# Statistics in Python

David Arroyo Menéndez

February 7, 2019

# Statistics: Summary

- Descriptives
- Manipulating Data
- Matplotlib
- Distributions
- Statistics Tests
- Logistic Regression
- Principal Component Analysis

# Descriptives

## Source!

```
$ python3 scipy-descriptives.py
```

# Manipulating Data

Pandas is for dataframes

```
$ python3 pandas/pandas-10min.py
$ python3 pandas/creating-dataframe.py
$ python3 pandas/creating-dataframa-from-arrays.py
$ python3 pandas/manipulating-data.py
$ python3 pandas/remove-rows-with-nan.py
$ python3 pandas/handle-missing-data.py
$ python3 pandas/data-analysis/pd-diabetes.py
```

Numpy is algebra is for arrays

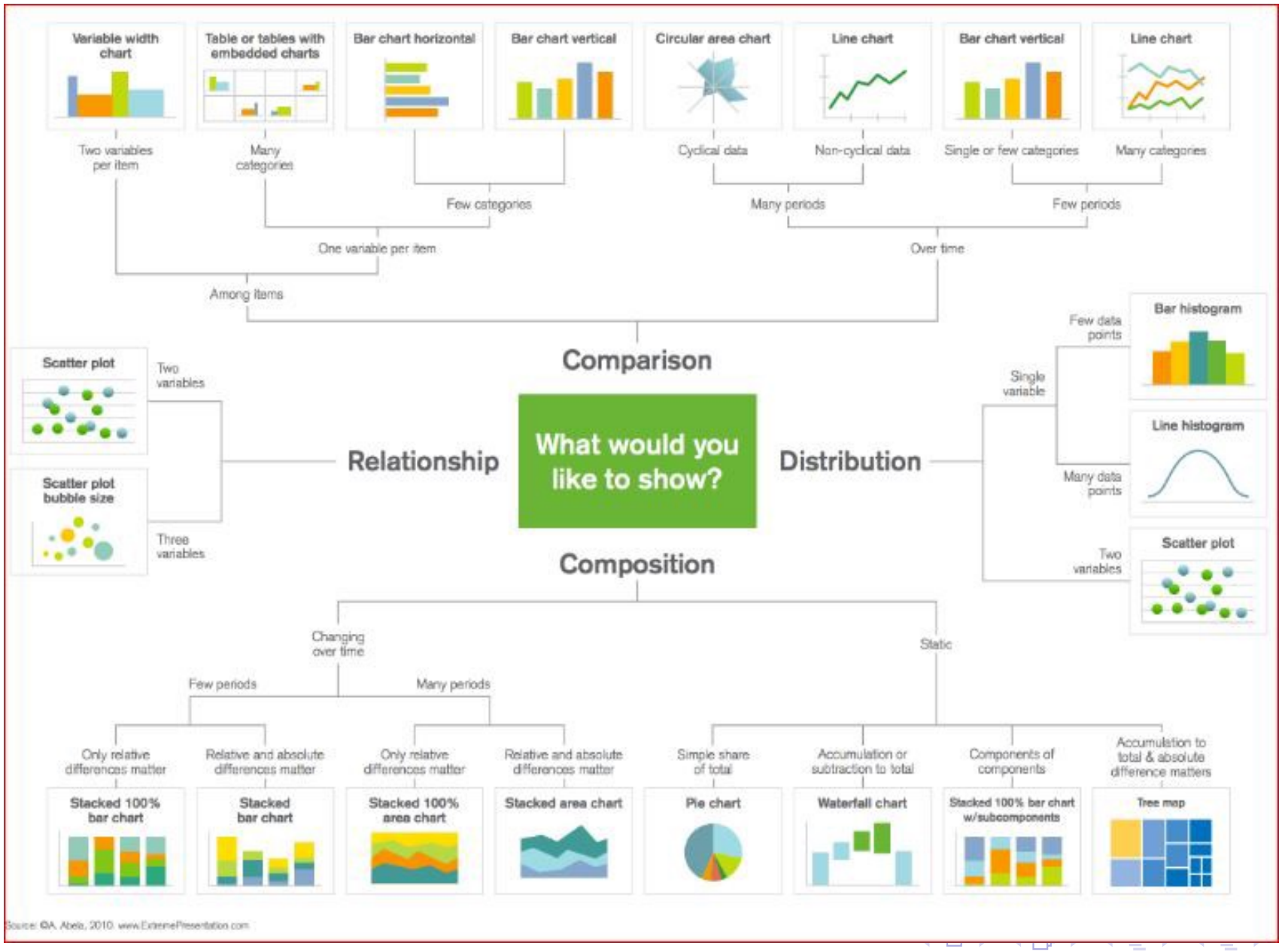
```
$ python3 numpy/reject-outliers.py
```

