

Data Science 1

Probability

Binomial Experiments

Part 2

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:

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

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 $\mathcal{S} = \{\text{Goal, No Goal}\}$.

Here are a few more:

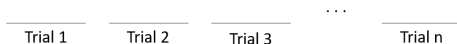
- $\mathcal{S} = \{\text{On, Off}\}$
- $\mathcal{S} = \{\text{Alive, Dead}\}$
- $\mathcal{S} = \{\text{Pregnant, Not Pregnant}\}$
- $\mathcal{S} = \{\text{Success, Failure}\}$

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

Binomial Experiment

Let's rethink the Binomial Experiment...

Since there are n trials that are independent each conducted so the probability of success is the same $P(S) = p$ and $P(F) = 1 - p$ then we should be able to rethink this.



By adding outcomes to each spaces



Group all the Successes together and all the Failures together



Binomial Experiment

Group all the Successes together and all the Failures together

$$\underline{S} \quad \underline{S} \quad \dots \quad \underline{S} \quad \underline{F} \quad \underline{F} \quad \dots \quad \underline{F}$$

Put in the probabilities of each Bernoulli outcome.

$$\underline{P(S)} \quad \underline{P(S)} \quad \dots \quad \underline{P(S)} \quad \underline{P(F)} \quad \underline{P(F)} \quad \dots \quad \underline{P(F)}$$

Include the fact that x is the number of success, so $n - x$ is the number of failures

$$\underbrace{\underline{P(S)} \quad \underline{P(S)} \quad \dots \quad \underline{P(S)}}_x \quad \underbrace{\underline{P(F)} \quad \underline{P(F)} \quad \dots \quad \underline{P(F)}}_{n-x}$$

Binomial Experiment

Include the fact that x is the number of success, so $n - x$ is the number of failures

$$\underbrace{\frac{P(S)}{\quad} \frac{P(S)}{\quad} \dots \frac{P(S)}{\quad}}_x \underbrace{\frac{P(F)}{\quad} \frac{P(F)}{\quad} \dots \frac{P(F)}{\quad}}_{n-x}$$

Simplify

$$P(S)^x P(F)^{n-x}$$

Now include the fact that $P(S) = p$ and $P(F) = 1 - p$.

$$p^x (1 - p)^{n-x}$$

And we are almost there...

Binomial Probability Distribution

Now include the fact that $P(S) = p$ and $P(F) = 1 - p$.

$$p^x(1 - p)^{n-x}$$

We just need to include the fact that we have *choose* where the x successes are from the n slots. This gives us:

$$f(x) = P(X = x) = \binom{n}{x} p^x(1 - p)^{n-x}$$

The above is the probability distribution for X the number of successes from n Bernoulli trials.

Notice that we didn't have to look at each outcome! And this can transfer to any Binomial Experiment.

Example

Go back to the goal example where X is the number of goals from three attempts. Before we used the equally likelihood approach which means $p = 1/2$ and three attempts gives $n = 3$. What is the probability we make 1 goal?

$$\begin{aligned}P(X = 1) = f(1) &= \binom{n}{x} p^x (1 - p)^{n-x} \\&= \binom{3}{1} \left(\frac{1}{2}\right)^1 \left(1 - \frac{1}{2}\right)^{3-1} \\&= 3 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)^2 \\&= 3 \left(\frac{1}{2}\right)^3 \\&= \frac{3}{8}\end{aligned}$$

Bernoulli Probability Distribution

Go back to the goal example where X is the number of goals from three attempts. What if equally likely is not reasonable and $p = 1/3$ and three attempts gives $n = 3$. What is the probability we make 1 goal?

$$\begin{aligned}P(X = 1) = f(1) &= \binom{n}{x} p^x (1 - p)^{n-x} \\&= \binom{3}{1} \left(\frac{1}{3}\right)^1 \left(1 - \frac{1}{3}\right)^{3-1} \\&= 3 \left(\frac{1}{3}\right) \left(\frac{2}{3}\right)^2 \\&= 3 \left(\frac{1}{3}\right) \left(\frac{4}{9}\right) \\&= \frac{12}{27} = \frac{4}{9}\end{aligned}$$

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 can be used to find probabilities in this complex sample space.