Week 11

Lecture 1

Probability Space and Basic Probability Calculus

Chapter 6: Probability

Outline

1. Discrete Random Variables
2. Continuous Random Variables
3. Maximum Likelihood and Advanced Topics
   This lecture focuses on the foundational concept of a **probability space**.

---

1. Experiment and Sample Space

Definition: Experiment

An experiment is a random process whose outcome is uncertain.
All probabilistic statements are defined relative to an experiment.

Definition: Sample Space

The sample space, denoted by Ω, is the set of all elementary outcomes of an experiment.

$$\Omega = \text{set of elementary outcomes}$$

Example 1: Single Coin Toss

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

Example 2: Single Die Throw

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

Cardinality:

$$|\Omega| = 6$$

Example 3: Double Coin Toss

$$\Omega = \{HH, HT, TH, TT\}$$

$$|\Omega| = 4$$

Example 4: Double Die Throw

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

$$|\Omega| = 36$$

Example 5: Card Picking Experiment

Pick one card from a standard 52 card deck.

$$|\Omega| = 52$$

2. Events as Sets

Definition: Event

An event is any subset of the sample space.

$$A \subseteq \Omega$$

The collection of all events of interest is denoted by $\mathcal{F}$.
For finite sample spaces:

$$\mathcal{F} = 2^{\Omega}$$

This is called the power set of Ω.
Cardinality:

$$|\mathcal{F}| = 2^{|\Omega|}$$

Example: Coin Toss

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

$$\mathcal{F} = \{\varnothing, \{H\}, \{T\}, \{H, T\}\}$$

3. Probability Space

A probability space is defined as:

$$(\Omega, \mathcal{F}, P)$$

Where:

- Ω is the sample space
- $\mathcal{F}$ is a sigma algebra of events
- $P : \mathcal{F} \to [0, 1]$ is a probability measure

4. Probability Measure

Definition

A probability measure is a function:

$$P : \mathcal{F} \to [0, 1]$$

that assigns a number between 0 and 1 to every event.

5. Axioms of Probability

Axiom 1

$$P(\Omega) = 1$$

Axiom 2

$$P(A) \geq 0 \quad \forall A \in \mathcal{F}$$

Axiom 3: Additivity

If $A_1, A_2, \ldots, A_n$ are disjoint events, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

6. Examples of Probability Measures

Example 1: Fair Coin

$$P(\varnothing) = 0$$

$$P(\{H\}) = P(\{T\}) = 0.5$$

$$P(\Omega) = 1$$

Additivity check:

$$P(\{H, T\}) = P(\{H\}) + P(\{T\}) = 1$$

Example 2: Equal Probability on Double Coin Toss

$$\Omega = \{HH, HT, TH, TT\}$$

Define:

$$P(A) = \frac{|A|}{4}$$

Properties:

$$P(\varnothing) = 0$$

$$P(\Omega) = 1$$

For example:

$$P(\{HH, HT, TT\}) = \frac{3}{4}$$

Example 3: Double Die Throw

$$P(A) = \frac{|A|}{36}$$

Example 4: Degenerate Probability

Define:

$$P(A) = \begin{cases} 1 & \text{if } HH \in A \\ 0 & \text{otherwise} \end{cases}$$

This corresponds to a deterministic experiment where outcome is always HH.

---

7. Basic Probability Calculus

Complement Rule

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

Union Rule

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

---

8. Inclusion Exclusion Principle

For three events:

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

General Form

For $A_1, \ldots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n)$$

---

9. Equal Probability Measure

Special case:

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

Probability reduces to counting.

---

10. Counting Example: Double Die Throw

Find probability that sum equals 6.
Favorable outcomes:

$$(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)$$

Total outcomes: 36

$$P(\text{sum} = 6) = \frac{5}{36}$$

11. Counting Example: Banana Letters

Letters: B, A, N, A, N, A
Randomly permute all 6 letters.

$$|\Omega| = 6!$$

Event: first three letters are A, A, A
Favorable arrangements:

$$3! \cdot 3!$$

Probability:

$$P(\text{AAA}) = \frac{3! \cdot 3!}{6!}$$

Summary

A probability space consists of:

$$(\Omega, \mathcal{F}, P)$$

where

- Ω is the sample space
- $\mathcal{F}$ is a collection of subsets of Ω
- P is a function mapping events to values in $[0, 1]$ satisfying the axioms of probability

This forms the foundation for all subsequent concepts in probability and machine learning.

****

Lecture 2

Conditioning, Independence and Bayes Rule

1. Conditional Probability

Definition

Let $A, B \subseteq \Omega$ with $P(B) > 0$.
The conditional probability of A given B is

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

Interpretation: Probability that A occurs given that B has occurred.

---

2. Conditional Probability Example: Double Dice Throw

Setup

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

$$P(A) = \frac{|A|}{36}$$

Example 1

Find

$$P(\text{sum} = 7 \mid \text{first} = 5)$$

Numerator

Event: first = 5 and sum = 7
Only outcome:

$$(5, 2)$$

$$P(\text{sum} = 7 \cap \text{first} = 5) = \frac{1}{36}$$

Denominator

First = 5:

$$(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)$$

$$P(\text{first} = 5) = \frac{6}{36}$$

Result

$$P(\text{sum} = 7 \mid \text{first} = 5) = \frac{1/36}{6/36} = \frac{1}{6}$$

Example 2

Find

$$P(\text{first} = 5 \mid \text{sum} = 8)$$

Numerator

First = 5 and sum = 8:

$$(5,3)$$

$$P = \frac{1}{36}$$

Denominator

Sum = 8:

$$(2,6), (3,5), (4,4), (5,3), (6,2)$$

$$P(\text{sum} = 8) = \frac{5}{36}$$

Result

$$P(\text{first} = 5 \mid \text{sum} = 8) = \frac{1/36}{5/36} = \frac{1}{5}$$

3. Total Probability Law

Definitions

Events $B_1, B_2, \ldots, B_n$ are:

Mutually Exclusive

$$B_i \cap B_j = \varnothing \quad \text{for } i \neq j$$

Exhaustive

$$B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$$

Total Probability Formula

For any event A:

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

Using conditioning:

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

---

4. Total Probability Example: Two Urn Experiment

Setup

Urn 1: 9 white, 1 black
Urn 2: 5 white, 5 black

Choose urn uniformly, then pick one ball.
Define:

- A: white ball picked
- B_1: urn 1 chosen
- B_2: urn 2 chosen

Since B_1, B_2 are mutually exclusive and exhaustive:

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2)$$

Compute:

$$P(B_1) = P(B_2) = \frac{1}{2}$$

$$P(A \mid B_1) = \frac{9}{10}$$

$$P(A \mid B_2) = \frac{5}{10}$$

Result:

$$P(A) = \frac{9}{10} \cdot \frac{1}{2} + \frac{5}{10} \cdot \frac{1}{2}$$

5. Independence

Definition

Events A and B are independent if

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

Equivalent form when $P(B) > 0$:

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

Independence Example: Double Dice

Let:

- A: sum = 7
- B: first = 5

$$P(A \cap B) = \frac{1}{36}$$

$$P(A) = \frac{6}{36}$$

$$P(B) = \frac{6}{36}$$

$$P(A)P(B) = \frac{6}{36} \cdot \frac{6}{36} = \frac{1}{36}$$

Thus A and B are independent.

