THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

AMA 2233

Data Analytics and Visualization

Yancheng Yuan

Department of Data Science and Artificial Intelligence

Recap of last lecture

Central Limit Theorem

Central Limit Theorem: Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with finite mean μ and finite non-zero variance σ^2, and let

$$\boldsymbol{S_n = X_1 + X_2 + \cdots + X_n}.$$

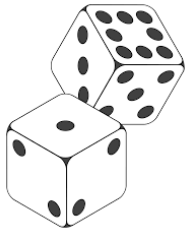
Then

$$\frac{\boldsymbol{S_n - n\mu}}{\sqrt{\boldsymbol{n\sigma^2}}} \xrightarrow{\boldsymbol{D}} \boldsymbol{N(0,1)} \quad \text{as } n \to \infty.$$

Remark

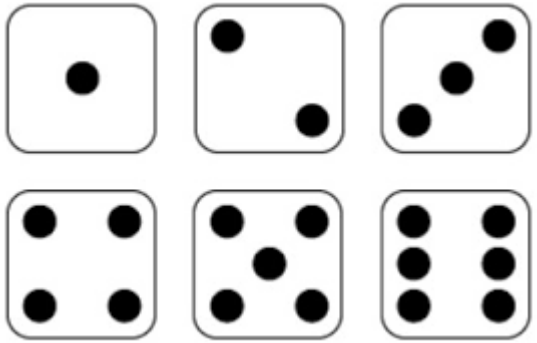
1. **No requirements on the distribution of X.**
2. **The variance σ^2 should be finite and non-zero.**
3. **The sample size n should be large enough.**

Example

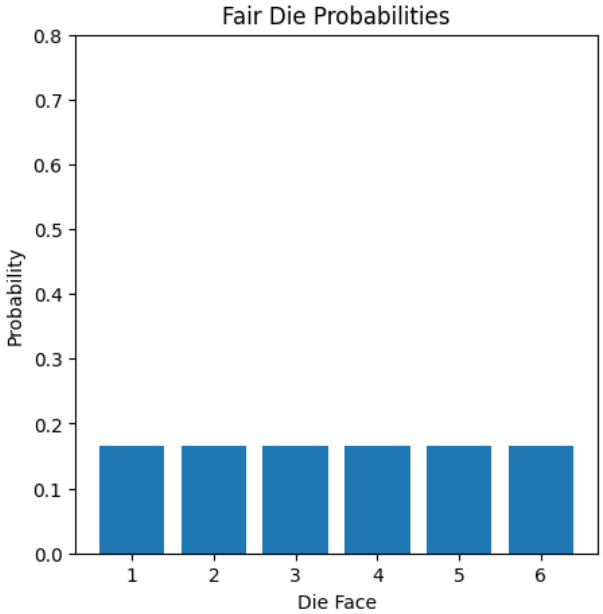


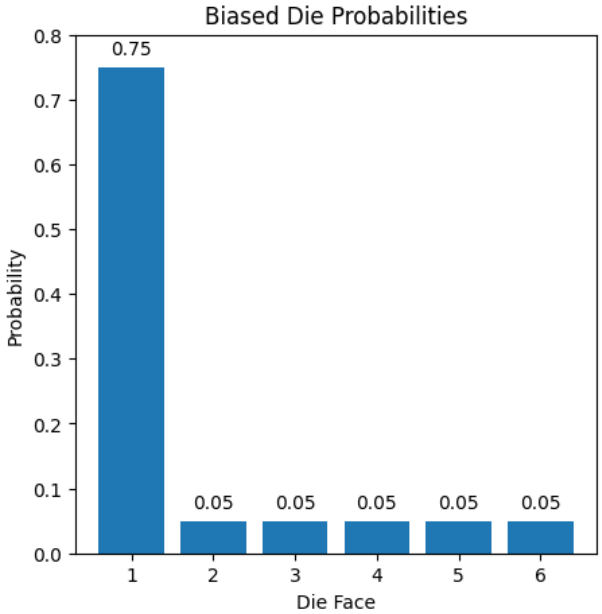
Dice Rolling: Rolling a fair and an unfair dice.

Possible outcomes:

