STAT 230 Probability

SEE TEXTBOOK FOR DETAILED EXAMPLES AND SOLUTIONS

Extracted from Course Notes

University of Waterloo

Winter 2026 Edition

Contents

Chapter 1

Introduction to Probability

1.1 Definitions of Probability

Key Concepts

[Motivation] We routinely encounter events with uncertain outcomes: coin flips, weather, stock prices, insurance claims. Although many phenomena may not be genuinely "random," observers often lack knowledge to predict outcomes with certainty. Understanding randomness is essential for decision-making under uncertainty.

Three Definitions of Probability:

1. **Classical Definition**: $P(\text{event}) = \dfrac{\text{number of ways event can occur}}{\text{number of outcomes in } S}$ (assuming equally likely outcomes)

2. **Relative Frequency Definition**: Probability as the limiting proportion of times the event occurs in a very long series of repetitions

3. **Subjective Probability Definition**: Probability as a measure of how sure the person making the statement is that the event will happen

Limitations of These Definitions

- **Classical**: "Equally likely" uses probability to define probability (circular!)
- **Relative Frequency**: We can never actually repeat indefinitely; impractical for many events
- **Subjective**: No rational basis for agreement; opinions vary

Solution: Treat probability as a mathematical system defined by axioms, assigning numerical values only for specific applications.

The mathematical approach requires:

- A **sample space** S of all possible outcomes
- A set of **events** (subsets of S) to which we can assign probabilities
- A **mechanism** for assigning probabilities (numbers between 0 and 1) to events

1.2 Mathematical Probability Models

Definition 1.2.1 (Sample Space)

A **sample space** S is a set of distinct outcomes for an experiment or process, with the property that in a single trial, one and only one of these outcomes occurs.

Remarks on Sample Spaces

- Sample spaces are not necessarily unique—different choices work for different purposes
- S is **discrete** if it consists of a finite or countably infinite set of points
- When possible, choose sample points that are "indivisible" (cannot be broken into smaller events)

Definition 1.2.2 (Event)

An **event** is a subset $A \subseteq S$. If the event is indivisible so it contains only one point (e.g., $A_1 = \{a_1\}$), we call it a **simple event**. An event A made up of two or more simple events is called a **compound event**.

Definition 1.2.3 (Probability Distribution on Discrete Sample Space)

Let $S = \{a_1, a_2, a_3, \ldots\}$ be a discrete sample space. Assign numbers (i.e., probabilities) $P(a_i)$, $i = 1, 2, 3, \ldots$, to the a_i's such that:

1. $0 \leq P(a_i) \leq 1$ for all i
2. $\sum_{\text{all } i} P(a_i) = 1$

The set of probabilities $\{P(a_i), i = 1, 2, 3, \ldots\}$ is called a **probability distribution** on S.

Intuition

The condition $\sum P(a_i) = 1$ reflects that when the experiment happens, *some* outcome in S must occur. The probability function $P(\cdot)$ provides a model encoding how frequently we expect to observe each outcome.

Definition 1.2.4 (Probability of an Event)

The probability $P(A)$ of an event A is the sum of the probabilities for all the simple events that make up A:

$$P(A) = \sum_{a \in A} P(a)$$

Definition 1.2.5 (Odds)

The **odds in favour** of an event A is:
$$\frac{P(A)}{1-P(A)}$$
The **odds against** the event is the reciprocal: $\dfrac{1-P(A)}{P(A)}$.

Converting Odds to Probability

If odds against event A are $m:1$, then $\dfrac{1-P(A)}{P(A)} = m$, which gives $P(A) = \dfrac{1}{m+1}$.

Example: If odds against a horse winning are 20:1, then $P(\text{win}) = \dfrac{1}{21}$.

Three-Step Method for Calculating Probability

[Problem-Solving Strategy]

1. **Specify** a sample space S
2. **Assign** a probability distribution to the simple events in S
3. **Calculate** $P(A)$ by adding probabilities of all simple events in A

Examples

Example 1.2.1 (Drawing a Card)

Draw one card from a standard well-shuffled deck of cards, comprised of 13 cards (2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A) in each of 4 distinct suits: diamonds ($\diamondsuit$), hearts ($\heartsuit$), spades ($\spadesuit$), and clubs ($\clubsuit$). Find the probability that the card drawn is a club.
Solution 1: Let $S = \{\diamondsuit, \heartsuit, \spadesuit, \clubsuit\}$. Then S has 4 points, with 1 of them being "club", so $P(\clubsuit) = \frac{1}{4}$. ■
Solution 2: Consider the sample space

$$S = \{2\diamondsuit, 3\diamondsuit, \ldots, A\diamondsuit, 2\heartsuit, 3\heartsuit, \ldots, A\heartsuit, 2\spadesuit, 3\spadesuit, \ldots, A\spadesuit, 2\clubsuit, 3\clubsuit, \ldots, A\clubsuit\}.$$

Then each of the 52 cards in S has probability $\frac{1}{52}$. If A denotes the event of interest, then $A = \{2\clubsuit, 3\clubsuit, \ldots, A\clubsuit\}$, and this event has 13 simple outcomes all with probability $\frac{1}{52}$. Therefore,

$$P(A) = \underbrace{\frac{1}{52} + \frac{1}{52} + \cdots + \frac{1}{52}}_{\text{13 terms}} = \frac{13}{52} = \frac{1}{4}. \quad ■$$

Remarks: A sample space is not necessarily unique. In Solution 1, $A =$ "the card is a club" is a simple event, but in Solution 2 it is a compound event. We assumed each simple event is equally probable—the only sensible choice here.

