

# Week 2: Advanced Probability Concepts

---

## 1. Introduction to Probability

---

Probability is a fundamental concept in statistics and data science, dealing with the likelihood of events occurring. It provides a mathematical framework for quantifying uncertainty. Understanding probability is crucial for making informed decisions, especially when dealing with data that involves randomness or incomplete information [1].

### Key Terminology

- **Experiment:** A process or observation that has an uncertain outcome.
- **Outcome:** A single possible result of an experiment.
- **Sample Space (S):** The set of all possible outcomes of an experiment.
- **Event (E):** A subset of the sample space, consisting of one or more outcomes.
- **Probability of an Event P(E):** A numerical value between 0 and 1, inclusive, representing the likelihood of an event occurring.  $P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$ .

## 2. Simple Probability Distributions

---

Probability distributions describe how probabilities are distributed over the values of a random variable. They can be broadly categorized into discrete and continuous distributions.

### 2.1. Discrete Probability Distributions

Discrete probability distributions deal with random variables that can take on a finite or countably infinite number of distinct values. The probability mass function (PMF) describes the probability of each possible value.

### 2.1.1. Bernoulli Distribution

The Bernoulli distribution models a single trial of an experiment with only two possible outcomes: success (usually denoted as 1) or failure (usually denoted as 0). The probability of success is  $p$ , and the probability of failure is  $1 - p$ .

**Parameters:** \*  $p$ : Probability of success ( $0 \leq p \leq 1$ )

**Probability Mass Function (PMF):**

$$P(X = k) = p^k(1 - p)^{1-k} \quad \text{for } k \in \{0, 1\}$$

**Mean (Expected Value):**  $E[X] = p$

**Variance:**  $Var(X) = p(1 - p)$

**Example:** Flipping a fair coin once. Success = Heads ( $k = 1$ ), Failure = Tails ( $k = 0$ ).  $p = 0.5$ .

### 2.1.2. Binomial Distribution

The binomial distribution models the number of successes in a fixed number of independent Bernoulli trials. It is applicable when there are only two outcomes for each trial, the number of trials is fixed, and the probability of success is constant for each trial.

**Parameters:** \*  $n$ : Number of trials \*  $p$ : Probability of success on a single trial

**Probability Mass Function (PMF):**

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k \in \{0, 1, \dots, n\}$$

Where  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  is the binomial coefficient, representing the number of ways to choose  $k$  successes from  $n$  trials.

**Mean (Expected Value):**  $E[X] = np$

**Variance:**  $Var(X) = np(1 - p)$

**Example:** If you flip a fair coin 10 times, what is the probability of getting exactly 7 heads? Here,  $n = 10$ ,  $p = 0.5$ ,  $k = 7$ .

$$P(X = 7) = \binom{10}{7} (0.5)^7 (0.5)^{10-7} = 120 \times (0.5)^7 \times (0.5)^3 = 120 \times 0.0078125 :$$

### 2.1.3. Poisson Distribution

The Poisson distribution models the number of events occurring in a fixed interval of time or space, given a constant average rate of occurrence and that these events occur independently of the time since the last event.

**Parameters:** \*  $\lambda$ : Average rate of events in the interval ( $\lambda > 0$ )

**Probability Mass Function (PMF):**

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{for } k \in \{0, 1, 2, \dots\}$$

**Mean (Expected Value):**  $E[X] = \lambda$

**Variance:**  $Var(X) = \lambda$

**Example:** A call center receives an average of 5 calls per hour. What is the probability of receiving exactly 3 calls in the next hour? Here,  $\lambda = 5$ ,  $k = 3$ .

$$P(X = 3) = \frac{e^{-5} 5^3}{3!} = \frac{0.006738 \times 125}{6} \approx 0.1404$$

## 2.2. Continuous Probability Distributions

Continuous probability distributions deal with random variables that can take on any value within a given range. The probability density function (PDF) describes the relative likelihood for the random variable to take on a given value.

### 2.2.1. Uniform Distribution

The uniform distribution describes a situation where all outcomes within a given interval are equally likely. It is often used as a baseline or when there is no prior information about the likelihood of different outcomes.

