# L15. Geometric and Negative Binomial Distribution

**Recall:** A Binomial random variable, $X$, measures the probability of obtaining $x$ successes in $n$ trials.

A defining feature of the a Binomial random variable is that the trials are what we qualify as *Bernoulli trials* (i.e. independent trials with constant probability of success,$p$.)

In this lecture, we examine two more distributions which also feature Bernoulli trials in their random experiment: the geometric distribution and the negative binomial distribution.

## Geometric Distribution

A geometric random variable counts the number of trials until the first success is observed.

**Definition 1: Geometric Distribution**

In a series of Bernoulli trials, the random variable, $X$, that equals the number of trials until the first success is a **geometric random variable** with parameter $0 < p < 1$ and

$$P(X = x) = (1-p)^{x-1}p \qquad x = 1, 2, \ldots$$

**Notation:** $X \sim \mathrm{Geom(x, p)}$

## Expected Value and Variance of $X \sim$ Geom$(x, p)$

**Theorem 1**

If $X \sim \mathrm{Geom}(x, p)$ then;

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$$

### Example 1

Schimpf-los is a German hotline that allows customers to release pent-up anger and aggression by swearing at telephone operators[a]. The service is available 24/7 and costs $ 1.49 per minute.

Assume that each of your calls to a popular service has a probability of 0.02 of connecting, that is, of not obtaining a busy signal. Assume that your calls are independent.

a. What is the probability that it takes three attempts before your call connects?

b. What is the probability that the first time your call connects is your 10th attempt?

c. What is the probability that it requires more than five calls for you to connect?

d. How many attempts are needed on average before a call is connected?

[a]https://www.reuters.com/article/us-germany-hotline/germans-blow-off-steam-with-swearing-hotline-idUSBRE86O15420120725

### Solution

Let $X$ = the number of attempts until the call is connected

$X \sim \text{Geom}(0.02)$

a. We connect on the third attempt

$$P(X=3) = (1-0.02)^2 \cdot (0.02) = 0.0192$$

b. Connect on the $10^{th}$ attempt

$$P(X=10) = (1-0.02)^9 \cdot (0.02) = 0.0167$$

c. More than five calls

$$\begin{aligned}
P(X>5) &= P(X=6) + P(X=7) + \cdots \\
&= 1 - P(X \le 5) \\
&= 1 - [P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)] \\
&= 1 - [0.02 + (0.98)\cdot(0.02) + (0.98)^2\cdot(0.02)] \\
&\qquad + (0.98)^3\cdot(0.02) + (0.98)^4\cdot(0.02)] \\
&= 1 - 0.0961 \\
&= 0.9039
\end{aligned}$$

d. $E(X) = \dfrac{1}{p} = \dfrac{1}{0.02} = 50\,\text{calls}$

### Example 2

In China, the driving exam consists of 100 multiple-choice type questions. One of the items is: "If you come across a road accident victim, whose intestines are lying on the road, should you pick them up and push them back in"[a] In case you are taking the exam, the answer is "no".

Suppose that the probability that an applicant for a driver's license will pass the written exam on any given try is 0.65.

a. What is the probability that an applicant will pass the exam on their fourth try?

b. What is the probability that it will take more than four tries to finally pass the exam?

c. On average, how many tries does it take before an applicant passes the exam? What is the standard deviation?

[a] 1http://www.chinesedrivingtest.com/test

### Solution

Let $X$ = the number of attempts on the exam until the applicant passes

$X \sim \text{Geom}(0.65)$

a. Pass on the fourth try

$$P(X = 4) = (1 - 0.65)^3 \cdot (0.65) = 0.0279$$

b. More than four tries

$$\begin{aligned} P(X > 4) &= P(X = 5) + P(X = 5) + P(X = 7) \cdots \\ &= 1 - P(X \leq 4) \\ &= 1 - [P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)] \\ &= 1 - [0.65 + (0.35) \cdot (0.65) + (0.35)^2 \cdot (0.65)] \\ &\qquad + (0.35)^3 \cdot (0.65)] \\ &= 1 - 0.985 \\ &= 0.015 \end{aligned}$$

c. $E(X) = \dfrac{1}{p} = \dfrac{1}{0.65} = 1.5384\,\text{tries}$

$$V(X) = \frac{1-p}{p^2} = \frac{1-0.65}{(0.65)^2} = 0.8284 \quad \Rightarrow \quad S(X) = 0.9102$$

**Example 3**

Heart failure is due to either natural occurrences 87% or outside factors 13%. Assume that causes of heart failure for the individuals are independent.

a. What is the probability that the fifth patient with heart failure who enters the emergency room is the first one due to outside factors?

b. What is the probability that exactly one patient in the next five who enters the emergency room is suffering from heart failure due to outside factors?

c. What is the mean number of heart failure patients with the condition due to natural causes who enter the emergency room before the first patient with heart failure from outside factors? What is the standard deviation?