Non Independence Example

Let:

- D: sum = 4

$$P(B \cap D) = 0$$

$$P(B) = \frac{1}{6}$$

$$P(D) = \frac{3}{36}$$

$$P(B)P(D) \neq 0$$

Thus B and D are not independent.

---

6. Bayes Rule

Derivation

From definition:

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

Using total probability with A and A^c:

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

Thus,

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

---

7. Bayes Rule Example: Urn Problem

Find:

$$P(B_1 \mid A)$$

Where:

- A: white ball picked
- B_1: urn 1 chosen

Given:

$$P(A \mid B_1) = \frac{9}{10}$$

$$P(A \mid B_2) = \frac{5}{10}$$

$$P(B_1) = P(B_2) = \frac{1}{2}$$

Using Bayes rule:

$$P(B_1 \mid A) = \frac{\frac{9}{10} \cdot \frac{1}{2}}{\frac{9}{10} \cdot \frac{1}{2} + \frac{5}{10} \cdot \frac{1}{2}} = \frac{9}{14}$$

8. Bayes Rule Example: COVID Testing

Let:

- C: person has COVID
- T: test positive

Given:

$$P(C) = \frac{1}{100}$$

$$P(T \mid C) = 0.9$$

$$P(T^c \mid C^c) = 0.9$$

Thus:

$$P(T \mid C^c) = 0.1$$

Apply Bayes rule:

$$P(C \mid T) = \frac{0.9 \cdot \frac{1}{100}}{0.9 \cdot \frac{1}{100} + 0.1 \cdot \frac{99}{100}}$$

Simplify:

$$= \frac{0.9}{0.9 + 9.9} \approx 0.1$$

Key Insight:

$$P(C \mid T) \neq P(T \mid C)$$

Even with a 90 percent accurate test, low prevalence leads to small posterior probability.

Summary

1. Conditional probability refines probability given information.
2. Total probability decomposes events over partitions.
3. Independence requires product rule.
4. Bayes rule flips conditioning.
5. Bayes rule is foundational for inference in machine learning.

****

Lecture 3

Random Variables

Definition

Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable is a function

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

If the range of X is finite or countable, then X is called a discrete random variable.
Since in all previous examples Ω was finite, any function defined on Ω is a discrete random variable.

Examples of Random Variables

Example 1: Double Dice Throw Experiment

Sample space:

$$\Omega = \{(a, b) : a, b \in \{1, 2, 3, 4, 5, 6\}\}$$

Define

$$X(a, b) = a$$

This random variable represents the first die result.

Example 2: Sum of Dice

Define

$$X(a, b) = a + b$$

This random variable represents the sum of the faces.

Example 3: Double Coin Toss

Sample space:

$$\Omega = \{HH, HT, TH, TT\}$$

Define X as the number of heads:

$$X(HH) = 2$$
$$X(HT) = 1$$
$$X(TH) = 1$$
$$X(TT) = 0$$

Probability Mass Function

For a discrete random variable X, the probability mass function is defined as

$$f_X : \mathbb{R} \to [0, 1]$$
$$f_X(x) = P(X = x)$$

More formally,

$$f_X(x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Properties

1. Non negativity:

$$0 \leq f_X(x) \leq 1$$

2. Total probability:

$$\sum_{x \in \text{Range}(X)} f_X(x) = 1$$

Proof:

$$\sum_x f_X(x) = \sum_x P(\{\omega : X(\omega) = x\})$$

The sets $\{\omega : X(\omega) = x\}$ are disjoint and their union equals Ω, hence

$$\sum_x f_X(x) = P(\Omega) = 1$$

Cumulative Distribution Function

The cumulative distribution function is defined as

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

Properties

1. $$0 \leq F_X(x) \leq 1$$

2. $$F_X(-\infty) = 0$$

3. $$F_X(\infty) = 1$$

Example: First Die Result

Under counting probability:

$$P(A) = \frac{|A|}{36}$$

Random variable:

$$X = \text{first die result}$$

PMF:

$$f_X(x) = \begin{cases} \frac{1}{6} & x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{otherwise} \end{cases}$$

CDF:
For $x < 1$:

$$F_X(x) = 0$$

For $k \in \{1, \ldots, 6\}$:

$$F_X(k) = \frac{k}{6}$$

For $x \geq 6$:

$$F_X(x) = 1$$

Example: Number of Heads

Double coin toss with

$$P(A) = \frac{|A|}{4}$$

Random variable: number of heads.
PMF:

$$P(X = 0) = \frac{1}{4}$$

$$P(X = 1) = \frac{1}{2}$$

$$P(X = 2) = \frac{1}{4}$$

CDF:

$$F_X(0) = \frac{1}{4}$$

$$F_X(1) = \frac{3}{4}$$

$$F_X(2) = 1$$

Example: Biased Dice

Suppose faces 1 to 5 are equally likely and face 6 is twice as likely.
Let

$$P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = C$$

$$P(X = 6) = 2C$$

Using normalization:

$$5C + 2C = 1$$

$$7C = 1$$

$$C = \frac{1}{7}$$

Hence,

$$P(X = i) = \frac{1}{7}, \quad i = 1, \ldots, 5$$

$$P(X = 6) = \frac{2}{7}$$

Indicator Random Variables

For an event A, define

$$X(\omega) = \mathbf{1}_A(\omega)$$

$$X(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

PMF:

$$P(X = 1) = P(A)$$

$$P(X = 0) = 1 - P(A)$$

Example: Double coin toss.
Let A be the event two heads occur.
Then

$$P(X = 1) = P(HH)$$

$$P(X = 0) = 1 - P(HH)$$

Functions of Random Variables

If X is a random variable and $g : \mathbb{R} \to \mathbb{R}$, then

$$Y = g(X)$$

is also a random variable.
Example:
Double dice throw.

$$X = \text{first die result}$$

Define

$$Y = |X - 2|$$

Compute PMF:

$$P(Y = 0) = P(X = 2) = \frac{1}{6}$$

$$P(Y = 1) = P(X = 1 \text{ or } X = 3) = \frac{2}{6}$$

Similarly for other values.

