Top of Form

Bottom of Form

**[Python for data scientists IV: intermediate statistics in Python](https://courses.thinkful.com/dsbc-python-4-v1/course/0)**

**[Basics of probability](https://courses.thinkful.com/dsbc-python-4-v1/checkpoint/1)**

**[Data distributions](https://courses.thinkful.com/dsbc-python-4-v1/checkpoint/2)**

**[Central limit theorem](https://courses.thinkful.com/dsbc-python-4-v1/checkpoint/3)**

[Checkpoint 3](https://courses.thinkful.com/dsbc-python-4-v1/checkpoint/3)

**Central limit theorem**

This checkpoint serves as a bridge between descriptive statistics and inferential. While the former can give vital information about the dataset at hand, the latter provides an almost-magical ability to infer estimates of an entire population given just a sample.

Conducting tests of inferential statistics and applying them to real-world scenarios is the focus of the next module. For now, we get to see what's powering these statistics "under the hood" with the law of large numbers and the central limit theorem.

For each, we'll cover the basic theory along with a demonstration in Python where we'll simulate tossing the roll of a die hundreds of times.

Chapter 3

**Probability Distributions**

A probability distribution specifies the relative likelihoods of all possible outcomes.

Random Variables

Formally, a random variable is a function that assigns a real number to each outcome in the probability space. Define your own discrete random variable for the uniform probability space on the right and sample to find the empirical distribution.

Click and drag to select sections of the probability space, choose a real number value, then press "Submit."

Top of Form



Bottom of Form

|  |  |
| --- | --- |
| Color | Value |
|  | 0 |
|  | 500 |

Sample from probability space to generate the empirical distribution of your random variable.

00.00.51.0

Sample Distribution

Reset

Discrete and Continuous

There are two major classes of probability distributions.

Discrete Continuous

A discrete random variable has a finite or countable number of possible values.

If XX is a discrete random variable, then there exists unique nonnegative functions, f(x)f(x) and F(x)F(x), such that the following are true:

P(X=x)P(X<x)=f(x)=F(x)P(X=x)=f(x)P(X<x)=F(x)

Choose one of the following major discrete distributions to visualize. The probability mass function f(x)f(x) is shown in **yellow** and the cumulative distribution function F(x)F(x) in **orange** (controlled by the slider).



Central Limit Theorem

The Central Limit Theorem (CLT) states that the sample mean of a sufficiently large number of i.i.d. random variables is approximately normally distributed. The larger the sample, the better the approximation.

Change the parameters αα and ββ to change the distribution from which to sample.

αα = 1.00   
ββ = 1.00

Choose the sample size and how many sample means should be computed (draw number), then press "Sample." Check the box to display the true distribution of the sample mean.

Sample size = 1   
Draws = 1

Theoretical

Sample

This visualization was adapted from Philipp Plewa's fantastic visualization of the [central limit theorem](https://bl.ocks.org/pmplewa/4120c2929ede7e336d9b55b760e496f6).

**Central Limit Theory — illustrated examples**

Let's take a look at an animated example for Central Limit Theorem.

Also be sure to check out the below video.

**Assignment**

For two of the following distributions:

* Normal
* Binomial
* Gamma
* Poisson

1. Create a line chart plotting the sample mean as more and more trials are conducted, up to 1,000 trials.
2. Randomly sample the mean 1,000 times and plot the resulting frequency of sample means as a histogram.

When you are finished compare your results to [**this**](https://github.com/Thinkful-Ed/data-201-assignment-solutions/blob/master/python_4_central_limit_theorem/answers_probability_central_limit_theory_drills.ipynb).

**Checkpoint**

Submit your ideas or a link to your work here and use it as a conversation starter during your next mentor session.

This checkpoint will not be graded, but is still required.

Your work

**07.23.20**

Top of Form

 \*\*bold\*\* \_italic\_ `code` > quote - bullet list

Preview

**Completed**