Data Science 1
Probability
Geometric Experiments

Edward L. Boone

Introduction

Items to remember:

Multiplication Rule: If E_1 can occur n_1 ways and E_2 can occur n_2 ways then the combined experiment can occur in $n_1 \times n_2$ ways.

Random Variables: How do I measure my outcomes and turn them into numbers?

$$X(s) : \mathcal{S} \to \mathbb{R}$$

Probability Distribution: How are probabilities assigned to the values of the random variable? $f(x) = P(X = x)$.

Bernoulli Experiment

A *Bernoulli* experiment is one where there are exactly two outcomes in the sample space.

We have already seen $\mathcal{S} = \{\text{Goal}, \text{No Goal}\}$.

Here are a few more:

- $\mathcal{S} = \{\text{On}, \text{Off}\}$
- $\mathcal{S} = \{\text{Alive}, \text{Dead}\}$
- $\mathcal{S} = \{\text{Pregnant}, \text{Not Pregnant}\}$
- $\mathcal{S} = \{\text{Success}, \text{Failure}\}$

Notice that there are only two possible outcomes and there is no "in between" on any of the states.

Geometric Experiment

The Geometric Experiment is very similar to the Binomial Experiment in that it is composed of independent *Bernoulli Trials*.

However, we are interested in the number of failures before the first success.

Here the sample space is easy to write:

$$\mathcal{S} = \{S, FS, FFS, FFFS, FFFFS, FFFFFS, FFFFFFS, ...\}$$

This makes the random variable easy as well:

$$X(s) = 0, 1, 2, 3, 4, 5, 6...$$

Geometric Distribution

Since the Geometric random variable is easy to define so is the probability distribution.

Let p be the probability of success (and hence $1-p$ be the probability of failure).

- $X=0$ corresponds to the outcome S.

$$P(S)=p$$

- $X=1$ corresponds to the outcome FS.

$$P(FS)=P(F)P(S)=(1-p)p$$

- $X=2$ corresponds to the outcome FFS.

$$P(FFS)=P(F)P(F)P(S)=(1-p)(1-p)p=(1-p)^2p$$

Geometric Distribution

By continuing this same logic we obtain the following probability distribution:

$$f(x) = P(X = x) = (1 - p)^x p, \quad x = 0, 1, 2, 3, 4, \ldots$$

Geometric Distribution

By continuing this same logic we obtain the following probability distribution:

$$f(x) = P(X = x) = (1 - p)^x p, \quad x = 0, 1, 2, 3, 4, \ldots$$

Example:

Suppose we know that Hassan loves basketball and when he shoots to score he is successful 14.3% of the time. What is the probability that it will take Hassan exactly 6 shots to score a goal?

Here $x = 6 - 1 = 5$ and $p = 0.143$ so...

$$\begin{aligned} f(5) = P(X = 5) = (1 - p)^x p &= (1 - p)^5 p \\ &= (1 - 0.143)^5 0.143 \\ &= 0.857^5 0.143 \approx 0.0661 \end{aligned}$$

Example

Asma is a real estate agent who is interested in selling residential properties. When she shows a property there appears to be a 19% chance that an offer will be made on the property. What is probability she will get an offer on the fourth showing?

Here $x = 4 = 3$ and $p = 0.19$ so...

$$\begin{aligned} f(3) = P(X = 3) = (1-p)^x p &= (1-p)^3 p \\ &= (1-0.19)^3 0.19 \\ &= 0.81^3 0.19 \approx 0.1009 \end{aligned}$$

Summary

The Geometric experiment and its associated random variable and probability distribution are simple but extremely important in understanding probability.

- It allows for the replication of the Bernoulli experiment.
- The sample space is quite large (countably infinite actually).
- The random variable is simply the number of failures until a success.
- The probability distribution can be used to find probabilities in this complex sample space.