P8107 Introduction to Mathematical Statistics

Quick Navigation

Table of Contents

Week 1: Introduction to Probability

1.1 Kolmogorov Axioms of Probability

Probability theory is built upon three fundamental axioms established by Russian mathematician Andrey Kolmogorov in 1933:

Axiom 1 (Non-negativity):

$$P(A) \geq 0$$

for any event A.

Axiom 2 (Normalization):

$$P(S) = 1$$

where S is the sample space.

Axiom 3 (Countable Additivity):
If $A_1, A_2, A_3, \ldots$ are disjoint events, then:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

1.2 Basic Probability Concepts

Sample Space S: The set of all possible outcomes of an experiment.

Event: Any subset of the sample space.

Probability Measure: A function that assigns probabilities to events, taking values in $[0, 1]$.

1.3 Fundamental Probability Rules

Addition Rule:

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

Complement Rule:

$$P(A^c) = 1 - P(A)$$

Mutually Exclusive Events:

If $A \cap B = \emptyset$, then:

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

1.4 Conditional Probability

Definition:

The conditional probability of event A given event B is:

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

provided that $P(B) > 0$.

Multiplication Rule:

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

Chain Rule:

For events $A_1, A_2, \ldots, A_n$:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

1.5 Independence

Two Events:

Events A and B are independent if and only if:

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

Equivalent Conditions:

- $P(A|B) = P(A)$ (when $P(B) > 0$)
- $P(B|A) = P(B)$ (when $P(A) > 0$)

Multiple Events:

Events $A_1, A_2, \ldots, A_n$ are mutually independent if for any subset $\{i_1, i_2, \ldots, i_k\}$:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

Week 2: Some Probability Laws

2.1 Law of Total Probability

Theorem:

Let $B_1, B_2, \ldots, B_n$ be a partition of the sample space S with $P(B_i) > 0$ for all i. Then for any event A:

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

Continuous Version:

If $B(\theta)$ has a continuous distribution over parameter θ:

$$P(A) = \int P(A|B(\theta)) \cdot f_B(\theta) \, d\theta$$

2.2 Bayes' Theorem

Basic Form:

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

Complete Form (with Law of Total Probability):

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A|B_j) \cdot P(B_j)}$$

Mathematical Derivation:

From the definition of conditional probability:

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

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

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

Therefore:

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

Terminology:

- $P(A)$: Prior probability
- $P(A|B)$: Posterior probability
- $P(B|A)$: Likelihood
- $P(B)$: Marginal probability

2.3 Conditional Independence

Definition:
Given event C with $P(C) > 0$, events A and B are conditionally independent if:

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

Equivalent Statement:

$$A \perp B|C \Leftrightarrow P(A|B, C) = P(A|C)$$

Important Property: Conditional independence does not imply unconditional independence, and vice versa.

2.4 Random Variables

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

$$X : S \to \mathbb{R}$$

Types:

- **Discrete Random Variable:** Takes values in a countable set
- **Continuous Random Variable:** Takes values in an uncountable set (typically an interval)

Cumulative Distribution Function (CDF):

$$F_X(x) = P(X \leq x)$$

Properties of CDF:

- Monotonically non-decreasing
- Right-continuous

- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to +\infty} F_X(x) = 1$

Week 3: Discrete Random Variables

3.1 Discrete Random Variables Fundamentals

Probability Mass Function (PMF):

$$p_X(x) = P(X = x)$$

Properties:

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

Expected Value:

$$E[X] = \sum_{\text{all } x} x \cdot p_X(x)$$

Variance:

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

Standard Deviation:

$$\sigma_X = \sqrt{\text{Var}(X)}$$

3.2 Binomial Distribution

Notation: $X \sim \text{Binomial}(n, p)$ or $X \sim B(n, p)$

Probability Mass Function:

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

where $k = 0, 1, 2, \ldots, n$ and $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Mathematical Derivation:

For n independent Bernoulli trials with success probability p, the probability of exactly k successes is:

- Choose k positions for successes: $\binom{n}{k}$ ways
- Probability of k successes: p^k
- Probability of $(n-k)$ failures: $(1-p)^{n-k}$

Expected Value and Variance:

$$E[X] = np$$

$$\mathrm{Var}(X) = np(1-p)$$

Moment Generating Function:

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

3.3 Geometric Distribution

Notation: $X \sim \mathrm{Geometric}(p)$

Probability Mass Function:

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots$$

Mathematical Derivation:

X represents the trial number of the first success:

- First $(k-1)$ trials are failures: $(1-p)^{k-1}$
- k-th trial is a success: p

Expected Value and Variance:

$$E[X] = \frac{1}{p}$$

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

Memoryless Property:

$$P(X > m + n | X > m) = P(X > n)$$

Moment Generating Function:

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t}$$

for $t < -\ln(1 - p)$

3.4 Hypergeometric Distribution

Notation: $X \sim \text{Hypergeometric}(N, K, n)$

Probability Mass Function:

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

where $\max(0, n - (N - K)) \leq k \leq \min(n, K)$

Mathematical Derivation:
Population of N items with K "success" items. Drawing n items without replacement, X is the number of success items:

- Ways to choose k successes from K: $\binom{K}{k}$
- Ways to choose $(n - k)$ failures from $(N - K)$: $\binom{N-K}{n-k}$
- Total ways to choose n items: $\binom{N}{n}$

Expected Value and Variance:

$$E[X] = n \cdot \frac{K}{N}$$

$$\text{Var}(X) = n \cdot \frac{K}{N} \cdot \left(1 - \frac{K}{N}\right) \cdot \frac{N - n}{N - 1}$$

Relationship to Binomial: As $N \to \infty$ and $\frac{K}{N} \to p$, hypergeometric approaches binomial.

3.5 Poisson Distribution

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

Probability Mass Function:

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

Mathematical Derivation:

Poisson distribution is the limit of binomial distribution as $n \to \infty$, $p \to 0$, but $np = \lambda$ remains constant:

$$\lim_{n \to \infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

Expected Value and Variance:

$$E[X] = \lambda$$

$$\text{Var}(X) = \lambda$$

Moment Generating Function:

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

Poisson Process Properties:

- Independent increments
- Stationary increments
- Rare events accumulation

Week 4: Moments, Moment Generating Functions, and Continuous Random Variables

4.1 Moments and Moment Generating Functions

k-th Raw Moment:

$$\mu_k' = E[X^k]$$

k-th Central Moment:

$$\mu_k = E[(X - \mu)^k], \quad \text{where } \mu = E[X]$$

Important Moments:

- First raw moment: $\mu_1' = E[X] = \mu$ (mean)
- Second central moment: $\mu_2 = E[(X - \mu)^2] = \text{Var}(X)$ (variance)
- Third central moment: $\mu_3 = E[(X - \mu)^3]$ (measure of skewness)
- Fourth central moment: $\mu_4 = E[(X - \mu)^4]$ (measure of kurtosis)

Moment Generating Function (MGF):

$$M_X(t) = E[e^{tX}]$$

Existence: MGF exists if there exists a positive constant a such that $M_X(t)$ is finite for all $t \in [-a, a]$.

Key Properties:

1. **Uniqueness:** MGF uniquely determines the distribution
2. **Moment Generation:** $\mu_k' = M_X^{(k)}(0)$ (k-th derivative at $t = 0$)
3. **Independence:** If X, Y independent, then $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

Taylor Expansion:

$$M_X(t) = E[e^{tX}] = E\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{E[X^k] \cdot t^k}{k!}$$

4.2 Continuous Random Variables Fundamentals

Probability Density Function (PDF):

$$f_X(x) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Relationship between CDF and PDF:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

$$f_X(x) = \frac{dF_X(x)}{dx} \quad \text{(at points of continuity)}$$

Probability Calculations:

$$P(a < X \le b) = \int_{a}^{b} f_X(x)\, dx = F_X(b) - F_X(a)$$

Expected Value:

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

Variance:

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

4.3 Uniform Distribution

Notation: $X \sim \text{Uniform}(a, b)$ or $X \sim U(a, b)$

Probability Density Function:

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

Cumulative Distribution Function:

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

Expected Value and Variance:

$$E[X] = \frac{a+b}{2}$$

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

Moment Generating Function Derivation:

$$M_X(t) = E[e^{tX}] = \int_a^b e^{tx} \cdot \frac{1}{b-a}\, dx$$

For $t \neq 0$:

$$M_X(t) = \frac{1}{b-a} \cdot \left[\frac{e^{tx}}{t}\right]_a^b = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

For $t = 0$:

$$M_X(0) = 1$$

Special Handling: At $t = 0$, the MGF has a removable discontinuity requiring limit evaluation.

