

1 Head Count

Note 15

Consider a coin with $\mathbb{P}[\text{Heads}] = 2/5$. Suppose you flip the coin 20 times, and define X to be the number of heads.

- (a) What is $\mathbb{P}[X = k]$, for some $0 \leq k \leq 20$?
- (b) Name the distribution of X and what its parameters are.
- (c) What is $\mathbb{P}[X \geq 1]$? Hint: You should be able to do this without a summation.
- (d) What is $\mathbb{P}[12 \leq X \leq 14]$?

Solution:

- (a) There are a total of $\binom{20}{k}$ ways to select k coins to be heads. The probability that the selected k coins to be heads is $(\frac{2}{5})^k$, and the probability that the rest are tails is $(\frac{3}{5})^{20-k}$. Putting this together, we have

$$\mathbb{P}[X = k] = \binom{20}{k} \left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{20-k}.$$

- (b) Since we have 20 independent trials, with each trial having a probability $2/5$ of success, $X \sim \text{Binomial}(20, 2/5)$.

- (c)

$$\mathbb{P}[X \geq 1] = 1 - \mathbb{P}[X = 0] = 1 - \left(\frac{3}{5}\right)^{20}.$$

- (d)

$$\begin{aligned} \mathbb{P}[12 \leq X \leq 14] &= \mathbb{P}[X = 12] + \mathbb{P}[X = 13] + \mathbb{P}[X = 14] \\ &= \binom{20}{12} \left(\frac{2}{5}\right)^{12} \left(\frac{3}{5}\right)^8 + \binom{20}{13} \left(\frac{2}{5}\right)^{13} \left(\frac{3}{5}\right)^7 + \binom{20}{14} \left(\frac{2}{5}\right)^{14} \left(\frac{3}{5}\right)^6. \end{aligned}$$

2 Head Count II

Note 19

Consider a coin with $\mathbb{P}[\text{Heads}] = 3/4$. Suppose you flip the coin until you see heads for the first time, and define X to be the number of times you flipped the coin.

- (a) What is $\mathbb{P}[X = k]$, for some $k \geq 1$?
- (b) Name the distribution of X and what its parameters are.
- (c) What is $\mathbb{P}[X > k]$, for some $k \geq 0$?
- (d) What is $\mathbb{P}[X < k]$, for some $k \geq 1$?
- (e) What is $\mathbb{P}[X > k \mid X > m]$, for some $k \geq m \geq 0$? How does this relate to $\mathbb{P}[X > k - m]$?
- (f) Suppose $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(q)$ are independent. Find the distribution of $\min(X, Y)$ and justify your answer.

Solution:

- (a) If we flipped k times, then we had $k - 1$ tails and 1 head, in that order, giving us

$$\mathbb{P}[X = k] = \frac{3}{4} \left(1 - \frac{3}{4}\right)^{k-1} = \frac{3}{4} \left(\frac{1}{4}\right)^{k-1}.$$

- (b) $X \sim \text{Geometric}(\frac{3}{4})$

- (c) If we had to flip *more than* k times before seeing our first heads, then our first k flips must have been tails, giving us

$$\mathbb{P}[X > k] = \left(1 - \frac{3}{4}\right)^k = \left(\frac{1}{4}\right)^k.$$

You can alternatively write as the sum $\sum_{i=k+1}^{\infty} \mathbb{P}[X = i] = \sum_{i=k+1}^{\infty} \frac{3}{4} * \left(\frac{1}{4}\right)^{i-1} = \frac{3}{4} * \left(\frac{1}{4}\right)^k * \frac{1}{1-1/4} = \left(\frac{1}{4}\right)^k$ using the formula for an infinite geometric sum

- (d) Notice $\mathbb{P}[X < k] = 1 - \mathbb{P}[X \geq k] = 1 - \mathbb{P}[X > k - 1]$ since X can only take on integer values. Along similar lines to the previous part, we then have

$$\mathbb{P}[X < k] = 1 - \mathbb{P}[X > k - 1] = 1 - \left(1 - \frac{3}{4}\right)^{k-1} = 1 - \left(\frac{1}{4}\right)^{k-1}.$$

- (e) By part (c), we have

$$\mathbb{P}[X > k \mid X > m] = \frac{\mathbb{P}[X > k \cap X > m]}{\mathbb{P}[X > m]} = \frac{\mathbb{P}[X > k]}{\mathbb{P}[X > m]} = \left(\frac{1}{4}\right)^{k-m}.$$

However, note that this is exactly $\mathbb{P}[X > k - m]$. The reason this makes sense is that if we want to compute the probability that the first heads occurs after k flips, and we know that the first

heads occurs after m flips, then the first m flips are tails. Thus, by the independence of the coin flips, the first m flips don't matter, and so we only need to compute the probability that the first heads occurs after $k - m$ flips. This is called the **memorylessness property** of the geometric distribution.

- (f) Let X be the number of coins we flip until we see a heads from flipping a coin with bias p , and let Y similarly be the number of coins we flip until we see a heads from flipping a coin with bias q .

Imagine we flip the bias p coin and the bias q coin at the same time. The minimum of the two random variables represents how many simultaneous flips occur before at least one head is seen.

The probability of not seeing a head at all on any given simultaneous flip is $(1 - p)(1 - q)$; this corresponds to a failure. This means that the probability that there will be a success on any particular trial is $1 - (1 - p)(1 - q) = p + q - pq$. Therefore, $\min(X, Y) \sim \text{Geometric}(p + q - pq)$.

Alternative 1: We can also solve this algebraically. The probability that $\min(X, Y) = k$ for some positive integer k is the probability that the first $k - 1$ coin flips for both X and Y were tails, and we get heads on the k th toss (this can come from either X or Y). Specifically, this occurs with probability

$$((1 - p)(1 - q))^{k-1} \cdot (p + q - pq)$$

We recognize this as the formula for a geometric random variable with parameter $p + q - pq$.

Alternative 2: 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. Let $Z = \min(X, Y)$. We can work with $\mathbb{P}[Z \geq k]$ rather than with $\mathbb{P}[Z = k]$; clearly the values $\mathbb{P}[Z \geq k]$ specify the values $\mathbb{P}[Z = k]$ since $\mathbb{P}[Z = k] = \mathbb{P}[Z \geq k] - \mathbb{P}[Z \geq (k + 1)]$, so it suffices to calculate them instead. We then get the following argument:

$$\begin{aligned} \mathbb{P}[Z \geq k] &= \mathbb{P}[\min(X, Y) \geq k] \\ &= \mathbb{P}[(X \geq k) \cap (Y \geq k)] \\ &= \mathbb{P}[X \geq k] \cdot \mathbb{P}[Y \geq k] && \text{since } X, Y \text{ are independent} \\ &= (1 - p)^{k-1} (1 - q)^{k-1} && \text{since } X, Y \text{ are geometric} \\ &= ((1 - p)(1 - q))^{k-1} \\ &= (1 - p - q + pq)^{k-1}. \end{aligned}$$

This is the tail probability of a geometric distribution with parameter $p + q - pq$, thus we can conclude that $Z \sim \text{Geom}(p + q - pq)$, which is the same result as before!

3 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.

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

- (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.

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

- (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.

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

- (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

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