PRACTICAL 5 (Continuous probability distribution)

This practical introduces the fundamental concepts of probability density functions (PDFs) and the normal distribution. It begins with verifying whether a function qualifies as a valid PDF, followed by practical applications for calculating probabilities. It also explores the properties of the normal distribution.

Part 1 : Checking if f(x) is a pdf

Check if the following function is a pdf,

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Steps:

1. Define the function
2. Check if the function is non negative on the domain
3. Check if the integral of the function over the domain is unity.

```
# Define the custom PDF
pdf_custom <- function(x) {
  if (x >= 0) {
    return((1/2) * exp(-x/2))
  } else {
    return(0)
  }
}


# Vectorize the PDF function, this is used to convert a scalar function into a vectorized
# function so that it can accept and process vectors as inputs.

pdf_custom <- Vectorize(pdf_custom)

# Generate a sequence of values for x
# You can choose an appropriate range based on the domain of your PDF
x_values <- seq(-10, 10, length.out = 1000)

# Evaluate the PDF over the sequence of x values
pdf_values <- pdf_custom(x_values)

# Check for non-negativity
# The all() function in R is used to check whether all elements of a logical vector are TRUE.
if (all(pdf_values >= 0)) {
  print("The function is non-negative over the specified domain.")
```

```
} else {
  print("The function is negative at some points in the specified domain.")
}
# Integrate the PDF from 0 to infinity
result <- integrate(pdf_custom, lower = 0, upper = Inf)

# Print the result
print(paste("The integral of the PDF from 0 to infinity is:", result$value))
```

Part 2 : finding probability

Suppose that in a certain region the daily rainfall (in inches) is a continuous RV X with p.d.f f(x) given by

$$f(x) = \frac{3}{4}(2x - x^2), 0 < x < 2 \;\; \& \;\; f(x) = 0 \; elsewhere$$

Find the probability that on a given day in this region, the rainfall is:

i. not more than 1 inch
ii. greater than 1.5 inches
iii. between 0.5 and 1.5 inches

```
# Define the PDF
pdf_rainfall <- function(x) {
  if (x >= 0 && x <= 2) {
  return(3/4 * (2*x - x^2))
  } else {
    return(0)
  }
}

# Vectorize the PDF function
pdf_rainfall <- Vectorize(pdf_rainfall)


# Integrate the PDF from 0 to 1
prob_not_more_than_1 <- integrate(pdf_rainfall, lower = 0, upper = 1)$value
print(paste("Probability that rainfall is not more than 1 inch:", prob_not_more_than_1))


# Integrate the PDF from 1.5 to 2
prob_greater_than_1_5 <- integrate(pdf_rainfall, lower = 1.5, upper = 2)$value
print(paste("Probability that rainfall is greater than 1.5 inches:", prob_greater_than_1_5))


# Integrate the PDF from 0.5 to 1.5
prob_between_0_5_and_1_5 <- integrate(pdf_rainfall, lower = 0.5, upper = 1.5)$value
```

```
print(paste("Probability that rainfall is between 0.5 and 1.5 inches:",
prob_between_0_5_and_1_5))
```

Part 3: Normal Distribution

Normal Distribution

The normal distribution, also known as the Gaussian distribution, is one of the most important probability distributions in statistics. It is symmetric, bell-shaped, and is defined by its mean (μ) and standard deviation (σ). The normal distribution describes many natural phenomena, such as heights, test scores, and measurement errors.

- ✓ Mean (μ): The average or central value of the distribution.
- ✓ Standard Deviation (σ): A measure of the spread or dispersion of the distribution.
- ✓ Z-score: A standardized value that indicates how many standard deviations a data point is from the mean.
- ✓ Standard Normal Distribution: A special case of the normal distribution where the mean is 0 and the standard deviation is 1.

R Functions for Normal Distribution

1. dnorm() - Probability Density Function
The dnorm() function returns the height of the probability density function for a given Z-score, mean, and standard deviation, using the syntax **dnorm(Z, mean, sd).**

2. pnorm() - Cumulative Distribution Function
The pnorm() function gives the cumulative probability up to a given Z-score, mean, and standard deviation, using the syntax **pnorm(Z, mean, sd).**

3. qnorm() - Quantile Function
The qnorm() function returns the Z-score (quantile) for a given cumulative probability, mean, and standard deviation, using the syntax **qnorm(Z, mean, sd).**

Examples Using R

Example 1: Calculating Cumulative Probability

- ✓ To calculate the standard normal cumulative probability for Z = 1, use pnorm(1).
- ✓ To calculate the cumulative probability for a normal distribution with a mean of 45 and a standard deviation of 10 at Z = 140, use pnorm(140, 45, 10).

Example 2: Z-transformation with Real Data
Let's use the women dataset, which contains the height and weight of women.

PRACTICAL 5 (Continuous probability distribution)

Start by loading the data and viewing the first few rows with

```
data("women")
head(women).
```

Calculate the mean and standard deviation of the height with

```
height.mean <- mean(women$height)
height.sd <- sd(women$height).
```

Find the probability of having height upto 68 cm

```
x <- 68
```

calculate the cumulative probability by specify the mean and standard deviation using

```
pnorm(x, mean = height.mean, sd = height.sd).
```

Visualization:Plotting the Normal Distribution

Define the sequence of x values with

```
x <- seq(-10, 10, by = .1).
```

Calculate the corresponding y values for a normal distribution

```
y <- dnorm(x, mean = 2.5, sd = 0.5).
```

Plot the normal distribution

```
plot(x, y, type="l", col="blue").
```

Standard Normal Distribution

Define the sequence of x values

```
x = seq(-4, 4, length = 1000).
```

Calculate the corresponding y values for a standard normal distribution

```
y = dnorm(x).
```

Plot the standard normal distribution

```
plot(x, y, type="l", col="red").
```

Excercise

1. If f(x) is a density function then find i) c, ii) p(1 < x < 2)

$$f(x) = \begin{cases} cx^2, & 0<x<3 \\ 0, & \text{otherwise} \end{cases}$$

2. For a continuous random variable x, following is the pdf . Find k, mean and variance.

$$f(x) = kx^2 e^{-x}, x \geq 0$$

PRACTICAL 5 (Continuous probability distribution)

3. For a continuous random variable x, following is the pdf. find P(1/4<x<3/4)

$$f(x) = \begin{cases} x^3, 0 \le x \le 1 \\ 3(2-x)^3, 1 \le x \le 2 \end{cases}$$

4. In a sample of 1000 cases, the mean of a certain test is 14 and standard deviation is 2.5. Assuming the distribution to be normal, find
   i. How many students score between 12 and 15?
   ii. How many score above 18?
   iii. How many score below 8?
   Write a R program for the above problem. Also plot the graph for each case.

5. In a male population of 1000, the mean height is 68.16 inches and standard deviation is 3.2 inches. How many men may be more than 6 feet (72 inches)? Write a R program for the above problem. Also plot the graph.