Example 1.2.2 (Coin Toss)

Toss a coin twice. Find the probability of getting exactly one head.
Solution 1 (Correct): Let $S = \{HH, HT, TH, TT\}$ and assume the simple events each have probability $\frac{1}{4}$. (Here, HT means head on the 1st toss and tail on the 2nd toss.) Since one head occurs for simple events HT and TH, the event of interest is $A = \{HT, TH\}$ and we get

$$P(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \quad \blacksquare$$

Solution 2 (Incorrect): Let $S = \{0 \text{ heads}, 1 \text{ head}, 2 \text{ heads}\}$ and assume each has probability $\frac{1}{3}$. Then $P(1 \text{ head}) = \frac{1}{3}$.
Why Solution 2 is wrong: We want a solution that reflects real-world relative frequency, not just mathematical consistency. The three outcomes in Solution 2 are NOT equally likely in actual experiments. The simple event $\{1 \text{ head}\}$ occurs more often than either $\{0 \text{ heads}\}$ or $\{2 \text{ heads}\}$ in repeated trials. If forced to use this sample space, we'd need: $P(0 \text{ heads}) = \frac{1}{4}$, $P(1 \text{ head}) = \frac{1}{2}$, $P(2 \text{ heads}) = \frac{1}{4}$.

Example 1.2.3 (Two Dice)

Roll a red die and a green die. Find $P(\text{total} = 5)$.
Solution: Let (x, y) represent getting x on the red die and y on the green die. The sample space S is:

$$\begin{array}{ccccc} (1,1) & (1,2) & (1,3) & \cdots & (1,6) \\ (2,1) & (2,2) & (2,3) & \cdots & (2,6) \\ \vdots & \vdots & \vdots & & \vdots \\ (6,1) & (6,2) & (6,3) & \cdots & (6,6) \end{array}$$

Each simple event is assigned probability $\frac{1}{36}$. For the event of interest:

$$A = \{(1,4), (2,3), (3,2), (4,1)\}$$

Therefore $P(A) = \frac{4}{36} = \frac{1}{9}$. $\blacksquare$
Follow-up: If the 2 dice were identical in colour, using distinguishable points (x, y) where $x \le y$ gives only 21 points. If we incorrectly assign equal probability $\frac{1}{21}$ to each, we get $P(A) = \frac{2}{21} \neq \frac{1}{9}$. The universe doesn't change frequencies based on dice color! The 21 points are not equally likely—$(1, 2)$ occurs twice as often as $(1, 1)$ in practice.

Important: Identical vs. Distinguishable Objects

When objects are identical (e.g., two dice of the same color), the points in a sample space treating them as identical are usually NOT equally likely.
Safe practice: Pretend objects can be distinguished even when they cannot. The laws of probability don't know whether you can tell the dice apart!

1.3 Counting in Uniform Probability Models

Key Concepts

[Uniform Distribution] When all n outcomes in S are equally likely, each has probability $\frac{1}{n}$. This is called a **uniform distribution**. If event A contains m outcomes, then $P(A) = \frac{m}{n}$.

Fundamental Counting Rules:

Addition Rule

If job 1 can be done in p ways and job 2 in q ways, then we can do **either** job 1 **OR** job 2 (but not both) in $p + q$ ways.

Multiplication Rule

If job 1 can be done in p ways and, for each of these, job 2 can be done in q ways, then we can do **both** job 1 **AND** job 2 in $p \times q$ ways.

OR $\leftrightarrow$ Addition, AND $\leftrightarrow$ Multiplication

This association between "OR"/"AND" and addition/multiplication occurs throughout probability. When solving problems, try to rephrase questions to identify implied ANDs and ORs.

Permutations and Combinations

Counting Formulas

Starting with n distinct objects:

1. **Permutations of all n objects:** $n! = n \times (n-1) \times \cdots \times 1$
2. **Permutations of k objects from n:**

$$n^{(k)} = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

3. **Arrangements with replacement:** n^k (each of k positions can be any of n objects)
4. **Combinations (subsets) of size k from n:**

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

Stirling's Approximation

For large n: $n! \approx n^n e^{-n}\sqrt{2\pi n}$
Error is less than 1% for $n \geq 8$.

When to Use Combinations vs. Permutations

- **Permutations**: Order matters (arrangements, sequences)
- **Combinations**: Order doesn't matter (subsets, groups, committees)

Examples

Example 1.3.1

Select two digits from $\{1, 2, 3, 4, 5\}$ with replacement. Find P(exactly one digit is even).
Solution: Note that the event of interest can be re-worded as: "The first digit is even AND the second is odd (this can be done in 2×3 ways) OR the first digit is odd AND the second is even (done in 3×2 ways)".
Since these are connected with "OR", we combine using the addition rule:

$$\text{Number of ways} = (2 \times 3) + (3 \times 2) = 12$$

Since the first digit can be chosen in 5 ways AND the second digit also in 5 ways, S contains $5 \times 5 = 25$ outcomes (via the multiplication rule), each with probability $\frac{1}{25}$.
Therefore:

$$P(\text{one digit is even}) = \frac{12}{25} \quad \blacksquare$$

Example 1.3.2

Letters a, b, c, d, e, f are arranged at random to form a six-letter word (using each letter once). Find P(2nd letter is e or f).
Solution: The sample space is $S = \{abcdef, abcdfe, \ldots, fedcba\}$, which has $6! = 720$ equally probable points.
Consider filling boxes $\square\square\square\square\square\square$ for the six positions. We can fill them in $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ ways.
For the event A = "2nd letter is e or f": Start with the constrained position (the 2nd box).