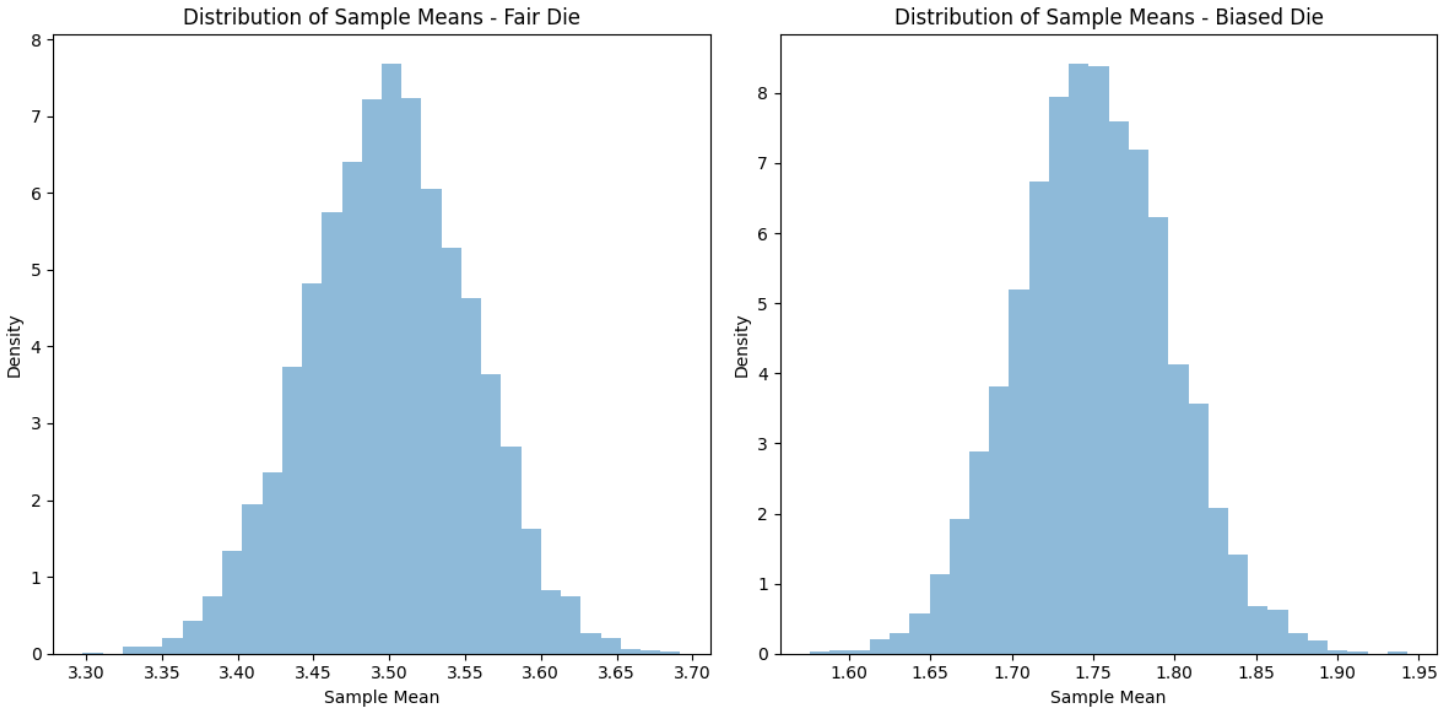


Two scenario

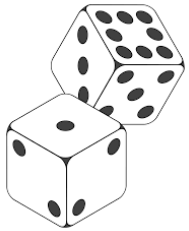




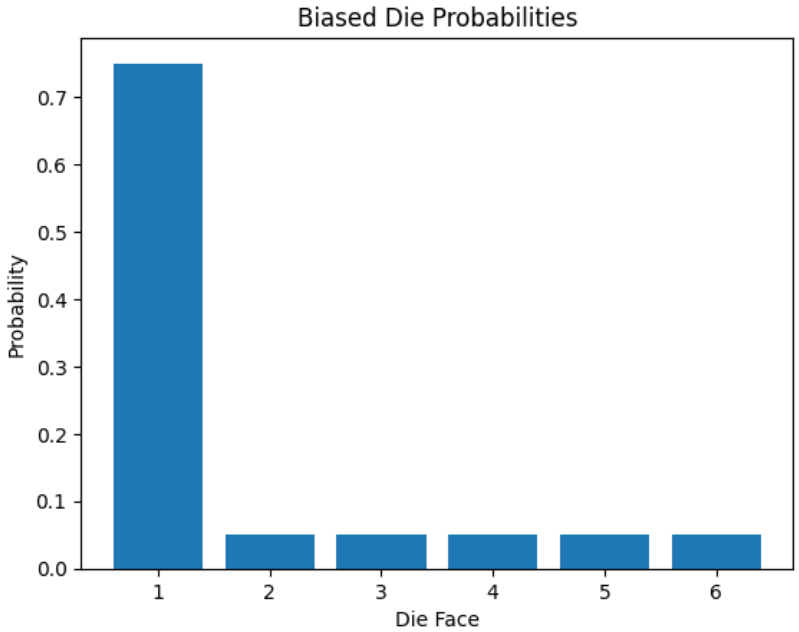
Results


Distribution of Sample Means - Fair Die
Density
Sample Mean
0
1
2
3
4
5
6
7
8
3.30
3.35
3.40
3.45
3.50
3.55
3.60
3.65
3.70
Distribution of Sample Means - Biased Die
Density
Sample Mean
0
1
2
3
4
5
6
7
8
1.60
1.65
1.70
1.75
1.80
1.85
1.90
1.95