Summary

A discrete random variable is a function from Ω to $\mathbb{R}$.
Key tools:

- Probability mass function
- Cumulative distribution function
- Indicator random variables
- Functions of random variables
  These form the foundation for further analysis including expectation, variance and advanced probabilistic models.

****

Lecture 4

Expectation and Variance

Expectation

Definition

Let $X : \Omega \to \mathbb{R}$ be a discrete random variable with probability mass function f_X.
The expectation of X is defined as

$$E[X] = \sum_{x \in \text{Range}(X)} x f_X(x)$$

This is a weighted average of all possible values of X.

Interpretation

Expectation represents the center of mass of the probability mass function.

Examples of Expectation

Example 1: Single Dice Throw

Let X be the face value of a fair dice.

$$f_X(x) = \frac{1}{6}, \quad x \in \{1, 2, 3, 4, 5, 6\}$$

Then

$$E[X] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Example 2: Double Dice Throw

Let X be the sum of two fair dice.
Possible values:

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

PMF:

$$f_X(2) = \frac{1}{36}, \quad f_X(3) = \frac{2}{36}, \ldots, f_X(7) = \frac{6}{36}, \ldots, f_X(12) = \frac{1}{36}$$

Then

$$E[X] = \sum_x x f_X(x) = 7$$

Example 3: Absolute Difference of Two Dice

Let $Y = |D_1 - D_2|$.
PMF:

$$f_Y(0) = \frac{6}{36}$$

$$f_Y(1) = \frac{10}{36}$$

$$f_Y(2) = \frac{8}{36}$$

$$f_Y(3) = \frac{6}{36}$$

$$f_Y(4) = \frac{4}{36}$$

$$f_Y(5) = \frac{2}{36}$$

Expectation:

$$E[Y] = \frac{6}{36}0 + \frac{10}{36}1 + \frac{8}{36}2 + \frac{6}{36}3 + \frac{4}{36}4 + \frac{2}{36}5$$

$$E[Y] = \frac{70}{36}$$

Conditional Expectation

Conditional PMF

Let $A \subseteq \Omega$ be an event.
The conditional PMF is

$$f_{X|A}(x) = P(X = x \mid A)$$

Conditional Expectation

$$E[X \mid A] = \sum_{x \in \text{Range}(X)} x f_{X|A}(x)$$

Example 1

Let X be sum of two dice and A be the event that first die equals 2.
Then

$$f_{X|A}(3) = \frac{1}{6}, \ldots, f_{X|A}(8) = \frac{1}{6}$$

Thus

$$E[X \mid A] = \frac{1}{6}(3 + 4 + 5 + 6 + 7 + 8) = 5.5$$

Example 2

Let A be the event that difference of dice equals 0.
Possible sums: $2, 4, 6, 8, 10, 12$ each with probability $\frac{1}{6}$.
Then

$$E[X \mid A] = \frac{1}{6}(2 + 4 + 6 + 8 + 10 + 12) = 7$$

Linearity of Expectation

For random variables X and Y:

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

For constant a:

$$E[aX] = aE[X]$$

Expectation is a linear operator.

Example: 10 Coin Toss Experiment

Let X be number of heads in 10 fair coin tosses.
Define indicator variables

$$X_i = \mathbf{1}\{\text{toss } i \text{ is head}\}$$

Then

$$X = \sum_{i=1}^{10} X_i$$

By linearity:

$$E[X] = \sum_{i=1}^{10} E[X_i]$$

Since

$$E[X_i] = P(\text{head}) = \frac{1}{2}$$

Therefore

$$E[X] = 10 \cdot \frac{1}{2} = 5$$

Variance

Definition

Variance of X:

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

Equivalent Formula

Expanding:

$$(X - E[X])^2 = X^2 + (E[X])^2 - 2XE[X]$$

Taking expectation:

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

Interpretation

Variance measures average squared deviation from the mean.
If X is constant:

$$\mathrm{Var}(X) = 0$$

Properties of Variance

For constant a:

$$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$$

For random variables X and Y:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

Variance is not linear.

Example: Variance of Dice Face

Let X be face of fair dice.

$$E[X] = 3.5$$

Compute

$$Y = (X - 3.5)^2$$

Values:

$$6.25, 2.25, 0.25, 0.25, 2.25, 6.25$$

Thus

$$\text{Var}(X) = \frac{1}{6}(6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25)$$

```
*********************************************************************************
```

Lecture 5

Multiple Random Variables, Independence and Covariance

Joint Distributions

Let $X, Y : \Omega \to \mathbb{R}$ be two random variables.

Joint Probability Mass Function

The joint PMF is defined as

$$f_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$$

In event notation,

$$f_{XY}(x, y) = \mathbb{P}\Big(\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\}\Big)$$

Joint Cumulative Distribution Function

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

Normalization Property

$$\sum_{x,y} f_{XY}(x,y) = 1$$

This holds since the sets

$$\{\omega : X(\omega) = x, Y(\omega) = y\}$$

are disjoint and their union equals Ω.

Marginal Distributions

The marginal PMFs are obtained by summing out the other variable.

Marginal of X

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

$$f_X(x) = \sum_y f_{XY}(x,y)$$

Marginal of Y

$$f_Y(y) = \sum_x f_{XY}(x,y)$$

Conditional Distributions

Conditional PMF of X given $Y = y$

$$f_{X|Y}(x|y) = \mathbb{P}(X = x \mid Y = y)$$

Using conditional probability,

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Similarly,

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

For any fixed y,

$$\sum_x f_{X|Y}(x|y) = 1$$

Independence of Random Variables

Definition via Joint Distribution

Random variables X and Y are independent if

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

for all x, y.

Equivalent Event-Based Definition

For all constants a, b,

$$\{X = a\} \text{ and } \{Y = b\}$$

are independent events.

