Theorem 4.1 (memoryless property) For $X \sim$ geometric(p) and any two nonnegative integers x and y,

$$P(X \geq x+y \,|\, X \geq x) = P(X \geq y).$$

Proof The conditional probability is

$$\begin{aligned}
P(X \geq x+y \,|\, X \geq x) &= \frac{P(X \geq x+y, X \geq x)}{P(X \geq x)} \\
&= \frac{P(X \geq x+y)}{P(X \geq x)} \\
&= \frac{(1-p)^{x+y}}{(1-p)^x} \\
&= (1-p)^y \\
&= P(X \geq y),
\end{aligned}$$

which proves the memoryless property. □

The memoryless property can be interpreted as follows. Consider a sequence of repeated, mutually independent, and identically distributed Bernoulli trials and a random variable X that is the number of failures before the first success. If you know that X is greater than or equal to x, then the distribution of the *remaining* number of Bernoulli trials before the first success has the same distribution as if the original x trials had never occurred. This interpretation is consistent with intuition. The previous history of the sequence of Bernoulli trials has no effect on the outcomes of future Bernoulli trials.

The memoryless property also has a geometric interpretation. Consider the probability mass function for a geometric(p) random variable from some value x to infinity. The sum of the mass values from x to infinity does not equal 1; rather it equals $(1-p)^x$. If each of the mass values from x to infinity is divided by $(1-p)^x$, then the resulting conditional probability mass function looks identical to the original geometric(p) probability mass function; it is just shifted to the right. No other discrete distribution has this property. For all other discrete probability distributions, the original (unconditional) probability mass function and the conditional probability mass function are not identical.

The moment generating function for a geometric(p) random variable X is

$$\begin{aligned}
M(t) &= E\left[e^{tX}\right] \\
&= \sum_{x=0}^{\infty} e^{tx} p(1-p)^x \\
&= p \sum_{x=0}^{\infty} \left(e^t(1-p)\right)^x \\
&= \frac{p}{1-(1-p)e^t}
\end{aligned}$$

for $(1-p)e^t < 1$ or $t < -\ln(1-p)$, which is required for the geometric series to converge. The moment generating function exists in a neighborhood about $t = 0$.

The population mean of a geometric(p) random variable can be found in three different ways. First, one can use the definition of the expected value:

$$\begin{aligned} E[X] &= \sum_{\mathcal{A}} x f(x) \\ &= \sum_{x=0}^{\infty} x p(1-p)^x \\ &= 1 \cdot p(1-p) + 2 \cdot p(1-p)^2 + 3 \cdot p(1-p)^3 + 4 \cdot p(1-p)^4 + \cdots . \end{aligned}$$

This non-trivial summation can be simplified by writing its terms in the following fashion:

$$\begin{array}{ccccccccccc} E[X] & = & p(1-p) & + & p(1-p)^2 & + & p(1-p)^3 & + & p(1-p)^4 & + & \cdots \\ & & & + & p(1-p)^2 & + & p(1-p)^3 & + & p(1-p)^4 & + & \cdots \\ & & & & & + & p(1-p)^3 & + & p(1-p)^4 & + & \cdots \\ & & & & & & & + & p(1-p)^4 & + & \cdots \\ & & & & & & & & & + & \ddots \end{array}$$

Each row is a geometric series with common multiplier $(1-p)$, so taking the row sums yields

$$E[X] = (1-p) + (1-p)^2 + (1-p)^3 + (1-p)^4 + \cdots$$

which is itself a geometric series with common multiplier $(1-p)$ that sums to

$$E[X] = \frac{1-p}{p}.$$

A second way to find the population mean of $X \sim$ geometric(p) is to use the moment generating function. The first derivative of the moment generating function with respect to t is

$$M'(t) = \frac{p(1-p)e^t}{\left(1-(1-p)e^t\right)^2}$$

for $|(1-p)e^t| < 1$. Using $t = 0$ as an argument yields the population mean

$$E[X] = M'(0) = \frac{1-p}{p}.$$

The third way to find the population mean of a geometric random variable is to use conditional expectation. This important topic and the derivation of the population mean of a geometric random variable will be presented in Chapter 6.

Using the moment generating function or conditioning, the population variance of a geometric(p) random variable X is

$$V[X] = E\left[(X-\mu)^2\right] = \frac{1-p}{p^2}.$$

The population skewness and kurtosis of a geometric(p) random variable X are

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{2-p}{\sqrt{1-p}} \qquad \text{and} \qquad E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{p^2-9p+9}{1-p}.$$

As $p \to 0$, the population skewness and kurtosis approach 2 and 9, respectively. As will be seen in the next chapter, these also happen to be the population skewness and kurtosis of a random variable having the *exponential* distribution.