4.4 Normal Distribution

Notation: $X \sim \text{Normal}(\mu, \sigma^2)$ or $X \sim N(\mu, \sigma^2)$

Probability Density Function:

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

Standard Normal Distribution: $Z \sim N(0, 1)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$\Phi(z) = \int_{-\infty}^z \phi(t)\, dt$$

Standardization:
If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

Expected Value and Variance:

$$E[X] = \mu$$

$$\text{Var}(X) = \sigma^2$$

Moment Generating Function Derivation:

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

Completing the Square:

Combine the exponential terms:

$$tx - \frac{(x-\mu)^2}{2\sigma^2} = -\frac{1}{2\sigma^2}\left[x^2 - 2x(\mu + \sigma^2 t) + \mu^2\right]$$

$$= -\frac{1}{2\sigma^2}\left[(x - (\mu + \sigma^2 t))^2 - (\mu + \sigma^2 t)^2 + \mu^2\right]$$

$$= -\frac{(x - (\mu + \sigma^2 t))^2}{2\sigma^2} + \mu t + \frac{\sigma^2 t^2}{2}$$

Final Result:

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

68-95-99.7 Rule:

- $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$
- $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$
- $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$

4.5 Important Theorems and Properties

Continuity Correction:

When approximating discrete distributions with continuous ones:

$$P(X = k) \approx P(k - 0.5 < Y < k + 0.5)$$

Linear Transformation Properties:

If $Y = aX + b$, then:

$$E[Y] = aE[X] + b$$

$$\text{Var}(Y) = a^2\text{Var}(X)$$

$$M_Y(t) = e^{bt} \cdot M_X(at)$$

Reproductive Property of Normal Distribution:

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Week 5: Continuous Random Variables and Chebyshev's Theorem

5.1 Gamma Distribution

Notation: $X \sim \mathrm{Gamma}(\alpha, \beta)$

Probability Density Function:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

where $\alpha > 0$ (shape parameter) and $\beta > 0$ (rate parameter).

Gamma Function:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

Properties:

- $\Gamma(n) = (n-1)!$ for positive integers n
- $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$
- $\Gamma(1/2) = \sqrt{\pi}$

Expected Value and Variance:

$$E[X] = \frac{\alpha}{\beta}$$

$$\text{Var}(X) = \frac{\alpha}{\beta^2}$$

Moment Generating Function:

$$M_X(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}, \quad t < \beta$$

Special Cases:

- **Exponential Distribution:** $\text{Gamma}(1, \beta) = \text{Exponential}(\beta)$
- **Chi-Square Distribution:** $\text{Gamma}(k/2, 1/2) = \chi^2(k)$

5.2 Beta Distribution

Notation: $X \sim \text{Beta}(\alpha, \beta)$

Probability Density Function:

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}, \quad 0 < x < 1$$

where $\alpha > 0$ and $\beta > 0$ are shape parameters.

Beta Function:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \int_0^1 t^{\alpha-1}(1 - t)^{\beta-1}dt$$

Expected Value and Variance:

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Special Cases:

- **Uniform Distribution:** $\text{Beta}(1, 1) = \text{Uniform}(0, 1)$
- **Symmetric about 1/2:** When $\alpha = \beta$

Connection to Order Statistics:
If $X_1, X_2, \ldots, X_n$ are i.i.d. $\text{Uniform}(0, 1)$, then the k-th order statistic has distribution

Beta$(k, n - k + 1)$.