Equivalent Expectation-Based Definition

For all functions g, h,

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

Proof of Expectation Form

$$\mathbb{E}[g(X)h(Y)] = \sum_{x,y} f_{XY}(x, y) g(x) h(y)$$

If X, Y are independent,

$$= \sum_{x,y} f_X(x) f_Y(y) g(x) h(y)$$

Rearranging,

$$= \left(\sum_x f_X(x) g(x) \right) \left(\sum_y f_Y(y) h(y) \right)$$

$$= \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

Sum of Independent Random Variables

Let $Z = X + Y$.
Then

$$f_Z(z) = \mathbb{P}(Z = z)$$

$$= \sum_x \mathbb{P}(X = x, Y = z - x)$$

$$= \sum_x f_{XY}(x, z - x)$$

If X, Y are independent,

$$f_Z(z) = \sum_x f_X(x) f_Y(z - x)$$

This operation is called convolution.

Conditional Expectation with Respect to a Random Variable

We have already defined

$$\mathbb{E}[X \mid A]$$

for event A.
For random variable Y, we define

$$\mathbb{E}[X \mid Y]$$

as a function of Y.
Example:
Let X be the sum of two dice and Y be the first die.
Then

$$\mathbb{E}[X \mid Y] = Y + 3.5$$

Thus conditional expectation is itself a random variable.

Covariance

Let $X, Y : \Omega \to \mathbb{R}$.

Definition

$$\mathrm{Cov}(X, Y) = \mathbb{E}\Big[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\Big]$$

Special case:

$$\mathrm{Cov}(X, X) = \mathrm{Var}(X)$$

Algebraic Simplification

Expanding,

$$(X - \mathbb{E}X)(Y - \mathbb{E}Y) = XY - (\mathbb{E}X)Y - X(\mathbb{E}Y) + \mathbb{E}X\mathbb{E}Y$$

Taking expectation,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$$

Independence Implies Zero Covariance

If X, Y are independent,

$$\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$$

Hence,

$$\mathrm{Cov}(X, Y) = 0$$

Thus independent random variables are uncorrelated.

Uncorrelated Does Not Imply Independent

Two random variables may satisfy

$$\mathrm{Cov}(X, Y) = 0$$

but still fail

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

Hence uncorrelated does not imply independent.

Interpretation of Covariance

- Positive covariance: variables increase or decrease together.
- Negative covariance: one increases while the other decreases.
- Zero covariance: no linear relationship.

Summary

1. Joint PMF generalizes single-variable PMF.

2. Marginals obtained by summation.
3. Conditionals obtained via ratio of joint to marginal.
4. Independence characterized by factorization of joint.
5. Sum of independent variables uses convolution.
6. Conditional expectation is a function of the conditioning variable.
7. Covariance measures linear dependence.
8. Independence implies zero covariance, but not vice versa.

```
**************************************************************************
```

Lecture 6

Standard Discrete Random Variables

We summarize four standard discrete random variables:

1. Bernoulli
2. Binomial
3. Poisson
4. Geometric
   Each distribution is defined via its probability mass function, expectation, and variance.

Bernoulli Distribution

Definition

A random variable X is distributed as Bernoulli with parameter p if

$$PX = 1 = p$$

$$PX = 0 = 1 - p$$

Notation:

$$X \sim Bernoullip$$

Parameter:

$$0 \leq p \leq 1$$

The parameter p is called the bias.

Expectation

$$EX = 1 \cdot p + 0 \cdot 1 - p = p$$

Variance

Using

$$VarX = EX^2 - EX^2$$

Since $X^2 = X$ for $X \in \{0, 1\}$,

$$EX^2 = EX = p$$

Thus,

$$VarX = p - p^2 = p1 - p$$

Binomial Distribution

Construction

Let

$$X_1, X_2, \ldots, X_n$$

be independent Bernoulli p random variables.
Define

$$X = \sum_{i=1}^{n} X_i$$

Then

$$X \sim Binomialnp$$

Probability Mass Function

For $k = 0, 1, \ldots, n,$

$$PX = k = \binom{n}{k} p^k 1 - p^{n-k}$$

Validity

$$\sum_{k=0}^{n} \binom{n}{k} p^k 1 - p^{n-k} = p + 1 - p^n = 1$$

by the binomial theorem.

Expectation

Using linearity of expectation:

$$EX = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} p = np$$

Variance

Since the X_i are independent:

$$VarX = \sum_{i=1}^{n} VarX_i$$

Each

$$VarX_i = p1 - p$$

Thus

$$VarX = np1 - p$$

Poisson Distribution

Definition

A random variable X is Poisson with parameter $\lambda > 0$ if

$$PX = k = e^{-\lambda} \frac{\lambda^k}{k!}$$

for $k = 0, 1, 2, \ldots$

Notation:

$$X \sim Poisson \lambda$$

Validity

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

Expectation

$$EX = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!}$$

Rewrite:

$$= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$

Cancel k:

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k-1!}$$

Factor λ:

$$= e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{k-1!}$$

Let $l = k - 1$:

$$= e^{-\lambda} \lambda \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} = e^{-\lambda} \lambda e^{\lambda} = \lambda$$

Thus,

$$EX = \lambda$$

Variance:

$$VarX = \lambda$$

Geometric Distribution

Definition

A random variable X is geometric with parameter p if

$$PX = k = 1 - p^{k-1}p$$

for $k = 1, 2, 3, \ldots$

Notation:

$$X \sim Geometricp$$

Support:

$$k = 1, 2, \ldots$$

Validity

$$\sum_{k=1}^{\infty} 1 - p^{k-1}p = p \sum_{k=1}^{\infty} 1 - p^{k-1}$$

Geometric series:

$$\sum_{k=1}^{\infty} 1 - p^{k-1} = \frac{1}{1 - 1 - p} = \frac{1}{p}$$

Thus,

$$p \cdot \frac{1}{p} = 1$$

Expectation

$$EX = \sum_{k=1}^{\infty} k 1 - p^{k-1}p$$

Rewrite:

$$= p \sum_{k=1}^{\infty} k 1 - p^{k-1}$$

Using differentiation of geometric series:

$$\sum_{k=1}^{\infty} k1 - p^{k-1} = \frac{1}{p^2}$$

Thus,

$$EX = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

Summary of Moments

Bernoulli p:

$$EX = p$$
$$VarX = p1 - p$$

Binomial np:

$$EX = np$$
$$VarX = np1 - p$$

Poisson λ:

$$EX = \lambda$$
$$VarX = \lambda$$

Geometric p:

$$EX = \frac{1}{p}$$

****

Lecture 7

Continuous Random Variables

Probability Space

A probability experiment is defined by the tuple

$$\Omega, \mathcal{F}, P$$

where

- Ω is the sample space
- $\mathcal{F}$ is a collection of events
- P is a probability measure

Sample Space

In the continuous setting, the sample space Ω is uncountable.
Example:

- Measuring the height of a randomly chosen person
- Waiting time for a bus
- Price of a house
- Thermal noise current in a wire

Sigma Algebra

The collection of events $\mathcal{F} \subseteq 2^\Omega$ satisfies:

1. $\Omega \in \mathcal{F}$
2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
3. If $A_1, A_2, \cdots \in \mathcal{F}$ then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

Probability Measure

The probability measure

$$P : \mathcal{F} \to \mathbb{R}_+$$

satisfies:

1. $PA \geq 0$
2. $P\Omega = 1$
3. For disjoint A_i,

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

Definition of Continuous Random Variable

A continuous random variable is a function

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

where both domain and range are uncountable.

