

CE49X: Statistical Analysis I

Core Topics in Descriptive Statistics, Probability, and Distributions

Dr. Eyuphan Koc

Department of Civil Engineering
Bogazici University

Fall 2025

Lecture Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions
- 4 Expectation and Variance Concepts
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions
- 4 Expectation and Variance Concepts
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Introduction to Descriptive Statistics

What is Descriptive Statistics?

- Methods for summarizing and describing data
- Transforms raw data into meaningful insights
- Foundation for inferential statistics and decision-making

Key Components:

- **Central Tendency:** Where is the center?
- **Spread/Variability:** How dispersed is the data?
- **Shape:** Is the distribution symmetric or skewed?
- **Visualization:** Graphical representations

Engineering Context

Analyzing material strengths, load measurements, experimental results

Measures of Central Tendency

Mean (Arithmetic Average):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Median: Middle value when data is ordered

- Robust to outliers
- Better for skewed distributions

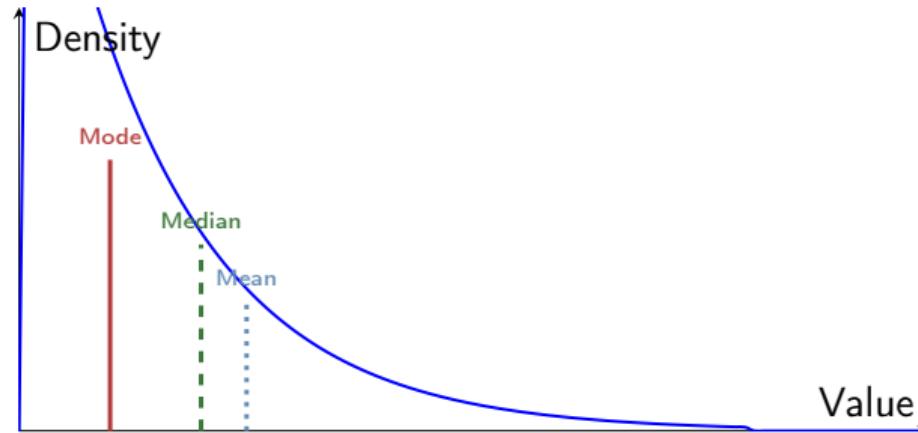
Mode: Most frequently occurring value

- Can be multiple modes (bimodal, multimodal)
- Useful for categorical data

When to Use Which?

Mean: Symmetric distributions; Median: Skewed data or outliers; Mode: Categorical data

Right-Skewed Distribution



Key Insight: In right-skewed distributions: Mode < Median < Mean

Measures of Spread

Range: $R = x_{\max} - x_{\min}$

- Simple but sensitive to outliers

Variance:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

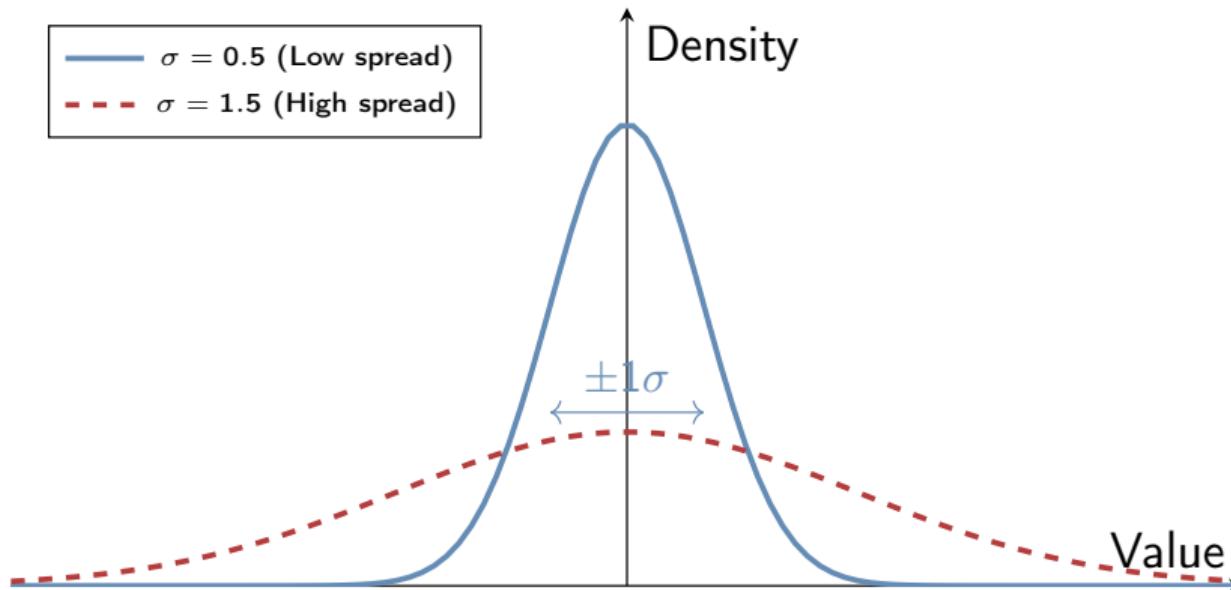
Standard Deviation: $s = \sqrt{s^2}$

- Same units as original data
- Most commonly used measure of spread

Interquartile Range (IQR): $IQR = Q_3 - Q_1$

- Robust to outliers
- Measures middle 50% of data

Comparing Distributions with Different Spreads



Key Insight: Larger σ means more spread, flatter distribution

Quantiles and Percentiles

Definition: The p -th percentile is a value below which $p\%$ of the data falls

Key Quantiles:

- Q_1 (25th percentile): First quartile
- Q_2 (50th percentile): Median
- Q_3 (75th percentile): Third quartile

Five-Number Summary:

Min, Q_1 , Median, Q_3 , Max

Engineering Example

For concrete strength data: Min=25 MPa, Q_1 =30 MPa, Median=35 MPa, Q_3 =40 MPa, Max=48 MPa

Shape Measures: Skewness and Kurtosis

Skewness: Measures asymmetry of distribution

$$\text{Skewness} = \frac{\mathbb{E}[(X - \mu)^3]}{\sigma^3}$$

- Positive skew: Long right tail (mean > median)
- Negative skew: Long left tail (mean < median)
- Zero skew: Symmetric (mean \approx median)

Kurtosis: Measures tail heaviness

$$\text{Kurtosis} = \frac{\mathbb{E}[(X - \mu)^4]}{\sigma^4}$$

- High kurtosis: Heavy tails, more outliers
- Low kurtosis: Light tails, fewer outliers
- Normal distribution: Kurtosis = 3

Data Visualization: Histograms

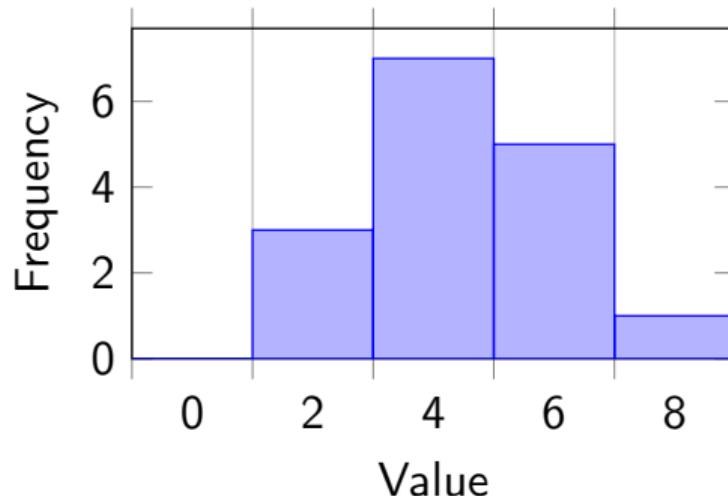
Histogram: Bar chart showing frequency distribution

