Data Science 1
Probability
Negative-Binomial 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.

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?

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

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.

Negative-Binomial Experiment

The Negative-Binomial Experiment is very similar to the Geometric Experiment except we are interested in X being the number of failures of of independent *Bernoulli Trials* before the k^{th} success However, we are interested in X the number of failures before the k^{th} success.

Here the sample space is not as easy to write since k is involved: Suppose $k = 3$

$$\mathcal{S} = \{SSS, FSSS, SFSS, SSFS, FFSSS, FSFSS, FSSFS, ...\}$$

This makes the random variable easy as well:

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

However we need to account for the different number of ways the Failures and Successes could present themselves.

Negative-Binomial Distribution

Since the Negative-Binomial random variable has the following probability distribution.

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

$$f(x) = P(X = x) = \binom{x+k-1}{x}(1-p)^x p^k, \quad x = 0, 1, 2, 3, \ldots$$

- There are $x+k-1$ places to put the x failures since the last place has to be a success.
- Hence the $\binom{x+k-1}{x}$ accounts for the places.
- There are several formulations of this distribution so be careful!

Example

Fatemeh is taking a proficiency exam on computer. The computer program is designed to keep offering questions until the student misses 5 questions. Given that for each question has 4 choices and the questions are independent, what is the probability that Fatemeh will answer exactly 9 questions if she guesses at each answer.

Here $x = 4$, $k = 5$ and $p = 0.25$ so...

$$
\begin{aligned}
f(4) = P(X = 4) &= \binom{x+k-1}{x}(1-p)^x p^k \\
&= \binom{4+5-1}{4}(1-0.25)^4 0.25^5 \\
&= \binom{8}{4}(0.75)^4 0.25^5 \\
&= 70(0.75)^4 0.25^5 \approx 0.0216
\end{aligned}
$$

Summary

The Negative-Binomial experiment and its associated random variable and probability distribution are a little more complex 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.