Probability Theory and Statistical Distributions

Data Science with Statistical Foundations

Learning Outcomes

After this lecture students should be able to

- Understand probability as a measure of uncertainty
- Apply addition and multiplication rules
- Compute conditional probabilities
- Apply Bayes' theorem to inference problems
- Understand discrete and continuous random variables
- Identify and use Binomial and Normal distributions
- Appreciate why probability is fundamental to Data Science

Why Probability in Data Science?

Almost every data science algorithm deals with uncertainty.
Examples:

- Spam email detection
- Medical diagnosis
- Weather forecasting
- Credit risk assessment
- Recommendation systems
- Fault detection

Probability provides a mathematical framework for making decisions under uncertainty.

Basic Terminology

Experiment
A repeatable process whose outcome is uncertain.

Examples:

- Tossing a coin
- Rolling a die
- Predicting customer purchase

Sample Space
Set of all possible outcomes.

$$S = \{H, T\}$$

$$S = \{1, 2, 3, 4, 5, 6\}$$

Events

An event is any subset of the sample space.
Example:
Rolling one die

$$A = \{\text{Even numbers}\} = \{2, 4, 6\}$$

$$B = \{\text{Numbers greater than } 3\} = \{4, 5, 6\}$$

Important operations

- Union $(A \cup B)$
- Intersection $(A \cap B)$
- Complement (A^c)

Axioms of Probability

For every event A

$$0 \leq P(A) \leq 1$$

$$P(S) = 1$$

For mutually exclusive events

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

These three axioms form the foundation of probability theory.

Addition Theorem

General case

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

Example

$$P(A) = 0.5, \quad P(B) = 0.6, \quad P(A \cap B) = 0.2$$

Then

$$P(A \cup B) = 0.9$$

The intersection is subtracted because it is counted twice.

Multiplication Rule

Definition

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

Similarly

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

This theorem is useful whenever events occur sequentially, e.g., drawing two cards without replacement.

Conditional Probability

Conditional probability measures probability after observing new information.

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

where

$$P(B) > 0$$

Examples

- Disease given positive test
- Rain given cloudy weather
- Purchase given advertisement clicked

Independent Events

Events are independent if

$$P(A|B) = P(A)$$

Equivalent condition

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

- Tossing two coins
- Rolling two independent dice

Counterexample: Drawing cards without replacement.

Bayes' Theorem

Bayes' theorem updates prior belief using observed evidence.

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

where

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Applications

- Machine learning
- Medical diagnosis
- Spam filtering
- Fault diagnosis
- Bayesian inference

Bayes' Theorem Example

Disease prevalence $P(D) = 0.01$
Test accuracy $P(+|D) = 0.99$
False positive $P(+|D^c) = 0.05$
Find $P(D|+)$
Solution

$$= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.05 \times 0.99}$$

Result $P(D|+) \approx 0.167$
Positive test does not imply high probability of disease.

Random Variables

Random Variable Maps outcomes to numerical values. Types

- Discrete
- Continuous

Examples, Discrete: Number of defective products.
Examples, Continuous: Temperature, Weight, Income, Height

Expected Value and Variance

Expected Value

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

Variance

$$Var(X) = E[(X - \mu)^2]$$

Measures

- Average outcome
- Spread of observations

These quantities summarize probability distributions.

Binomial Distribution

Used when

- Fixed number of trials
- Two outcomes
- Independent trials
- Constant probability

Probability Mass Function

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Examples

- Number of heads
- Number of defective products
- Number of customers responding

Properties of Binomial Distribution

Parameters

$$n, \; p$$

Mean

$$E(X) = np$$

Variance

$$Var(X) = np(1-p)$$

Typical Shape

- Symmetric when $p = 0.5$
- Skewed otherwise

Normal Distribution

The most widely used statistical distribution.
Probability Density Function

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

Characteristics

- Bell-shaped
- Symmetric
- Mean $=$ Median $=$ Mode
- Defined by μ and σ

Why is the Normal Distribution Important?

Many real-world measurements approximately follow a normal distribution.
Examples

- Human height
- Measurement errors
- Sensor noise
- Examination scores
- Manufacturing tolerances

Central Limit Theorem explains why the normal distribution appears frequently.

Standard Normal Distribution

Standardization

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

Properties

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

Applications

- Hypothesis testing
- Confidence intervals
- Outlier detection
- Machine learning feature scaling

Binomial vs Normal Distribution

Feature	Binomial	Normal
Type	Discrete	Continuous
Parameter	n, p	μ, σ
Values	Integers	Real numbers
Applications	Success counts	Measurements
Examples	Coin tosses	Height

Probability Distributions in Data Science

Distribution	Typical Applications
Bernoulli	Binary classification
Binomial	Number of successes
Poisson	Arrival counts
Normal	Measurement errors
Exponential	Waiting times
Uniform	Random simulation

Key Takeaways

- Probability quantifies uncertainty.
- Addition and multiplication rules are fundamental.
- Conditional probability enables reasoning with evidence.
- Bayes' theorem updates beliefs using observations.
- Binomial models discrete counts.
- Normal distribution models continuous measurements.
- These concepts underpin machine learning, AI, and statistical inference.

Python: Generating Random Numbers

Python provides powerful libraries for probability and statistics.

Import the required libraries

```
import numpy as np
import matplotlib.pyplot as plt
```

Common functions

Function	Purpose
np.random.rand()	Uniform random numbers
np.random.randint()	Random integers
np.random.binomial()	Binomial distribution
np.random.normal()	Normal distribution
np.random.choice()	Random sampling

These functions are extensively used for simulation and synthetic data generation.