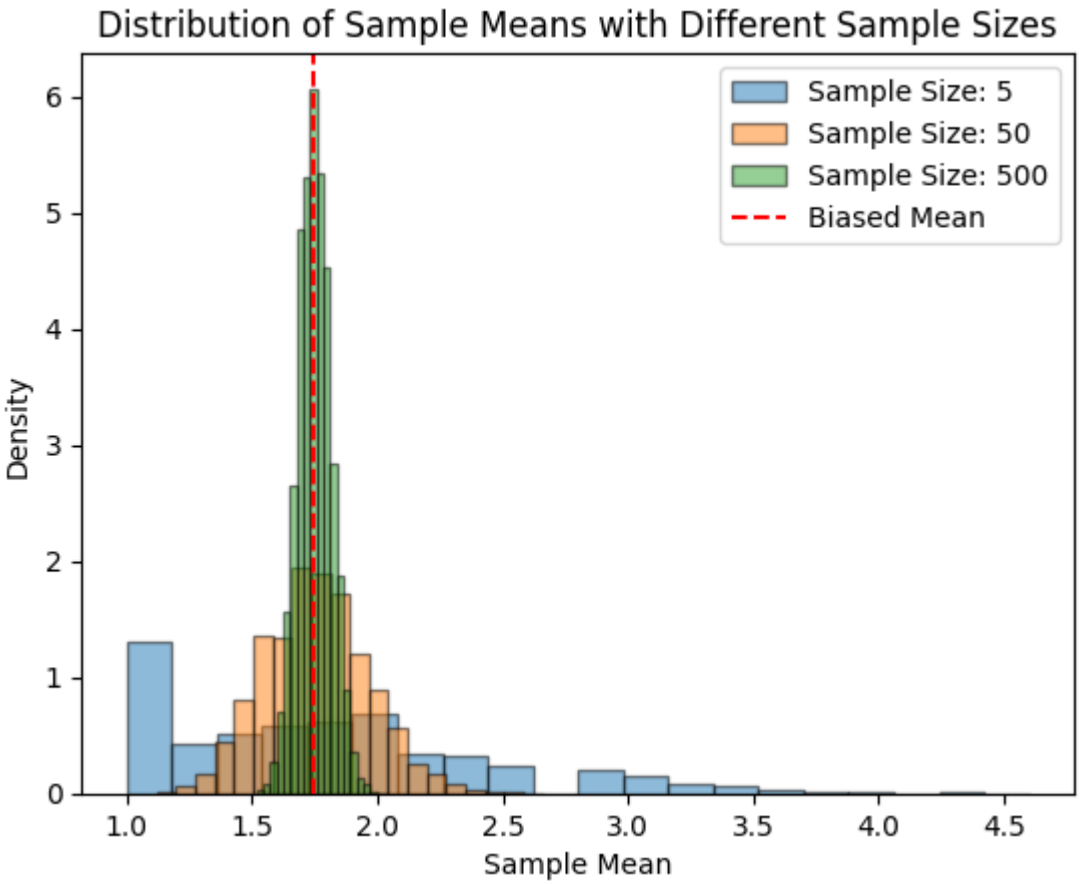
Example



Dice Rolling: Rolling a biased dice.



Results


Distribution of Sample Means with Different Sample Sizes
Sample Size: 5
Sample Size: 50
Sample Size: 500
Biased Mean
Density
Sample Mean
0
1
2
3
4
5
6
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5

Agenda

1. High-dimension integration and Monte-Carlo method
2. The convergence rate of MC method
3. Visualization and applications

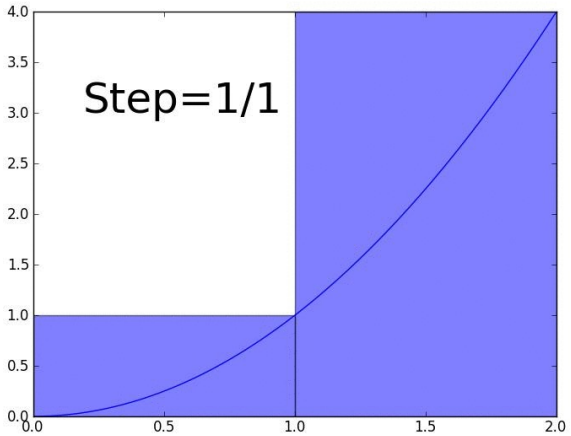
Integration and its connection to probability

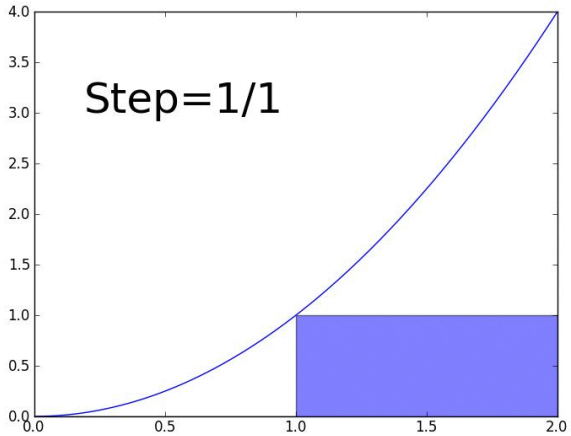
Integration of 1D function

Riemann integral : Let $f: R \to R$ be an integrable function, we define the Riemann integral of f as

$$\int_a^b f(x)dx = \lim_{\Delta \to 0} \sum_{i=1}^{n} f(x_i)\Delta_i,$$

where $a = x_0 < x_1 < \cdots < x_n = b, \Delta_i = x_i - x_{i-1}, \Delta = \max_i \Delta_i.$



