MM 225: AI and Data Science

Continuous probabilities

M P Gururajan and Hina A Gokhale

August 10, 2023

Lecture 3: Continuous probabilities

Outline

Monte Carlo

Monte Carlo casino

By AwOiSoAk KaOsIoWa, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=54618166

Ulam and Nuemann

Ulam: By Los Alamos National laboratory - Attribution, https://commons.wikimedia.org/w/index.php?curid=26069369

Neumann: By LANL - http://www.lanl.gov/history/atomicbomb/images/NeumannL.GIF (archive copy), Attribution, https://commons.wikimedia.org/w/index.php?curid=3429594

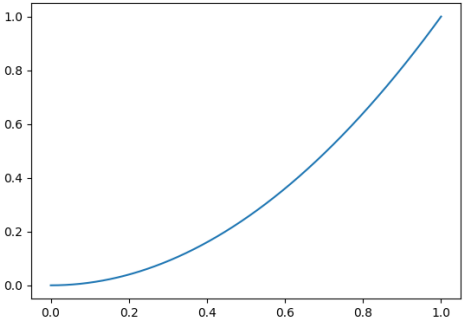
Area under a curve

What is the area under the curve $y = x^2$ between 0 and 1?

$$\int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3} \tag{1}$$

Evaluating integrals using Monte Carlo (MC) method

$y = x^2$ embedded in unit square



MCArea.py

```
import matplotlib.pyplot as plt
import numpy as np
import random
N = 10000
A = 0.0
X = []
Y = []
```

MCArea.py

```
for i in range(N):
    x = random.random()
    y = random.random()
    if(i%10 == 0):
        X.append(x)
        Y.append(y)
    if(y < x*x):
        A = A + 1
Area = A/N
print(Area)
```

MCArea.py

```
XX = np.linspace(0,1,1000)
plt.plot(X,Y,'o')
plt.plot(XX,XX*XX,'r-')
plt.show()
```

Result

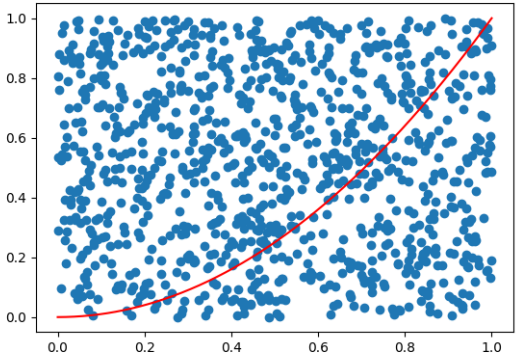


Figure: 33.57% of the points lie below the red line. Hence, we get the answer 0.3357 compared to analytical solution of 0.3333.

Comment

- Coin toss or throw of a die: outcomes are discrete and countable
- Dart throwing at the board: the sample space is the position of the dart
- Sample space: not discrete or countably infinite!
- Probability of dart at any point in the given area – has to be given as a function of position
- m in the discrete case: has to be replaced by a continuous function $f(x, y)$
- Discuss in detail in the subsequent lecture

Evaluating π

Idea

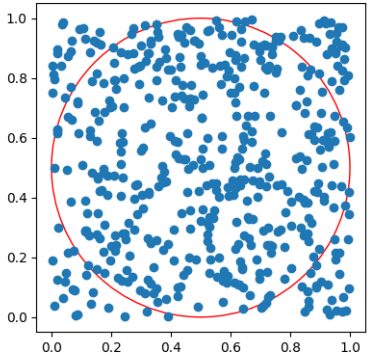


Figure: Given the circle is inscribed insde the unit square, $\frac{\pi}{4}\%$ of the points are expected lie inside the red circle.

MCPiEval.py

```
import numpy as np
import random
N = 10000
A = 0.0
for i in range(N):
    x = random.random()
    y = random.random()
    if(((x-0.5)*(x-0.5)+(y-0.5)*(y-0.5)) < 0.25):
        A = A + 1
Pi = 4.*(A/N)
print(Pi)
```

Answers: 3.116, 3.1432, 3.1356

Home assignment

- How to plot the circle and the data points on the same figure?
- How to change the aspect ratio of the figure?

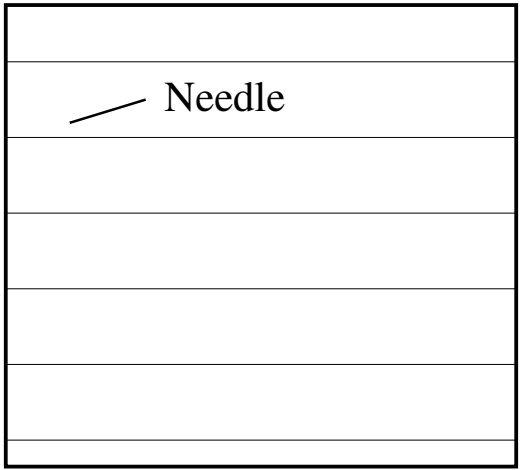
Evaluating π: Buffon's needle method

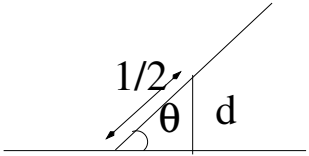
Idea

- Table with parallel lines unit length apart
- Throw a needle of unit length at random
- Does the needle lie across one of the lines?
- L: direction of needle; θ: angle needle makes with the parallel line; d: distance from centre of the needle to the nearest line
- Needle lies across the nearest line if and only if

