# Probability with Coding

Karthikeya Hanu Prakash Kanithi

IIT Hyd

October 18, 2023

# Objectives

- Generate standard normal random variable using C
- Solve our problem statement

# Python vs C

- Python already has inbuilt libraries through which we can generate standard normal random variables. But In C we have to generate them using random numbers...
- Here is where we will learn about Box-muller transforms

# Box-Muller Transform

- Suppose U1 and U2 are independent samples chosen from the uniform distribution on the unit interval (0, 1). Let

$$Z_0 = R\cos(\Theta) = \sqrt{-2\ln U_1}\cos(2\pi U_2) \tag{1}$$

- and

$$Z_1 = R\sin(\Theta) = \sqrt{-2\ln U_1}\cos(2\pi U_2) \tag{2}$$

- Then $Z_0$ and $Z_1$ are independent random variables with a standard normal distribution.
- So, now we will generate $Z_0$ using C Code as given below

# Proof

- Let $X$ and $Y$ be independent standard normal variables

$$X, Y \sim \mathcal{N}(0,1) \quad and \quad X \perp Y \tag{3}$$

- The joint pdf of $X$ and $Y$ is given by

$$\begin{aligned} f_{XY}(X, Y) &= f(x)f(y) \qquad (4) \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} . \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} \qquad (5) \\ &= \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} \qquad (6) \end{aligned}$$

# Proof

- The relationship between Cartesian coordinates $(x, y)$ and polar coordinates $(r, \theta)$ is as follows

$$x = r\cos\theta \tag{7}$$
$$y = r\sin\theta \tag{8}$$

- Change $f_{XY}(x, y)$ to polar coordinates :

$$f_{XY}(x, y)dxdy = f_{R\theta}(r, \theta)drd\theta \tag{9}$$

- i.e.,

$$f_{R\theta}(r, \theta) = f_{XY}(x, y)\frac{dxdy}{drd\theta} = f_{XY}(x, y)\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| \tag{10}$$

# Proof

- where J is the Jacobian

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \tag{11}$$

$$= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r \tag{12}$$

- For $r \geq 0$ and $\theta \in [0, 2\pi)$, we have

$$f_{R\theta}(r,\theta)drd\theta = \frac{1}{2\pi}e^{\frac{-(r^2)}{2}}rdrd\theta \tag{13}$$

- Now we change the variable from $(r,\theta)$ to $(r^2,\theta)$;

$$rdr = \frac{1}{2}dr^2 \tag{14}$$

# Proof

- Now it can be written as,

$$
\begin{aligned}
f_{R\theta}(r,\theta)drd\theta &= f_{R^2\theta}(r^2,\theta)dr^2d\theta && (15)\\
&= \frac{1}{2\pi}e^{\frac{-(r^2)}{2}}\frac{1}{2}dr^2d\theta && (16)\\
&= \left(\frac{1}{2}e^{\frac{-(r^2)}{2}}dr^2\right)\left(\frac{1}{2}d\theta\right) && (17)\\
&= f_{R^2}(r^2)dr^2 f_\theta(\theta)d\theta && (18)
\end{aligned}
$$

- from the above equation, we can say that

$$R^2 \perp \theta \text{ (}i.e., R^2 \text{ and } \theta \text{ are independent)} \quad (19)$$

# Proof

- Generate, $\theta \sim Unif(0, 2\pi)$
- Generate, $V \sim Exp(\lambda = \frac{1}{2})$ (i.e., $V = R^2$) and compute

$$R = \sqrt{V} \tag{20}$$

- Compute

$$X = R\cos\theta \tag{21}$$

$$Y = R\sin\theta \tag{22}$$

$$\tag{23}$$

- where, $X, Y$ are i.i.d in $\sim \mathcal{N}(0, 1)$

# Proof

- Then

$$\theta = 2\pi U_1 \tag{24}$$
$$V = -2log(U_2) \tag{25}$$

- We can prove $V = -2log(U_2)$ using the c.d.f definition of the exponential distribution

# Proof for $V = -2log(U_2)$ I

- Let $X$ be a random variable following an exponential distribution with rate parameter $\lambda = \frac{1}{2}$, denoted as $X \sim \text{Exp}\left(\frac{1}{2}\right)$. The cumulative distribution function (CDF) of the exponential distribution is given by:

$$F(x) = 1 - e^{-\frac{x}{2}} \tag{26}$$

- Now, suppose we have a random variable $U$ following a uniform distribution in the interval $[0, 1]$, denoted as $U \sim \text{U}(0, 1)$. The CDF of the uniform distribution is simply:

$$F_U(u) = u, \text{ for } 0 \leq u \leq 1 \tag{27}$$

# Proof for $V = -2log(U_2)$ II

- We can use the probability integral transform to express the exponential random variable $X$ in terms of the uniform random variable $U$:

$$F(x) = F_U(u) \tag{28}$$
$$1 - e^{-\frac{x}{2}} = u \tag{29}$$

- Now, solve for $x$:

$$e^{-\frac{x}{2}} = 1 - u \tag{30}$$
$$-\frac{x}{2} = \ln(1 - u) \tag{31}$$
$$x = -2\ln(1 - u) \tag{32}$$

# Proof for $V = -2log(U_2)$ III

- So, $X \sim -2\ln(1 - U)$ for $U \sim \mathrm{U}(0, 1)$. This expression represents the exponential random variable $X$ in terms of a uniform random variable $U$.

