2.6 The Geometric Distribution

The geometric distribution is similar to the binomial distribution in the sense that it concerns a collection of identical and independent trials, but from a different perspective. The geometric distribution is best motivated by the following example:

Example 2.6.1. Suppose you play the following game: You roll a six sided die repeatedly, until you see a six for the first time. Your score S is the number of rolls it took until you saw a six for the first time. We will compute $\mathbb{P}(S = i) = f_S(i)$ for some values of i.

Firstly we look at $\mathbb{P}(S = 1)$, in this case this means we rolled the dice once and it immediately showed a six. Therefore $\mathbb{P}(S = 1) = 1/6$.

Now suppose we look at $\mathbb{P}(S = 2)$. In this case the first dice roll has to not be a six, (as otherwise $S = 1$), while the second roll needs to be a six. Therefore we have that:

$$\mathbb{P}(S = 2) = \mathbb{P}(\{\text{First roll is not a six}\} \cap \{\text{The second roll is a six}\}).$$

These events are both independent, (knowing the outcome of one dice roll does not affect the other), so we may apply Lemma 1.7.2:

$$\mathbb{P}(S = 2) = \mathbb{P}(\{\text{First roll is not a six}\}) \times \mathbb{P}(\{\text{The second roll is a six}\}) = \frac{5}{6} \times \frac{1}{6} = \frac{5}{36}.$$

We can then continue this pattern on for larger values of i. By following the same approach above, show that $\mathbb{P}(S = 3) \approx 0.116$.

Essentially the geometric distribution describes a sequence of identical trials that are either success or fail, with some fixed probability p. Then the geometric random variable S is the number of trials we have to do, until we see a success for the first time.

Definition 2.6.1. Let $0 \leq p \leq 1$ be a probability. Then we say a random variable $X \sim \text{Geom}(p)$ if it has the following distribution function, for all $i \geq 1$:

$$\mathbb{P}(X = i) = p \times (1 - p)^{i-1}.$$

We remark that the distribution describes the process we undertook in Example 2.6.1. That is for a geometric variable X to be equal to i, it must have had precisely $i - 1$ failures, and then a success. The probability of success is given by p, hence the probability of failure is $1-p$. Therefore by Lemma 1.7.2, the probability of having $i-1$ failures in a row is $(1-p)^{i-1}$, and then the probability of a success is p. Hence $\mathbb{P}(X = i) = (1 - p)^{i-1}p$. As with other distributions we state its expectation:

Lemma 2.6.1. Suppose $X \sim \text{Geom}(p)$ is a geometric random variable with success probability p. Then we have that:

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

Example 2.6.2. Suppose a staff member is working on the front till of a retail store. Customers come to them one by one, and are looking to either make a purchase or a refund. On average 90% of customers make a purchase, while the remaining customers make a refund, independently of all other customers.

(i) What is the expected number of customers that the staff member will see until someone will ask for a refund.

(ii) What is the probability that the 11^{th} customer of the day will be the first to ask for a refund?

We start by defining a suitable random variable, let R be the number of customers seen until the first refund occurs. Then as 90% of customers make purchases, this means that the probability a given customer makes a refund is 0.1. Therefore $R \sim \text{Geom}(0.1)$. We now apply Lemma 2.6.1 to see that:

$$\mathbb{E}[R] = \frac{1}{0.1} = 10.$$

Therefore the staff member should expect the 10^{th} customer to be the first to ask for a refund.

For the second part we are looking for $\mathbb{P}(R = 11)$, therefore we apply the definition of geometric distribution (Definition 2.6.1), to see that:

$$\mathbb{P}(R = 11) = 0.1 \times (1 - 0.1)^{10} \approx 0.0387.$$