Components:

- Bins: Intervals that group data
- Height: Frequency or density
- Shows shape of distribution

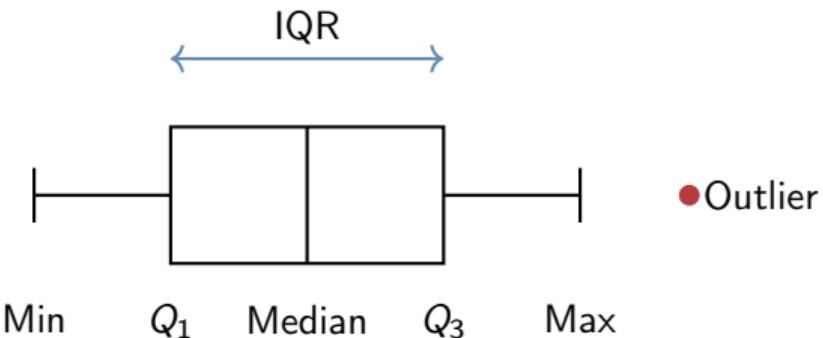
Key Insights:

- Central tendency
- Spread
- Skewness
- Outliers
- Modality



Data Visualization: Boxplots

Boxplot (Box-and-Whisker Plot): Displays five-number summary



Advantages:

- Clearly shows median and quartiles
- Identifies outliers
- Compact representation
- Easy comparison between groups

Engineering Application: Material Testing

Example: Concrete Compressive Strength

Testing 50 concrete cylinders yields (in MPa):

$$\bar{x} = 35.2, \quad s = 3.8, \quad \text{Median} = 35.5$$

$$Q_1 = 32.5, \quad Q_3 = 37.8, \quad \text{Skewness} = -0.15$$

Interpretation:

- Mean strength around 35 MPa
- Standard deviation of 3.8 MPa indicates moderate variability
- Near-symmetric distribution (skewness ≈ 0)
- IQR = 5.3 MPa shows middle 50% spread

Engineering Decision:

- Check if mean exceeds design specification
- Assess if variability is acceptable for quality control
- Identify any outliers indicating quality issues

Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions
- 4 Expectation and Variance Concepts
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Probability: The Logic of Uncertainty

Why Probability?

- Engineering systems operate under uncertainty
- Loads, material properties, environmental conditions vary
- Probability provides rigorous framework for quantifying uncertainty

Key Concepts:

- **Sample Space (S):** Set of all possible outcomes
- **Event (A):** Subset of sample space
- **Probability ($\mathbb{P}(A)$):** Measure of likelihood

Probability Axioms (Kolmogorov)

1. $0 \leq \mathbb{P}(A) \leq 1$ for any event A
2. $\mathbb{P}(S) = 1$ (something must happen)
3. If A_1, A_2, \dots are disjoint, then $\mathbb{P}(\bigcup A_i) = \sum \mathbb{P}(A_i)$

Sample Spaces and Events

Example 1: Quality Control

Testing a component: $S = \{\text{Pass}, \text{Fail}\}$

Event A: Component passes, $\mathbb{P}(A) = 0.95$

Example 2: Die Roll

Rolling a die: $S = \{1, 2, 3, 4, 5, 6\}$

Event A: Even number, $A = \{2, 4, 6\}$, $\mathbb{P}(A) = 3/6 = 0.5$

Example 3: Load Measurement

$S = [0, \infty)$ (all non-negative real numbers)

Event A: Load exceeds 100 kN, $A = (100, \infty)$

Operations on Events:

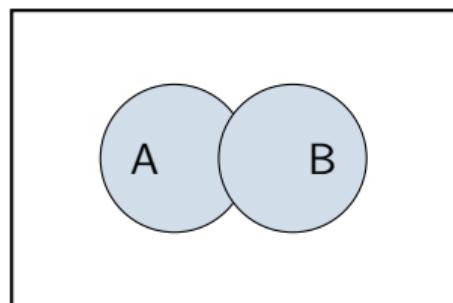
- Union: $A \cup B$ (A or B occurs)
- Intersection: $A \cap B$ (Both A and B occur)
- Complement: A^c (A does not occur), $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Visualizing Set Operations

[shrink=5]

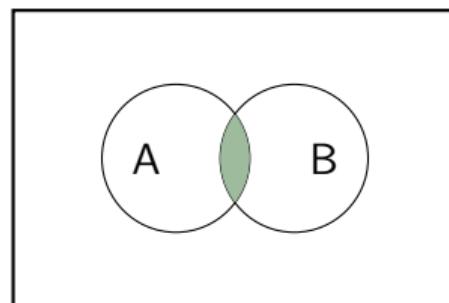
Venn Diagrams for Events

Union: $A \cup B$



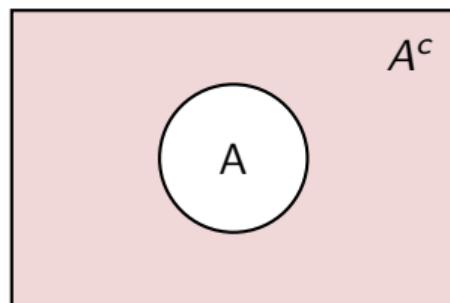
(A or B)

Intersection: $A \cap B$



(Both A and B)

Complement: A^c



(Not A)

Conditional Probability

Definition: Probability of A given that B has occurred

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

Intuition: Update probability based on new information

- B is the new sample space
- $A \cap B$ is the favorable outcomes within B

Example: Quality Control

- $\mathbb{P}(\text{Defective}) = 0.05$
- $\mathbb{P}(\text{Detected} | \text{Defective}) = 0.90$
- $\mathbb{P}(\text{Detected} | \text{Good}) = 0.02$

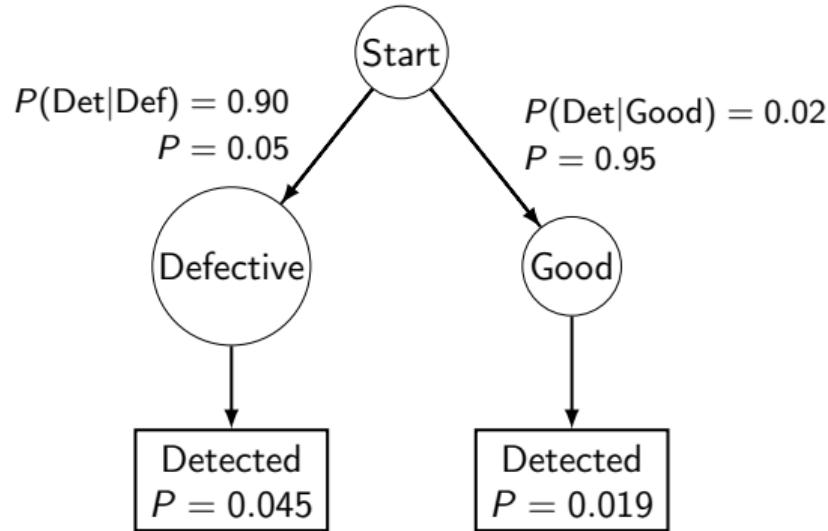
What is $\mathbb{P}(\text{Detected and Defective})$?

$$\mathbb{P}(\text{Detected} \cap \text{Defective}) = \mathbb{P}(\text{Detected} | \text{Defective}) \cdot \mathbb{P}(\text{Defective}) = 0.90 \times 0.05 = 0.045$$

Probability Tree Diagram

[shrink=5]

Quality Control Example



Total Probability: $\mathbb{P}(\text{Detected}) = 0.045 + 0.019 = 0.064$

Independence

Definition: Events A and B are independent if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Equivalently: $\mathbb{P}(A | B) = \mathbb{P}(A)$ (knowing B doesn't change probability of A)

Examples of Independence:

- Two separate coin flips
- Strengths of components from different production batches
- Earthquakes in different regions

Examples of Dependence:

- Weather on consecutive days
- Strength of adjacent concrete sections (same pour)
- Multiple tests on the same specimen

Common Mistake

Disjoint events are NOT independent! If $A \cap B = \emptyset$ and $\mathbb{P}(A), \mathbb{P}(B) > 0$, then $\mathbb{P}(A | B) = 0 \neq \mathbb{P}(A)$

Law of Total Probability

Setup: Partition sample space into events B_1, B_2, \dots, B_n where:

- $B_i \cap B_j = \emptyset$ for $i \neq j$ (disjoint)
- $\bigcup_{i=1}^n B_i = S$ (exhaustive)

Law of Total Probability:

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A | B_i) \mathbb{P}(B_i)$$

Example: Two Suppliers

Factory gets components from Supplier 1 (60%) and Supplier 2 (40%).