# C-code I

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <time.h>

double sn() {
    double u1, u2;
    while (1) {
        u1 = ((double)rand() / RAND_MAX);
        if (u1 > 0 && u1 < 1)
            break;}
    while (1) {
        u2 = ((double)rand() / RAND_MAX);
        if (u2 > 0 && u2 < 1)
            break;}
```

# C-code II

```
    double z1 = sqrt(-2 * log(u1)) * cos(2 *
        M_PI * u2);
    return z1;
}

int main() {
    // Seed the random number generator
    srand(time(0));

    int numSamples = 100000; // You can change
        this to the desired number of samples
    FILE *file = fopen("uni.dat", "w");

    if (file == NULL) {
```

# C-code III

```
        printf("Unable to open file for writing.\n");
        return 1;
    }

    for (int i = 0; i < numSamples; i++) {
        double sample = sn();
        fprintf(file, "%lf\n", sample);
    }

    fclose(file);

    return 0;
}
```

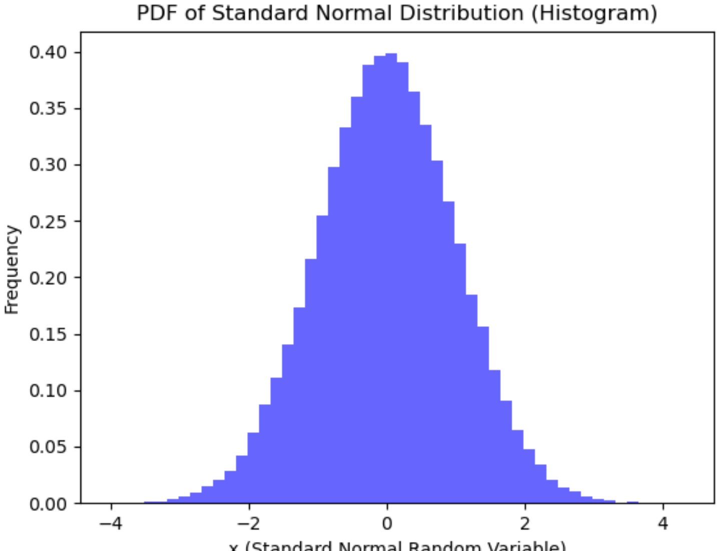
# Python-code

```python
import numpy as np
import matplotlib.pyplot as plt

# Read data from the "uni.dat" file
x = np.genfromtxt("uni.dat")

# Create a histogram plot of the PDF
plt.hist(x, bins=50, density=True, alpha=0.6, color='b', label='PDF (Histogram)')
plt.xlabel('x (Standard Normal Random Variable)')
plt.ylabel('Frequency')
plt.title('PDF of Standard Normal Distribution (Histogram)')
plt.savefig('/home/sayyam/KHP/figs/figure1.png')
```

# Python-code


PDF of Standard Normal Distribution (Histogram)
Frequency
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−2
0
2
4
x (Standard Normal Random Variable)

## Problem Statement

Let $\phi(.)$ denote the cumulative distribution function of a standard normal random variable. If the random variable $X$ has the cumulative distribution function

$$F(x) = \begin{cases} \phi(x), & x < -1 \\ \phi(x+1), & x \geq -1 \end{cases} \tag{33}$$

then which one of the following statements is true?

1. $P(X \leq -1) = \frac{1}{2}$
2. $P(X = -1) = \frac{1}{2}$
3. $P(X < -1) = \frac{1}{2}$
4. $P(X \leq 0) = \frac{1}{2}$

# C-code I

```
#include <stdio.h>
#include <math.h>
#define M_PI 3.14159265358979323846
double cdf(double x) {
    if (x < -1) {
        return 0.5 * (1.0 + erf(x / sqrt(2.0)));
    }
    else {
        return 0.5 * (1.0 + erf(x+1 / sqrt(2.0)));
    }
}

double pdf(double x) {
```

# C-code II

```
    if (x < -1) {
        return 1.0 / (sqrt(2.0 * M_PI)) * exp(-x * x / 2.0);
    } else {
        return 1.0 / (sqrt(2.0 * M_PI)) * exp(-(x + 1) * (x + 1) / 2.0);
    }
}

int main() {
    int num = 10000;
    double x_max = 4.0;
    double x_min = -4.0;
    double step = (x_max - x_min) / num;

```

# C-code III

