STAT 3011 Discussion 015

Introduction to R: Week 5

Talha Hamza
University of Minnesota

Spring 2025

Probability Formulas

General Addition Property of Probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Independence Rule:
If events A and B are independent, then:

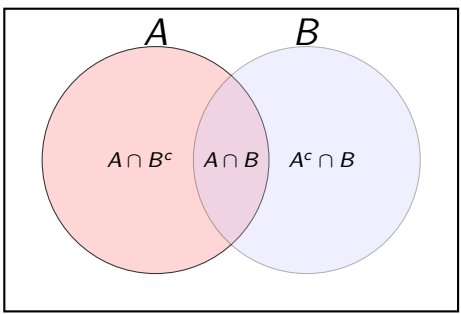
$$P(A \cap B) = P(A) \cdot P(B)$$

Conditional Probability Formula:
Read as Probability of event A occuring GIVEN that B has already occurred

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{where } P(B) > 0$$

Partitioning of Probability

Using a Venn Diagram to Visualize the General Addition Rule:



Partitioning Rule:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

Discrete Probability Distribution

- A discrete probability distribution lists all possible values a random variable can take and their associated probabilities.
- The sum of all probabilities must be 1.

Example: Dice Game

- Let X = winnings (\$) from a game where you roll a fair six-sided die.
- Payout rule:
  - Roll a 6 $\rightarrow$ Win \$10.
  - Roll a 4 or 5 $\rightarrow$ Win \$5.
  - Roll 1, 2, or 3 $\rightarrow$ Win \$0.

X (Winnings in \$)	P(X)
0	$3/6 = 0.50$
5	$2/6 = 0.33$
10	$1/6 = 0.17$

Expected Value

- The expected value E(X) represents the **long-run average outcome.**
- Formula:

$$E(X) = \sum xP(X = x)$$

Example: Dice Game

X (Winnings in $)	P(X)
0	$3/6 = 0.50$
5	$2/6 = 0.33$
10	$1/6 = 0.17$

Calculation:

$$E(X) = (0 \times 0.50) + (5 \times 0.33) + (10 \times 0.17)$$

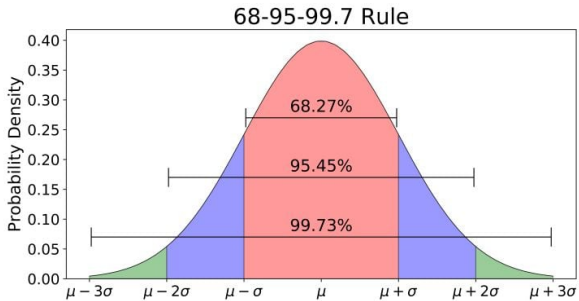
$$= 0 + 1.65 + 1.67 = \$3.33$$

Understanding the Standard Normal Distribution

- The **normal distribution** is a continuous probability distribution defined by:
  - **Mean** (μ): The center of the distribution.
  - **Standard deviation** (σ): Measures data spread.
- The **standard normal distribution** is a special case where:
  - $\mu = 0$, $\sigma = 1$
  - Denoted as: $Z \sim N(0, 1)$

The 68-95-99.7 Rule (Empirical Rule)

- This rule describes data distribution in a normal curve:
  - **68%** of data falls within $\pm 1\sigma$.
  - **95%** falls within $\pm 2\sigma$.
  - **99.7%** falls within $\pm 3\sigma$.
- Useful for estimating probabilities and detecting outliers.



Questions? Let's Discuss!