Geometric Random Variables

A random variable X is a geometric random variable with parameter α if its pmf is of the form

$$p(x) = \alpha \cdot (1-\alpha)^{x-1}$$

We write $X \sim \text{Geo}(\alpha)$ to denote this. If independent trials, each with probability α of success, are performed, and X is the number of trials required to see a success, then X is a geometric random variable. We have shown this is a valid pmf, and also that $E(X) = \frac{1}{\alpha}$ and $V(X) = \frac{1-\alpha}{\alpha^2}$. *Replacement means independence.*

Example

Suppose that there are 10 red balls and 40 white balls in an urn. Suppose that a ball is drawn, with replacement, until a red ball is drawn. What is the probability of it taking exactly 4 draws to obtain a red ball? What is the probability of it taking at least 4 draws to obtain a red ball?

Let X be the number of draws required to get a red ball. $X \sim \text{Geo}(0.2)$, $p(x) = 0.2 \cdot (0.8)^{x-1}$.

$$\begin{aligned}
\Pr(X = 4) &= 0.2 \cdot (0.8)^{x-1} \\
&= 0.2 \cdot (0.8)^3 \\
&= 0.2 \cdot 0.512 \\
&= 0.1024
\end{aligned}$$

$$\begin{aligned}
\Pr(X \geq 4) &= 1 - \Pr(X \leq 3) \\
&= 1 - (\Pr(X=1) + \Pr(X=2) + \Pr(X=3)) \\
&= 1 - \left(0.2 \cdot (0.8)^{1-1} + 0.2 \cdot (0.8)^{2-1} + 0.2 \cdot (0.8)^{3-1}\right) \\
&= 1 - 0.488 \\
&= 0.512
\end{aligned}$$

Negative Binomial Random Variables

Negative binomial random variables are extensions of geometric random variables, except X is the number of trials required to see r successes instead of just one success. A random variable X is said to be a negative binomial random variable with parameters (r, α) if its pmf is of the form:

$$p(x) = \binom{x-1}{r-1} \cdot \alpha^r \cdot (1-\alpha)^{x-r}$$

We write $X \sim \text{NB}(r, \alpha)$ to denote this. Note: The following are equivalent $X \sim \text{NB}(1, \alpha)$ and $X \sim \text{Geo}(\alpha)$. Further, if $X_i \sim \text{Geo}(\alpha)$ $(i = 1, 2, \ldots, r)$ is a family of independent geometric random variables, then the sum $X = \sum_{i=1}^{r} X_i$ is negative binomial with parameters (r, α).

If $X \sim \text{NB}(r, \alpha)$, then the mean and variance are:

$$E(X) = \frac{r}{\alpha} \qquad \text{and} \qquad V(X) = \frac{r \cdot (1 - \alpha)}{\alpha^2}$$

Example

Throw a die until four 1s are observed and let X be the number of throws required to see the four 1s. Then X is a negative binomial random variable with parameters $\left(4, \frac{1}{6}\right)$. What is the probability of it taking exactly 20 throws to do this? What is the average number of throws required to do this? What is the variance of the number of throws required to do this?

We know that $X \sim \text{NB}\left(4, \frac{1}{6}\right)$.

$$\begin{aligned} \Pr(X = 20) &= \binom{20-1}{4-1} \cdot \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^{20-4} \\ &= \binom{19}{3} \cdot \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^1 6 \\ &= (969)(0.0007716)(0.054088) \end{aligned}$$

$$E(X) = \frac{4}{(1/6)} = 24$$

$$V(X) = \frac{4(1 - (1/6))}{(1/6)^2} = 120$$

Example

Suppose that there are 100 red tickets in a box of 10 000 tickets. You must pay \$5 to draw a ticket, which is then put back in the box. If you draw a red ticket, you earn a token, and 5 tokens can be redeemed for a prize worth \$2000. Should you play this game in aim of winning the prize?

