About

This notebook illustrates the Central Limit Theorem (CLT) in action, using two large samples: the first from the standard uniform distribution, the second from a Poisson distribution. A histogram of the actual sample distribution of the mean is plotted alongside the approximated sample distribution of the mean, as set out by the CLT.

Background

The **Central Limit Theorem** states that

> If $X_1, X_2 \ldots, X_n$ are n independent random observations from a population with mean μ and finite variance σ^2, then for large n, the distribution of their mean $\overline{X}_n$ is approximately normal with mean μ and variance σ^2/n, such that:
>
> $$\overline{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right).$$
>
> Reference: *M248 Handbook, pp.12.*

Note that we use the estimated standard error of the mean, rather than the standard error.[1]

Imports

```
import scipy.stats as stats
from util.clt import clt
```

Illustration 1

This uses samples generated from the **standard uniform distribution,** $U(0, 1)$. As expected, the two plots closely match each other.

```
clt(a_dist=stats.uniform(loc=0, scale=1),
    n=30,
    N=10000,
    bins=50)
```

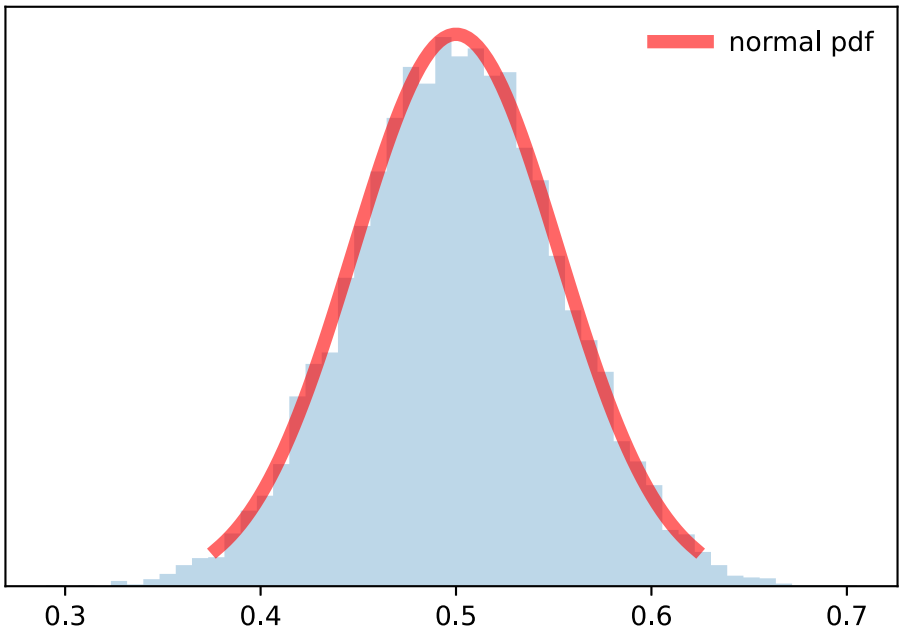
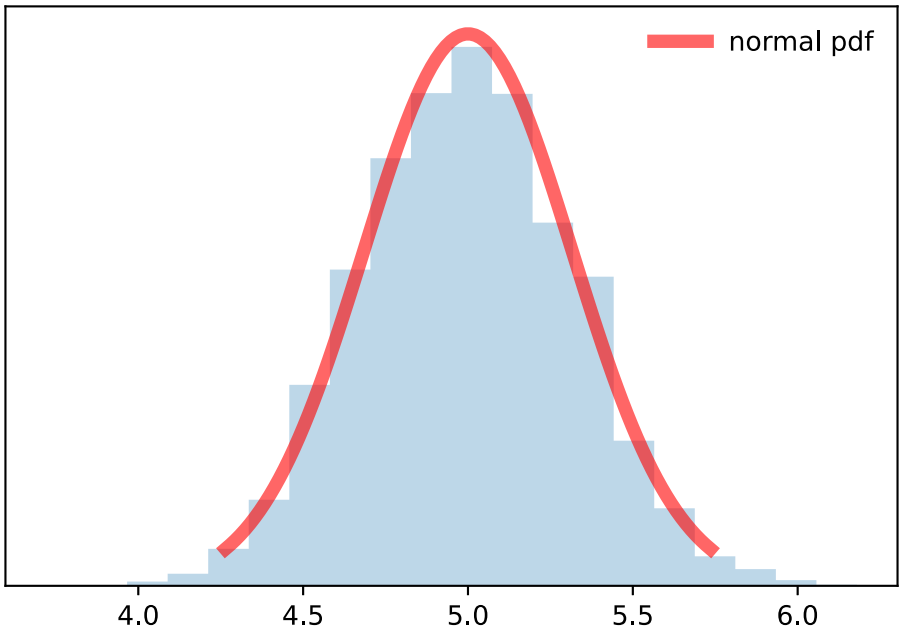


Illustration 2

This uses samples generated from the **Poisson distribution,** $\text{Poisson}(5)$. Again the CLT provides a good approximation of the distribution.

```
clt(a_dist=stats.poisson(mu=5),
    n=50,
    N=10000,
    bins=20)
```



Function

This is a transclusion of the function `clt()` imports, header, and docstring, for illustration purposes.

```
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import norm


def clt(a_dist: object, n: int, N: int, bins: int) -> None:
    '''
    Generates N random samples of size n from the distribution a_dist.
    Calulate's each sample's mean, and plots the sampling
    distribution of the mean as a histogram.

    The approximate sampling distribution of the mean (by the CLT) uses
    parameters mu, and the estimated standard error.

    The parameter bins controls the number of bins used in the
    histogram.
    Changing this parameter will make the histogram less accurate,
    but improves the plot's fidelity for discrete distributions.
```

1. The estimated standard deviation of the sampling distribution of the mean, using the sample standard deviation as opposed to the population standard deviation, M140 Handbook. ↩