Defect rates: Supplier 1 (2%), Supplier 2 (5%).

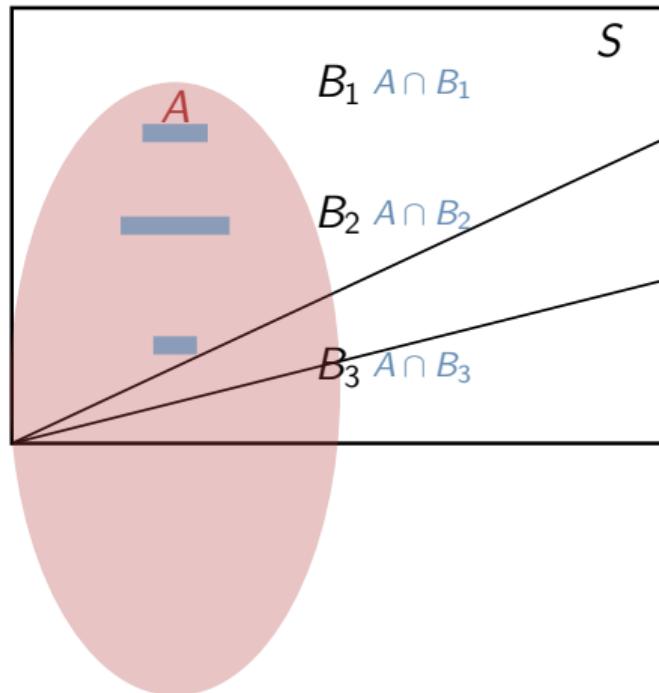
What is overall defect rate?

$$\mathbb{P}(D) = \mathbb{P}(D | S_1) \mathbb{P}(S_1) + \mathbb{P}(D | S_2) \mathbb{P}(S_2)$$

$$= 0.02 \times 0.6 + 0.05 \times 0.4 = 0.012 + 0.020 = 0.032 = 3.2\%$$

Visualizing Law of Total Probability

Partitioning the Sample Space



$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3)$$

Bayes' Theorem

Bayes' Theorem: Updates probabilities based on new evidence

$$\begin{aligned}\mathbb{P}(B_i | A) &= \frac{\mathbb{P}(A | B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A | B_i)\mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A | B_j)\mathbb{P}(B_j)}\end{aligned}$$

Terminology:

- $\mathbb{P}(B_i)$: Prior probability (before observing A)
- $\mathbb{P}(B_i | A)$: Posterior probability (after observing A)
- $\mathbb{P}(A | B_i)$: Likelihood

Example: Diagnostic Testing

Test for structural damage: Sensitivity = 95%, Specificity = 90%

Prevalence of damage = 5%

If test is positive, what is probability of actual damage?

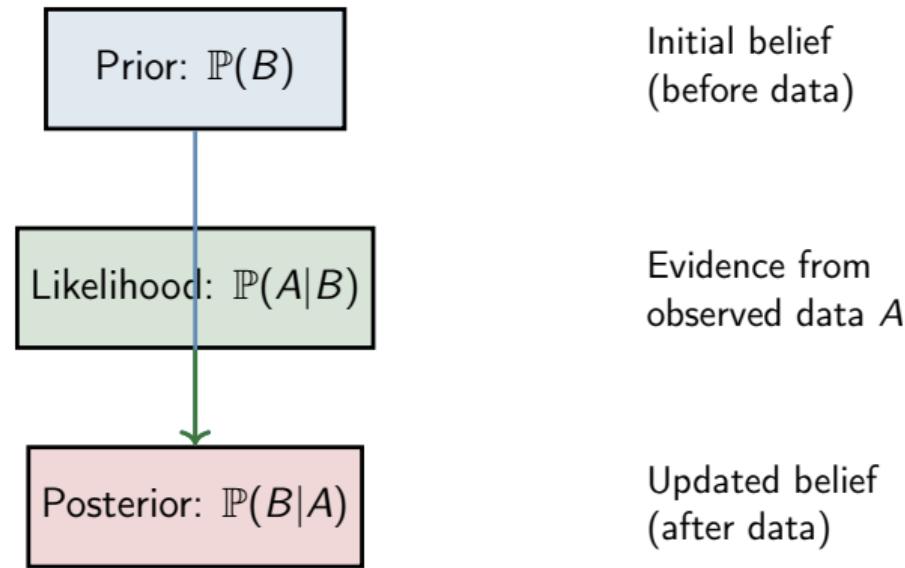
$$\mathbb{P}(\text{Damage} | +) = \frac{0.95 \times 0.05}{0.95 \times 0.05 + 0.10 \times 0.95} = \frac{0.0475}{0.1425} \approx 0.333$$

Only 33% probability despite positive test! (Low prevalence effect)

Visualizing Bayes' Theorem

[shrink=5]

The Bayesian Update Process



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

Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions**
- 4 Expectation and Variance Concepts
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Random Variables

Definition: A random variable X is a function from sample space S to real numbers \mathbb{R} : $X : S \rightarrow \mathbb{R}$

Types:

- **Discrete:** Takes countable values (often integers)
- **Continuous:** Takes values in an interval

Why Random Variables?

- Assign numerical values to outcomes
- Enable mathematical operations and analysis
- Bridge between probability and statistics

Examples

- X = Number of defective items in a batch (discrete)
- Y = Time until failure of a component (continuous)
- Z = Load on a bridge (continuous)

Probability Mass Function (PMF)

Definition: For discrete random variable X , the PMF is:

$$p_X(x) = \mathbb{P}(X = x)$$

Properties:

- ① $p_X(x) \geq 0$ for all x
- ② $\sum_{\text{all } x} p_X(x) = 1$

Cumulative Distribution Function (CDF):