- Fill the 2nd box: 2 ways (e or f)
- Fill the 1st box: 5 ways (any remaining letter)
- Fill remaining 4 boxes: $4! = 24$ ways

Number of outcomes in $A = 2 \times 5 \times 24 = 240$.
Since we have a uniform probability model:

$$P(A) = \frac{240}{720} = \frac{1}{3} \quad \blacksquare$$

Key insight: Start counting from the constrained position! If we started with the 1st box (6 ways), the number of ways to fill the 2nd box would depend on whether we used e or f in the 1st box.

Example 1.3.3 (Passwords with Replacement)

A 4-digit password from $\{0, 1, \ldots, 9\}$ with replacement. Find probabilities for:
A = "all even digits", B = "all digits unique", C = "contains at least one 2"
Solution: The sample space is $S = \{0000, 0001, \ldots, 9999\}$, which has $10^4 = 10000$ equally probable points.
(a) For $A = \{0000, 0002, \ldots, 8888\}$: We can select each digit in 5 ways (even digits: 0, 2, 4, 6, 8). So there are 5^4 outcomes in A:

$$P(A) = \frac{5^4}{10^4} = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

(b) For $B = \{0123, 0124, \ldots, 9876\}$: Select 1st digit in 10 ways, 2nd in 9 ways (no repeat), 3rd in 8 ways, 4th in 7 ways. So there are $10 \times 9 \times 8 \times 7 = 10^{(4)}$ outcomes in B:

$$P(B) = \frac{10^{(4)}}{10^4} = \frac{5040}{10000} = \frac{63}{125}$$

(c) For C: Use the complement $\bar{C}$ = "no 2s". Each digit can be any of 9 values (not 2), so $|\bar{C}| = 9^4$. Thus $|C| = 10^4 - 9^4$:

$$P(C) = \frac{10^4 - 9^4}{10^4} = 1 - \left(\frac{9}{10}\right)^4 = \frac{3439}{10000} \quad \blacksquare$$

Example 1.3.4 (Passwords without Replacement)

A 4-digit password from $\{0, 1, \ldots, 9\}$ without replacement. Find probabilities for:
A = "all even digits", B = "begins or ends with 1", C = "contains a 2"
Solution: The sample space is $S = \{0123, 0132, \ldots, 9876\}$, which has $10^{(4)} = 10 \times 9 \times 8 \times 7 = 5040$ equally probable points.
(a) For $A = \{0246, 0248, \ldots, 8642\}$: Take $5^{(4)}$ arrangements of length 4 using only even digits $\{0, 2, 4, 6, 8\}$:

$$P(A) = \frac{5^{(4)}}{10^{(4)}} = \frac{5 \times 4 \times 3 \times 2}{10 \times 9 \times 8 \times 7} = \frac{120}{5040} = \frac{1}{42}$$

(b) For $B = \{1023, 0231, \ldots, 9871\}$: There are 2 positions for digit 1 (first or last). For each choice, fill remaining 3 positions in $9^{(3)}$ ways:

$$P(B) = \frac{2 \times 9^{(3)}}{10^{(4)}} = \frac{2 \times 504}{5040} = \frac{1008}{5040} = \frac{1}{5}$$

(c) For C: Use the complement. Removing 2, there are $9^{(4)}$ passwords without a 2:

$$P(C) = \frac{10^{(4)} - 9^{(4)}}{10^{(4)}} = 1 - \frac{9^{(4)}}{10^{(4)}} = 1 - \frac{9 \times 8 \times 7 \times 6}{10 \times 9 \times 8 \times 7} = 1 - \frac{3}{5} = \frac{2}{5} \quad \blacksquare$$

Example 1.3.5 (Committee Selection)

From 6 third-year and 7 fourth-year students, a committee of 5 is randomly formed. Among the third-year students is one named Roger. Find probabilities for:
A = "Roger is included", B = "all fourth-year", C = "at most 4 third-year"
Solution: Label third-year students $T_1, \ldots, T_6$ (with T_1 = Roger) and fourth-year students $F_1, \ldots, F_7$. The sample space consists of all committees (subsets) of size 5 from 13 students, so $|S| = \binom{13}{5} = 1287$.
(a) For A: Roger (T_1) must be in the subset, leaving 4 spots to fill from the remaining 12 students:

$$P(A) = \frac{\binom{12}{4}}{\binom{13}{5}} = \frac{495}{1287} = \frac{12! \cdot 5! \cdot 8!}{4! \cdot 8! \cdot 13!} = \frac{5}{13}$$

(b) For B: Choose 5 from the 7 fourth-year students:

$$P(B) = \frac{\binom{7}{5}}{\binom{13}{5}} = \frac{21}{1287} = \frac{7}{429}$$

(c) For C: The complement $\bar{C}$ = "all 5 are third-year". Choose 5 from 6 third-year students:

$$P(C) = 1 - \frac{\binom{6}{5}}{\binom{13}{5}} = 1 - \frac{6}{1287} = \frac{1281}{1287} = \frac{427}{429} \quad \blacksquare$$

Example 1.3.6 (Balls from a Box)

