DATA SCIENCE FOR ECONOMISTS

ECON 220 LAB

Jafet Baca-Obando

Week 8, More on probability distributions– 10/17/2025

Outline

01 Uniform distribution

02 Normal distribution

03 Binomial distribution

Importing required libraries

```python
import numpy as np
import pandas as pd
import scipy.stats as stats
import seaborn as sns
```

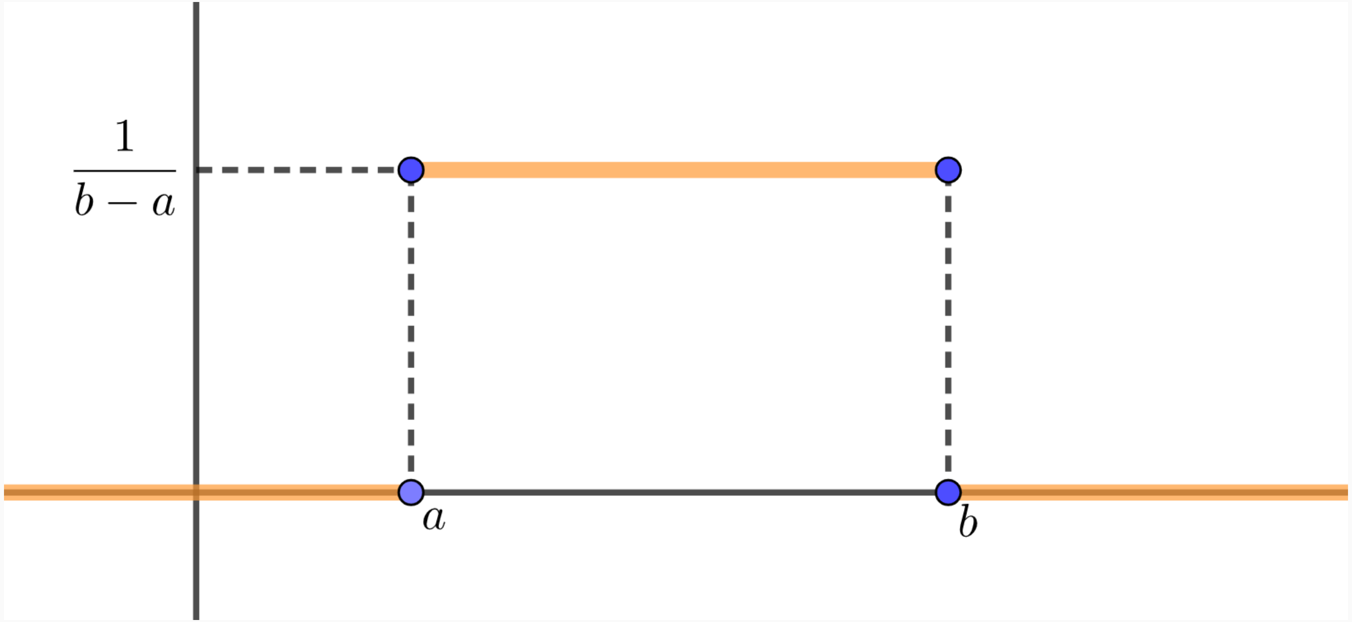
✓ 0.0s

Python

The uniform distribution

- We'll examine the **continuous** uniform distribution on $[a, b]$.
- This distribution assigns equal probability density to every point in the interval.
- Its **probability density function** (PDF) is $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, 0 otherwise.

The uniform distribution



Generating random draws from a uniform distribution

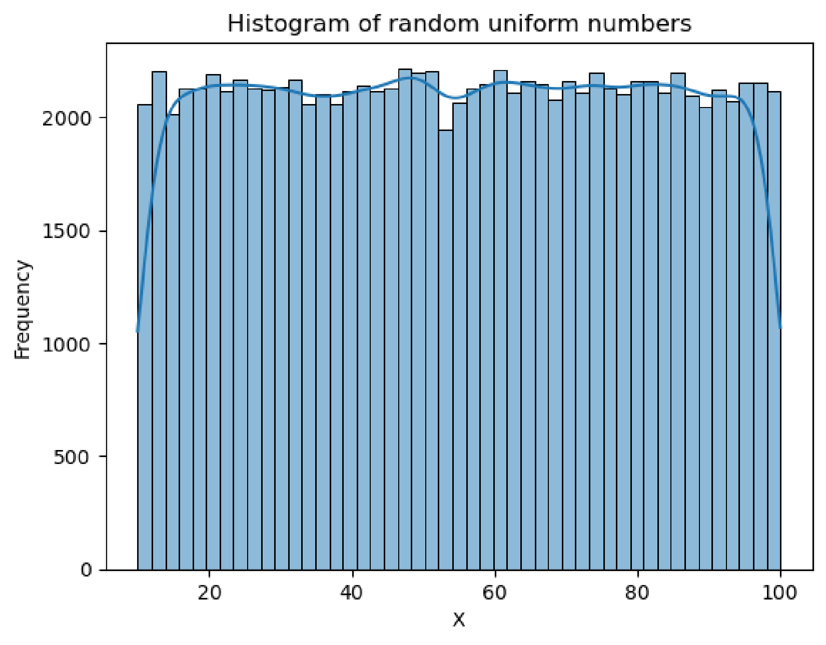
```python
# Generate vector of 30 random numbers between 10 and 100:
# Use: stats.uniform.rvs(size, a, b-a)
np.random.seed(123) # Seed in effect for one cell
u2 = stats.uniform.rvs(size=100000, loc=10, scale=100-10)
u2
```