**Parameters:** \*  $a$ : Minimum value \*  $b$ : Maximum value ( $a < b$ )

**Probability Density Function (PDF):**

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

**Mean (Expected Value):**  $E[X] = \frac{a+b}{2}$

**Variance:**  $Var(X) = \frac{(b-a)^2}{12}$

**Example:** A random number generator produces numbers between 0 and 1. The probability of any specific number is 0, but the probability of a number falling within a certain range is proportional to the length of that range.

### 2.2.2. Normal (Gaussian) Distribution

The normal distribution is one of the most important and widely used continuous probability distributions. It is characterized by its bell-shaped curve and symmetry around its mean. Many natural phenomena follow a normal distribution, and it is central to the Central Limit Theorem.

**Parameters:** \*  $\mu$ : Mean (location parameter) \*  $\sigma$ : Standard deviation (scale parameter) ( $\sigma > 0$ )

**Probability Density Function (PDF):**

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}$$

**Mean (Expected Value):**  $E[X] = \mu$

**Variance:**  $Var(X) = \sigma^2$

**Example:** Heights of adult males, measurement errors, blood pressure readings. The empirical rule states that for a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, 95% within two, and 99.7% within three.

## 3. Conditional Probability

---

Conditional probability is the probability of an event occurring given that another event has already occurred. It is a fundamental concept for understanding relationships between events and is crucial for Bayesian inference and machine learning algorithms.

## Definition

The conditional probability of event A occurring given that event B has occurred is denoted as  $P(A|B)$  and is calculated as:

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

Where: \*  $P(A \cap B)$  is the probability of both events A and B occurring (their intersection). \*  $P(B)$  is the probability of event B occurring.

## Independent Events

Two events A and B are **independent** if the occurrence of one does not affect the probability of the other. In this case:

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

And consequently, the probability of both events occurring is:

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

## Dependent Events

If two events are not independent, they are **dependent**. The occurrence of one event influences the probability of the other.

## Bayes' Theorem

Bayes' Theorem is a powerful tool for updating probabilities based on new evidence. It describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

This can be expanded using the law of total probability for  $P(B)$ :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Where  $A^c$  is the complement of event A (i.e., A does not occur).

**Example:** A medical test for a disease has a 99% accuracy rate (i.e.,  $P(\text{Positive}|\text{Disease}) = 0.99$  and  $P(\text{Negative}|\text{No Disease}) = 0.99$ ). The prevalence of the disease in the population is 0.1% ( $P(\text{Disease}) = 0.001$ ). If a person tests positive, what is the probability that they actually have the disease?

Let  $D$  = Disease,  $D^c$  = No Disease,  $T+$  = Test Positive,  $T-$  = Test Negative.

We want to find  $P(D|T+)$ .

Given:  $P(T+|D) = 0.99$   $P(T-|D^c) = 0.99 \implies P(T+|D^c) = 1 - 0.99 = 0.01$   $P(D) = 0.001 \implies P(D^c) = 1 - 0.001 = 0.999$

Using Bayes' Theorem:

$$P(D|T+) = \frac{P(T+|D)P(D)}{P(T+|D)P(D) + P(T+|D^c)P(D^c)}$$

$$P(D|T+) = \frac{0.99 \times 0.001}{(0.99 \times 0.001) + (0.01 \times 0.999)}$$

$$P(D|T+) = \frac{0.00099}{0.00099 + 0.00999} = \frac{0.00099}{0.01098} \approx 0.09016$$

This means even with a positive test, the probability of actually having the disease is only about 9%, highlighting the importance of understanding conditional probabilities and base rates.

## References

- 
- [1] Khan Academy. (n.d.). *Basic probability*. Retrieved from <https://www.khanacademy.org/math/statistics-probability/probability-library> [2] Stat Trek. (n.d.). *Binomial Distribution*. Retrieved from <https://stattrek.com/probability/binomial-distribution.aspx> [3] Investopedia. (n.d.). *Poisson Distribution*. Retrieved from <https://www.investopedia.com/terms/p/poisson-distribution.asp> [4] Simply Psychology. (n.d.). *Normal Distribution*. Retrieved from <https://www.simplypsychology.org/normal-distribution.html> [5] Towards Data Science. (n.d.). *Understanding Conditional Probability*. Retrieved from <https://towardsdatascience.com/understanding-conditional-probability-a-data-science-perspective-e2311232230>