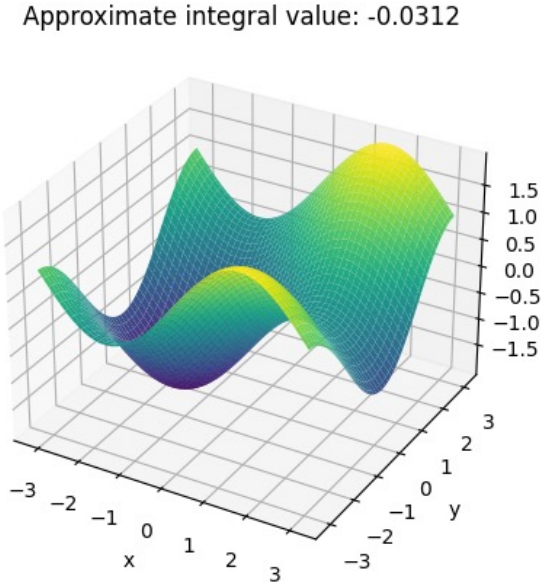


Example: $f(x) = x^2$.

Integration of d dimensional function

Riemann integral : Let $f: R^d \to R$ be an integrable function, we define the Riemann integral of f as

$$\iiint f(x_1, \cdots, x_d)\, dx_1 \cdots dx_d = \iint \left(\int_{a_d}^{b_d} f(x_1, \cdots, x_d)\, dx_d \right) dx_1 \cdots dx_{d-1}.$$



Challenges

1. **Curse of dimensionality**.
   Need $O(\frac{1}{\epsilon^d})$ grid points to get an ϵ partition.

2. An explicit formula may not be available.

Connection to probability

Riemann integral: Let $f: R \to R$ be an integrable function, we have

$$\int_a^b f(x)dx = (b - a)E[f(X)],$$

where X is a uniformly distributed random variable on [a, b].

Connection to probability

Riemann integral: Let $f: R^d \to R$ be an integrable function, we have

$$\iiint_V f(x_1, \cdots, x_d) dx_1 \cdots dx_d = vol(V) E[f(X)]$$

where X is a uniformly distributed random variable on V.

MC method

Monte Carlo Method for estimating integral

$$\iiint_V f(x_1, \cdots, x_d) dx_1 \cdots dx_d$$

Step 1: Set N.

Step 2: Random sample N points uniformly on V.

Step 3: Estimate $vol(V)$.

Step 4: Compute $vol(V) * \frac{f(X_1) + \cdots + f(X_N)}{N}$.

Convergence rate of MC method
(Will derive it during the lecture)

Convergence rate

Let $f: R^d \to R$ be an integrable function and $V \in R^d$ is the unit cube, we have

$$\boldsymbol{E\left[\left(\int_V f(x)dx - \frac{1}{N}\sum_{i=1}^{N} f(X_i)\right)^2\right] \leq O(\frac{1}{N}).}$$

Remark: The convergence rate is independent of d !!!

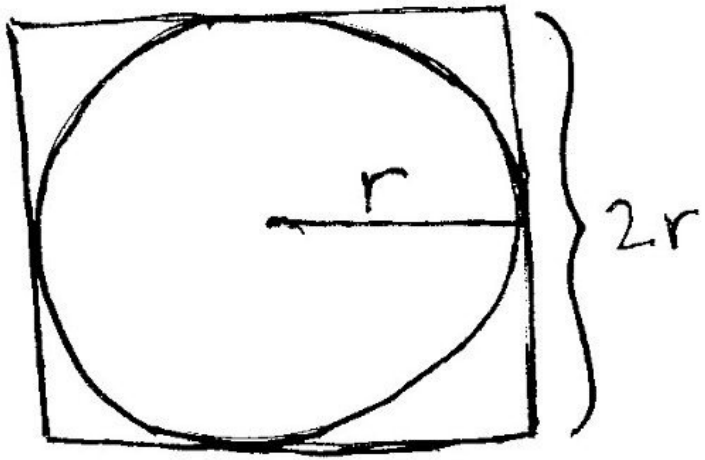
Visualization and Applications

Estimate π with Monte Carlo method

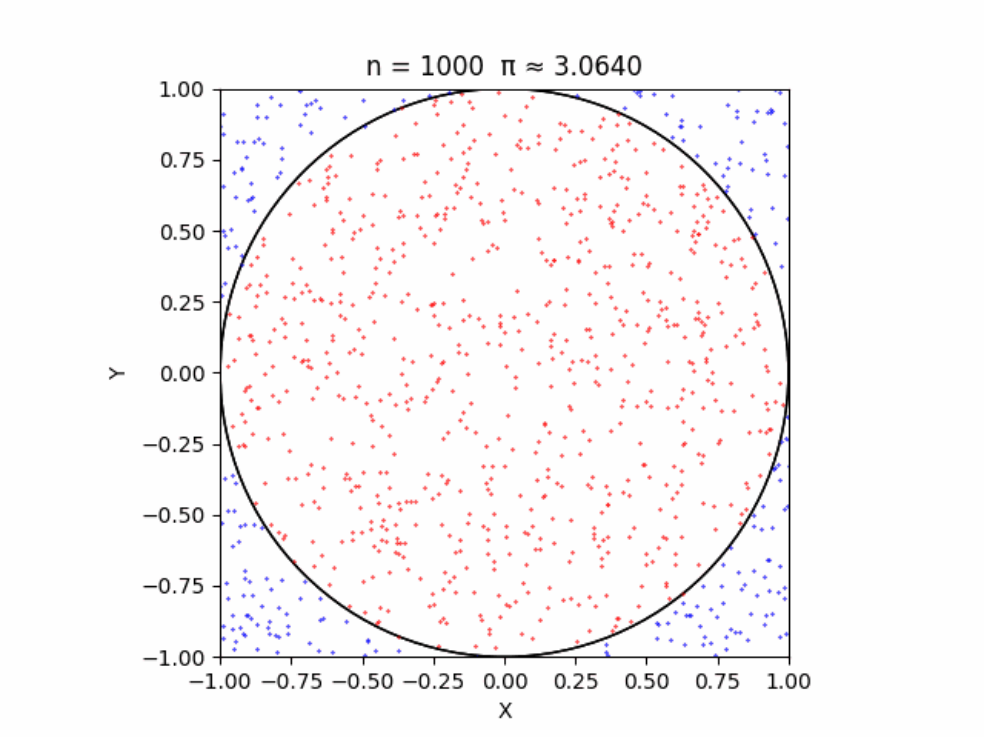
π: It is a classical mathematical question to estimate the value of π, which is the area of the unit disk. We know that