$$\frac{d}{\sin\theta} < \frac{1}{2} \tag{2}$$

Buffon's experiment: schematic





Evaluating π

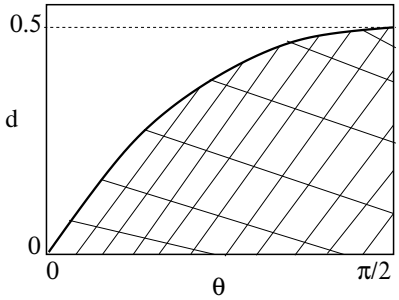


Figure: Area of the rectangle is $\frac{\pi}{4}$. Area of the shaded portion is $\int_0^{\frac{\pi}{2}} \frac{1}{2} \sin\theta d\theta = \frac{1}{2}$. Hence, probability that the needle lies across the line is $\frac{\frac{1}{2}}{\frac{\pi}{4}} = \frac{2}{\pi}$.

Algorithm

- Choose a random number d between 0 and 0.5
- Choose a random angle θ between 0 and $\frac{\pi}{2}$
- Count number of times $d < \frac{1}{2}\sin\theta$
- The fraction of times the $d < \frac{1}{2}\sin\theta$ is equal to $\frac{2}{\pi}$

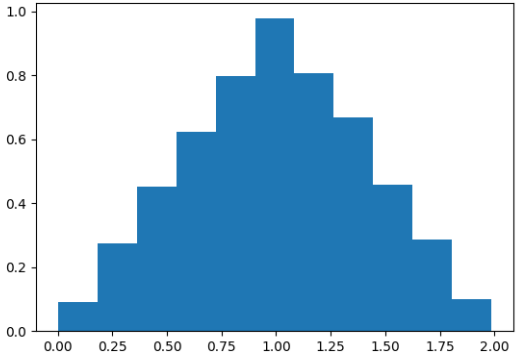
Home assignment: Write a python script to evaluate π by carrying out Buffon's needle experiment on the computer. What is the error? How does it change with the number of needle throws?

Summing random numbers

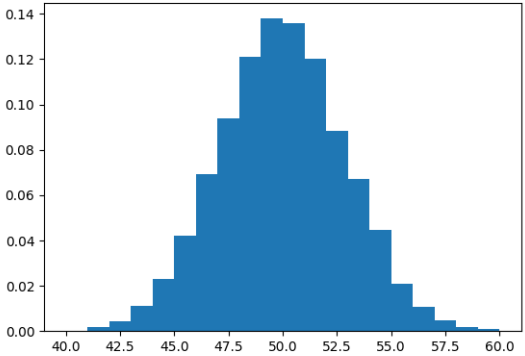
A couple of questions

- **Summing two**
  Let X be the sum of two random real numbers between [0,1]. How is X distributed?
- **Summing 100**
  Let X be the sum of 100 random real numbers between [0,1]. How is X distributed?

Sum of two random numbers



Sum of hundred random numbers

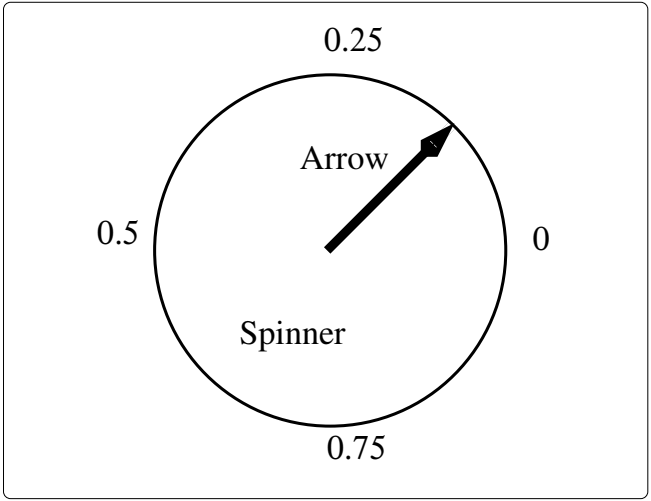


Home assignment

Write a python script which reproduces the figures shown in the previous slides. How does the results change if the number of experiments are 10, 100, 1000, and 10000?

Continuous density function

Spinner



Random variable

- Experiment: spin the disk
- Outcome: how far from zero does the arrow come to rest?
- Let P be the probability
- $P(0 \leq X \leq 1) = 1$; arrow comes to rest somwewhere on the circle
- $P(0 \leq X < \frac{1}{2}) = P(\frac{1}{2} \leq X < 1) = \frac{1}{2}$: arrow coming to rest in the upper half is equally likely to it coming to rest on the lower half
- In general, $P(c \leq X < d) = d - c$ for all c, d
- Compare with $P(E) = \sum_{\omega \in E} m(\omega)$
- $P(E) = \int_E f(x)dx$ with $f(x) = 1$
- $f(x)$: density function

THANK YOU!!!