Box with 3 red, 4 white, 3 green balls. Sample 4 without replacement. Find probabilities for: A = "2 red balls", B = "2 red, 1 white, 1 green", C = "2 or more red"
Solution: Label balls 1–10 (1,2,3 = red; 4,5,6,7 = white; 8,9,10 = green). Sample space consists of all subsets of size 4, so $|S| = \binom{10}{4} = 210$.
(a) For A: Choose 2 red from 3 red balls, then 2 from 7 non-red:

$$P(A) = \frac{\binom{3}{2}\binom{7}{2}}{\binom{10}{4}} = \frac{3 \times 21}{210} = \frac{63}{210} = \frac{3}{10}$$

(b) For B: Choose 2 red from 3, then 1 white from 4, then 1 green from 3:

$$P(B) = \frac{\binom{3}{2}\binom{4}{1}\binom{3}{1}}{\binom{10}{4}} = \frac{3 \times 4 \times 3}{210} = \frac{36}{210} = \frac{6}{35}$$

(c) For C: Outcomes have either exactly 2 or exactly 3 red balls:

$$P(C) = \frac{\binom{3}{2}\binom{7}{2} + \binom{3}{3}\binom{7}{1}}{\binom{10}{4}} = \frac{63+7}{210} = \frac{70}{210} = \frac{1}{3} \quad \blacksquare$$

Common Error: Don't use $\binom{3}{2}\binom{8}{2}$ for part (c)—this overcounts! For example, picking red balls $\{1, 2\}$ then $\{3, 4\}$ gives subset $\{1, 2, 3, 4\}$, but so does picking $\{1, 3\}$ then $\{2, 4\}$. Always break "at least" into exact cases.

Common Counting Error

For events like "at least...", "more than...", break into mutually exclusive pieces with exact values, then add. Don't combine constraints in one step—this often leads to overcounting!

Section 1.3 Problems

1. Three students pick sections (4 available) at random. Find P(all same), P(all different), P(no one picks sectio
2. Canadian postal codes: 3 letters alternating with 3 digits. Find P(all letters same), P(digits all even or all odd
3. Binary sequence of length 10 chosen uniformly. Find P(exactly 5 zeros).
4. **Birthday Problem:** Among r people, find P(no shared birthday) for $r = 20, 40, 60$.

Chapter 2

Probability Rules and Conditional Probability

2.1 Use of Sets

Key Concepts

[Set Operations for Events]

- **Union** $A \cup B$: "A or B" (at least one occurs)
- **Intersection** $A \cap B = AB$: "A and B" (both occur)
- **Complement** $\bar{A}$: "not A" (all outcomes in S but not in A)
- **Empty event** $\emptyset = \bar{S}$: no outcomes, $P(\emptyset) = 0$

De Morgan's Laws

1. $\overline{A \cup B} = \bar{A} \cap \bar{B}$
2. $\overline{A \cap B} = \bar{A} \cup \bar{B}$

Interpreting De Morgan's Laws

"Not (A or B)" means "neither A nor B" = "not A and not B"
"Not (A and B)" means "at least one doesn't happen" = "not A or not B"

2.2 Addition Rules for Unions of Events

Rule 1 (Normalization)

$P(S) = 1$

Rule 2 (Boundedness)

For any event A: $0 \leq P(A) \leq 1$

Rule 3 (Monotonicity)

If $A \subseteq B$, then $P(A) \leq P(B)$

Rule 4a (Union of Two Events)

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

Intuition for Rule 4a

In $P(A) + P(B)$, outcomes in AB are counted twice. Subtracting $P(AB)$ corrects this double-counting.

Rule 4b (Union of Three Events)

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(AB) - P(AC) - P(BC) \\ & + P(ABC) \end{aligned}$$

Inclusion-Exclusion Principle

The pattern generalizes: add single events, subtract pairs, add triples, subtract quadruples, etc. For n events:

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

Definition 2.2.1 (Mutually Exclusive Events)

Events A and B are **mutually exclusive** (or **disjoint**) if $AB = \emptyset$.
More generally, $A_1, \ldots, A_n$ are mutually exclusive if $A_i A_j = \emptyset$ for all $i \neq j$.

Rule 5 (Union of Mutually Exclusive Events)

If $A_1, \ldots, A_n$ are mutually exclusive:

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

Countable Additivity

Rule 5 extends to countably infinite collections: if $\{A_i\}_{i=1}^{\infty}$ are mutually exclusive, then

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

Rule 6 (Complement Rule)

$P(A) = 1 - P(\bar{A})$

When to Use the Complement

Use when "at least one" or "not all" appears. It's often easier to count what you *don't* want. Example: $P(\text{at least one 6 in three dice}) = 1 - P(\text{no 6s}) = 1 - (5/6)^3$

Examples

Example 2.2.1

In a standard deck of 52 cards, two suits (diamonds, hearts) are red and two (spades, clubs) are black. The cards J, Q, K are "face" cards. If one card is randomly drawn, find $P(\text{red or face})$.
Solution: Let $A =$ event of drawing a red card, $B =$ event of drawing a face card.
Using the sample space of 52 cards:

$$P(A) = P(\{2\diamondsuit, 3\diamondsuit, \ldots, A\diamondsuit, 2\heartsuit, 3\heartsuit, \ldots, A\heartsuit\}) = \frac{26}{52} = \frac{1}{2}$$

$$P(B) = P(\{J\diamondsuit, Q\diamondsuit, K\diamondsuit, J\heartsuit, Q\heartsuit, K\heartsuit, J\spadesuit, Q\spadesuit, K\spadesuit, J\clubsuit, Q\clubsuit, K\clubsuit\}) = \frac{12}{52} = \frac{3}{13}$$

$$P(AB) = P(\{J\diamondsuit, Q\diamondsuit, K\diamondsuit, J\heartsuit, Q\heartsuit, K\heartsuit\}) = \frac{6}{52} = \frac{3}{26}$$