# Matplotlib

You can display statistics with matplotlib

```
$ python3 barchart_demo.py
$ python3 boxplot-example2.py
$ python3 boxplot-example.py
$ python3 colorbar_basics.py
$ python3 image_demo.py
$ python3 pie_features.py
$ python3 plot_3D.py
$ gimp surface3d_frontpage.png &
$ python3 pyplot_text.py
$ python3 scatter-example.py
$ python3 stackplot_demo.py
$ python3 subplot.py
$ python3 unicode_minus.py
```

# Matplotlib. What would you like to show?

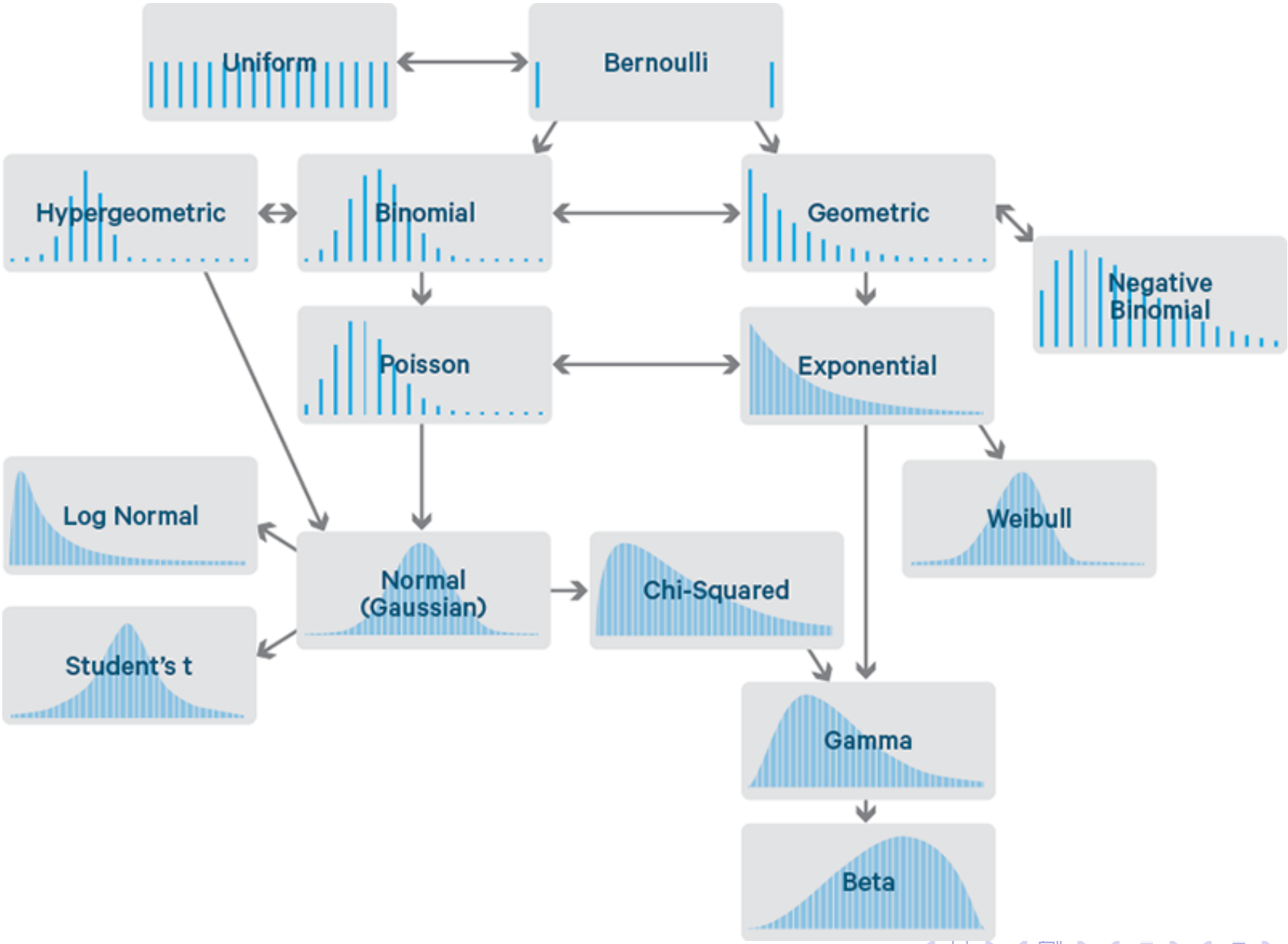


# Distibutions in Statistics

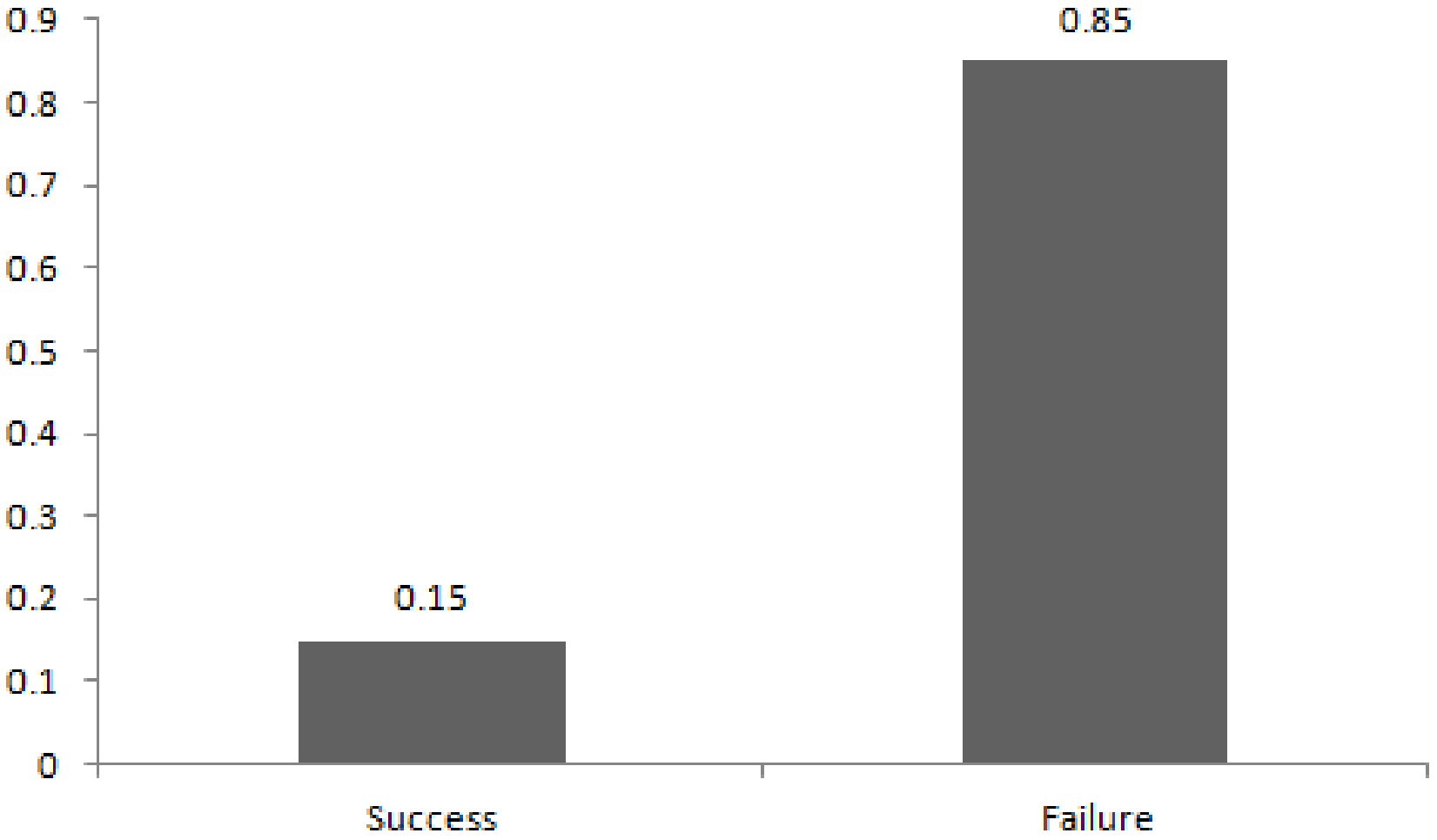
An histogram trends to be a continuos line in a table, we can draw a distribution with this trend.

```
$ python3 bernoulli.py
$ python3 plot_normal.py
$ python3 poisson.py
$ python3 binomial.py
$ python3 exponential-distribution.py
```

