# Probability Distributions & Simulation

# Probability distributions

A **probability distribution** provides a comprehensive descri ption of how probabilities are assigned to various outcomes of a random experiment with all the probabilities summing upto 1.

**Two major types:**

**Discrete distributions**: take countable values (e.g., Binomial, Poisson)

**Continuous distributions**: take values over an interval (e.g., Normal)

R provides a rich set of built-in functions for probability, simulation, and visualization.

# Probability mass function and probability density function

Probability mass function is the probability distribution of a discrete random variable, and provides the possible values and their associated probabilities denoted by $p_X(\text{x})=\text{P}(\text{X}=\text{x})$.

Probability density function is the probability density function of a continuous random variable and is denoted by $f(x)$. The probability is given by the integral of the pdf over that range.

# Binomial distribution

Binomial distribution is a discrete distribution with parameters n and p

n- no of independent Bernoulli trials

p-probability of success

Binomial distribution models no of successes and the probability mass function is

$P(X = x) = nC_x p^x q^{n-x}, x = 0,1,2, \dots n$

# Examples

1. A fair coin is tossed **10 times**. Find the probability of getting **exactly 6 heads**.
2. A die is thrown 12 times. Find the probability of getting exactly 4 sixes.
3. If 40% of emails are spam, find the probability that exactly 5 out of 10 emails are spam.

# Code with examples

$P(X = x)$

dbinom(x=3 , size=10 , prob=0.4)

x denotes no of sucesses, size=size of the sample, prob is the probability of success

$P(X \leq x)$ : to find cumulative probability

sum(dbinom(0:3, 10 , 0.4))

pbinom(3, 10,0.4)

sum(dbinom(0:10, size = 10, prob = 0.4))

pbinom(10, 10,0.4) to check sum of prob is 1

To find the value of x when probability is given

qbinom(0.95, size = 10, prob = 0.4)

## Generates random samples from Binomial distribution

rbinom(n = 10, size = 10, prob = 0.4)  generates 10 random values with probability of success 0.4 in between 0 to 10(size of the distribution)

x<-rbinom(n = 10, size = 10, prob = 0.4) (assigning to x)

x

# Plotting in Binomial distribution

```
x <- 0:10

y <- dbinom(x, size = 10, prob = 0.4)

barplot(y, names.arg = x, main = "Binomial Distribution (n=10, p=0.4)", xlab = "x", ylab = "P(X = x)")

plot(x,y,main="Binomial",xlab = "x", ylab = "P(X = x)")
```

# Comparison between exact & simulated value

dbinom(3, 10, 0.4)----Probability of getting exactly 3 successes out of 10 trials.

Assuming x is generated using rbinom

x==3 creates a logical vector

mean(x == 3) proportion of times the value 3 occurs in the simulation

First one is the exact probability and the second one is the approximate probability from simulation

# Simulation using rbinom

```
n <- 1000
x <- rbinom(n, size=10, prob=0.4)
table(x)/n
```

# Poisson distribution

The **Poisson distribution** is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time with constant mean.

The probability mass function is

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, x = 0,1,2,3, \ldots$$

$\lambda = np$ is the mean of the distribution, e = 2.71828…

n is very large

p is very small

$\lambda = np$ is finite and a constant

Mean=Variance= $\lambda = np$

# Examples

- Number of accidents at a traffic junction per day
- Number of defects in a roll of fabric
- Number of customers arriving at a bank per minute

# Coding with examples

There are 4 standard functions in R for Poisson

dpois(x,lambda)---- for $P(X = x)$

ppois(x,lamba)----for $P(X \leq x)$

dpois(3, lambda = 2)

ppois(3, lambda = 2) or sum(dpois(0:3,lambda=2))

qpois(0.95,2)

# Generating values

rpois(5, lambda = 2) – generates 5 values

set.seed(123) ---- This gives the same output everytime (We can use any integer inside seed)

rpois(5, lambda = 2)

x <- rpois(10, lambda = 2)

X

Note: rpois command is used for simulation (generating values)

# Plotting in Poisson distribution

```
x <- 0:10

prob <- dpois(x, lambda = 3)

barplot(prob, names.arg = x,  xlab = "Number of occurrences (x)",  ylab = "Probability", main = "Poisson Distribution (λ = 3)")  # creates a barplot
```

**To create an usual plot**

```
plot(x,prob, xlab = "Number of occurrences (x)", ylab = "Probability",main = "Poisson Distribution (λ = 3)")
```

**To create a histogram**

```
plot(x, prob, type = "h", lwd = 2, xlab = "x",ylab = "P(X = x)",
    main = "Poisson Distribution (λ = 3)")
points(x, prob, pch=16)
```

**To create a line plot**

```
plot(x, prob,  type = "l",  lwd = 2, xlab = "x", ylab = "P(X = x)",
main = "Poisson Distribution (λ = 3)")
points(x, prob, pch = 16)
```

**To plot only points**

```
plot(x, prob,type = "p", pch = 16,xlab = "x",ylab = "P(X = x)",
main = "Poisson Distribution (λ = 3)")
#lwd is linewidth
```

# Normal distribution

Normal distribution is a continuous distribution

with pdf $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$

$\mu$ is the mean of the distribution

$\sigma$ is the standard deviation of the distribution

**Key properties:**

- Symmetric, bell-shaped curve
- Mean = Median = Mode
- Total area under curve = 1

# Standard normal variate

**Definition**

If

$$X \sim N(\mu, \sigma^2)$$

then the **standard normal variate** $Z$ is defined as

$$Z = \frac{X - \mu}{\sigma}$$

Here,

- Mean of $Z = 0$
- Standard deviation of $Z = 1$

So,

$$Z \sim N(0, 1)$$

Suppose the heights of students are normally distributed with $\mu = 170$ cm and $\sigma = 10$ cm.