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{t \leq x} p_X(t)$$

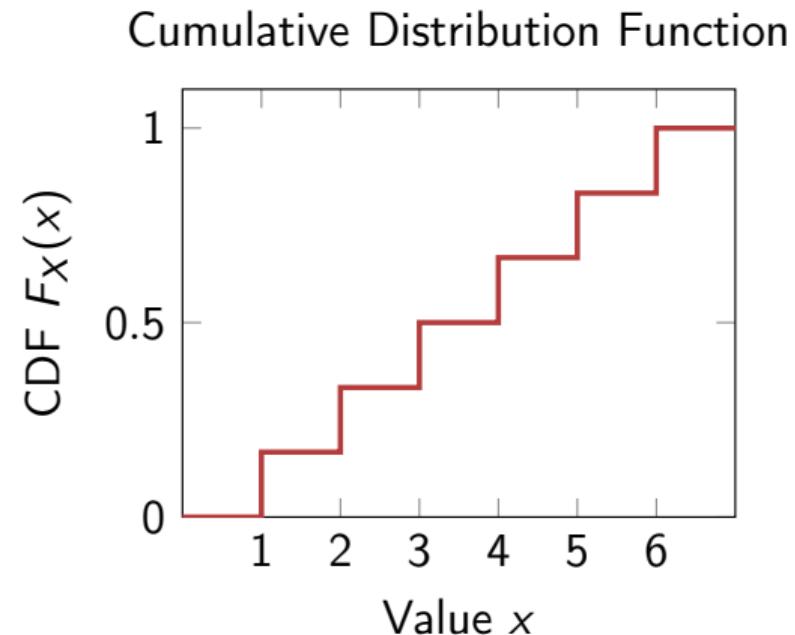
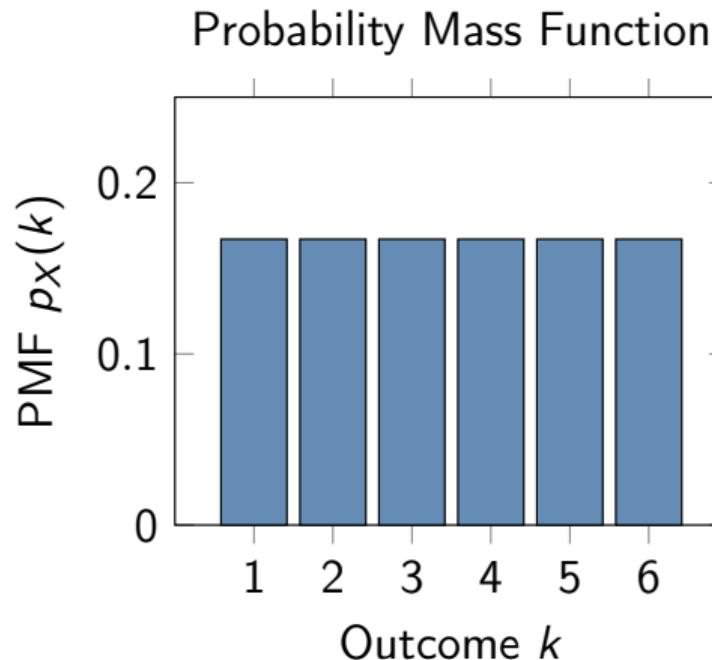
Properties: Non-decreasing, right-continuous, $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$

Example: Rolling a Die

$X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$: $p_X(k) = 1/6$ for $k = 1, \dots, 6$

$$F_X(3.5) = \mathbb{P}(X \leq 3.5) = \mathbb{P}(X \in \{1, 2, 3\}) = 3/6 = 0.5$$

Example: Fair Die Roll



Bernoulli Distribution

Definition: Models a single trial with two outcomes (success/failure)

$$X \sim \text{Bernoulli}(p)$$

PMF:

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

Or compactly: $p_X(x) = p^x(1 - p)^{1-x}$ for $x \in \{0, 1\}$

Properties:

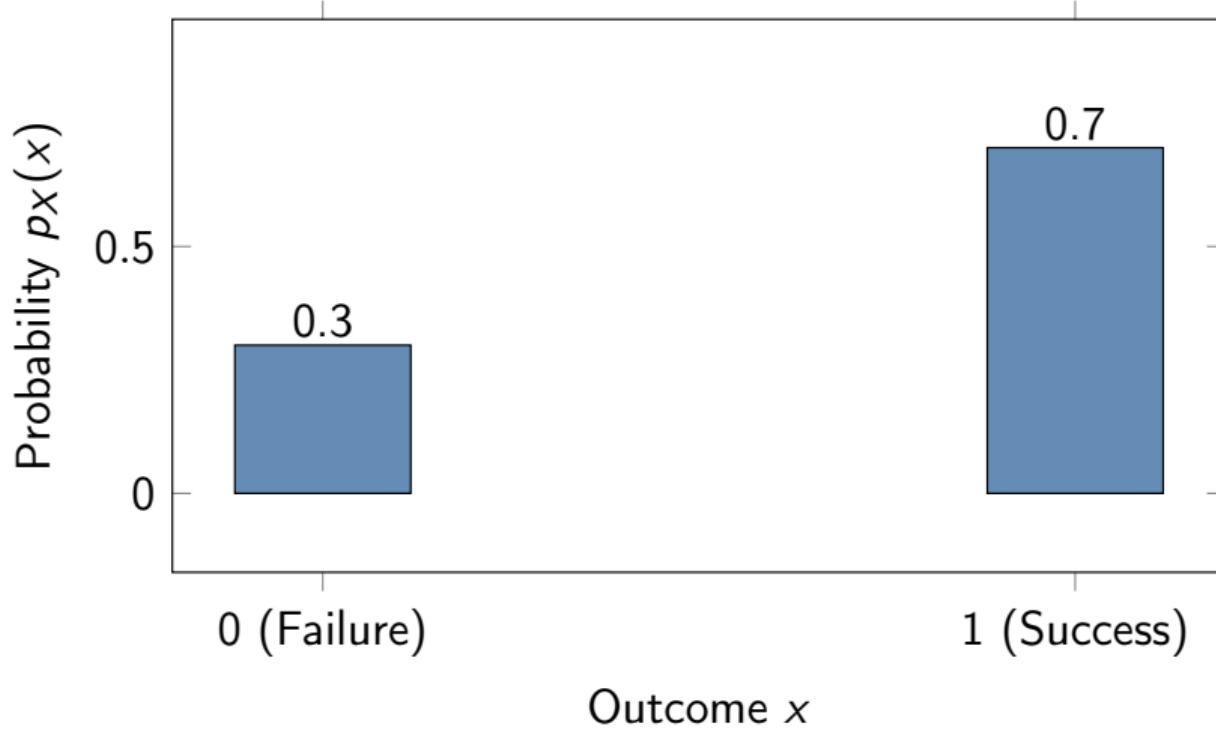
- $\mathbb{E}[X] = p$
- $\text{Var}(X) = p(1 - p)$

Engineering Examples

- Component passes inspection ($p = 0.95$)
- Weld is defective ($p = 0.02$)
- Structure survives earthquake ($p = 0.98$)

Visualizing Bernoulli Distribution

PMF: Bernoulli($p = 0.7$)



Binomial Distribution

Story: Number of successes in n independent Bernoulli(p) trials

$$X \sim \text{Binomial}(n, p)$$

PMF:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Properties:

- $\mathbb{E}[X] = np$
- $\text{Var}(X) = np(1 - p)$

Example

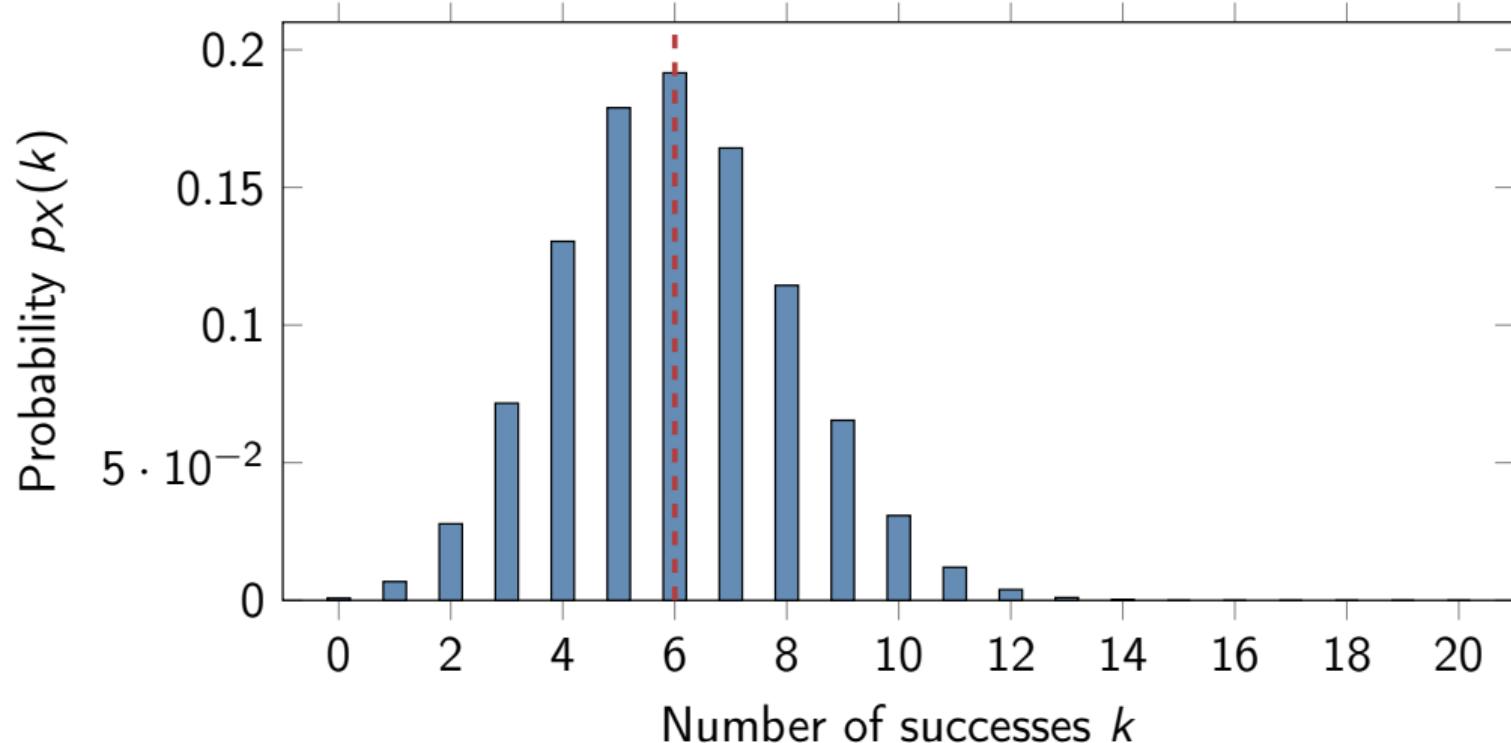
Test 20 components, each passes with probability 0.95

$X \sim \text{Bin}(20, 0.95)$, $\mathbb{E}[X] = 20 \times 0.95 = 19$

$\mathbb{P}(X = 20) = (0.95)^{20} \approx 0.358$

Visualizing Binomial Distribution

PMF: Binomial($n = 20, p = 0.3$)



Hypergeometric Distribution

Story: Sampling without replacement from finite population

Population: N total objects, K are successes; Draw n objects without replacement; X = successes in sample

$$X \sim \text{HGeom}(N, K, n)$$

PMF:

$$p_X(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Properties:

- $\mathbb{E}[X] = n \cdot \frac{K}{N}$
- $\text{Var}(X) = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$

Example

Batch of 100 parts, 5 defective. Sample 10 parts.

$$X \sim \text{HGeom}(100, 5, 10), \mathbb{E}[X] = 10 \times \frac{5}{100} = 0.5$$

Poisson Distribution

Story: Number of rare events in fixed time/space interval

$$X \sim \text{Poisson}(\lambda)$$

where $\lambda > 0$ is the rate parameter (average number of events)

PMF:

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

Properties:

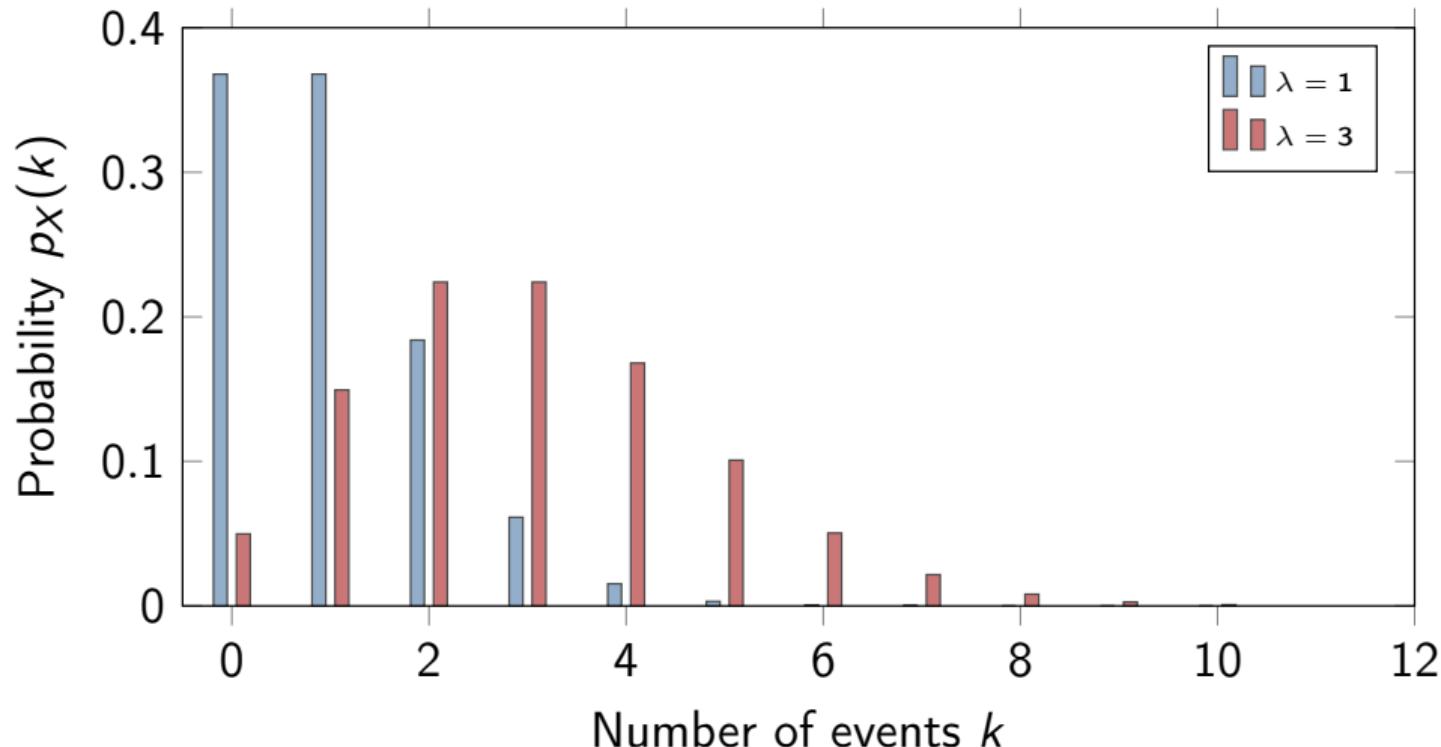
- $\mathbb{E}[X] = \lambda$
- $\text{Var}(X) = \lambda$ (unique property!)
- Approximates Binomial when n large, p small, $np = \lambda$ moderate

Engineering Examples

- Number of trucks crossing bridge per hour ($\lambda = 50$)
- Number of accidents per year at intersection ($\lambda = 3$)
- Number of defects per 100m of pipeline ($\lambda = 0.5$)

Visualizing Poisson Distribution

Comparing Poisson PMFs for Different λ



Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions
- 4 Expectation and Variance Concepts**
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Expectation: The Average Value

Definition: Expected value (mean) of discrete random variable X :

$$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$$

For continuous X with PDF f_X :

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

Interpretation:

- Long-run average value
- Center of mass of distribution
- "Typical" value (but may never actually occur!)