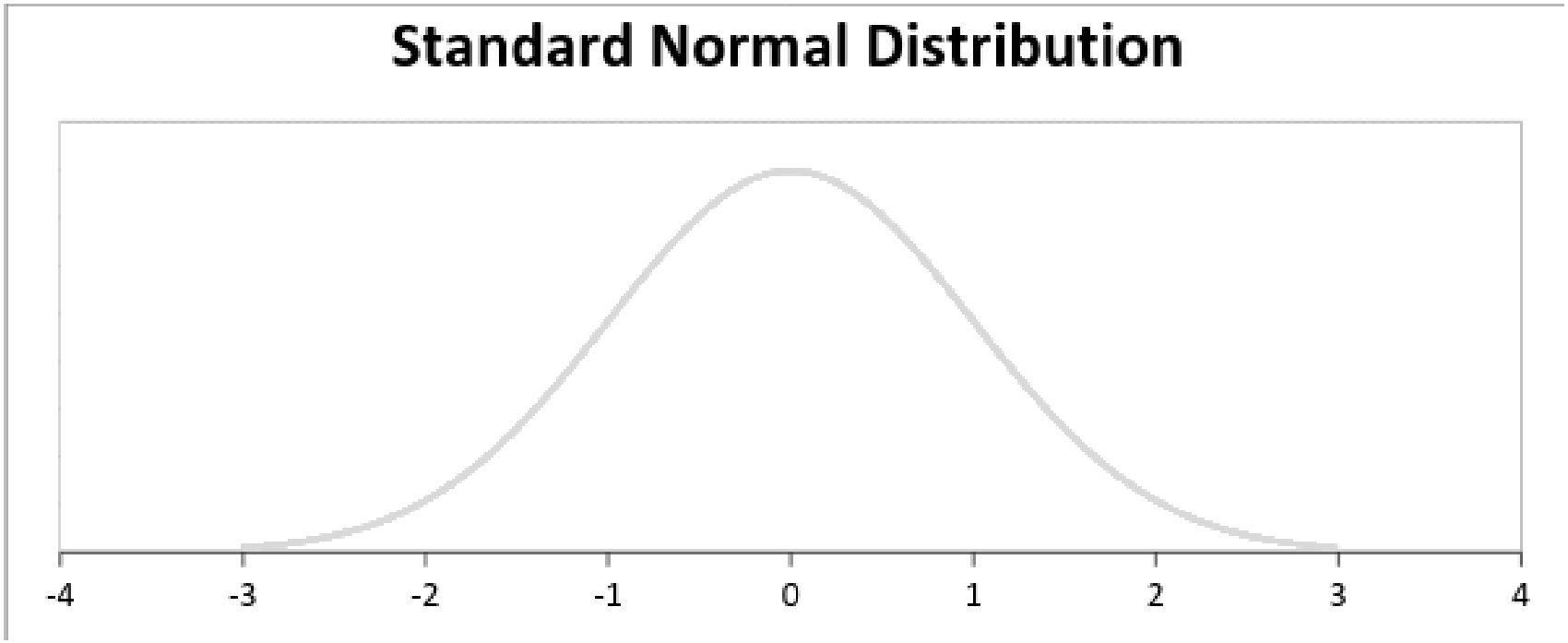
# Distributions in Statistics (II)



