i] = 1 \cdot \frac{1}{5} + 0 \cdot \frac{4}{5} = \frac{1}{5}.$$

Then the expected total number of tickets you receive, by linearity of expectation, is

$$3\,\mathbb{E}[A_1] + \cdots + 3\,\mathbb{E}[A_{10}] + 4\,\mathbb{E}[B_1] + \cdots + 4\,\mathbb{E}[B_{20}] = 10\left(3 \cdot \frac{1}{3}\right) + 20\left(4 \cdot \frac{1}{5}\right) = 26.$$

Note that $10\left(3 \cdot \frac{1}{3}\right)$ and $20\left(4 \cdot \frac{1}{5}\right)$ matches the expression directly gotten using the expected value of a binomial random variable.

(b) There are $1,000,000 - 4 + 1 = 999,997$ places where "book" can appear, each with a (non-independent) probability of $1/26^4$ of happening. If A is the random variable that tells how many times "book" appears, and A_i is the indicator variable that is 1 if "book" appears starting at the ith letter, then

$$\begin{aligned}\mathbb{E}[A] &= \mathbb{E}[A_1 + \cdots + A_{999,997}] \\ &= \mathbb{E}[A_1] + \cdots + \mathbb{E}[A_{999,997}] \\ &= \frac{999,997}{26^4} \approx 2.19.\end{aligned}$$

4 Balls in Bins

Note 15 You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(a) What is $\mathbb{E}[X_i]$?

(b) What is the expected number of empty bins?

(c) Define a collision to occur when a ball lands in a nonempty bin (if there are n balls in a bin, count that as $n-1$ collisions). What is the expected number of collisions?

Solution:

(a) We will use linearity of expectation. Note that the expectation of an indicator variable is just the probability the indicator variable = 1. (Verify for yourself that is true).

$$\begin{aligned}\mathbb{E}[X_i] &= \mathbb{P}[\text{ball 1 falls into bin } i] + \mathbb{P}[\text{ball 2 falls into bin } i] + \cdots + \mathbb{P}[\text{ball } k \text{ falls into bin } i] \\ &= \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n}.\end{aligned}$$

(b) Let I_i be the indicator variable denoting whether bin i ends up empty. This can happen if and only if all the balls end in the remaining $n-1$ bins, and this happens with a probability of $\left(\frac{n-1}{n}\right)^k$. Hence the expected number of empty bins is

$$\mathbb{E}[I_1 + \ldots + I_n] = \mathbb{E}[I_1] + \ldots + \mathbb{E}[I_n] = n\left(\frac{n-1}{n}\right)^k$$

(c) The number of collisions is the number of balls minus the number of occupied bins, since the first ball of every occupied bin is not a collision.

$$\begin{aligned}\mathbb{E}[\text{collisions}] &= k - \mathbb{E}[\text{occupied bins}] = k - n + \mathbb{E}[\text{empty locations}] \\ &= k - n + n\left(1 - \frac{1}{n}\right)^k\end{aligned}$$