Using Rule 4a:

$$P(\text{red or face}) = P(A \cup B) = P(A) + P(B) - P(AB) = \frac{1}{2} + \frac{3}{13} - \frac{3}{26} = \frac{8}{13} \quad \blacksquare$$

Example 2.2.2 (Language Classes)

An elementary school offers Russian, French, and German classes to 100 students. Given: 26 in Russian, 29 in French, 17 in German; 12 in R∩F, 6 in R∩G, 4 in F∩G, 2 in all three. Find P(no language class).
Solution: Let R, F, G denote enrollment in Russian, French, German respectively. We have:

$$\begin{aligned} P(R) &= 0.26, & P(F) &= 0.29, & P(G) &= 0.17 \\ P(RF) &= 0.12, & P(RG) &= 0.06, & P(FG) &= 0.04 \\ & & P(RFG) &= 0.02 & & \end{aligned}$$

Using Rule 4b (inclusion-exclusion):

$$\begin{aligned} P(R \cup F \cup G) &= P(R) + P(F) + P(G) - P(RF) - P(RG) - P(FG) + P(RFG) \\ &= 0.26 + 0.29 + 0.17 - 0.12 - 0.06 - 0.04 + 0.02 \\ &= 0.52 \end{aligned}$$

By De Morgan's Law: $\overline{R \cup F \cup G} = \bar{R} \cap \bar{F} \cap \bar{G}$
Therefore:

$$P(\text{no language class}) = P(\bar{R} \cap \bar{F} \cap \bar{G}) = 1 - P(R \cup F \cup G) = 1 - 0.52 = 0.48 \quad \blacksquare$$

Example 2.2.3 (Three Dice)

Three fair six-sided dice are rolled. Calculate P(at least one 6).
Solution 1 (Direct): Let A_i = "6 occurs on die i" for $i = 1, 2, 3$. Sample space $S = \{(1,1,1), (1,1,2), \ldots, (6,6,6)\}$ has $6^3 = 216$ points.
Note that:

$$\begin{aligned} A_1A_2 &= \{(6,6,1), (6,6,2), \ldots, (6,6,6)\} \quad \text{(6 outcomes)} \\ A_1A_3 &= \{(6,1,6), (6,2,6), \ldots, (6,6,6)\} \quad \text{(6 outcomes)} \\ A_2A_3 &= \{(1,6,6), (2,6,6), \ldots, (6,6,6)\} \quad \text{(6 outcomes)} \\ A_1A_2A_3 &= \{(6,6,6)\} \quad \text{(1 outcome)} \end{aligned}$$

By Rule 4b:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{6}{216} - \frac{6}{216} - \frac{6}{216} + \frac{1}{216} = \frac{91}{216} \end{aligned}$$

Solution 2 (Complement): The complement is "no 6 on any die":

$$\overline{A_1 \cup A_2 \cup A_3} = \{(1,1,1), (1,1,2), \ldots, (5,5,5)\}$$

This has $5^3 = 125$ outcomes. By Rule 6:

$$P(A_1 \cup A_2 \cup A_3) = 1 - \frac{125}{216} = \frac{91}{216} \quad \blacksquare$$

2.3 Dependent and Independent Events

Definition 2.3.1 (Independent Events)

Events A and B are **independent** if and only if:

$$P(AB) = P(A)P(B)$$

If not independent, they are **dependent**.

Independence vs. Mutual Exclusivity

Common misconception: Mutually exclusive events are NOT independent (unless one has probability 0).
If A and B are mutually exclusive with $P(A) > 0$ and $P(B) > 0$:

$$P(AB) = 0 \neq P(A)P(B) > 0$$

Intuition: If A happens, B definitely doesn't happen—they're highly dependent!

Definition 2.3.2 (Mutual Independence)

Events $A_1, A_2, \ldots, A_n$ are **mutually independent** if for all subsets $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$:

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

Checking Independence for Three Events

For A, B, C to be mutually independent, ALL of these must hold:

$$\begin{aligned} P(AB) &= P(A)P(B) \\ P(AC) &= P(A)P(C) \\ P(BC) &= P(B)P(C) \\ P(ABC) &= P(A)P(B)P(C) \end{aligned}$$

Pairwise independence does NOT imply mutual independence!

Example 2.3.1

Fair coin tossed twice. Define: A = head on 1st toss, B = heads on both tosses, C = head on 2nd toss. Determine which pairs are independent.

Solution: Sample space $S = \{HH, HT, TH, TT\}$ with each outcome having probability $\frac{1}{4}$.
Calculate probabilities:

$$P(A) = P(\{HH, HT\}) = \frac{1}{2} \qquad P(C) = P(\{HH, TH\}) = \frac{1}{2}$$
$$P(B) = P(\{HH\}) = \frac{1}{4}$$

And intersections:

$$P(AB) = P(\{HH\}) = \frac{1}{4} \qquad P(AC) = P(\{HH\}) = \frac{1}{4}$$

A and B: Check if $P(AB) = P(A)P(B)$:

$$P(A)P(B) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8} \neq \frac{1}{4} = P(AB)$$