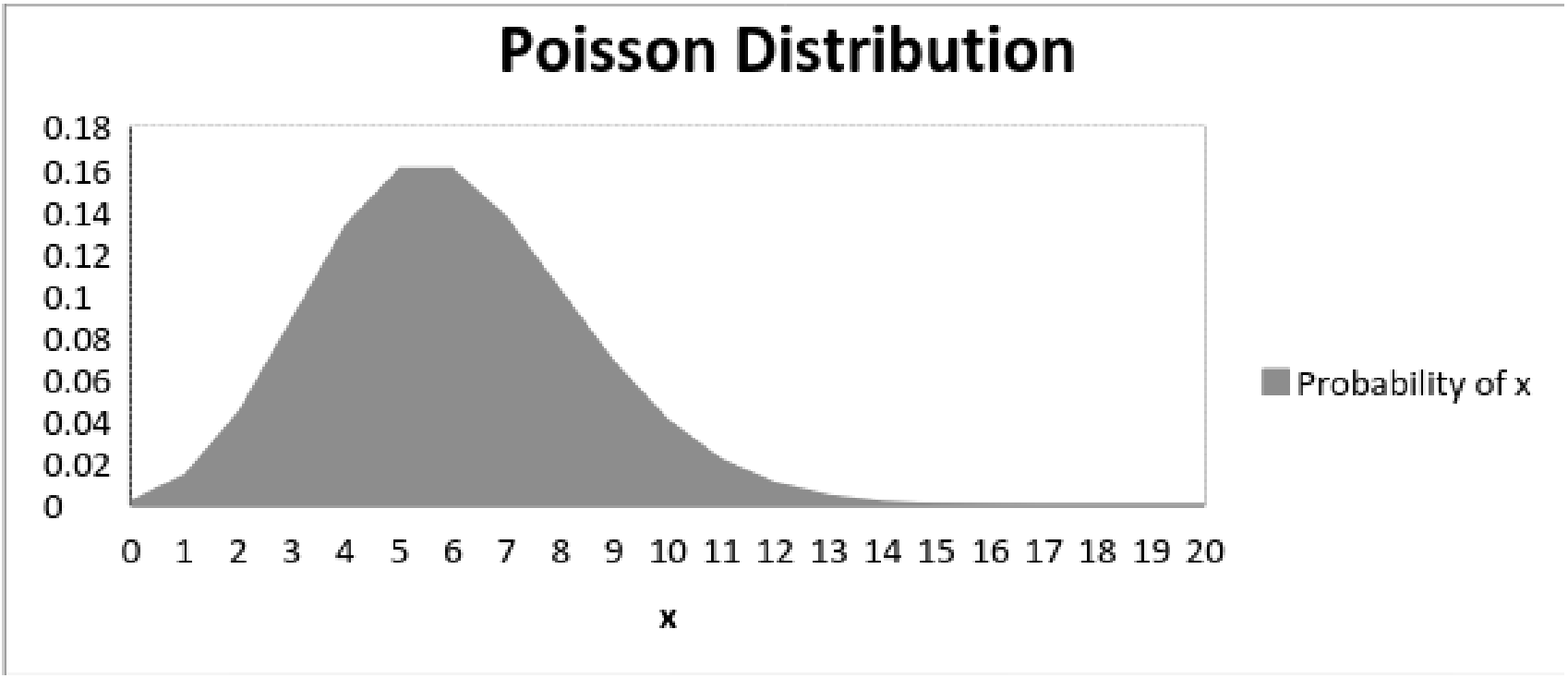
# Bernoulli Distribution



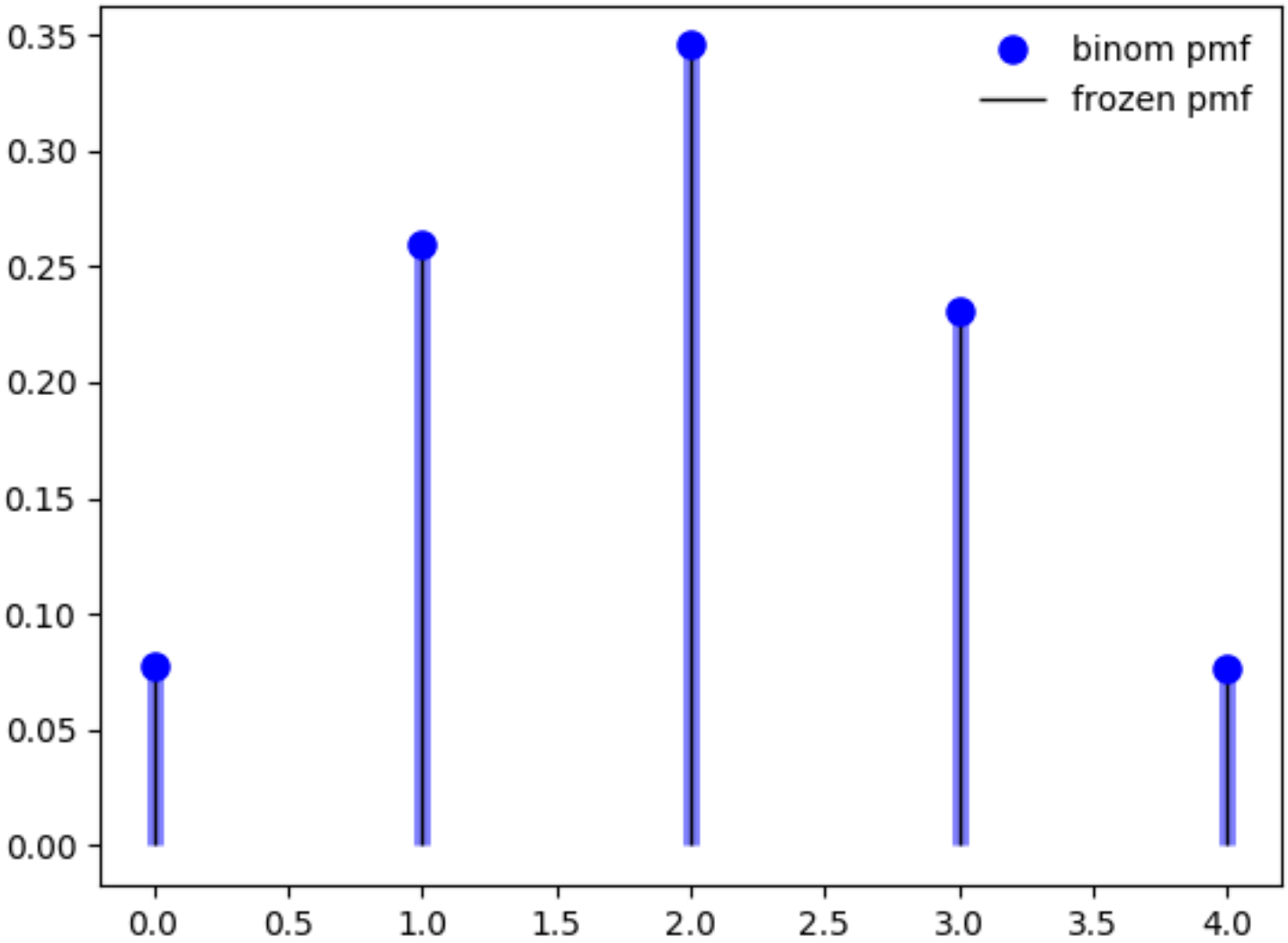
# Normal Distribution



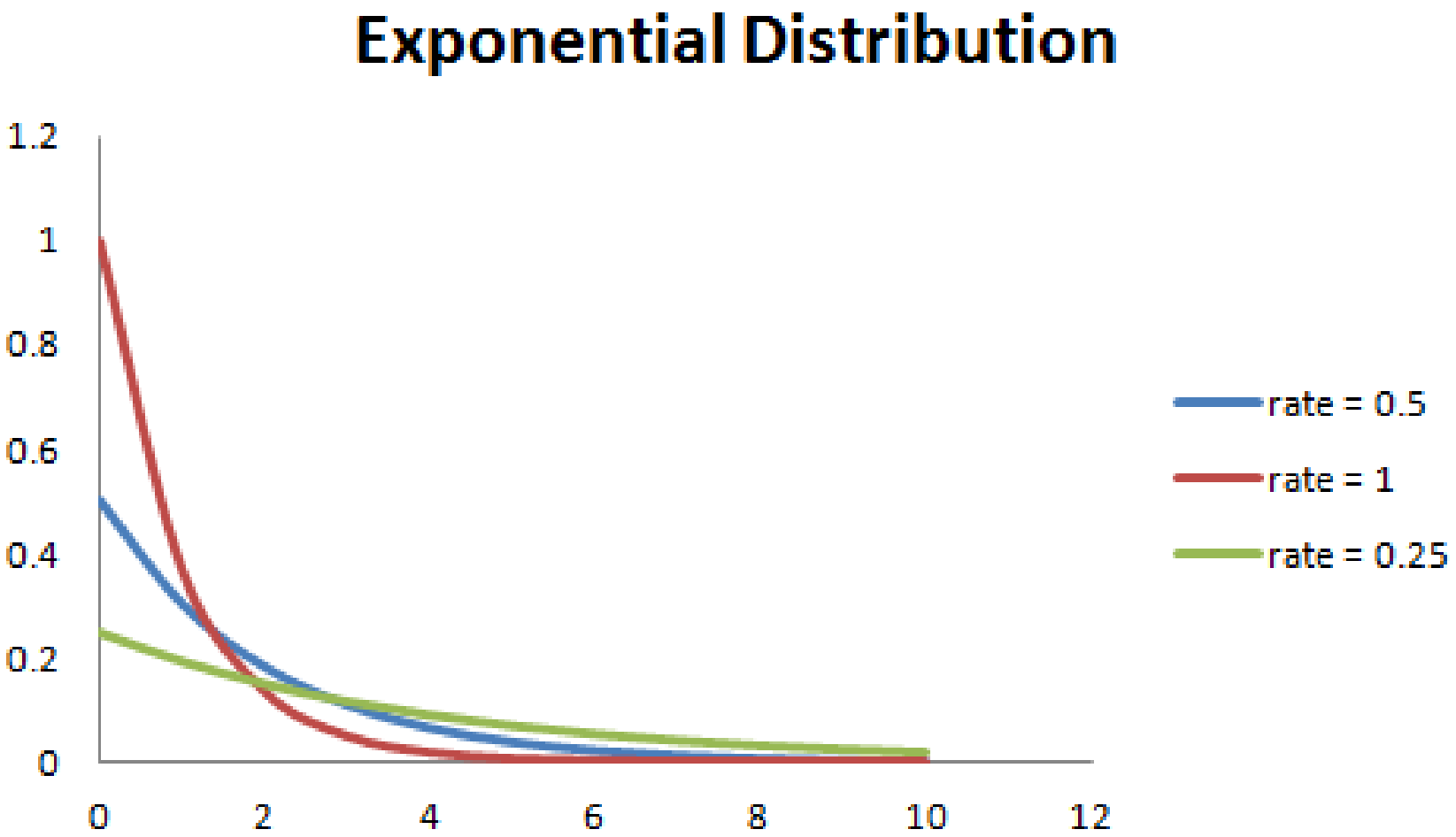
# Poisson Distribution



# Binomial Distribution



# Exponential Distribution



# Moments in a Distribution

| Moment number | Name | Measure of | Formula |
|---|---|---|---|
| 1 | Mean | Central tendency | $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$ |
| 2 | Variance (Volatility) | Dispersion | $\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$ |
| 3 | Skewness | Symmetry (Positive or Negative) | $Skew = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{(X_i - \bar{X})}{\sigma} \right]^3$ |