✓ 0.0s Open 'u2' in Data Wrangler Python

```
array([72.6822267 , 35.75254015, 30.41663082, ..., 88.097556  ,
       60.96019593, 65.16046407])
```

- **loc**: lower bound of the distribution
- **scale**: width of the distribution

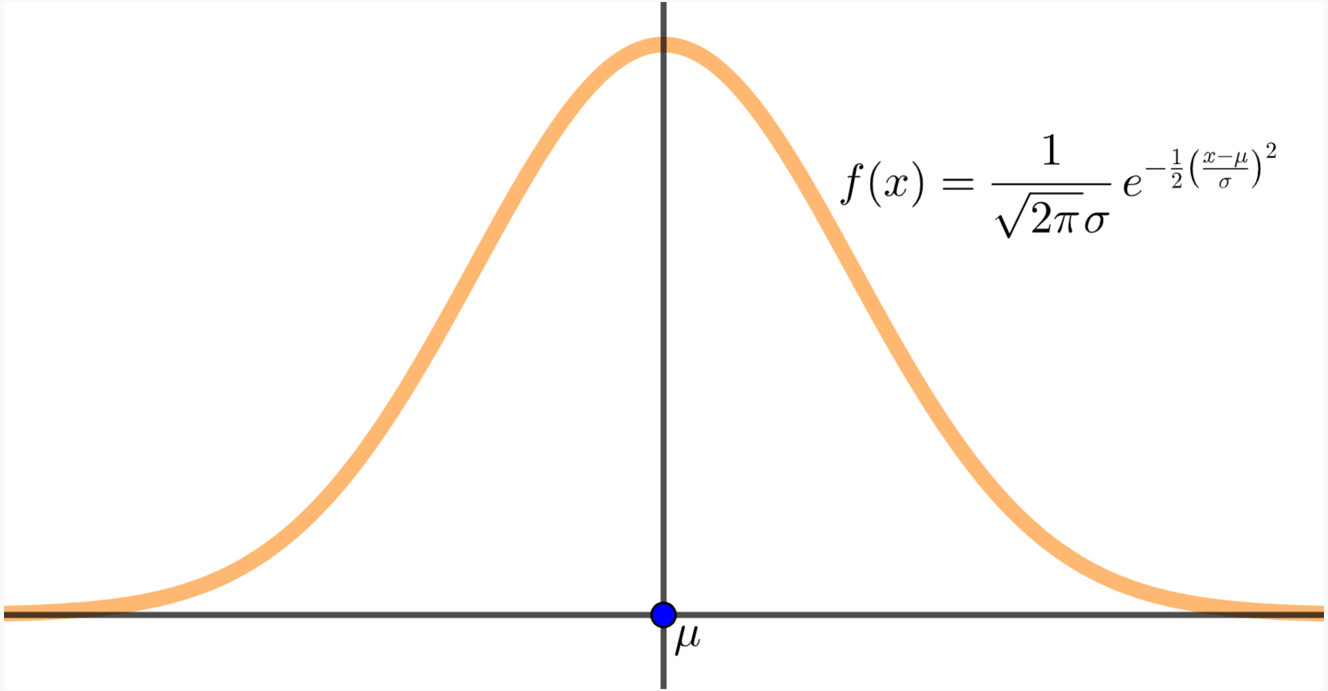
Histogram and density of uniform draws



The normal distribution

- A random variable X follows a normal distribution $\mathcal{N}\left(\mu, \sigma^2\right)$ if its PDF is given by: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.
- μ: mean, σ^2: variance
- The normal distribution is continuous, symmetric, and bell-shaped.
  - Also, centered at the mean μ, with spread determined by the standard deviation σ.
- Examples: Height, shoe size, ACT scores typically follow a normal distribution.