$\Rightarrow$ **A and B are dependent.**
A and C: Check if $P(AC) = P(A)P(C)$:

$$P(A)P(C) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = P(AC)$$

$\Rightarrow$ **A and C are independent.** ■

Example 2.3.3

If A and B are independent events, show that $\bar{A}$ and B are also independent.
Proof: Since $B = AB \cup \bar{A}B$ and these two events are disjoint (mutually exclusive):

$$P(B) = P(AB) + P(\bar{A}B)$$

Solving for $P(\bar{A}B)$:

$$\begin{aligned} P(\bar{A}B) &= P(B) - P(AB) \\ &= P(B) - P(A)P(B) \quad \text{(by independence of } A \text{ and } B\text{)} \\ &= P(B)(1 - P(A)) \\ &= P(B)P(\bar{A}) \end{aligned}$$

Since $P(\bar{A}B) = P(\bar{A})P(B)$, we conclude that $\bar{A}$ and B are independent. ■
Corollary: By similar arguments, if A and B are independent, then so are:

- A and $\bar{B}$
- $\bar{A}$ and $\bar{B}$

Example 2.3.4 (Independent Trials)

A pseudo random number generator produces independent digits from $\{0, 1, \ldots, 9\}$ uniformly.
(a) Find P(first 5 digits are all odd).
(b) Find P(9 first occurs on the 10th trial).
Solution:
(a) Let A_i = "digit i is odd" for $i = 1, 2, \ldots, 5$. There are 5 odd digits $\{1, 3, 5, 7, 9\}$, so $P(A_i) = \frac{5}{10} = \frac{1}{2}$ for each i.
Since selections are independent:

$$P(A_1A_2A_3A_4A_5) = P(A_1)P(A_2)P(A_3)P(A_4)P(A_5) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

(b) Let B_i = "digit i is 9". We have $P(B_i) = \frac{1}{10} = 0.1$ and $P(\bar{B}_i) = \frac{9}{10} = 0.9$.
For 9 to first occur on trial 10, the first 9 digits must not be 9, and the 10th must be 9:

$$P(\bar{B}_1\bar{B}_2 \cdots \bar{B}_9 B_{10}) = (0.9)^9 \times 0.1 \approx 0.0387 \quad \blacksquare$$

2.4 Conditional Probability and Product Rules

Definition 2.4.1 (Conditional Probability)

The **conditional probability** of A given B is:

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

Intuition for Conditional Probability

Given that B occurred, we "zoom in" on B as our new sample space. $P(A|B)$ is the proportion of B that also belongs to A.

Independence and Conditional Probability

If A and B are independent:

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

So A and B are independent iff knowing B occurred doesn't change the probability of A.

Rule 7a (Product Rule for Two Events)

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

Rule 7b (Product Rule for Three Events)

$$P(ABC) = P(A)P(B|A)P(C|AB)$$

Memorizing Product Rules

Imagine events unfolding chronologically: $P(ABC) = P(A) \times P(B \text{ given } A) \times P(C \text{ given } A \text{ and } B)$

Rule 8 (Law of Total Probability)

Let $\{A_1, \ldots, A_n\}$ partition S (mutually exclusive, union $= S$), with all $P(A_i) > 0$. Then for any event B:

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

Intuition for Law of Total Probability

Break B into pieces based on which A_i occurred. Calculate probability of B within each piece, then combine.
Tree diagrams are useful: branches represent A_i's, then B or $\bar{B}$. Multiply along paths, add across paths ending at B.

Rule 9 (Bayes' Rule)

Let $\{A_1, \ldots, A_n\}$ partition S with all $P(A_i) > 0$. For any B with $P(B) > 0$:

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Interpreting Bayes' Rule

- $P(A_j) =$ **prior** belief about A_j
- $P(B|A_j) =$ **likelihood** of observing B if A_j is true
- $P(A_j|B) =$ **posterior** belief after observing B

Bayes' Rule "reverses" conditioning: we know $P(B|A_i)$ and want $P(A_j|B)$.

Examples

Example 2.4.1

Fair coin tossed three times. Find $P(\text{exactly one head} \mid \text{at least one head})$.
Solution: Sample space: $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.
Let $A =$ "exactly one head" and $B =$ "at least one head".

$$A = \{HTT, THT, TTH\} \Rightarrow P(A) = \frac{3}{8}$$
$$B = S \setminus \{TTT\} \Rightarrow P(B) = 1 - P(\{TTT\}) = 1 - \frac{1}{8} = \frac{7}{8}$$

Since $A \subseteq B$ (if exactly one head occurs, then at least one head occurs):

$$P(AB) = P(A) = \frac{3}{8}$$

By the definition of conditional probability:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{3/8}{7/8} = \frac{3}{7} \quad \blacksquare$$

Example 2.4.4 (Insurance Classes)

In an insurance portfolio: 10% are class 1 (high risk, claim prob. 0.15), 40% are class 2 (medium risk, claim prob. 0.05), 50% are class 3 (low risk, claim prob. 0.02). Find $P(\text{claim in a given year})$.
Solution: Define events: $A_i =$ "policy is class i" for $i = 1, 2, 3$, and $B =$ "policy has a claim".
Given information:

$$\begin{aligned} P(A_1) &= 0.1, & P(B|A_1) &= 0.15 \\ P(A_2) &= 0.4, & P(B|A_2) &= 0.05 \\ P(A_3) &= 0.5, & P(B|A_3) &= 0.02 \end{aligned}$$

Since A_1, A_2, A_3 form a partition of the sample space, apply the Law of Total Probability:

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \\ &= (0.1)(0.15) + (0.4)(0.05) + (0.5)(0.02) \\ &= 0.015 + 0.02 + 0.01 \\ &= 0.045 \quad \blacksquare \end{aligned}$$

Example 2.4.5 (HIV Testing)

An HIV blood test has: 2% false negative rate, 0.5% false positive rate. Assume 0.04% of Canadian males have HIV. Find $P(\text{has HIV} \mid \text{positive test})$.