For continuous random variables:

$$PX = x = 0$$

for every real number x.

Probability Density Function

Definition

The probability density function is

$$f_X x = \lim_{dx \to 0} \frac{PX \in x, x + dx}{dx}$$

Interpretation:

$$PX \in x, x + dx = f_X x \, dx$$

Properties of PDF

1.
$$f_X x \geq 0$$

2.
$$\int_{-\infty}^{\infty} f_X x \, dx = 1$$

The density may exceed 1.

Cumulative Distribution Function

Definition

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

Relationship with PDF:

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

Properties of CDF

1.
$$F_X - \infty = 0$$

2.
$$F_X \infty = 1 \$\$3.\$ F_X x \$ is nondecreasing.$$

Example 1: Uniform Waiting Time

Let

$$X = \text{waiting time for bus}$$

Assume arrival is uniform between two buses 15 minutes apart.
Support:

$$0 \leq x \leq 15$$

PDF

Since density is constant:

$$f_X x = \begin{cases} c & 0 \leq x \leq 15 \\ 0 & \text{otherwise} \end{cases}$$

Using normalization:

$$\int_0^{15} c \, dx = 1$$

$$15c = 1$$

$$c = \frac{1}{15}$$

Thus,

$$f_X x = \begin{cases} \frac{1}{15} & 0 \le x \le 15 \\ 0 & \text{otherwise} \end{cases}$$

CDF

For $x < 0$:

$$F_X x = 0$$

For $0 \le x \le 15$:

$$F_X x = \int_0^x \frac{1}{15} dt = \frac{x}{15}$$

For $x > 15$:

$$F_X x = 1$$

Example 2: Linear Density

Let

$$f_X x = \begin{cases} \frac{x}{2} & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Validity Check

Compute total integral:

$$\int_0^2 \frac{x}{2} dx = \frac{1}{2} \int_0^2 x dx = \frac{1}{2} \cdot \frac{x^2}{2} \Big|_0^2 = \frac{1}{2} \cdot \frac{4}{2} = 1$$

Hence valid PDF.

CDF

For $x < 0$:

$$F_X x = 0$$

For $0 \leq x \leq 2$:

$$F_X x = \int_0^x \frac{t}{2} dt = \frac{1}{2} \cdot \frac{x^2}{2} = \frac{x^2}{4}$$

For $x > 2$:

$$F_X x = 1$$

****

Lecture 8

Conditional PDF

Probability Space Setup

An experiment is given by the tuple
$\Omega, \mathcal{F}, P$
Let $X : \Omega \to \mathbb{R}$ be a continuous random variable.
Let $A \subseteq \Omega$ be an event with $PA > 0$.

Definition

The conditional probability density function of X given A is

$$f_{X|A} x = \frac{PX \in x \text{ to } x + dx \mid A}{dx}$$

This is the density of X after incorporating the information that A has occurred.
Using conditional probability,

$$PX \in x \text{ to } x + dx \mid A = \frac{PX \in x \text{ to } x + dx \cap A}{PA}$$

Hence,

$$f_{X|A} x = \frac{f_X x}{PA} \quad \text{whenever } x \text{ satisfies } A$$

and 0 otherwise.

Example: Conditional Density

Given Density

Let

$$f_X(x) = \begin{cases} \frac{x}{2} & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

This is a valid density since

$$\int_0^2 \frac{x}{2}\,dx = \frac{1}{2} \cdot \frac{x^2}{2}\Big|_0^2 = 1$$

Event

Let

$$A = \{X > 1\}$$

Compute

$$P(A) = \int_1^2 \frac{x}{2}\,dx = \frac{x^2}{4}\Big|_1^2 = 1 - \frac{1}{4} = \frac{3}{4}$$

Conditional Density

For $x < 1$:

$$f_{X|A}(x) = 0$$

For $1 \le x \le 2$:

$$f_{X|A}(x) = \frac{f_X(x)}{P(A)} = \frac{\frac{x}{2}}{\frac{3}{4}} = \frac{4x}{6} = \frac{2x}{3}$$

For $x > 2$:

$$f_{X|A}(x) = 0$$

Thus,

$$f_{X|A}(x) = \begin{cases} 0 & x \le 1 \\ \frac{2x}{3} & 1 \le x \le 2 \\ 0 & x > 2 \end{cases}$$

The shape remains linear on the interval but is rescaled so that

$$\int_1^2 f_{X|A}(x)\,dx = 1$$

Functions of Random Variables

Definition

If X is a random variable and

$$Y = gX$$

then Y is also a random variable.
To find the density of Y,

$$f_Y y = \frac{PY \in y \text{ to } y + dy}{dy}$$

Example 1: Linear Transformation

Given

Let X be uniform on -1 to 1:

$$f_X x = \begin{cases} \frac{1}{2} & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let

$$Y = \frac{X}{2}$$

Derivation

$$f_Y y = \frac{PX \in 2y \text{ to } 2y + 2dy}{dy}$$

If $2y$ lies outside -1 to 1, density is zero.
Thus y must lie in $-\frac{1}{2}$ to $\frac{1}{2}$.
Within this range,

$$PX \in 2y \text{ to } 2y + 2dy = \frac{1}{2} \cdot 2dy$$

Hence,

$$f_Y y = 1 \quad \text{for} \quad -\frac{1}{2} \leq y \leq \frac{1}{2}$$

So Y is uniform on $-\frac{1}{2}$ to $\frac{1}{2}$.

Example 2: Absolute Value Transformation

Let

$$Y = |X|$$

with X uniform on -1 to 1.

Since both x and $-x$ map to the same y, density doubles on positive side.

