Probability

Probability

Probability is a measure of the likelihood that an event will occur. It ranges from 0 to 1, where 0 indicates impossibility and 1 indicates certainty.

```
# Define the sample space and favorable outcomes
sample_space <- 1:6
favorable_outcomes <- c(5, 6)

# Calculate the probability
probability <- length(favorable_outcomes) / length(sample_space)
print(probability)
```

```
## [1] 0.3333333
```

Event

An event is a set of outcomes of an experiment to which a probability is assign.

```
total_outcomes <- 52  # Total cards in a deck
favorable_outcomes <- 26  # Red cards (hearts + diamonds) which is an event.
```

Conditional Probability

Conditional probability is the probability of an event occurring given that another event has already occurred.It is denoted as (P(A|B)) and is calculated as:

P(A|B) = P(A B) / P(B).

where (P(A B)) is the probability of both events A and B occurring.

```
# Total cards in the deck
total_cards <- 52

# Define the events
face_cards <- 12  # 3 face cards (J, Q, K) in each suit, total of 12
red_face_cards <- 6  # 3 face cards each in hearts and diamonds (red suits)

# Probability of B (drawing a face card)
prob_B <- face_cards / total_cards

# Probability of A   B (drawing a red face card)
```

```
prob_A_and_B <- red_face_cards / total_cards

# Conditional probability P(A|B)
prob_A_given_B <- prob_A_and_B / prob_B
print(prob_A_given_B)
```

```
## [1] 0.5
```

Union and Intersection of Events

1. **Union**:The union of two events A and B,denoted (AUB),is the event that either A or B or both occur. Its probability is given by:

P(A B) = P(A) + P(B) − P(A B).

```
# Define events
A <- c(2, 4, 6)  # Even numbers
B <- c(5, 6)     # Numbers greater than 4
sample_space <- 1:6

# Calculate probabilities
P_A <- length(A) / length(sample_space)
P_B <- length(B) / length(sample_space)
P_A_and_B <- length(intersect(A, B)) / length(sample_space)

# Union probability
P_A_or_B <- P_A + P_B - P_A_and_B
print(P_A_or_B)
```

```
## [1] 0.6666667
```

2. **Intersection**:The intersection of two events A and B, denoted (A B),is the event that both A and B occur.Its probability is:

P(A B) = P(A) + P(B) − P(AUB).

```
# Intersection probability
P_A_and_B <- length(intersect(A, B)) / length(sample_space)
print(P_A_and_B)
```

```
## [1] 0.1666667
```

Complement

The complement of an event A, denoted (A') or (A^c),is the event that A does not occur. Its probability is:

P(A) = 1 − P(A) .

```r
# Complement probability
P_A_complement <- 1 - P_A
print(P_A_complement)
```

```
## [1] 0.5
```

Random Variable

A random variable is a variable that takes on numerical values determined by the outcome of a random phenomenon.

```r
# Outcomes and probabilities
x <- c(1, 2, 3, 4)
p <- c(0.1, 0.2, 0.4, 0.3)
```

Discrete Random Variable

A discrete random variable takes on a countable number of distinct values. Examples include the number of heads in a series of coin flips.

1. Cumulative Distribution Function (CDF)

The CDF of a discrete random variable (X) is:

```r
# CDF calculation
cdf <- cumsum(p)
cdf_table <- data.frame(x, CDF = cdf)
print(cdf_table)
```

```
##   x CDF
## 1 1 0.1
## 2 2 0.3
## 3 3 0.7
## 4 4 1.0
```

2. Mean (Expected Value)

The mean of (X) is (E(X) = sum x_i / length(x_i)).For discrete distribution we have:

```r
# Mean calculation
mean_X <- sum(x * p)
print(mean_X)
```

```
## [1] 2.9
```

3. Variance

The variance of (X) is (Var(X) = E[(X - E(X))^2]).

```r
# Variance calculation
variance_X <- sum(x^2 * p) - mean_X^2
print(variance_X)
```

```
## [1] 0.89
```

Continuous Random Variable

A continuous random variable takes on an infinite number of possible values.

1. Cumulative Distribution Function (CDF)

The CDF of a continuous random variable (X) is:

```r
# Parameters for the uniform distribution
a <- 0  # Lower bound
b <- 1  # Upper bound

# Define the PDF (for uniform distribution between 0 and 1)
pdf <- function(x) dunif(x, min = a, max = b)

# Define the CDF (for uniform distribution between 0 and 1)
cdf <- function(x) punif(x, min = a, max = b)

# CDF at x = 0.5
cdf_value <- cdf(0.5)
print(cdf_value)
```

```
## [1] 0.5
```

2. Mean (Expected Value)

The mean of (X) is (E(X) = integral of x*f(x)dx),where (f(x)) is the probability density function (PDF).

```r
# Mean of the uniform distribution
mean_X <- (a + b) / 2
print(mean_X)
```

```
## [1] 0.5
```

3. Variance

The variance of (X) is (Var(X) = E[(X-E(X))^2]).

```r
# Variance of the uniform distribution
variance_X <- (b - a)^2 / 12
print(variance_X)
```

```
## [1] 0.08333333
```

Shape, Skewness, and Modality

1. **Shape** of a distribution can be visualized using histograms or density plots in R.

```
# Install and load e1071 package if not installed
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.4.2
```

```
# Example data: Normally distributed data
data <- rnorm(1000)

# Calculate skewness
skewness_value <- skewness(data)
print(skewness_value)
```

```
## [1] 0.01032179
```

2. **Skewness** measures the asymmetry of the probability distribution.In R, you can calculate skewness using the skewness() function from the e1071 package.

```
# Calculate kurtosis
kurtosis_value <- kurtosis(data)
print(kurtosis_value)  # Output: near 3 for normal distribution
```

```
## [1] -0.1531115
```

3. **Modality** refers to the number of peaks in a distribution.A distribution can be unimodal (one peak), bimodal (two peaks), etc.

```
library(mclust)
```

```
## Warning: package 'mclust' was built under R version 4.4.2
```

```
## Package 'mclust' version 6.1.1
## Type 'citation("mclust")' for citing this R package in publications.
```

```
# Fit a Gaussian mixture model (GMM) to estimate modality
gmm_model <- Mclust(data)
print(gmm_model$G)
```

```
## [1] 1
```