Example: Die Roll

$$X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$$

$$\mathbb{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$$

Properties of Expectation

Linearity of Expectation: For any constants a, b and random variables X, Y :

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Key Points:

- Works for ANY random variables (independent or not!)
- Most useful property for calculations
- Extends to any finite number of variables

Example

$X \sim \text{Bin}(100, 0.3)$ represents number of defective items.

Write $X = X_1 + X_2 + \dots + X_{100}$ where $X_i \sim \text{Bern}(0.3)$

By linearity:

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_{100}] = 100 \times 0.3 = 30$$

No need to sum over all possible values!

Non-Linearity

$\mathbb{E}[X^2] \neq (\mathbb{E}[X])^2$ in general!

Law of the Unconscious Statistician (LOTUS)

LOTUS: To find $\mathbb{E}[g(X)]$ for function g :

$$\mathbb{E}[g(X)] = \sum_x g(x) \cdot p_X(x) \quad (\text{discrete})$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx \quad (\text{continuous})$$

Key Insight: Don't need to find distribution of $g(X)$!

Example

$X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$. Find $\mathbb{E}[X^2]$:

$$\begin{aligned}\mathbb{E}[X^2] &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} \\ &= \frac{1 + 4 + 9 + 16 + 25 + 36}{6} = \frac{91}{6} \approx 15.17\end{aligned}$$

Applications:

- Computing variance: $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
- Finding moments: $\mathbb{E}[X^n]$
- Expected value of transformations

Variance: Measuring Spread

Definition: Variance measures average squared deviation from mean

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Computational Formula:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Standard Deviation: $\text{SD}(X) = \sqrt{\text{Var}(X)}$ (same units as X)

Properties:

- $\text{Var}(aX + b) = a^2\text{Var}(X)$ (constants: b shifts, a scales)
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ if X, Y independent
- $\text{Var}(X) \geq 0$ always

Example: Die Roll

$X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\}$, $\mathbb{E}[X] = 3.5$, $\mathbb{E}[X^2] = 15.17$

$$\text{Var}(X) = 15.17 - (3.5)^2 = 15.17 - 12.25 = 2.92$$

Applications: Binomial and Geometric

Binomial Variance

$X \sim \text{Bin}(n, p)$: Write $X = X_1 + \dots + X_n$, $X_i \sim \text{Bern}(p)$ independent

$$\mathbb{E}[X] = np \text{ (by linearity)}$$

$$\begin{aligned}\text{Var}(X) &= \text{Var}(X_1) + \dots + \text{Var}(X_n) \text{ (by independence)} \\ &= n \cdot p(1 - p) = np(1 - p)\end{aligned}$$

Geometric Distribution

$X \sim \text{Geom}(p)$: Number of trials until first success

PMF: $\mathbb{P}(X = k) = (1 - p)^{k-1} p$ for $k = 1, 2, 3, \dots$

$$\mathbb{E}[X] = \frac{1}{p} \text{ (e.g., if } p = 0.1, \text{ expect 10 trials)}$$

$$\text{Var}(X) = \frac{1-p}{p^2}$$

Memoryless property: $\mathbb{P}(X > n + k \mid X > n) = \mathbb{P}(X > k)$

Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions
- 4 Expectation and Variance Concepts
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Continuous Random Variables

Definition: Random variable X is continuous if it can take any value in an interval

Probability Density Function (PDF): $f_X(x)$

- NOT a probability! $f_X(x)$ can be > 1
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$
- $\mathbb{P}(X = x) = 0$ for any specific value x

Properties of PDF:

- ① $f_X(x) \geq 0$ for all x
- ② $\int_{-\infty}^{\infty} f_X(x) dx = 1$

CDF: $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt$

Relationship: $f_X(x) = \frac{d}{dx} F_X(x)$

Uniform Distribution

Uniform Distribution on $[a, b]$: $X \sim \text{Unif}(a, b)$

All values equally likely in the interval

PDF:

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

CDF:

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$

Properties:

- $\mathbb{E}[X] = \frac{a+b}{2}$ (midpoint)
- $\text{Var}(X) = \frac{(b-a)^2}{12}$

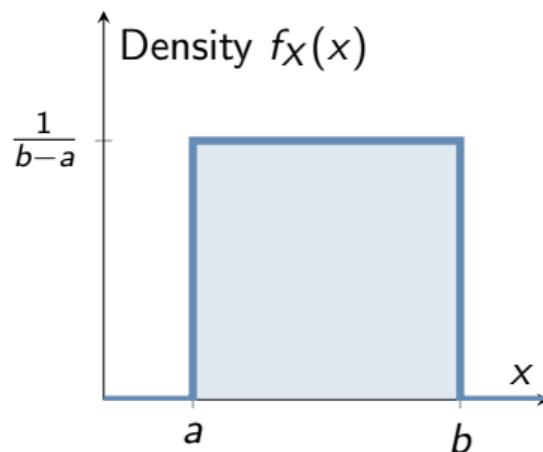
Example

Random position on 10m beam: $X \sim \text{Unif}(0, 10)$

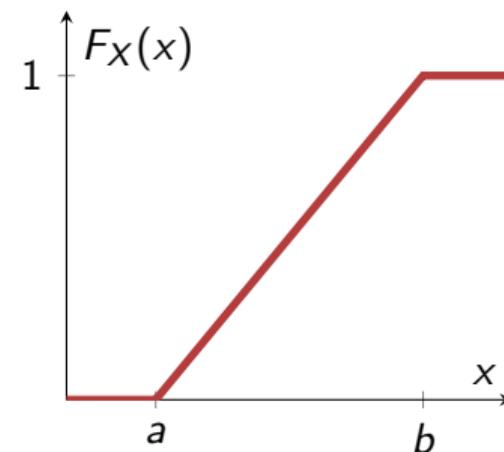
Visualizing Uniform Distribution

PDF and CDF for Uniform($a = 1$, $b = 4$)

Probability Density Function



Cumulative Distribution Function



Key: All values in $[a, b]$ equally likely; constant PDF

Normal Distribution

Normal (Gaussian) Distribution: $X \sim \mathcal{N}(\mu, \sigma^2)$

Most important distribution in statistics!

PDF:

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

Properties:

- $\mathbb{E}[X] = \mu$ (location parameter)
- $\text{Var}(X) = \sigma^2$ (scale parameter)
- Symmetric around μ
- Bell-shaped curve

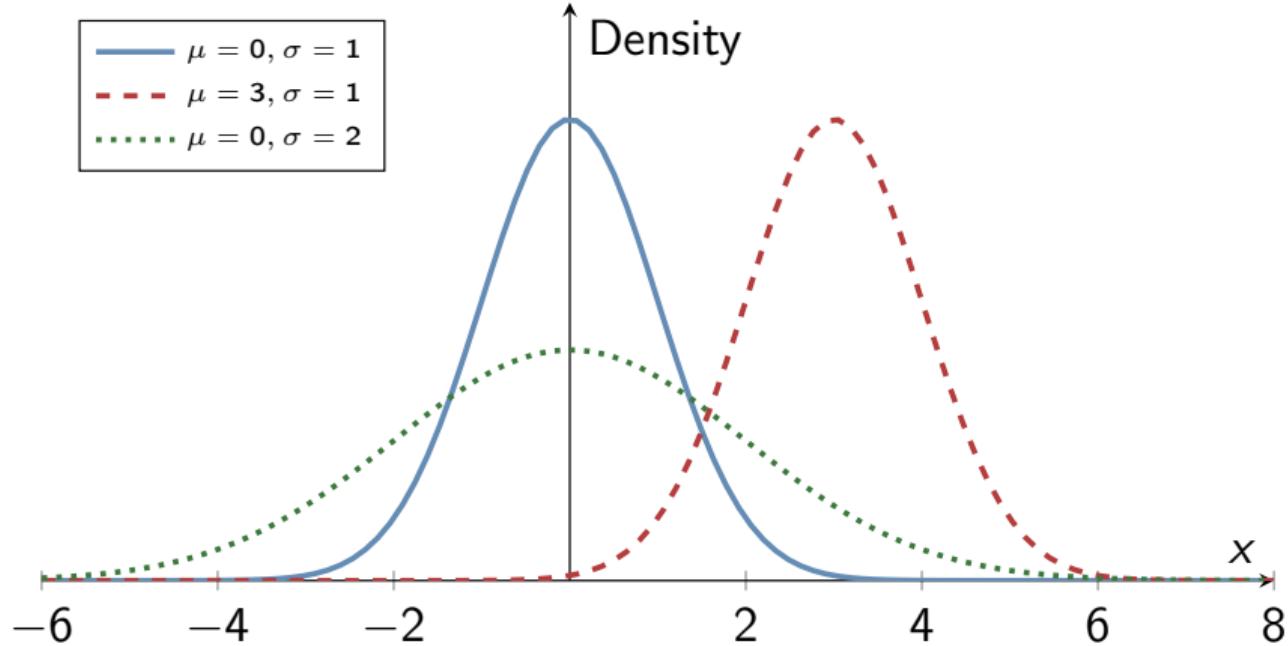
Standard Normal: $Z \sim \mathcal{N}(0, 1)$