Thus,

$$f_Y y = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Check normalization:

$$\int_0^1 1 dy = 1$$

Key Observations

1. Conditional density rescales original density over restricted region.
2. Density values can exceed 1 as long as total integral equals 1.
3. Functions of random variables are themselves random variables.
4. Linear scaling compresses or stretches support.
5. Non one to one mappings such as absolute value combine contributions from multiple regions.

****

Lecture 9

Expectation of Continuous Random Variables

Definition

Let $X : \Omega \to \mathbb{R}$ be a continuous random variable with density $f_X x$.

The expectation of X is defined as

$$EX = \int_{-\infty}^{\infty} x f_X x \, dx$$

This is the continuous analogue of

$$EX = \sum_x x f_X x$$

Interpretation:
Expectation is the center of mass of the density function.

Properties of Expectation

Linearity

For random variables X and Y,

$$EX + Y = EX + EY$$

Expectation of a Function

If $Y = gX$, then

$$EY = \int_{-\infty}^{\infty} gx \, f_X x \, dx$$

This avoids computing the density of Y.

Example 1: Uniform Distribution

Let X be uniform on a to b:

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

Expectation

$$EX = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

Example 2: Triangular Density

Let

$$f_X x = \begin{cases} \frac{x}{2} & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Expectation

$$EX = \int_0^2 x \frac{x}{2} dx = \frac{1}{2} \int_0^2 x^2 dx$$

$$= \frac{1}{2} \frac{x^3}{3} \Big|_0^2 = \frac{1}{2} \frac{8}{3} = \frac{4}{3}$$

Variance

Definition

Variance is

$$VarX = EX - EX^2$$

More commonly written as

$$VarX = EX^2 - EX^2$$

Variance measures spread around the mean.
Standard deviation is

$$\sqrt{VarX}$$

Properties of Variance

1. Non-negativity

$$VarX \ge 0$$

Equality holds iff X is constant.
2. Scaling

$$VaraX = a^2 VarX$$

3. Non-linearity
In general,

$$VarX + Y \ne VarX + VarY$$

Variance of Uniform Distribution

Let X be uniform on a to b.

We compute

$$EX^2 = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b$$

$$= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

We already know

$$EX = \frac{a+b}{2}$$

Thus

$$VarX = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4}$$

After simplification,

$$VarX = \frac{(b-a)^2}{12}$$

Conditional Expectation

Definition

Given event A,

$$EX \mid A = \int_{-\infty}^{\infty} x f_{X|A} x dx$$

Law of Total Expectation

If A and A^c partition the sample space,

$$EX = EX \mid A \, PA + EX \mid A^c \, PA^c$$

Example: Bus Waiting Time

Arrival time uniform between 7:10 and 7:30.
Bus every 15 minutes.
Let X be waiting time.
Define event

$$A = \text{arrival before 7:15}$$

Then

$$PA = \frac{5}{20} = \frac{1}{4}$$

Conditional Expectations

If A occurs, waiting time uniform on 0 to 5:

$$EX \mid A = 2.5$$

If A^c occurs, waiting time uniform on 0 to 15:

$$EX \mid A^c = 7.5$$

Total Expectation

$$EX = \frac{1}{4} \cdot 2.5 + \frac{3}{4} \cdot 7.5$$
$$= 6.25$$

Key Observations

1. Continuous expectation replaces sums by integrals.
2. Variance measures spread and scales quadratically.
3. Law of total expectation simplifies complex computations.
4. Conditioning can dramatically simplify expectation calculations.

****

Lecture 10

Multiple Continuous Random Variables

Joint Density Function

Let X and Y be two continuous random variables defined on the same probability space. The joint probability density function is defined as

$$f_{XY}(x,y) = \frac{P\big(X \in x \text{ to } x + dx,\ Y \in y \text{ to } y + dy\big)}{dx\,dy}$$

This is a two dimensional density. Probability of a small rectangle is

$$P\big(X \in x \text{ to } x + dx,\ Y \in y \text{ to } y + dy\big) = f_{XY}(x,y)\,dx\,dy$$

Properties:

1. Non negativity

$$f_{XY}(x,y) \geq 0$$

2. Total probability equals 1

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$$

Joint Cumulative Distribution Function

The joint CDF is defined as

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

Properties:

$$F_{XY}(-\infty, -\infty) = 0$$
$$F_{XY}(\infty, \infty) = 1$$

Monotonicity:
If $x_1 \leq x_2$ and $y_1 \leq y_2$, then

$$F_{XY}(x_1, y_1) \leq F_{XY}(x_2, y_2)$$

It is non decreasing in each argument.

Example: Height and Weight

Let

- X = height in meters

- Y = weight in kilograms

Define joint density

$$f_{XY}(x, y) = \begin{cases} 0 & \text{if } x \notin 1 \text{ to } 2 \\ 0 & \text{if } y \notin 50x - 20 \text{ to } 50x + 20 \\ c & \text{otherwise} \end{cases}$$

Support is the parallelogram:

- $1 \leq x \leq 2$
- $50x - 20 \leq y \leq 50x + 20$

To determine c, use normalization:

$$\int_1^2 \int_{50x-20}^{50x+20} c \, dy \, dx = 1$$

Inner integral:

$$\int_{50x-20}^{50x+20} dy = 40$$

Thus

$$\int_1^2 40c \, dx = 40c$$

Hence

$$40c = 1$$

$$c = \frac{1}{40}$$

Marginal Densities

Marginal of X

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

For $x \in 1$ to 2:

$$f_X(x) = \int_{50x-20}^{50x+20} \frac{1}{40} \, dy$$

$$= \frac{1}{40} \cdot 40 = 1$$

Thus

$$f_X(x) = \begin{cases} 1 & x \in 1 \text{ to } 2 \\ 0 & \text{otherwise} \end{cases}$$

So X is uniform on 1 to 2.

Marginal of Y

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

Must consider cases based on y:

- $y < 30 : 0$
- $30 \leq y < 70$: increasing support
- $70 \leq y < 80$: full overlap
- $80 \leq y \leq 120$: decreasing support
- $y > 120 : 0$

Each case is obtained by solving

$$50x - 20 \leq y \leq 50x + 20$$

for x, then integrating $\frac{1}{40}$ over that interval.
Sanity check:

$$\int_{-\infty}^{\infty} f_Y(y)\, dy = 1$$

Conditional Densities

Conditional of Y given X

Definition:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

For $x \in 1$ to 2:
Since $f_X(x) = 1$,

$$f_{Y|X}(y|x) = f_{XY}(x, y)$$

Thus

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{40} & y \in 50x - 20 \text{ to } 50x + 20 \\ 0 & \text{otherwise} \end{cases}$$

So conditional on height x, weight is uniform on

$$50x - 20 \text{ to } 50x + 20$$

Independence of Random Variables

Random variables X and Y are independent if

$$f_{XY}(x, y) = f_X(x)\, f_Y(y)$$

Equivalent condition:
For all functions g and h,

$$E\big[g(X)h(Y)\big] = E[g(X)]\, E[h(Y)]$$

Independence Check in Height Weight Example

Consider point $(1.9, 70)$.
From joint:

$$f_{XY}(1.9, 70) = 0$$

But

$$f_X(1.9) \neq 0$$
$$f_Y(70) \neq 0$$

Thus

$$f_{XY}(x, y) \neq f_X(x)f_Y(y)$$

Hence X and Y are not independent.

