Data Science 1
Probability
Binomial Experiments
Part 1

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.

Combinations: How many ways can I pull k objects from n distinct objects:

$$_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

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

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

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.

Binomial Experiment

A *Binomial* experiment is where there are n *independent* trials of a Bernoulli experiment.

Here *independent* means that the outcome of one of the Bernoulli experiments has no influence on any of the other Bernoulli experiments.

Replicating experiments is always a good idea. This is just taking a Bernoulli experiment and replicating it n times.

Binomial Example

Suppose our Bernoulli experiment is $\mathcal{S} = \{\text{Goal}, \text{No Goal}\}$. and we want to replicate 3 goal attempts in a way that one does not influence the other.

Then our sample space for the Binomial experiment will be:

$$\mathcal{S} = \left\{ \begin{array}{cc} GGG & GGN \\ GNG & NGG \\ NGN & NNG \\ GNN & NNN \end{array} \right\}$$

Notice we have $2^3 = 8$ outcomes.

Now that we have the sample space we need the random variable...

Bernoulli Random Variable

The Binomial Random Variable, X, will total the number of goals (G) in each outcome

$$\mathcal{S} = \left\{ \begin{array}{cc} GGG & GGN \\ GNG & NGG \\ NGN & NNG \\ GNN & NNN \end{array} \right\}$$

$$X(s) = \text{ number of Gs in three trials}$$

- $X(NNN) = 0$ since no goals were observed.
- $X(NNG) = X(NGN) = X(GNN) = 1$ since one goal was observed.
- $X(NGG) = X(GNG) = X(GGN) = 2$ since two goals were observed.
- $X(GGG) = 3$ since three goals were observed.

Bernoulli Random Variable

Now we need to determine a probability distribution for X.

If we use the equally likely approach then we simply count up the number of outcomes that match the random variable and divide by the total number of outcomes.

- $f(0) = P(X = 0) = \frac{1}{8}$ since there is one outcome of 8 that has no goals.
- $f(1) = P(X = 1) = \frac{3}{8}$ since there are three of the 8 outcomes that have only one goal.
- $f(2) = P(X = 2) = \frac{3}{8}$ since there are three of the 8 outcomes have exactly two goals.
- $f(3) = P(X = 3) = \frac{1}{8}$ since only one outcome from the 8 has all goals.

Note: Technically we would put $f(x) = 0$, otherwise to account for the fact that any other value of X would have a probability of 0 as there are no outcomes that would match it.

Bernoulli Probability Distribution

We can rewrite this to clean it up a bit.

$$f(x) = P(X = x) = \begin{cases} \frac{1}{8} & \text{if } x = 0 \\ \frac{3}{8} & \text{if } x = 1 \\ \frac{3}{8} & \text{if } x = 2 \\ \frac{1}{8} & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

We will learn to simplify this even more in the next video.

Summary

The Binomial 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. If there are n trials then the number of outcomes is 2^n which can grow large very quickly.
- The random variable is simply the number of successes the outcome has.
- The probability distribution is a bit more complex and requires us to examine each outcome (for now).