$$P(\{x \in C\}) = \frac{\pi}{4} = E[1_C(x)],$$

where $1_C(x)$ is the indicator function of the unit circle in the square.



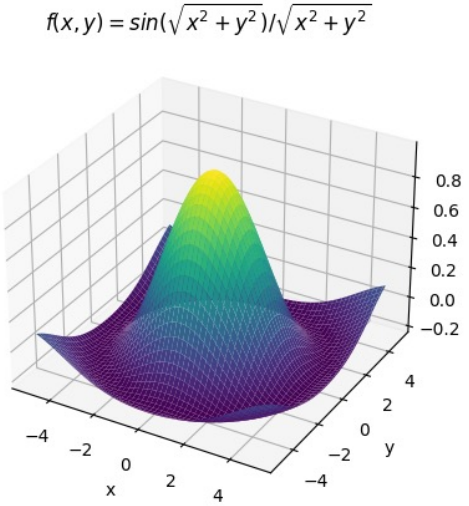
Estimate π with Monte Carlo method



Estimate the Integration

Example: $f(x,y) = \frac{\sin\left(\sqrt{x^2+y^2}\right)}{\sqrt{x^2+y^2}}$, estimate

$$\int_{-5}^{5}\int_{-5}^{5} f(x,y)\,dxdy\,.$$



Estimate the Integration



```python
def monte_carlo_integration(f, x_min, x_max, y_min, y_max, num_samples):
    x_samples = np.random.uniform(x_min, x_max, num_samples)
    y_samples = np.random.uniform(y_min, y_max, num_samples)
    f_samples = f(x_samples, y_samples)
    integral_approx = (x_max - x_min) * (y_max - y_min) * np.mean(f_samples)
    return integral_approx
```

Estimate the Integration



```python
def f(x, y):
    r = np.sqrt(x**2 + y**2)
    return np.sin(r) / r

x_min, x_max = -5, 5
y_min, y_max = -5, 5
num_samples = 500000

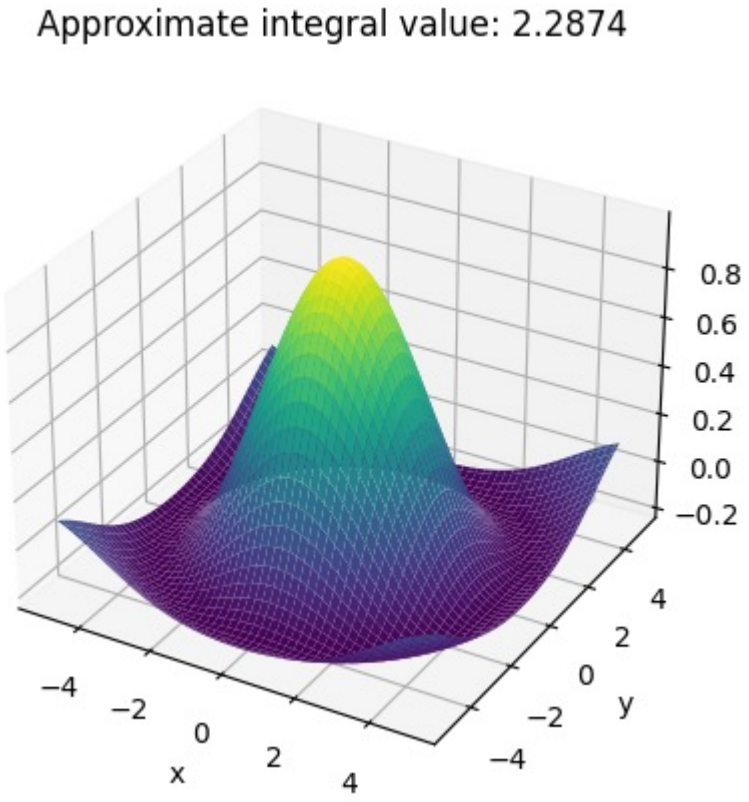
integral_approx = monte_carlo_integration(f, x_min, x_max, y_min, y_max, num_samples)

x = np.linspace(x_min, x_max, 100)
y = np.linspace(y_min, y_max, 100)
X, Y = np.meshgrid(x, y)
Z = f(X, Y)

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.plot_surface(X, Y, Z, cmap='viridis')
ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_title(f"Approximate integral value: {integral_approx:.4f}")

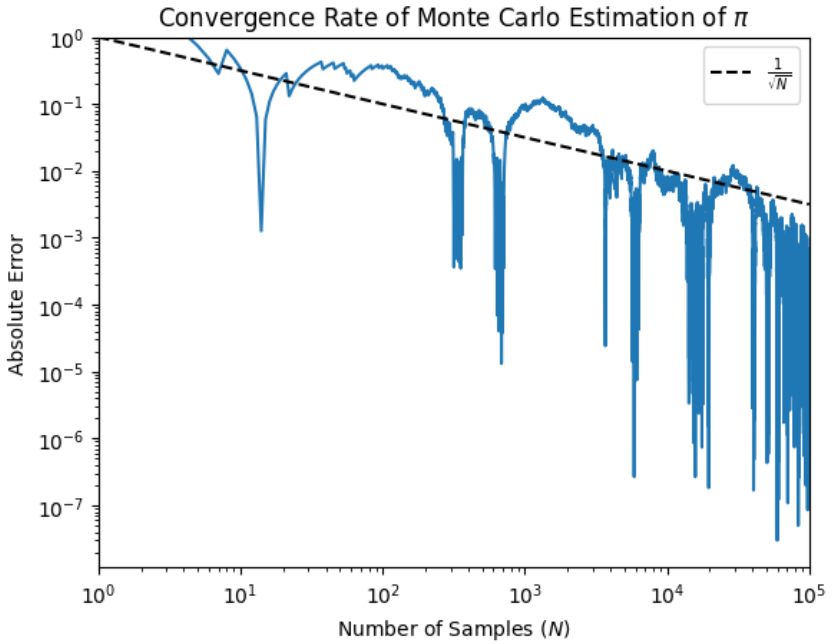
plt.show()
```