The normal distribution



Drawing numbers from a normal distribution

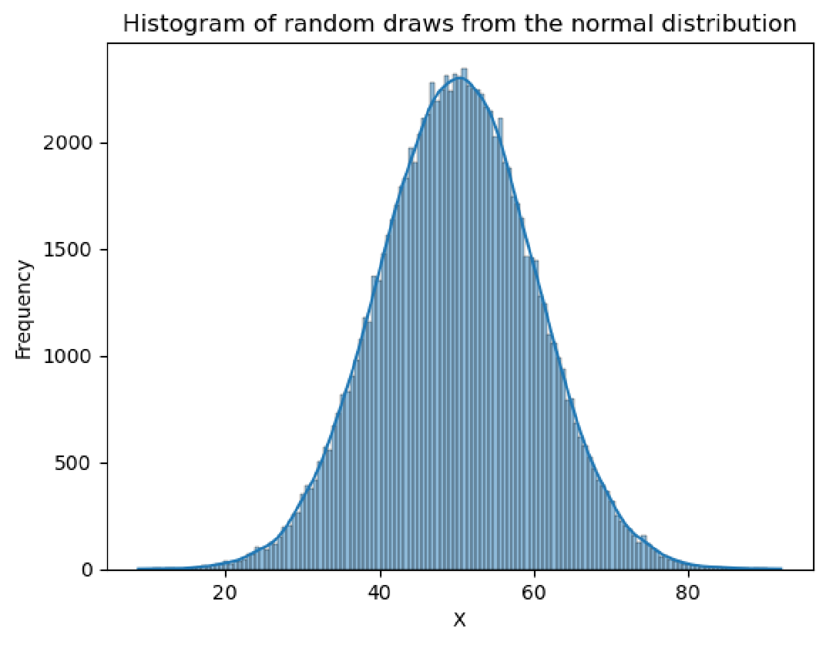
```python
# Generate 20 random numbers from Norm(50, 10^2)
np.random.seed(578)
n1 = stats.norm.rvs(size=100000, loc=50, scale=10)
n1
```

Python

```
array([46.38190268, 56.01681305, 42.68337069, ..., 48.84427816,
       42.8974724 , 67.90048878])
```

- **loc**: mean the distribution
- **scale**: standard deviation of the distribution

Histogram and density of normally distributed draws



The binomial distribution

- Gives the probability of observing k successes in n independent trials, each with probability p of success.
  - Example: If we toss a fair dice 5 times, the probability of getting 3 exactly 2 times is given by a binomial distribution.
- It's a **discrete** distribution.
- Has **probability mass function** (PMF):

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, k = 0,1,2, \dots, n$$

- Assumptions: (i) A fixed number of independent trials, n; (ii) constant probability of success, p, across trials.

The binomial distribution

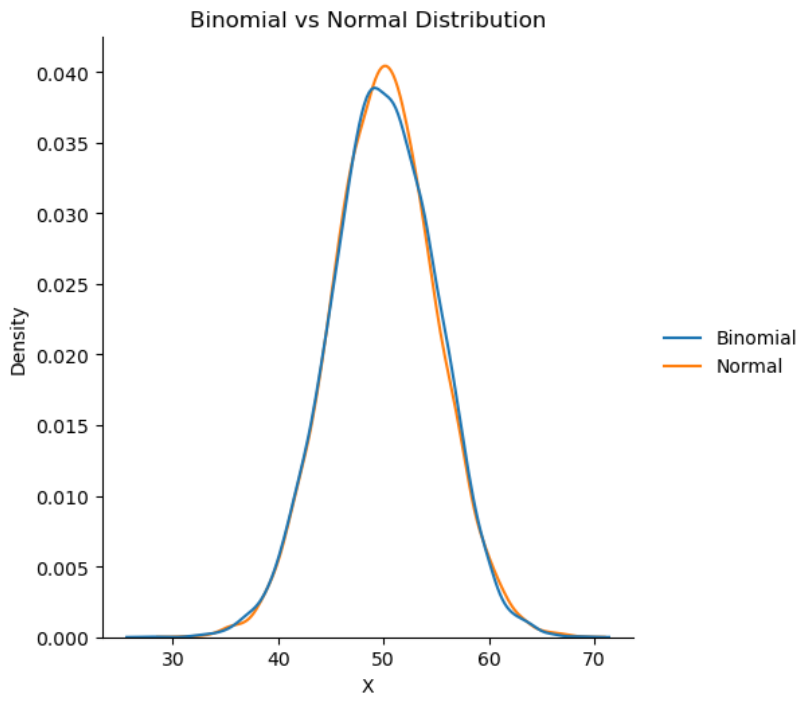
- Key property:
  - When $np \geq 10$ and $n(1-p) \geq 10$, the binomial distribution $Binomial(n, p)$ can be approximated by a normal distribution $\mathcal{N}(np, np(1-p))$ and vice versa.

```python
p = 0.5
n = 100
sc = (n*p*(1-p))**0.5
data_binom = stats.binom.rvs(size=10000, n=n, p=p)
data_norm  = stats.norm.rvs(size=10000, loc=n*p, scale=sc)


data = pd.DataFrame({'Binomial':data_binom, 'Normal':data_norm})
plot_approx = sns.displot(data, kind='kde')
plot_approx.set(title="Binomial vs Normal Distribution", xlabel="X", ylabel="Density")
```

✓ 0.2s Python

The binomial distribution



Recap

- We obtained random draws using the uniform, normal, and binominal distributions.
- We created histograms and density approximations.
- We approximated the normal distribution using the binomial distribution.

To-do list

- **Complete Data Exercise 5**
  - Upload Jupyter notebook (.ipynb file) and HTML file on **October 19**
- **Complete Data Exercise 6**
  - Upload Jupyter notebook (.ipynb file) and HTML file on **October 26**