Solution: Let A = "selected male has HIV" and B = "blood test is positive".
Given:

$$\begin{aligned} P(B|A) &= 0.98 \quad \text{(true positive rate} = 1 - 0.02) \\ P(B|\bar{A}) &= 0.005 \quad \text{(false positive rate)} \\ P(A) &= 0.0004, \quad P(\bar{A}) = 0.9996 \end{aligned}$$

First, find $P(B)$ using the Law of Total Probability:

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(\bar{A})P(B|\bar{A}) \\ &= (0.0004)(0.98) + (0.9996)(0.005) \\ &= 0.000392 + 0.004998 = 0.00539 \end{aligned}$$

Apply Bayes' Rule:

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(B)} = \frac{(0.0004)(0.98)}{0.00539} \\ &= \frac{0.000392}{0.00539} = \frac{196}{2695} \approx 0.0727 \end{aligned}$$

Interpretation: If a randomly selected male tests positive, there is only about a 7.3% chance he actually has HIV! This counterintuitive result occurs because the disease is rare, so even a low false positive rate produces many false positives relative to true positives. ■

Chapter 3

Univariate Discrete Probability Distributions

3.1 Discrete Random Variables

[Motivation] Random variables provide a numerical description of experimental outcomes, allowing us to use algebra and calculus to manipulate probability models.

Definition 3.1.1 (Random Variable)

A **random variable** X is a function $X : S \to \mathbb{R}$ that assigns a real number to each point in the sample space S.

Definition 3.1.2 (Probability Mass Function)

The **probability mass function (pmf)** of a discrete random variable X is:

$$f(x) = P(X = x) \quad \text{for } x \in A$$

where A is the range of X.
Properties: $f(x) \geq 0$ for all x, and $\sum_{x \in A} f(x) = 1$.

Definition 3.1.3 (Cumulative Distribution Function)

The **cumulative distribution function (cdf)** of X is:

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u) \quad \text{for } x \in \mathbb{R}$$

Properties:

1. $0 \leq F(x) \leq 1$
2. F is non-decreasing
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
4. For discrete X: F is a step function with jumps at values in the range

Relationship Between pmf and cdf

- $F(x) = \sum_{u \leq x} f(u)$ (cdf from pmf)
- $f(x) = F(x) - F(x^-)$ where $F(x^-) = \lim_{u \to x^-} F(u)$ (pmf from cdf)
- $P(a < X \leq b) = F(b) - F(a)$

3.2 Functions of Random Variables

[Key Idea] If X is a random variable and g is a function, then $Y = g(X)$ is also a random variable.

To find the pmf of Y:

$$f_Y(y) = P(Y = y) = P(g(X) = y) = \sum_{\{x : g(x) = y\}} f_X(x)$$

3.3 Expectation of a Random Variable

Definition 3.3.1 (Expected Value / Mean)

The **expected value** (or **mean**) of a discrete random variable X with range A and pmf $f(x)$ is:

$$\mu = E(X) = \sum_{x \in A} x \cdot f(x)$$

provided the sum converges absolutely.

Intuition for Expected Value

$E(X)$ is the "center of mass" of the probability distribution—the long-run average value of X in repeated experiments.
If X represents winnings in a game, $E(X)$ is your average winnings per play over many plays.

Theorem 3.3.1 (Expectation of a Function)

For any function g:

$$E(g(X)) = \sum_{x \in A} g(x) \cdot f(x)$$

Note: You don't need to find the pmf of $Y = g(X)$ first!

Theorem 3.3.2 (Linearity of Expectation)

For constants $c_1, \ldots, c_n$ and functions $g_1, \ldots, g_n$:

$$E\left(\sum_{i=1}^{n} c_i g_i(X)\right) = \sum_{i=1}^{n} c_i E(g_i(X))$$

Corollary 3.3.1

For constants a and b:

$$E(aX + b) = aE(X) + b$$

Definition 3.3.2 (Variance)

The **variance** of X is:

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

Intuition for Variance

Variance measures the average squared distance from the mean—how "spread out" the distribution is.
$\mathrm{Var}(X) = 0$ iff X is constant (always equals μ).

Definition 3.3.3 (Standard Deviation)

The **standard deviation** is:

$$\sigma = \mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

Theorem 3.3.3 (Variance Formulas)

$$\text{Var}(X) = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$$
$$\text{Var}(X) = E(X(X-1)) + \mu - \mu^2$$

The second formula is useful when the pmf has $x!$ in the denominator.

Theorem 3.3.4 (Variance of Linear Function)

For constants a and b:

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$
$$\text{SD}(aX + b) = |a|\text{SD}(X)$$

Note: Adding a constant doesn't change variance; multiplying scales variance by the square.

3.4 Moment Generating Functions

Definition 3.4.1 (Moment Generating Function)

The **moment generating function (mgf)** of X is:

$$M(t) = E(e^{tX}) = \sum_{x \in A} e^{tx} f(x)$$

provided this is finite for t in some interval $(-a, a)$ around 0.

Why "Moment Generating"?

The mgf "generates" moments through differentiation. Note $M(0) = E(e^0) = 1$ always.

Definition 3.4.2 (Moments)

The n**th moment** of X is $E(X^n)$.

- 1st moment = mean (location)
- 2nd moment = used for variance (spread)
- Higher moments describe shape

Theorem 3.4.1 (Uniqueness)

If $M_X(t) = M_Y(t)$ for all t in some interval around 0, then X and Y have the same distribution.