Estimate the Integration



Verify the $O(1/\sqrt{N})$ rate

Example: $Estimate\ \pi.$



Verify the $O(1/\sqrt{N})$ rate

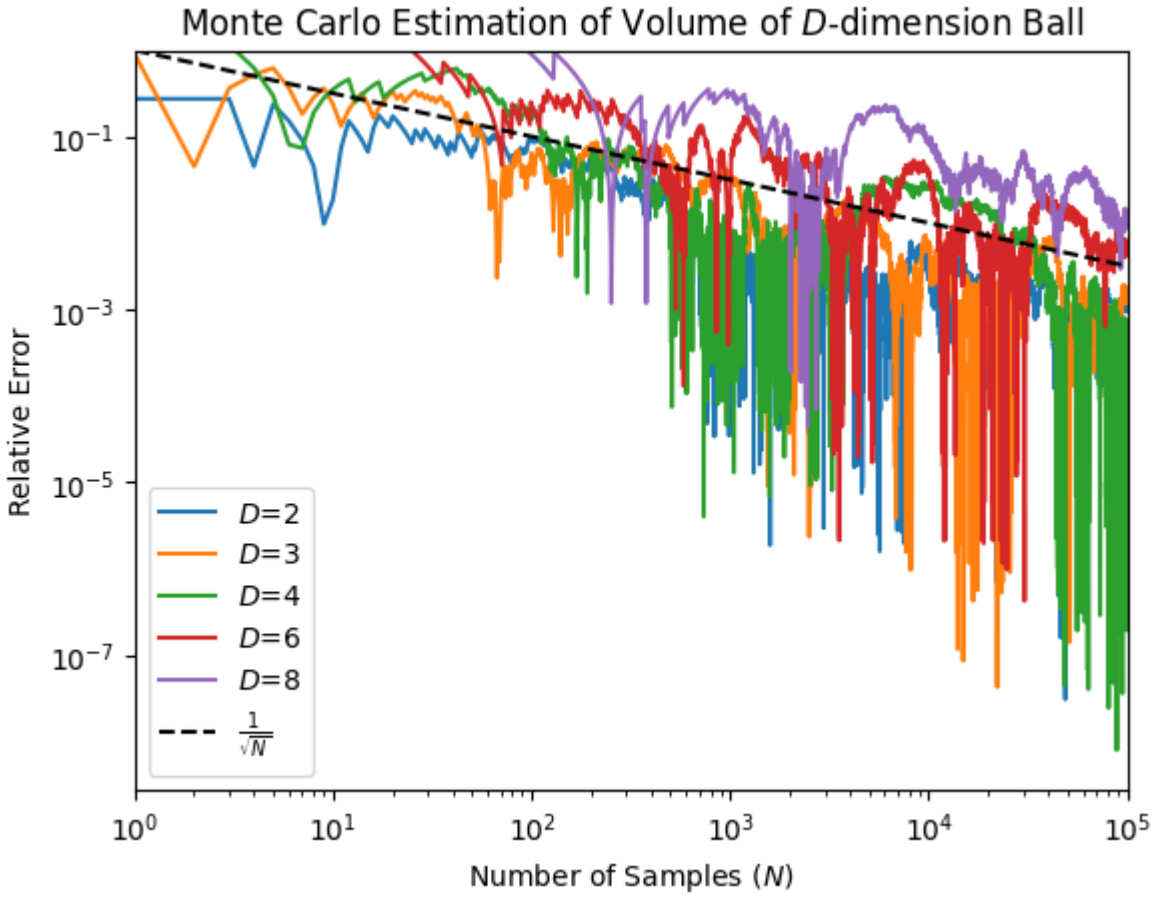
Example: Estimate the volume of a unit ball in D-dimensional space, which has the closed form:

$$V = \frac{\pi^{D/2}}{\Gamma(\frac{D}{2}+1)},$$

Where Γ is Euler's gamma function.