Let X be the number of draws to get 5 red tickets, so $X \sim \text{NB}(5, 0.01)$.

$$E(X) = \frac{5}{0.01} = 500$$

On average, it takes 500 draws to get 5 red tickets. Since in each draw you pay \$5, you would expect to pay:

$$5 \cdot E(X) = 5 \cdot 500 = 2500$$

So, you should not play this game to win the prize, meaning, on average, you will lose \$500.

Continuous Random Variables

We say that $X : \Omega \to \mathbb{R}$ is a (real-valued) continuous random variable if there is a non-negative function $f : \mathbb{R} \to [0, \infty]$ such that, for any set $A \subseteq \mathbb{R}$ of real numbers,

$$\Pr(X \in A) = \int_A f(x)\ dx$$

For such data, f is called the probability density function (pdf) of X. The cdf of a continuous rv with pdf f is the area function from calculus:

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(y)\ dy$$

Recall: The Fundamental Theorem of Calculus: $\frac{d}{dx} F(x) = f(x)$.

By the definition of probability, we have:

$$\Pr(X \in \mathbb{R}) = \int_{\mathbb{R}} f(x)\ dx = 1$$

and by the Fundamental Theorem of Calculus,

$$\Pr(X \in [a, b]) = \Pr(a \leq X \leq b) = \int_a^b f(x)\ dx = F(b) - F(a)$$

Example

Let X be the age of a randomly chosen Canadian. Assume (unreasonably) that X is a continuous random variable whose pdf is constant on $[0,50]$ and linearly decreases to 0 on $(50,100]$ (nobody in this model lives past 100). What is the probability of choosing someone whose age is between 40 and 50?

X is a continuous random variable.

$$f(x) = \begin{cases} C & \text{if } x \in [0,50] \\ \text{lin. decr.} & \text{if } x \in (50,100] \\ 0 & \text{otherwise} \end{cases}$$

At $x = 50$, $f(50) = C$, at $x = 100$, $f(100) = 0$. The slope is $m = \frac{0-C}{100-50} = -\frac{C}{50}$. So, $f(x) = C - \frac{C}{50} \cdot (x-50)$ for $x \in (50,100]$. Now,

$$f(x) = \begin{cases} C & \text{if } x \in [0,50] \\ C - \frac{C}{50} \cdot (x-50) & \text{if } x \in (50,100] \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)\ dx \\ &= \int_{-\infty}^{0} f(x)\ dx + \int_{0}^{50} f(x)\ dx + \int_{50}^{100} f(x)\ dx + \int_{100}^{\infty} f(x)\ dx \\ &= 0 + \int_{0}^{50} C\ dx + \int_{50}^{100} -\frac{C}{50}x + 2C\ dx + 0 \\ &= Cx\Big|_{x=0}^{x=50} + \left(-\frac{C}{100}x^2 + 2Cx\right)\Big|_{x=50}^{x=100} \\ &= (50C - 0C) + \left(\left(-\frac{10000C}{100} + 200C\right) - \left(-\frac{2500}{100}C + 100C\right)\right) \\ &= 50C - 100C + 200C - 75C \\ &= 75C \\ C &= \frac{1}{75} \end{aligned}$$

$$f(x) = \begin{cases} \frac{1}{75} & \text{if } x \in [0,50] \\ -\frac{1}{3750}x + \frac{2}{75} & \text{if } x \in (50,100] \\ 0 & \text{otherwise} \end{cases}$$

Continued on the next page.

Example : (Continued)

$$
\begin{aligned}
\Pr\left(40 \leq X \leq 50\right) &= \int_{40}^{50} f\left(x\right)\ dx \\
&= \int_{40}^{50} \frac{1}{75}\ dx \\
&= \frac{1}{75}x\Big|_{x=40}^{x=50} \\
&= \frac{50}{75} - \frac{40}{75} \\
&= \frac{10}{75} \\
&= \frac{2}{25} = 1.\overline{33}
\end{aligned}
$$