| 4 | Kurtosis | Shape (Tall or flat) | $Kurt = \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{(X_i - \bar{X})}{\sigma} \right]^4$ |

Where X is a random variable having N observations (i = 1,2,…,N).

# Statistics Tests / Choice Models

To see a result from a hipothesis you can use tests:

```
$ python scipy-special-tests.py
$ python discrete-choice-models.py
$ python pearson.py # for testing non-correlation
$ python fisher.py
```

# Multivariate Statistics. Choosing a model

| Dependent Variable | Explained Variable | Parametric |
| --- | --- | --- |
| 1 quantitative variable | One variable cualitative with two levels | ANOVA o |

# Logistic Regression

## Scikit is your friend

```
$ python3 scikit/logistic-regression/logistic-function.py
$ python3 scikit/logistic-regression/data-using-pandas.py
```

# Principal Component Analysis

## Scikit is your friend

```
$ python3 scikit/logistic-regression/logistic-function.py
$ python3 scikit/logistic-regression/data-using-pandas.py
```

# Montecarlo Statistics Methods

It's a statistic game where the players is betting.

```
$ python3 statistics/montecarlo/bettor.py
$ python3 statistics/montecarlo/doublebettor.py
$ python3 statistics/montecarlo/bettor-statistics.py
$ python3 statistics/montecarlo/dalambert.py
```