```
    FILE *outfile = fopen("uni.dat", "w");

    for (int i = 0; i < num; i++) {
        double x = x_min + step * i;
        double sample = pdf(x);
        double sample1 = cdf(x);
        fprintf(outfile, "%lf %lf %lf\n", x,
            sample, sample1);
    }

    fclose(outfile);

    return 0;
}
```

# Python-code I

```
import numpy as np
import matplotlib.pyplot as plt

# Read data from the "uni.dat" file
data = np.genfromtxt("uni.dat")

x = data[:, 0]  # Use correct index for columns (0 for x)
pdf = data[:, 1]  # Use correct index for columns (1 for pdf)
cdf = data[:, 2]

# Create a histogram plot of the PDF
plt.figure()
plt.plot(x, pdf)
```

# Python-code II

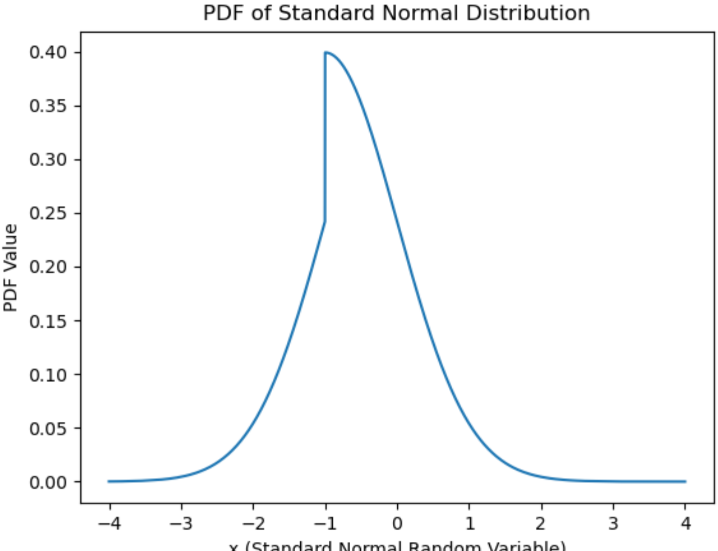
```
plt.xlabel('x (Standard Normal Random Variable)')
plt.ylabel('PDF Value')
plt.title('PDF of Standard Normal Distribution')
plt.savefig('/home/mayank/EE23010/KHP/figs/figure2.png')
plt.show()

plt.figure()
plt.plot(x, cdf)
plt.xlabel('x (Standard Normal Random Variable)')
plt.ylabel('PDF Value')
```

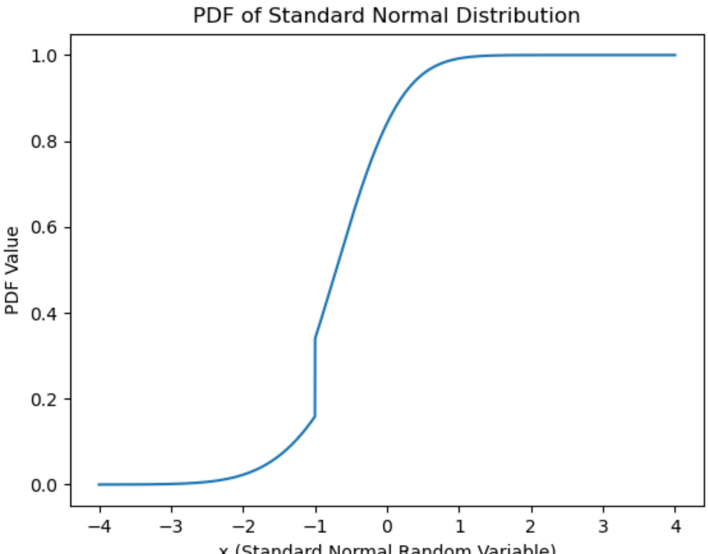
# Python-code III

```
plt.title('PDF of Standard Normal Distribution
    ')
plt.savefig('/home/mayank/EE23010/KHP/figs/
    figure3.png')
plt.show()
```

# PDF of X


PDF of Standard Normal Distribution
PDF Value
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−3
−2
−1
0
1
2
3
4
x (Standard Normal Random Variable)

# CDF of X


PDF of Standard Normal Distribution
1.0
0.8
0.6
0.4
0.2
0.0
PDF Value
−4
−3
−2
−1
0
1
2
3
4
x (Standard Normal Random Variable)