**Solution**

a. Want the first patient with heart failure due to outside factors to be the fifth patient that enters the ER. So we are counting the number of trials until we see the first success.

Let $X$ = the number of patients who arrive at the ER until the first one who presents with heart failure due to outside factors

$X \sim \text{Geom}(0.13)$

$P(X = 5) = (1 - 0.13)^4 \cdot (0.13) = 0.0745$

b. Want exactly one patient in the next five to present with heart failure due to outside factors. So we are counting the number of successes in a fixed number of trials.

Let $Y$ = the number of patients who arrive at the ER with heart failure due to outside factors.

$Y \sim \text{Binom}(5, 0.13)$

$P(Y = 1) = C_1^5(0.13) \cdot (1 - 0.13)^4 = 0.3724$

c. Let $Z$ = number of patients who have heart failure due to natural causes.

Since we want the first occurrence, $Z \sim \text{Geom}(0.87)$

$$E(Z) = \frac{1}{p} = \frac{1}{0.87} = 1.1494$$

$$V(Z) = \frac{1-p}{p^2} = \frac{1-0.87}{(0.87)^2} = 0.1718 \quad \Rightarrow \quad S(X) = 0.4144$$

## Negative Binomial Distribution

A natural generalization of the geometric distribution is the distribution of a random variable $X$ representing the number of Bernoulli trials necessary for the $r^{th}$ success to occur, where $r$ is a given positive integer.

**Definition 2: Negative Binomial Distribution**

In a series of Bernoulli trials, the random variable, $X$, that equals the number of trials until the $r^{th}$ success is a **negative binomial random variable** with parameter $0 < p < 1$ and

$$P(X = x) = C_{r-1}^{x-1}(1-p)^{x-r}p^r \qquad x = r, r+1, r+2, \ldots$$

**Notation:** $X \sim \text{NegBinom}(\text{r}, \text{p})$

**Remark**
The Geometric Distribution is a special case of the Negative Binomial with $r = 1$

## Expected Value and Variance of $X \sim$ NegBinom$(x, p)$

**Theorem 2**

If $X \sim \text{NegBinom}(r, p)$ then;

$$E(X) = \frac{r}{p} \qquad Var(X) = \frac{r(1-p)}{p^2}$$

**Example 4**

Alice is required to sell candy bars to raise money for the 6th grade field trip. There are thirty houses in the neighbourhood, and she is not supposed to return home until five candy bars have been sold. At each house, there is a 0.4 probability of selling one candy bar. Assume that sales at each house is independent.

a. What's the probability of selling the last candy bar at the $11^{th}$ house?

b. What's the probability of Alice finishing on or before the $8^{th}$ house?

**Solution**

Let $X$ = the number of houses that Alice visits until she sells her last candy bar

$X \sim \text{NegBinom}(5, 0.4)$

a. $P(X = 11) = C_{5-1}^{11-1} \cdot (1-0.4)^{11-5} \cdot (0.4)^5 = 0.1003$

b. $P(X \le 8)$

$$\begin{aligned} P(X \le 8) &= P(5 \le X \le 8) \\ &= P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) \\ &= C_4^4 (1 - 0.4)^0 \cdot (0.4)^5 + C_4^5 (1 - 0.4) \cdot (0.4)^5 \\ &\qquad + C_4^6 (1 - 0.4)^2 \cdot (0.4)^5 + C_4^7 (1 - 0.4)^3 \cdot (0.4)^5 \\ &= 0.1737 \end{aligned}$$

**Example 5**

Assume that the probability that a camera passes the water resistance test is 0.8 and that the cameras perform independently. Determine the following:

a. Probability that the second failure occurs on the tenth camera tested.

b. Probability that the second failure occurs in tests of four or fewer cameras.

c. Expected number of cameras tested to obtain the third failure.

**Solution**

Let $X$ =number of cameras tested until the second failure

$$X \sim \text{NegBinom}(2, 0.2)$$

a. $P(X = 10) = C_{2-1}^{10-1} (1 - 0.8)^8 \cdot (0.2)^2 = 0.0604$

b. $P(X \le 4)$

$$\begin{aligned} P(X \le 4) &= P(X = 2) + P(X = 3) + P(X = 4) \\ &= C_1^1 (1 - 0.2)^0 \cdot (0.2)^2 + C_1^2 (1 - 0.2) \cdot (0.2)^2 \\ &\qquad + C_1^3 (1 - 0.2)^2 \cdot (0.2)^2 \\ &= 0.1808 \end{aligned}$$

c. $E(X) = \dfrac{r}{p} = \dfrac{3}{0.2} = 15\,\text{cameras}$

### Example 6

If the probability is 0.75 that a person will believe a rumour about the transgressions of a certain politician, find the probabilities that:

a. The eighth person to hear the rumour will be the fifth person to believe it.

b. The fifteenth person to hear the rumour will be the tenth person to believe it.

c. The tenth person to hear the rumour is the first person to believe it.

d. Of the next ten people who hear the rumour, eight or nine of them will believe it.