Summary

- Joint density generalizes one dimensional density to higher dimensions.
- Marginals obtained by integrating out other variables.

- Conditionals obtained via ratio of joint to marginal.
- Independence characterized by factorization of joint density.
- Continuous multi variable theory parallels discrete case with integrals replacing sums.

****

Lecture 11

Sum of Independent Random Variables

Let X and Y be independent continuous random variables.
Define

$$Z = X + Y$$

Convolution Formula

If X and Y are independent with densities f_X and f_Y, then the density of Z is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

This operation is called **convolution**.

Example: $X, Y \sim \text{Uniform}(0, 1)$ **Independent**

Then

$$f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly for $f_Y(y)$.
Since $0 \leq X, Y \leq 1$, we have

$$0 \leq Z \leq 2$$

Outside $[0, 2]$, $f_Z(z) = 0$.

Case 1: $0 \leq z \leq 1$

$$f_Z(z) = \int_0^1 f_Y(z - x)dx$$

$f_Y(z - x)$ is nonzero only when

$$0 \le z - x \le 1$$

For $0 \le z \le 1$, this implies

$$0 \le x \le z$$

Thus

$$f_Z(z) = \int_0^z 1dx = z$$

Case 2: $1 \le z \le 2$

Now $f_Y(z - x)$ is nonzero when

$$0 \le z - x \le 1$$

This implies

$$z - 1 \le x \le 1$$

Hence

$$f_Z(z) = \int_{z-1}^1 1dx = 2 - z$$

Final Density

$$f_Z(z) = \begin{cases} z & 0 \le z \le 1 \\ 2 - z & 1 \le z \le 2 \\ 0 & \text{otherwise} \end{cases}$$

This is a triangular density.

Maximum of Independent Random Variables

Let

$$Z = \max(X, Y)$$

CDF Method

$$F_Z(z) = P(Z \leq z)$$

Since $Z \leq z$ iff both $X \leq z$ and $Y \leq z$,

$$F_Z(z) = P(X \leq z, Y \leq z)$$

If X and Y are independent,

$$F_Z(z) = F_X(z)F_Y(z)$$

Density obtained by differentiation:

$$f_Z(z) = \frac{d}{dz}F_Z(z)$$

Example: $X, Y \sim \mathrm{Uniform}(0, 1)$

CDF of X:

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

Hence

$$F_Z(z) = \begin{cases} 0 & z \leq 0 \\ z^2 & 0 \leq z \leq 1 \\ 1 & z \geq 1 \end{cases}$$

Density:

$$f_Z(z) = \begin{cases} 2z & 0 \leq z \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Minimum of Independent Random Variables

Let

$$Z = \min(X, Y)$$

CDF Derivation

$$F_Z(z) = P(Z \leq z)$$

$Z \leq z$ iff at least one of X or Y is $\leq z$:

$$F_Z(z) = P(X \le z \cup Y \le z)$$

Using De Morgan's law:

$$F_Z(z) = 1 - P(X > z, Y > z)$$

Independence gives

$$F_Z(z) = 1 - P(X > z)P(Y > z)$$

Since

$$P(X > z) = 1 - F_X(z)$$

we obtain

$$F_Z(z) = 1 - (1 - F_X(z))(1 - F_Y(z))$$

Density obtained by differentiation.
For $X, Y \sim \text{Uniform}(0, 1)$,

$$F_Z(z) = \begin{cases} 0 & z \le 0 \\ 1 - (1 - z)^2 & 0 \le z \le 1 \\ 1 & z \ge 1 \end{cases}$$

Differentiating,

$$f_Z(z) = \begin{cases} 2(1 - z) & 0 \le z \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Covariance and Correlation

Covariance

For random variables X and Y,

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

Using linearity,

$$\text{Cov}(X, Y) = E[XY] - EX\,EY$$

Correlation Coefficient

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

Properties:

- $-1 \leq \rho_{X,Y} \leq 1$
- $\rho = 0$ implies uncorrelated
- Independence implies uncorrelated

Independence Implies Uncorrelated

If X and Y are independent,

$$E[XY] = EX\,EY$$

Thus

$$\mathrm{Cov}(X, Y) = 0$$

Uncorrelated Does Not Imply Independence

Let

$$X \sim \mathrm{Uniform}(-1, 1)$$

Define

$$Y = X^2$$

Then

$$E[XY] = E[X^3]$$

Since X^3 is odd over symmetric interval,

$$E[X^3] = 0$$

Also

$$EX = 0$$

Hence

$$E[XY] = EX\,EY$$

So X and Y are uncorrelated.
However, $Y = X^2$ implies dependence.

Random Vectors and Covariance Matrix

Let

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

The covariance matrix is the $n \times n$ matrix

$$\mathrm{Cov}(X) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{pmatrix}$$

Properties:

- Symmetric matrix
- Diagonal entries are variances
- Off diagonal entries are covariances

****

Lecture 12

Transformed Random Variables

Let W and X be continuous random variables with joint density $f_{WX}(w, x)$.
Define new random variables

$$Y = g(W, X)$$

$$Z = h(W, X)$$

We seek the joint density $f_{YZ}(y, z)$.

Invertible Transformations

The transformation is said to be invertible if there exist functions q and r such that

$$W = q(y, z)$$

$$X = r(y, z)$$

That is, from Y, Z we can uniquely recover W, X.
If such functions do not exist, the transformation is not invertible.

Change of Variables Formula

For an invertible transformation,

$$f_{YZ}(y, z) = f_{WX}(q(y, z), r(y, z)) \left| \det J(y, z) \right|$$

where the Jacobian matrix is

$$J(y, z) = \begin{pmatrix} \frac{\partial q}{\partial y} & \frac{\partial q}{\partial z} \\ \frac{\partial r}{\partial y} & \frac{\partial r}{\partial z} \end{pmatrix}$$

and

$$\det J = \frac{\partial q}{\partial y} \frac{\partial r}{\partial z} - \frac{\partial q}{\partial z} \frac{\partial r}{\partial y}$$

The absolute value appears because densities scale by area.