$$\Gamma(n) = (n-1)!, \qquad \Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\times\left(n-\frac{3}{2}\right)\times\cdots\times\frac{1}{2}\times\pi^{1/2}.$$

Verify the $O(1/\sqrt{N})$ rate



Application in financial forecasting

Geometric Brownian Motion (GBM) model: The GBM model assumes that the future is independent of the past. If this model is applied in the context of asset pricing, it assumes that the past price information has already been incorporated into the current price and the price movement in the future is conditionally independent of any price movements in the past.

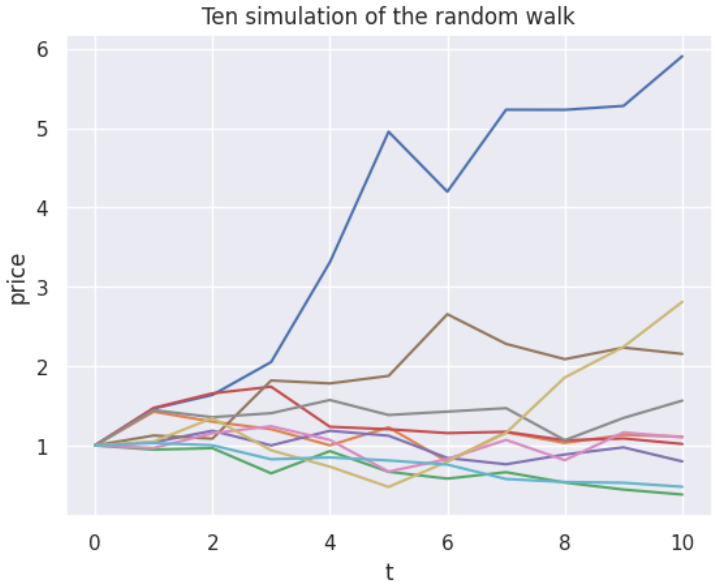
$$S_t = S_{t-1} \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)dt - \sigma\sqrt{dt}Z_t\right)$$

where $Z_t \sim N(0,1)$, μ is the drift rate which determines the growth in the expected value of the random process, σ is the volatility of the drift rate.

Application in financial forecasting

Settings:

$$S_0 = 1, \mu = 0.05, \sigma = 0.2.$$



Application in financial forecasting

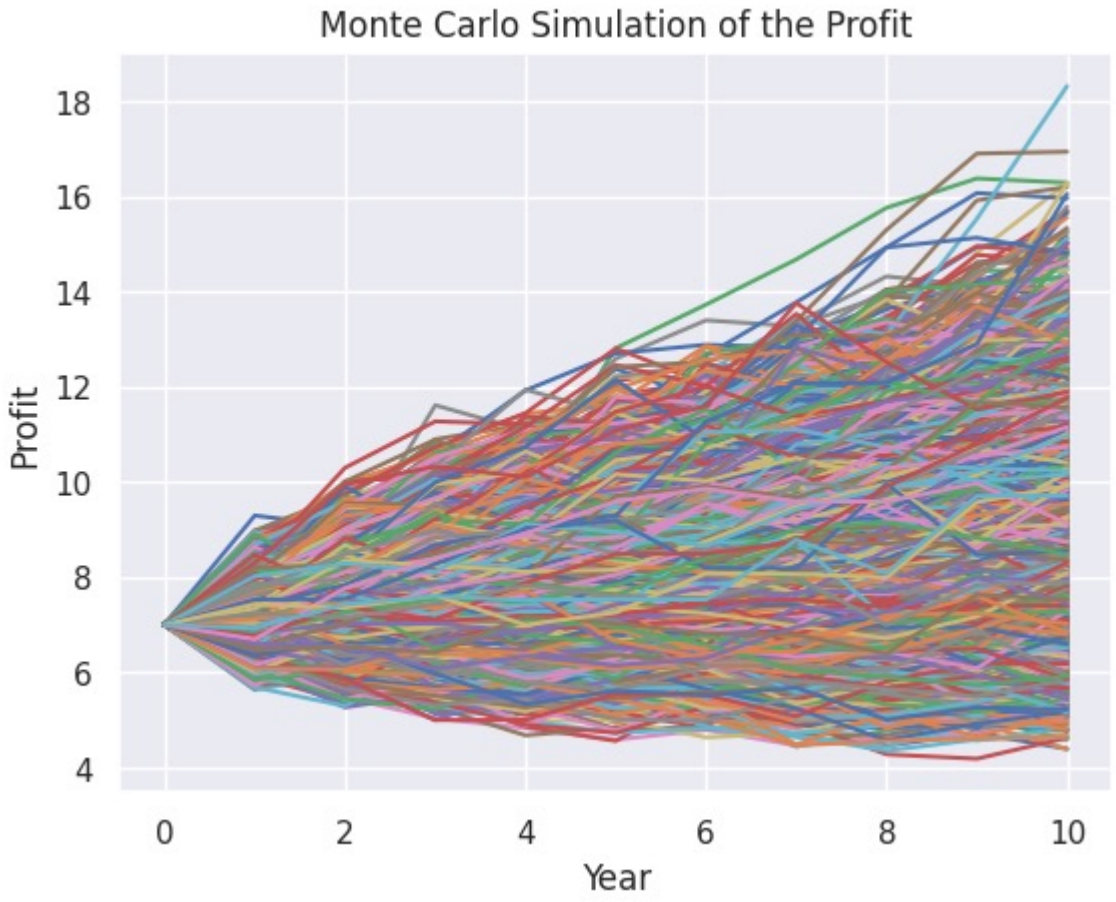
Model settings: **Profit = Revenue — Fixed Cost — Variable Cost**.

