Geometric and Poisson Distributions

Note 18

First, let's recap some key terms!

Random Variable: A random variable X is a function from $\Omega \to \mathbb{R}$, mapping the possible outcomes to real numbers. Note that this function itself is not random; the *outcomes* are random. We define

$$\mathbb{P}[X = k] = \mathbb{P}[\{\omega \in \Omega : X(\omega) = k\}].$$

Distribution of a random variable: the set of all $(k, \mathbb{P}[X = k])$, describing the probability of attaining each value of the random variable.

And now, the focus of today's discussion:

Geometric Distribution: $X \sim \text{Geometric}(p)$; X represents the number of independent trials until the first success (including the success), where p is the probability of success in each trial.

$$\mathbb{P}[X = k] = p(1-p)^{k-1}, \qquad \mathbb{E}[X] = \frac{1}{p}, \qquad \text{Var}(X) = \frac{1-p}{p^2}.$$

Poisson Distribution: $X \sim \text{Poisson}(\lambda)$; X represents the number of occurrences of an event in one unit of time, if on average there are λ occurrences in one unit of time. The distribution is described by the following:

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad \mathbb{E}[X] = \lambda, \qquad \text{Var}(X) = \lambda.$$

Further, if $X \sim \text{Poisson}(\lambda_x)$ and $Y \sim \text{Poisson}(\lambda_y)$ are independent, then $X + Y \sim \text{Poisson}(\lambda_x + \lambda_y)$.

1 Head Count II

Note 18

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 $k \geq 1$? Express your answer in terms of k. (Do not just copy down a formula—re-derive it yourself!)

(b) What is the name of the distribution of X, and what are its parameters?

(c) What is $\mathbb{P}[X > k]$, for some $k \geq 0$? (You should not have any summations.)

(d) What is $\mathbb{P}[X < k]$, for some $k \geq 1$? (You should not have any summations.)

(e) What is $\mathbb{P}[X > k \mid X > m]$, for some $k \geq m \geq 0$? Show that your answer is equal to $\mathbb{P}[X > k - m]$. Why do we call this the memoryless property?

(f) Suppose $Y \sim \text{Geometric}(p)$ and $Z \sim \text{Geometric}(q)$ are independent. Find the distribution of $\min(Y, Z)$ and justify your answer.

Hint: consider flipping two coins (with $\mathbb{P}[\text{Heads}] = p$ and $\mathbb{P}[\text{Heads}] = q$ respectively) simultaneously.

2 Unreliable Servers

Note 18

A Google competitor owns a warehouse consisting of a very large number of servers (a server farm). On any given day, every server in the farm may go down with the same probability, independently of all other servers and of what happens on other days. On average, 4 servers go down in the cluster per day.

(a) What is an appropriate distribution to model the number of servers that crash on any given day for a certain cluster? (Give the name and parameter(s) of the distribution.)

(b) Compute the expected value and variance of the number of crashed servers on a given day for a certain cluster.

(c) Compute the probability that strictly less than 3 servers crashed on a given day for a certain cluster.

(d) Compute the probability that at least 3 servers crashed on a given day for a certain cluster.

(e) Compute the probability that exactly 6 servers crashed over a given two-day period for a certain cluster.

(f) After a recent software update, servers never crash except when rebooting. Every morning each server is rebooted and each time a server is rebooted, it has a 5% chance of crashing. Which distribution can we use to model the day of a given computer's first crash?

(g) Compute the expected day of a given computer's first crash.

(h) Five of these upgraded computers are connected to create a server farm; these computers are rebooted at the same time every morning. What is the expected day of the cluster's first crash?

3 Shuttles and Taxis at Airport

Note 18

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) Write in terms of a Poisson distribution, the distribution of taxis between 00:00 and 00:01:

(b) 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 vehicles (shuttles and taxis) that arrive between times 00:00 and 00:20?

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

(d) 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?

(e) 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?