5.3 Chi-Square Distribution

Notation: $X \sim \chi^2(k)$ where k is the degrees of freedom

Probability Density Function:

$$f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad x > 0$$

Expected Value and Variance:

$$E[X] = k$$

$$\text{Var}(X) = 2k$$

Moment Generating Function:

$$M_X(t) = (1 - 2t)^{-k/2}, \quad t < \frac{1}{2}$$

Construction from Normal Distribution:
If $Z_1, Z_2, \ldots, Z_k$ are independent $N(0, 1)$ random variables, then:

$$\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$$

Reproductive Property:
If $X_1 \sim \chi^2(k_1)$ and $X_2 \sim \chi^2(k_2)$ are independent, then:

$$X_1 + X_2 \sim \chi^2(k_1 + k_2)$$

5.4 Chebyshev's Theorem

Statement:
For any random variable X with finite mean μ and variance σ^2, and for any $k > 1$:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Equivalent Form:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Mathematical Proof:

Let $Y = (X - \mu)^2$. By Markov's inequality:

$$P(Y \geq (k\sigma)^2) \leq \frac{E[Y]}{(k\sigma)^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

Since $\{|X - \mu| \geq k\sigma\} = \{(X - \mu)^2 \geq k^2\sigma^2\}$, the result follows.

Applications:

- Provides distribution-free bounds on tail probabilities
- Useful when the exact distribution is unknown
- Foundation for the Weak Law of Large Numbers

Example Applications:

- For $k = 2$: At least 75% of data falls within 2 standard deviations
- For $k = 3$: At least 89% of data falls within 3 standard deviations

5.5 Markov's Inequality

Statement:

For any non-negative random variable X and any $a > 0$:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Mathematical Proof:

$$E[X] = \int_0^\infty x f_X(x) dx \geq \int_a^\infty x f_X(x) dx \geq a \int_a^\infty f_X(x) dx = a \cdot P(X \geq a)$$

Generalized Form:

For any random variable X and any non-decreasing function g with $g(x) \geq 0$:

$$P(X \geq a) \leq \frac{E[g(X)]}{g(a)}$$

Connection to Chebyshev's Theorem:

Chebyshev's inequality is a special case of Markov's inequality applied to $Y = (X - \mu)^2$.

Applications:

- Tail bound estimation
- Convergence analysis
- Concentration inequalities

Week 6: Joint, Marginal, and Conditional Probability Distributions

6.1 Joint Probability Distributions

Discrete Case:
For discrete random variables X and Y, the joint probability mass function is:

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

Properties:

- $p_{X,Y}(x, y) \geq 0$ for all (x, y)
- $\sum_x \sum_y p_{X,Y}(x, y) = 1$

Continuous Case:
For continuous random variables X and Y, the joint probability density function satisfies:

$$f_{X,Y}(x, y) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$$

Joint Cumulative Distribution Function:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

Relationship between CDF and PDF:

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

Probability Calculations:

$$P(a < X \leq b, c < Y \leq d) = \int_{c}^{d} \int_{a}^{b} f_{X,Y}(x, y)\, dx\, dy$$

6.2 Marginal Distributions

Discrete Case:

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

Continuous Case:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

Marginal CDFs:

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y)$$

Expected Values:

$$E[X] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X,Y}(x, y)\, dx\, dy = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

6.3 Conditional Distributions

Discrete Case:

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)}, \quad \text{provided } p_X(x) > 0$$

Continuous Case:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \text{provided } f_X(x) > 0$$

Conditional Expectation:

$$E[Y|X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\, dy$$

Law of Total Expectation:

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

Conditional Variance:

$$\mathrm{Var}(Y|X = x) = E[Y^2|X = x] - (E[Y|X = x])^2$$

Law of Total Variance:

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y|X)] + \mathrm{Var}(E[Y|X])$$

6.4 Independence of Random Variables

Definition:
Random variables X and Y are independent if and only if:

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \quad \text{for all } x, y$$

Equivalent Conditions:

For Discrete Variables:

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y) \quad \text{for all } x, y$$

For Continuous Variables:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \quad \text{for all } x, y$$

Conditional Independence:

$$f_{Y|X}(y|x) = f_Y(y) \quad \text{and} \quad f_{X|Y}(x|y) = f_X(x)$$

Properties of Independent Variables:

- $E[XY] = E[X] \cdot E[Y]$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

- $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$
- $\text{Cov}(X, Y) = 0$ (but zero covariance doesn't imply independence)

6.5 Transformations of Bivariate Distributions

Linear Transformations:

If (X, Y) has joint density $f_{X,Y}(x, y)$ and we define:

$$U = aX + bY + c$$

$$V = dX + eY + f$$

where $J = |ae - bd| \neq 0$ (the Jacobian), then:

$$f_{U,V}(u, v) = \frac{1}{|J|} f_{X,Y}(x(u, v), y(u, v))$$

General Transformation Method:

For transformations $U = g_1(X, Y)$ and $V = g_2(X, Y)$:

1. Find the inverse transformation: $X = h_1(U, V)$, $Y = h_2(U, V)$
2. Compute the Jacobian: $J = \left| \frac{\partial(x,y)}{\partial(u,v)} \right|$
3. Apply the transformation formula: $f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v)) \cdot |J|$

Sum and Difference:

For independent X and Y:

- $U = X + Y$: $f_U(u) = \int_{-\infty}^{\infty} f_X(x) f_Y(u - x) dx$ (convolution)
- $V = X - Y$: $f_V(v) = \int_{-\infty}^{\infty} f_X(x) f_Y(x - v) dx$

Bivariate Normal Distribution:

$$(X, Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

Joint PDF:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left(-\frac{Q}{2(1 - \rho^2)} \right)$$

where:

$$Q = \frac{(x - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x - \mu_1)(y - \mu_2)}{\sigma_1\sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2}$$

Properties:

- Marginal distributions: $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$
- Independence: X and Y are independent if and only if $\rho = 0$
- Linear combinations are normal: $aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$

Summary and Key Points

Core Conceptual Framework

1. **Probability Space Triple:** $(S, \mathcal{F}, P)$
2. **Probability Axioms:** Non-negativity, normalization, countable additivity
3. **Conditional Probability & Independence:** Foundation for Bayes' theorem
4. **Random Variables:** Bridge between probability spaces and real numbers
5. **Distribution Theory:** Unified framework for discrete and continuous distributions
6. **Moment Theory:** Numerical characterization of distributions

Essential Formula Reference

Probability Foundations:

- **Bayes' Theorem:** $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- **Law of Total Probability:** $P(A) = \sum_i P(A|B_i)P(B_i)$

Discrete Distributions:

- **Binomial Distribution:** $P(X = k) = \binom{n}{k}p^k(1 - p)^{n-k}$
- **Poisson Distribution:** $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
- **Geometric Distribution:** $P(X = k) = (1 - p)^{k-1}p$

Continuous Distributions:

- **Normal Distribution:** $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$
- **Gamma Distribution:** $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$

- **Beta Distribution:** $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
- **Chi-Square Distribution:** $f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$

Moment Generating Functions:

- **Normal MGF:** $M_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})$
- **Gamma MGF:** $M_X(t) = \left(\frac{\beta}{\beta-t}\right)^{\alpha}$
- **Chi-Square MGF:** $M_X(t) = (1 - 2t)^{-k/2}$

Inequalities and Theorems:

- **Chebyshev's Theorem:** $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$
- **Markov's Inequality:** $P(X \geq a) \leq \frac{E[X]}{a}$

Joint Distributions:

- **Joint PDF Relationship:** $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$
- **Marginal PDF:** $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$
- **Conditional PDF:** $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
- **Law of Total Expectation:** $E[Y] = E[E[Y|X]]$
- **Law of Total Variance:** $\mathrm{Var}(Y) = E[\mathrm{Var}(Y|X)] + \mathrm{Var}(E[Y|X])$

Study Recommendations

1. **Master Fundamental Concepts:** Build strong foundation in probability axioms and basic calculations
2. **Distribution Mastery:** Understand when and how to apply discrete and continuous distributions
3. **MGF Techniques:** Practice deriving and using moment generating functions for various distributions
4. **Inequality Applications:** Learn to apply Chebyshev's and Markov's inequalities for tail bounds
5. **Joint Distribution Analysis:** Master techniques for working with multiple random variables
6. **Transformation Methods:** Practice Jacobian calculations and bivariate transformations
7. **Real-World Applications:** Connect theoretical concepts to practical statistical problems
8. **Problem Decomposition:** Learn to break complex probability problems into manageable steps

Weekly Learning Progression

- **Weeks 1-2:** Probability foundations and fundamental laws
- **Weeks 3-4:** Single random variable theory (discrete and continuous)
- **Weeks 5-6:** Advanced distributions and multivariate analysis

This summary covers the first six weeks of Columbia University's P8107 course syllabus, incorporating modern probability theory and mathematical statistics standards from the Wackerly, Mendenhall, and Scheaffer textbook.

Navigation Links

Good luck with your studies!