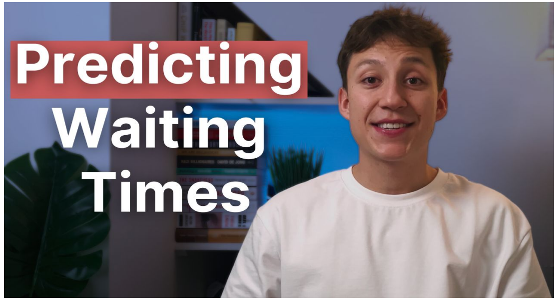
Geometric Distribution

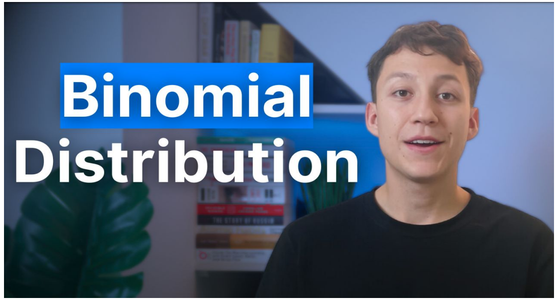
Egor Howell

Intuition

The Geometric distribution is a discrete probability distribution that infers the probability of the number of Bernoulli trials we need before we get a success. A Bernoulli trial is when an individual event has only two outcomes: success or failure with a certain fixed probability.

The Geometric distribution is often referred to as the discrete version of the Exponential distribution.





Intuition

There are actually two different types of the Geometric distribution:

- The first is the number of trials required to get a success.
- The second is the number of failures before the success.

The first one is referred to as the shifted Geometric distribution.

In reality, either can be used but the distinction just needs to be clear from the outset to ensure consistency of results.

We will use the shifted version as I feel like it it easier to work with mathematically and intuitively.

Mathematical Detail

The probability mass function looks like:

$$P(X = n) = (1 - p)^{n-1}\, p$$

Equation generated by author in LaTeX.

Where ***p*** is the probability of success and ***n*** is the number of events it took to get the success.

The expected value is given by:

$$E[X] = \frac{1}{p}$$

Equation generated by author in LaTeX.

Example

What is the probability of rolling a ***4*** on a regular ***6-sided*** die on the ***5th*** roll?

- ***n= 6***
- ***p = 1/6***

$$P(X = 5) = \left(1 - \frac{1}{6}\right)^{5-1} \frac{1}{6} = 0.08036$$

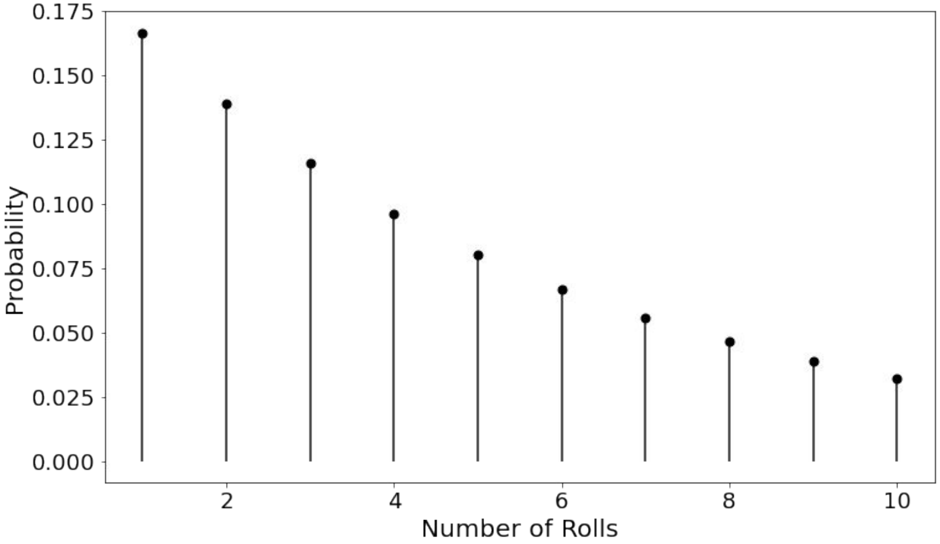
Equation generated by author in LaTeX.

The expected value is:

$$E[X] = 6$$

Equation generated by author in LaTeX.

Plots




```python
# Import packages
from scipy.stats import geom
import matplotlib.pyplot as plt

# Probability and Number of Trials
n = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
p = 1/6

# Generate the PMF
dist = geom.pmf(X, p)

# Plot the distribution
plt.figure(figsize=(12, 7))
plt.scatter(n, dist, linewidth=2, color='black')
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)
plt.ylabel('Probability', fontsize=20)
plt.xlabel('Number of Rolls', fontsize=20)
plt.vlines(n, 0, dist, colors='black', linewidth=2, alpha=0.7)
plt.savefig('plot.png')
plt.show()
```

We observe that the probability of rolling a 4 exponentially decreases as the number of rolls increases. This makes sense, as it is very unlikely that our first 4 will happen on the 100th roll.

Thanks

Geometric Distribution Simply Explained

A simple description and uses of the Geometric distribution

Egor Howell
Published in Towards Data Science · 4 min read · Aug 9, 2022

146

Photo by Moritz Kindler on Unsplash

@egorhowell

Newsletter

Dishing The Data