Find the standard normal variate for a student of height **185 cm**.

$$Z = \frac{185 - 170}{10} = 1.5$$

This means the height is **1.5 standard deviations above the mean**.

# Examples

**Problem 1**

The mean and standard deviation of a normal distribution are 50 and 10 respectively.

Find $P(X < 60)$.

**Problem 2**

If $X \sim N(100, 15^2)$, find the value of $P(X > 130)$.

**Problem 3**

The heights of students are normally distributed with mean 165 cm and standard deviation 8 cm.

Find the probability that a student's height lies between 160 cm and 175 cm.

# Coding with examples

There are 4 standard functions in R for Normal distribution

dnorm(55, mean = 50, sd = 10)—it gives the value of f(55)

pnorm(60, mean = 50, sd = 10)—gives the cumulative prob till x=60 (ie) $P(X \leq x)$ left tail probability

qnorm(0.8413447,mean = 50, sd = 10)– it gives the value of x when cumulative prob is known

Right tail probability (ie) $P(X > x)$

1-pnorm(60, mean = 50, sd = 10)

(or)

pnorm(60, mean = 50, sd = 10, lower.tail = FALSE)

To find P(40<X<60)

pnorm(60, mean = 50, sd = 10) - pnorm(40, mean = 50, sd = 10)

# Generating values

rnorm(5,mean=0,sd=1)—generates 5 random values in $-\infty < x < \infty$.

set.seed(123) will generate same set of values

rnorm(5,mean=0,sd=1)

# Plotting in normal distribution

**Basic normal curve**

```
x <- seq(-4, 4, length = 100)
y <- dnorm(x, mean = 0, sd = 1)
plot(x, y, type = "p",  pch = 16,  xlab = "x", ylab = "f(x)",
    main = "Normal Distribution (Points)")
```

**Plot using points**

```
plot(x, y, type = "p",  pch = 16, xlab = "x", ylab = "f(x)",
    main = "Normal Distribution (Points)")
```

**Plot using both**

```
plot(x, y, type = "b", pch = 16, lwd = 2, xlab = "x", ylab =
"f(x)",main = "Normal Distribution (Line + Points)")
```

# Plotting Normal distribution from simulated data

**To plot histogram**

```
set.seed(123)
data <- rnorm(1000, mean = 50, sd = 10)
hist(data, probability = TRUE,  xlab = "x",  main = "Histogram of Normal Data")
curve(dnorm(x, mean = mean(data), sd = sd(data)),
      add = TRUE, lwd = 2)
```

```
x <- seq(-10, 10, length = 200)
plot(x, dnorm(x, 0, 1), type = "l", lwd = 2,
     xlab = "x", ylab = "Density",
     main = "Normal Distribution for Different
Parameters")
```

# Code to draw a normal curve and to shade P(40<X<60)

```
mu <- 50;sigma <- 10;a <- 40;b <- 60
x <- seq(mu - 4*sigma, mu + 4*sigma, length = 1000)
y <- dnorm(x, mean = mu, sd = sigma)
plot(x, y, type = "l", lwd = 2, main = "Normal Distribution with Shaded Region", xlab = "x", ylab = "Density")
x_shade <- seq(a, b, length = 500)
y_shade <- dnorm(x_shade, mean = mu, sd = sigma)
polygon(c(a, x_shade, b),c(0, y_shade, 0),col = "lightblue", border = NA)
abline(v = c(a, b), col = "red", lwd = 2, lty = 2)
prob <- pnorm(b, mu, sigma) - pnorm(a, mu, sigma)
prob
```

# Questions for practice- Binomial and Poisson distributions

1. A coin is tossed 8 times. Let X be the number of heads. Find a) $P(X = 5)$ b) $P(X \geq 6)$
2. The probability that a student passes an exam is 0.7. If 5 students appear, find P(exactly 3 pass) and P(at least 4 pass)

3. Write an R program to find and plot the binomial distribution for n=10, p=0.5.

4. The number of typing errors per page follows a Poisson distribution with mean 2. Find the probability that a page contains a) exactly 1 error b) at least 3 errors.

5. Write an R program to find P(2<X≤ for a Poisson distribution with mean 4.

# Questions for practice-Normal distribution

1. The heights of students are normally distributed with mean 165 cm and standard deviation 6 cm.
   Find the probability that a randomly chosen student has height
   (a) less than 160 cm
   (b) between 160 cm and 175 cm.

2. Let $X \sim N(60, 8^2)$. Find $P(52 < X < 68)$.

3. For a standard normal variable Z, find

*a)* $P(-1.5 < Z < 2.0)$

*b)* $P(Z > 1.96)$

4. Write an R program to generate normal dat,compute probabilities and visualize shaded regions for different values of a and b.