

The Geometric and Negative Binomial Distributions - Lab

Introduction

In this lab, you'll practice your knowledge on the geometric and negative binomial distributions!

Objectives

You will be able to:

- Understand and explain the Negative Binomial Distribution and its uses
- Understand and explain Geometric Distributions and their uses

Quick Recap

Recall that the binomial distribution describes the probability of a success or a failure outcome in an experiment that is repeated multiple times.

In fact, a binomial distribution describes a repeated bernoulli experiment. A bernoulli experiment can be seen as one trial where there is a known success rate p . Examples:

- Rolling a dice once where success is defined as throwing a 5 or higher, the probability of success is 1/3.
- Shooting at a basketball rink where the success probability is 70%.
- etc.

The binomial distribution then has 2 parameters: n and p , where n is the number of independent experiments.

The binomial and hypergeometric distribution describe the number of successes in a fixed number, independent repetitions of Bernoulli experiments. The negative binomial and geometric distribution describe the number of independent repetitions to a *fixed* number of *successes*. The geometric distribution is a special case of the negative-binomial distribution.

To be more specific, the geometric distribution is the probability for which trial the first successful Bernoulli experiment will occur on. The more general version of this is the negative binomial distribution which is the probability for which trial the n th success will occur on.

1. Let's take a look at the classic coin flipping case.

A) What's the probability that the first success occurs on the 1st trial? (Yes its that intuitive.)

0.5

B) What's the probability that the first success occurs on the 2nd trial?

(Hint: Think of all the possible scenarios: Either the success occurred on the first trial, the success occurred on the second trial, or there still was no success. These three scenarios should encompass all possible scenarios and thus have a total probability of 1. Calculate the probability that the first trial was successful. Calculate the probability that neither the first nor second trial was successful. It should now be straightforward to calculate the final scenario: that the second trial was the original success.)

```
In [ ]: #Probability success 1st trial = 0.5
#Probability of no success on 1st and 2nd trial: (0.5)(0.5)= 0.25
#Since Probability_Success_Trial1 + Probability_No_success_Trial_1_and_2 + Probability_Success_Trial_2 = 1:
# 0.5 + 0.25 + Probability_Success_Trial_2 = 1
Probability_Success_Trial_2 = 0.25
```

C) What's the probability that the first success occurs on the 3rd trial? The 5th?

```
In [ ]: #Better thought of as probabilities of all failures up to that point times probability of success on that particular trial.
#3rd_trial = (0.5)**3 = .125
#5th_trial = (.5)**5
```

2. Geometric Function

Now write a probability distribution function for a random variable y . The function should take in the probability, p of the success of an individual Bernoulli experiment. (In our previous coin flipping example, $p=0.5$.)

```
In [1]: def geometric_dist(y,p):
    """y is a discrete random variable. It should be an integer that is greater then zero.
    p is the probability of a success for the Bernoulli experiment to be conducted.
    This function should return the probability that the first successful Bernoulli experiment will occur on the yth trial."""
    #This outline could further help students.
    probability_failures_all_previous = (1-p)**(y-1)
    probability_success_this_trial = p
    overall_prob = probability_failures_all_previous * probability_success_this_trial
    prob = overall_prob#The probability that the first successful bernoulli experiment occurs on the yth trial.
    return prob
```

3. Product Failures

Assume that the probability of a product working is 95%. Before shipping the products, the manufacturer is checking each product for defects. What is the probability that the 10th product checked is the first defective one found?

```
In [2]: #Code and answer here  
prob = geometric_dist(10, .95)  
print(prob)
```

1.8554687500000146e-12

4. Product Failures take 2

In many cases, a manufacturer might only test a sample of the products for defaults. Assuming 95% of the products do indeed work, what is the probability that in testing 20 units, that none will be defective?

```
In [3]: #This is easier to solve simply with binomial distrubution. Important for students to see that.  
.95**20
```

Out[3]: 0.3584859224085419

```
In [6]: #Forced use of geometric function: cumulative complements to success happening on any of first 20 iterations.  
prob_no_failures = 1  
for i in range(1,21):  
    prob_first_failure_now = geometric_dist(i, .05)  
    prob_no_failures -= prob_first_failure_now  
print(prob_no_failures)
```

0.35848592240854255

5. Consumer Profiling

A previous sample showed that 70% of U.S. shoppers prefer to buy groceries in store as compared to online. Calculate the probability that the 6th person interviewed is the first to prefer to buy groceries online.

```
In [8]: #Code and answer here  
prob = geometric_dist(6, .7)  
print(prob)
```

0.0017010000000000011

6. Consumer Profiling 2

What is the probability that at least 6 people have to be interviewed before finding someone who prefers to buy groceries online? (Assuming the statistic is true.)

```
In [9]: #Essentially same problem as 4 but new context and worded differently.  
#Asking for probability that first 5 all prefer to buy groceries online  
prob = .7**5  
print(prob)
```

0.16806999999999994

```
In [11]: #Another, more roundabout solution:  
prob_no_online_groceries = 1  
for i in range(1,6):  
    prob_first_online_grocery_now = geometric_dist(i, .3)  
    prob_no_online_groceries -= prob_first_online_grocery_now  
print(prob_no_online_groceries)
```

0.16807000000000005

Summary

Awesome, you've now learned about many use cases for the geometric and negative binomial distribution!