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

CDF denoted $\Phi(z)$, used for calculations

Visualizing Normal Distribution

Effect of Parameters μ and σ

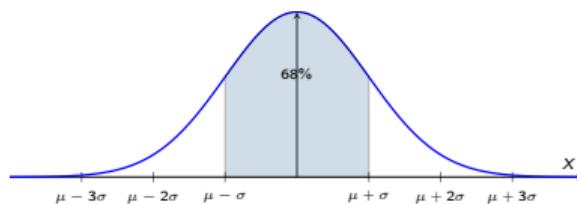


Key Insights: μ shifts location, σ controls spread

Normal Distribution: The 68-95-99.7 Rule

Empirical Rule: For $X \sim \mathcal{N}(\mu, \sigma^2)$:

- $\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$ (68%)
- $\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$ (95%)
- $\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$ (99.7%)



Example

Steel yield strength: $X \sim \mathcal{N}(250, 15^2)$ MPa

About 95% of specimens have strength between $250 - 30 = 220$ and $250 + 30 = 280$ MPa

Exponential Distribution

Exponential Distribution: $X \sim \text{Expo}(\lambda)$

Models time until an event occurs (waiting times, lifetimes)

PDF:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

CDF:

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

Properties:

- $\mathbb{E}[X] = \frac{1}{\lambda}$ (average waiting time)
- $\text{Var}(X) = \frac{1}{\lambda^2}$
- Memoryless: $\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$
- Related to Poisson: If events follow $\text{Poisson}(\lambda)$, inter-arrival times follow $\text{Expo}(\lambda)$

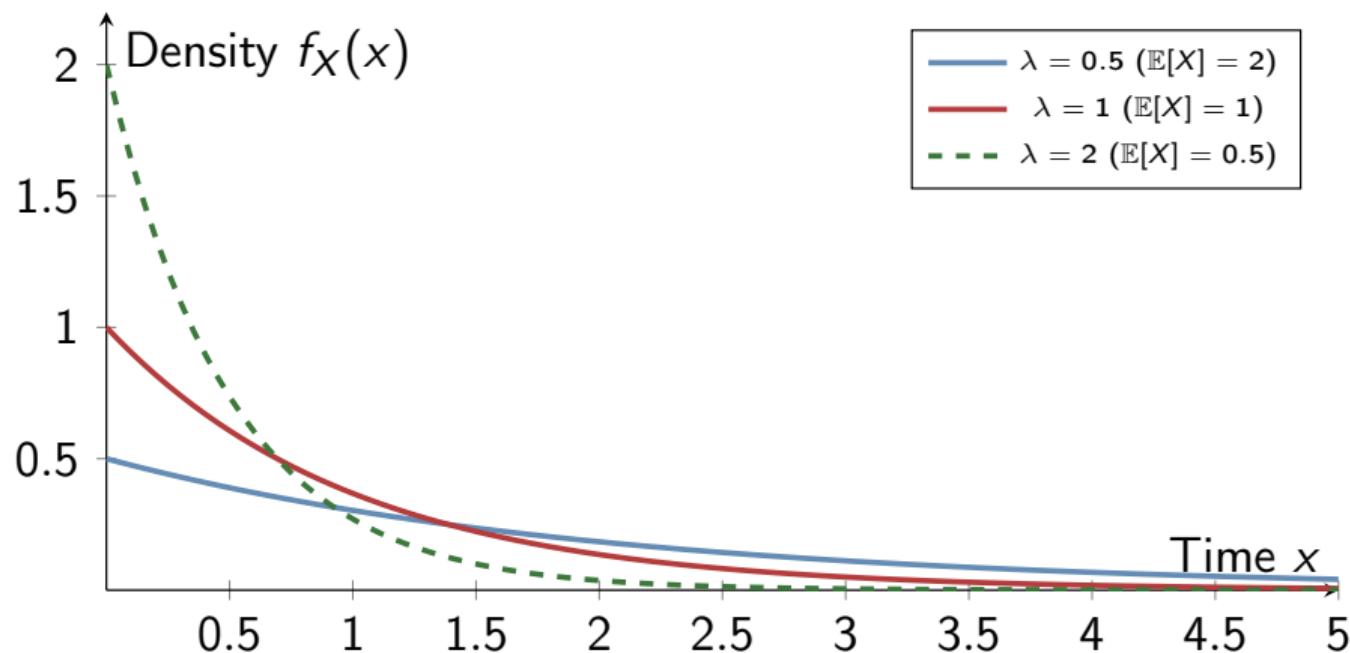
Example

Time until component failure: $X \sim \text{Expo}(0.01)$ hours

$\mathbb{E}[X] = 100$ hours, $\mathbb{P}(X > 100) = e^{-1} \approx 0.368$

Visualizing Exponential Distribution

PDF for Different Rate Parameters λ



Memoryless Property: Past doesn't affect future waiting time

Continuous RVs: Expected Value and Variance

Expected Value:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

LOTUS for Continuous:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$$

Variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx$$

All properties from discrete case carry over:

- Linearity: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$
- If X, Y independent: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions
- 4 Expectation and Variance Concepts
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Definition of Moments

Raw Moments (about origin):

$$\mu'_n = \mathbb{E}[X^n]$$

- $\mu'_1 = \mathbb{E}[X]$ (mean)
- $\mu'_2 = \mathbb{E}[X^2]$
- $\mu'_3 = \mathbb{E}[X^3]$
- $\mu'_n = \mathbb{E}[X^n]$

Central Moments (about mean μ):

$$\mu_n = \mathbb{E}[(X - \mu)^n]$$

- $\mu_1 = 0$ (always)
- $\mu_2 = \text{Var}(X)$ (variance)
- μ_3 related to skewness
- μ_4 related to kurtosis

Intuition: Moments are summary statistics that capture key features of distribution

Mean and Variance (1st and 2nd Moments)

Mean (1st Raw Moment):

$$\mu = \mathbb{E}[X] = \mu'_1$$

- Center of distribution
- Expected value, average

Variance (2nd Central Moment):

$$\sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mu'_2$$

Using computational formula:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mu'_2 - (\mu'_1)^2$$

- Spread of distribution
- Average squared deviation from mean
- Units: (original units)²

Standard Deviation: $\sigma = \sqrt{\text{Var}(X)}$ (same units as X)

Skewness (3rd Moment)