Assume the fixed cost is 2 and the revenue follows a GBM model with

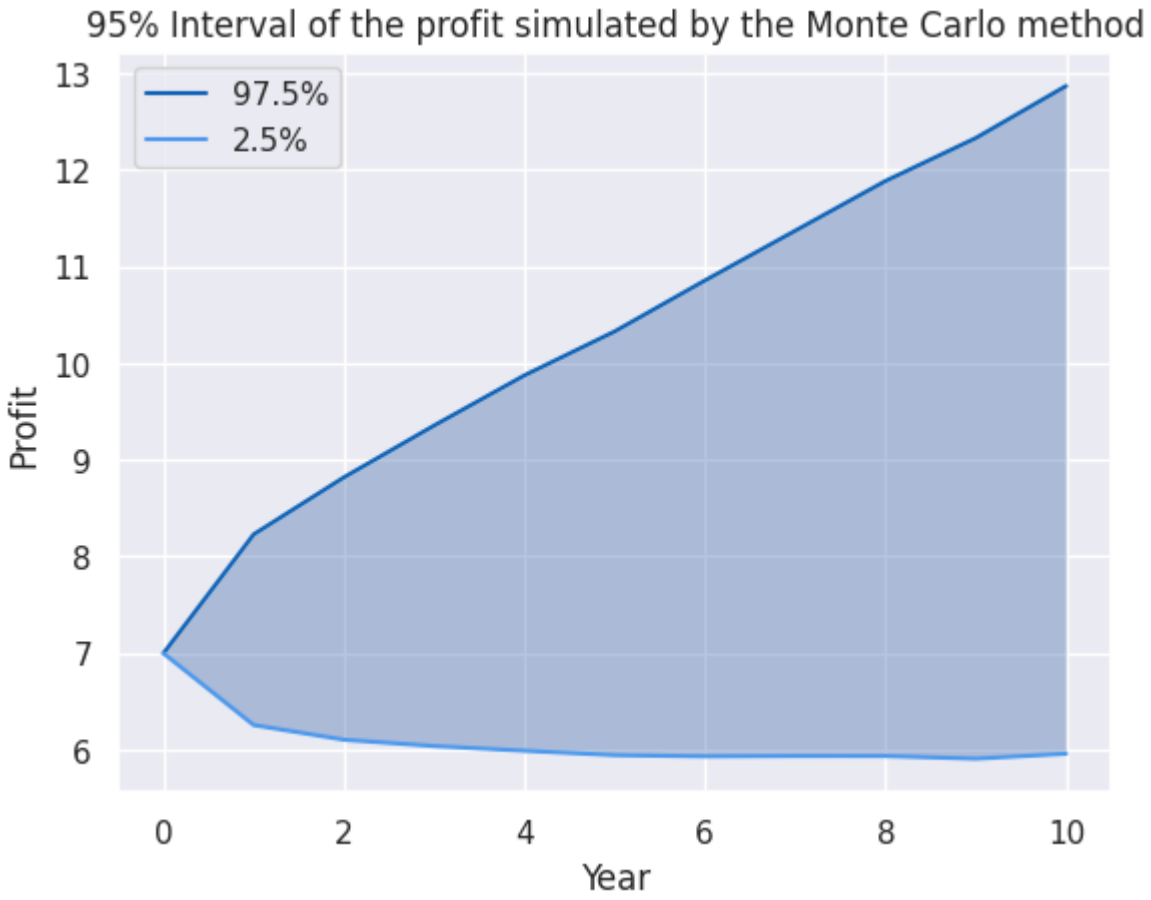
$$S_0 = 10\ \mu = 0.02, \sigma = 0.05,$$

and the variable cost is 10% of the revenue.

Application in financial forecasting



Application in financial forecasting



Summary

1. Monte-Carlo method

2. The convergence rate of the Monte Carlo method

3. Visualization and applications

Reading Materials:

1. https://en.wikipedia.org/wiki/Monte_Carlo_method

2. The original paper of Monte Carlo Method: Metropolis, N., & Ulam, S. (1949). The monte carlo method. Journal of the American statistical association, 44(247), 335-341.


Thank You!