### Solution

a. Let $X$ = of people who hear the rumour until the person who hears it, is the fifth person who believes it

$X \sim \text{NegBinom}(5, 0.75)$

$$P(X = 8) = C_{5-1}^{8-1}\,(1 - 0.75)^3 \cdot (0.75)^5 = 0.1298$$

b. Let $Y$ = of people who hear the rumour until the person who hears it, is the tenth person who believes it

$Y \sim \text{NegBinom}(10, 0.75)$

$$P(X = 10) = C_{10-1}^{15-1}\,(1 - 0.75)^5 \cdot (0.75)^{10} = 0.1109$$

c. Let $Z$ = of people who hear the rumour until the person who hears it, is the first person who believes it

$Z \sim \text{Geom}(0.75)$

$$P(Z = 10) = (1 - 0.75)^9 \cdot (0.75) = 2.86 \times 10^{-9}$$

d. Let $K$ = of people believe the rumour

$K \sim \text{Binom}(10, 0.75)$

$$\begin{aligned} P(K = 8 \cup K = 9) &= P(K = 8) + P(K = 9) \\ &= C_8^{10}\,(0.75)^8 \cdot (1 - 0.8)^2 + C_9^{10}\,(0.75)^9 \cdot (1 - 0.8) \\ &= 0.2816 + 0.1877 \\ &= 0.4693 \end{aligned}$$

**Example 7**

Roll two fair dice repeatedly. If the sum is at least 10, then you win.

a. What is the probability that you start by winning 3 times in a row?

b. What is the probability that after rolling the pair of dice 5 times you win exactly 3 times?

c. What is the probability that the first time you win is before the tenth roll (of the pair), but after the fifth?

**Solution**

a. $P(\text{win}) = \dfrac{6}{36} = \dfrac{1}{6}$

Let $E$ = be the event that we win three times in a row.

$$P(E) = [P(\text{win})]^3 = \frac{1}{6^3}$$

b. Let $X$ = the number of times that we win.

$X \sim \text{Binom}\left(5, \frac{1}{6}\right)$

$$P(X = 3) = C_3^5 \left(\frac{1}{6}\right)^3 \cdot \left(1 - \frac{1}{6}\right)^2 = \frac{250}{6^5}$$

c. Let $Y$ = number of games until we win for the first time.

$Y \text{ Geom}\left(\frac{1}{6}\right)$

$$\begin{aligned}
P(6 \le Y \le 9) &= P(Y = 6) + P(Y = 7) + P(Y = 8) + P(Y = 9) \\
&= \left(\frac{5}{6}\right)^5 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^6 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^7 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^8 \cdot \frac{1}{6} \\
&= \frac{5^5}{6^6}\left[1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^3\right] \\
&= 0.2081
\end{aligned}$$

**Example 8**

Two fair 6-sided dice are rolled, repeatedly. The first die is rolled until the first 3 comes up. The second die is rolled until the first 6 comes up. What is the probability that the two dice are rolled the same number of times?

**Solution**

Let $N_1$ = number or rolls until $1^{st}$ die shows 3 for the first time
$N_2$ = number or rolls until $2^{nd}$ die shows 6 for the first time

$N_1 \sim \text{Geom}\left(\frac{1}{6}\right)$
$N_2 \sim \text{Geom}\left(\frac{1}{6}\right)$

$$\begin{aligned}
P(N_1 = N_1) &= \sum_{k=1}^{\infty} P(N_1 = N_2 \mid N_1 = k) \cdot P(N_1 = k) \\
&= \sum_{k=1}^{\infty} P(N_2 = k \mid N_1 = k) \cdot P(N_1 = k) \\
&= \sum_{k=1}^{\infty} P(N_2 = k) \cdot P(N_1 = k) \quad ; \quad \text{Independence} \\
&= \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^{k-1} \left(\frac{1}{6}\right) \\
&= \frac{1}{36} \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{2k-2} \\
&= \frac{1}{36} \sum_{k=1}^{\infty} \left(\frac{25}{36}\right)^{k-1} \quad ; \quad \text{Convergent G-Series} \\
&= \frac{1}{36} \cdot \frac{1}{1 - (25/36)} \\
&= \frac{1}{11}
\end{aligned}$$