Skewness: Measures asymmetry of distribution

$$\text{Skewness} = \frac{\mathbb{E}[(X - \mu)^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3}$$

Interpretation:

- Skewness = 0: Symmetric
- Skewness > 0: Right-skewed (long right tail)
- Skewness < 0: Left-skewed (long left tail)



Engineering Relevance:

- Extreme loads often right-skewed
- Material strengths may be left-skewed
- Affects reliability calculations

Kurtosis (4th Moment)

Kurtosis: Measures tail heaviness / peakedness

$$\text{Kurtosis} = \frac{\mathbb{E}[(X - \mu)^4]}{\sigma^4} = \frac{\mu_4}{\sigma^4}$$

Excess Kurtosis: Kurtosis - 3 (so Normal has excess kurtosis 0)

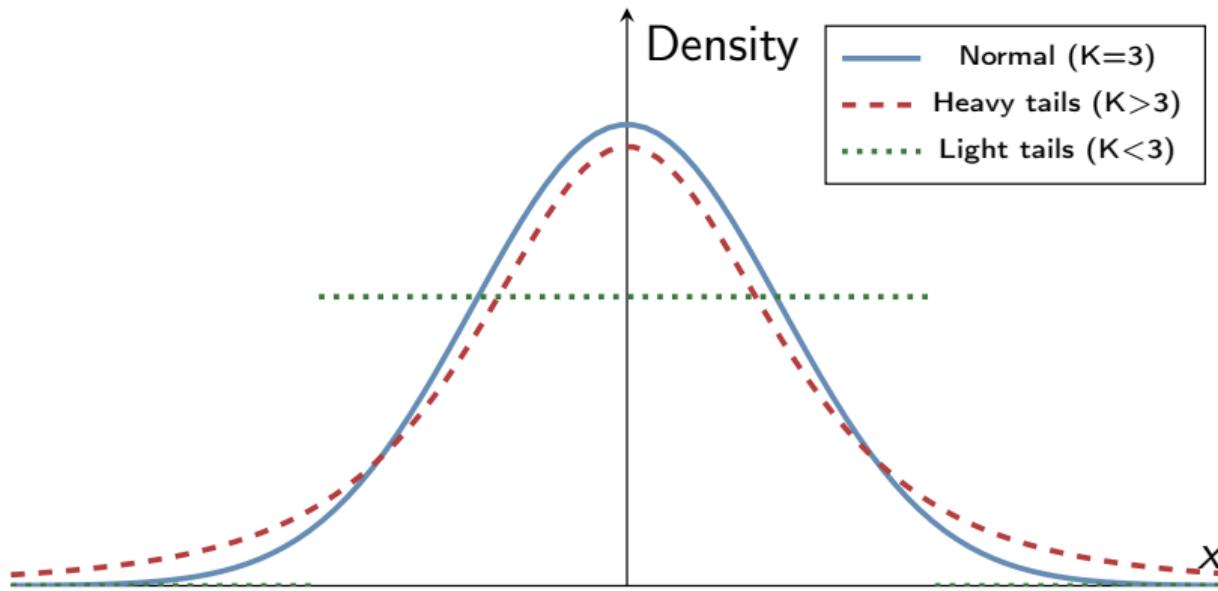
Interpretation:

- Normal distribution: Kurtosis = 3
- Kurtosis > 3 (leptokurtic): Heavy tails, more outliers
- Kurtosis < 3 (platykurtic): Light tails, fewer outliers

Engineering Relevance:

- High kurtosis: Greater risk of extreme events
- Important for safety and reliability analysis
- Affects probability of exceeding design thresholds

Comparing Distributions with Different Kurtosis



Key: High kurtosis means more probability in tails (extreme values)

Moment Generating Functions (MGFs)

Definition: For random variable X , $M_X(t) = \mathbb{E}[e^{tX}]$ (exists if finite for t near 0)

Why "Moment Generating"?

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] = \mathbb{E}\left[1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \dots\right] \\ &= 1 + t\mathbb{E}[X] + \frac{t^2}{2!}\mathbb{E}[X^2] + \frac{t^3}{3!}\mathbb{E}[X^3] + \dots \end{aligned}$$

Taking derivatives at $t = 0$:

$$\mathbb{E}[X^n] = \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}$$

Key Properties:

- Uniquely determines distribution
- $M_{aX+b}(t) = e^{bt} M_X(at)$
- If X, Y independent: $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

MGF Examples

Bernoulli(p)

$$M_X(t) = (1 - p) + pe^t$$

Binomial(n, p)

$$M_X(t) = ((1 - p) + pe^t)^n$$

Note: MGF of sum of n independent Bernoullis!

Poisson(λ)

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Normal(μ, σ^2)

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Using MGFs to Find Moments

Exponential(λ)

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

Example: Poisson(λ)

$$M_X(t) = e^{\lambda(e^t - 1)}$$

First moment (mean):

$$M'_X(t) = \lambda e^t \cdot e^{\lambda(e^t - 1)}$$

$$\mathbb{E}[X] = M'_X(0) = \lambda e^0 \cdot e^0 = \lambda$$

Second moment:

$$M''_X(t) = [\lambda e^t]^2 e^{\lambda(e^t - 1)} + \lambda e^t e^{\lambda(e^t - 1)}$$

$$\mathbb{E}[X^2] = M''_X(0) = \lambda^2 + \lambda$$

Variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

Poisson has unique property: mean = variance!

Summary of Key Distributions

Distribution	Parameters	Mean	Variance
Bernoulli	p	p	$p(1 - p)$
Binomial	n, p	np	$np(1 - p)$
Poisson	λ	λ	λ
Geometric	p	$1/p$	$(1 - p)/p^2$
Uniform	a, b	$(a + b)/2$	$(b - a)^2/12$
Normal	μ, σ^2	μ	σ^2
Exponential	λ	$1/\lambda$	$1/\lambda^2$

Remember:

- Choose distribution based on problem context
- Discrete: Count data (Bernoulli, Binomial, Poisson)

Outline

- 1 Descriptive Statistics
- 2 Probability Review
- 3 Discrete Random Variables and Distributions
- 4 Expectation and Variance Concepts
- 5 Continuous Random Variables and Distributions
- 6 Moments and Summary Measures
- 7 Conclusion

Summary: Statistical Analysis Core Topics

What We Covered:

- ① **Descriptive Statistics:** Mean, median, variance, skewness, kurtosis, visualization
- ② **Probability Review:** Sample spaces, conditional probability, Bayes' theorem
- ③ **Discrete Distributions:** Bernoulli, Binomial, Hypergeometric, Poisson
- ④ **Expectation & Variance:** Linearity, LOTUS, computational methods
- ⑤ **Continuous Distributions:** Uniform, Normal (68-95-99.7 rule), Exponential
- ⑥ **Moments & MGFs:** Summary measures, skewness, kurtosis, moment generating functions

Key Takeaways:

- Statistical methods provide tools to understand and quantify uncertainty
- Choose appropriate distributions based on problem context
- Expectation and variance are fundamental summary statistics
- MGFs provide powerful technique for deriving moments

Next Steps and Applications

Where to Go from Here:

- **Joint Distributions:** Multiple random variables, correlation
- **Sampling Distributions:** Sample mean, Central Limit Theorem
- **Statistical Inference:** Hypothesis testing, confidence intervals
- **Regression:** Modeling relationships between variables

Engineering Applications:

- Reliability analysis and design
- Quality control and process monitoring
- Risk assessment and decision-making
- Data-driven modeling and prediction

Practice Recommendation

Work through examples with real engineering data using Python (NumPy, Pandas, SciPy, Matplotlib) to solidify these concepts!

Thank You!

Questions?

Dr. Eyuphan Koc
Department of Civil Engineering
Bogazici University
eyuphan.koc@bogazici.edu.tr