

# Probability distributions: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2019

## Syntax

- Using the probability mass function from SciPy:

```
from scipy import linspace
from scipy.stats import binom
outcome_counts = linspace(0,30,31)
dist = binom.pmf(outcome_counts,30,0.39)
```

- Using the probability mass function from SciPy:

```
from scipy import linspace
from scipy.stats import binom
outcome_counts = linspace(0,30,31)
dist = binom.cdf(outcome_counts,30,0.39)
```

- Getting the summ of all the probabilities to the left of  $k$ , including  $k$ :

```
left = binom.cdf(k,N,p)
```

- Getting the sum of all the probabilities to the right of  $k$ :

```
right = 1 - left
```

## Concepts

- Binomial probabilities are the chance of a certain outcome happening in a sequence.
- One way to visualize binomials is a binomial distribution. Given  $N$  events, it plots the probabilities of getting different numbers of successful outcomes. The binomial distribution parameters are:
  - $N$ : The total number of events.
  - $p$ : The probability of the outcome we're interested in seeing.
- Formula for binomial probability:

- The probability mass function (pmf) gives us the probability of each occurring, and takes in the following parameters:
  - $x$  : The list of outcomes.
  - $n$  : The total number of events.
  - $p$  : The probability of the outcome we're interested in seeing.
- A probability distribution can only tell us which values are likely, and how likely they are.
- We can calculate the expected probability of a probability distribution using  $\sum x_i p_i$ , where  $n$  is the total number of events, and  $p$  is the probability of the outcome we're interested in seeing.
- The formula for standard deviation, or a measure of how much the values vary from the mean, of a probability distribution is  $\sqrt{\sum x_i^2 p_i - (\sum x_i p_i)^2}$  where  $n$  is the total number of events,  $p$  is the probability of the outcome we're interested in seeing, and  $1-p$  is the probability of the outcome not happening.
- The cumulative density function is the probability that  $x$  or less events will occur.
- The z-score is the number of standard deviations away from the mean, and used to find the percentage of values to the left or right.
- We can calculate the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) using the following formulas:
  - $\mu = \sum x_i p_i$
  - $\sigma = \sqrt{\sum x_i^2 p_i - (\sum x_i p_i)^2}$
- We can figure out the z-score of a value using the following formula:
  - $z\text{-score} = \frac{x - \mu}{\sigma}$

## Resources

- [Probability Mass Function](#)
- [SciPy Documentation](#)
- [Documentation for scipy.stats.binom](#)
- [Cumulative Density Function](#)