Example

Assume

$$W \sim \text{Uniform}(0, 1)$$

$$X \sim \text{Uniform}(0, 1)$$

with W and X independent.
Define

$$Y = W + X$$

$$Z = W - X$$

Inverse Transformation

Solve for W and X:

$$W = \frac{Y + Z}{2}$$

$$X = \frac{Y - Z}{2}$$

Thus

$$q(y, z) = \frac{y + z}{2}$$

$$r(y, z) = \frac{y - z}{2}$$

Jacobian

Compute partial derivatives:

$$\frac{\partial q}{\partial y} = \frac{1}{2}$$

$$\frac{\partial q}{\partial z} = \frac{1}{2}$$

$$\frac{\partial r}{\partial y} = \frac{1}{2}$$

$$\frac{\partial r}{\partial z} = -\frac{1}{2}$$

Hence

$$J = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

Determinant:

$$\det J = \frac{1}{2}\left(-\frac{1}{2}\right) - \frac{1}{2}\left(\frac{1}{2}\right) = -\frac{1}{2}$$

Thus

$$|\det J| = \frac{1}{2}$$

Joint Density of Y, Z

Since W and X are independent,

$$f_{WX}(w, x) = f_W(w) f_X(x)$$

Each equals 1 on $[0, 1]$ and 0 otherwise.

Thus

$$f_{YZ}(y, z) = f_W\left(\frac{y + z}{2}\right) f_X\left(\frac{y - z}{2}\right) \cdot \frac{1}{2}$$

Support

The density is nonzero when

$$0 \le \frac{y + z}{2} \le 1$$

and

$$0 \le \frac{y - z}{2} \le 1$$

These inequalities define the region

$$0 \le y \le 2$$
$$-y \le z \le y$$
$$y - 2 \le z \le 2 - y$$

The intersection forms a diamond-shaped region with vertices

$$(0, 0), (1, 1), (2, 0), (1, -1)$$

Final Density

$$f_{YZ}(y, z) = \begin{cases} \frac{1}{2} & \text{inside the diamond region} \\ 0 & \text{otherwise} \end{cases}$$

Observations

- Even though W and X are independent, Y and Z are not.
- The support is not rectangular.
- Change of variables provides an exact density transformation rule.
- The Jacobian determinant accounts for area distortion under transformation.

Lecture 13

Uniform Distribution

Definition

A continuous random variable X is said to be uniformly distributed on the interval a to b, written as

$$X \sim \text{Uniform}(a, b),$$

if its probability density function is

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

Expectation and Variance

The mean of X is

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

The variance of X is

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

Exponential Distribution

Definition

A continuous random variable X is said to be exponentially distributed with parameter $\lambda > 0$, written as

$$X \sim \text{Exp}(\lambda),$$

if its probability density function is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Validity of Density

Check normalization:

$$\int_0^\infty \lambda e^{-\lambda x}\, dx = \lambda\left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty = \lambda \cdot \frac{1}{\lambda} = 1.$$

Cumulative Distribution Function

For $x \geq 0$,

$$F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda u}\, du = 1 - e^{-\lambda x}.$$

For $x < 0$,

$$F_X(x) = 0.$$

Memoryless Property

For $a > b \geq 0$,

$$P(X \geq a \mid X \geq b) = P(X \geq a - b).$$

Proof:

$$P(X \geq a \mid X \geq b) = \frac{P(X \geq a)}{P(X \geq b)} = \frac{e^{-\lambda a}}{e^{-\lambda b}} = e^{-\lambda(a-b)} = P(X \geq a - b).$$

Expectation

Compute

$$E[X] = \int_0^\infty x \lambda e^{-\lambda x}\, dx.$$

Using integration by parts,

$$E[X] = \frac{1}{\lambda}.$$

Minimum of Independent Exponentials

Let

$$X \sim \mathrm{Exp}(\lambda), \quad Y \sim \mathrm{Exp}(\tau),$$

with X and Y independent.
Define

$$Z = \min(X, Y).$$

CDF of Z:

$$F_Z(z) = 1 - P(X > z, Y > z) = 1 - P(X > z)P(Y > z).$$

Thus,

$$F_Z(z) = 1 - e^{-\lambda z}e^{-\tau z} = 1 - e^{-(\lambda+\tau)z}.$$

Therefore,

$$Z \sim \operatorname{Exp}(\lambda + \tau).$$

Density:

$$f_Z(z) = \begin{cases} (\lambda + \tau)e^{-(\lambda+\tau)z}, & z \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Normal Distribution

Standard Normal

A random variable Z is said to follow a standard normal distribution if

$$Z \sim N(0, 1),$$

with density

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right).$$

Normalization

Let

$$A = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx.$$

Then

$$A^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy.$$

Switch to polar coordinates:

$$x = r\cos\theta, \quad y = r\sin\theta,$$

with Jacobian r.
Thus,

$$A^2 = \int_0^{2\pi} \int_0^\infty e^{-r^2/2} r \, dr \, d\theta.$$

Let $u = r^2/2$, so $du = r \, dr$:

$$A^2 = \int_0^{2\pi} d\theta \int_0^\infty e^{-u} \, du = 2\pi.$$

Hence,

$$A = \sqrt{2\pi}.$$

General Normal Distribution

Let

$$Z \sim N(0, 1),$$

and define

$$X = \sigma Z + \mu.$$

Then

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

Using change of variables,

$$f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x - \mu}{\sigma}\right).$$

Thus,

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

We write

$$X \sim N(\mu, \sigma^2).$$

Mean and Variance

Using linearity of expectation,

$$E[X] = E[\sigma Z + \mu] = \mu.$$

Variance:

$$\mathrm{Var}(X) = \mathrm{Var}(\sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2.$$

Summary

- $X \sim \mathrm{Uniform}(a, b)$:

$$f_X(x) = \frac{1}{b-a}, \quad E[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

- $X \sim \mathrm{Exp}(\lambda)$:

$$f_X(x) = \lambda e^{-\lambda x}, \quad E[X] = \frac{1}{\lambda}.$$

- $X \sim N(\mu, \sigma^2)$:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad E[X] = \mu, \quad \mathrm{Var}(X) = \sigma^2.$$

****