Theorem 3.4.2 (Moments from MGF)

$$E(X^n) = M^{(n)}(0) = \frac{d^n}{dt^n}M(t)\Big|_{t=0}$$

Theorem 3.4.3 (MGF of Linear Transformation)

If $Y = aX + b$, then:

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

3.5 Special Discrete Probability Distributions

3.5.1 Discrete Uniform Distribution

Discrete Uniform Distribution $X \sim$ DU(a, b)

X takes values $\{a, a+1, \ldots, b\}$ with equal probability.
PMF: $f(x) = \frac{1}{b-a+1}$ for $x = a, a+1, \ldots, b$
Mean: $E(X) = \frac{a+b}{2}$
Variance: $\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$
Example: Fair die roll: $X \sim \text{DU}(1, 6)$

3.5.2 Binomial Distribution

Binomial Distribution $X \sim$ Bin(n, p)

X = number of successes in n independent Bernoulli trials, each with success probability p.
PMF: $f(x) = \binom{n}{x}p^x(1-p)^{n-x}$ for $x = 0, 1, \ldots, n$
Mean: $E(X) = np$
Variance: $\text{Var}(X) = np(1-p)$
MGF: $M(t) = (pe^t + 1 - p)^n$

Checking Binomial Assumptions

1. Fixed number n of trials
2. Each trial has only two outcomes (success/failure)
3. Probability p is constant across trials
4. Trials are independent

3.5.3 Hypergeometric Distribution

Hypergeometric Distribution $X \sim \mathbf{HG}(N, r, n)$

From N objects (r successes, $N - r$ failures), sample n without replacement. X = number of successes.

PMF: $f(x) = \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$

Range: $x = \max(0, n - N + r), \ldots, \min(r, n)$

Mean: $E(X) = \dfrac{nr}{N}$

Variance: $\mathrm{Var}(X) = \dfrac{nr(N - r)(N - n)}{N^2(N - 1)}$

Binomial Approximation to Hypergeometric

When N is large and n is small relative to N, the hypergeometric is well-approximated by $\mathrm{Bin}(n, r/N)$.
Intuition: With replacement or without makes little difference when sampling a small fraction of a large population.

3.5.4 Geometric Distribution

Geometric Distribution $X \sim \mathbf{Geo}(p)$

X = number of failures before the first success in independent Bernoulli trials.

PMF: $f(x) = (1 - p)^x p$ for $x = 0, 1, 2, \ldots$

CDF: $F(x) = 1 - (1 - p)^{x+1}$ for $x \geq 0$

Mean: $E(X) = \dfrac{1 - p}{p}$

Variance: $\mathrm{Var}(X) = \dfrac{1 - p}{p^2}$

MGF: $M(t) = \dfrac{p}{1 - (1 - p)e^t}$ for $t < \ln(1 - p)^{-1}$

Memoryless Property (Informal)

The geometric distribution has no "memory": $P(X > s + t | X > s) = P(X > t)$.
If you've already had s failures, the distribution of additional failures until success is the same as starting fresh.

3.5.5 Negative Binomial Distribution

Negative Binomial Distribution $X \sim \mathbf{NB}(k, p)$

X = number of failures before the kth success.
PMF: $f(x) = \binom{x+k-1}{x} p^k (1-p)^x$ for $x = 0, 1, 2, \ldots$
Mean: $E(X) = \dfrac{k(1-p)}{p}$
Variance: $\mathrm{Var}(X) = \dfrac{k(1-p)}{p^2}$
Special case: $\mathrm{NB}(1, p) = \mathrm{Geo}(p)$

Binomial vs. Negative Binomial

- **Binomial**: Fixed n trials, random number of successes
- **Negative Binomial**: Fixed k successes, random number of trials

3.5.6 Poisson Distribution

Poisson Distribution $X \sim \mathbf{Poi}(\mu)$

Models counts of rare events; limit of $\mathrm{Bin}(n, p)$ as $n \to \infty$, $p \to 0$, $np = \mu$ fixed.
PMF: $f(x) = \dfrac{\mu^x e^{-\mu}}{x!}$ for $x = 0, 1, 2, \ldots$
Mean: $E(X) = \mu$
Variance: $\mathrm{Var}(X) = \mu$ (mean = variance!)
MGF: $M(t) = e^{\mu(e^t - 1)}$

Poisson Process Conditions

The Poisson distribution arises when counting events over time/space satisfying:

1. **Independence**: Non-overlapping intervals are independent
2. **Individuality**: At most one event in a very short interval
3. **Homogeneity**: Events occur at a uniform rate λ

If these hold, events in time t follow $\mathrm{Poi}(\lambda t)$.

Chapter 4

Multivariate Discrete Probability Distributions

4.1 Basic Terminology and Techniques

Joint PMF

For discrete random variables X and Y:

$$f(x,y) = P(X = x, Y = y)$$

Properties: $0 \le f(x,y) \le 1$ and $\displaystyle\sum_{\text{all }(x,y)} f(x,y) = 1$

Marginal PMFs

$$f_X(x) = \sum_{\text{all } y} f(x,y) \quad \text{and} \quad f_Y(y) = \sum_{\text{all } x} f(x,y)$$

Intuition

Marginal distributions "ignore" one variable by summing over all its values. In a table, sum rows for f_X, sum columns for f_Y.

Definition 4.1.1 (Independent Random Variables)

X and Y are **independent** iff:

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

Definition 4.1.2 (Mutual Independence)

$X_1, \ldots, X_n$ are **independent** iff:

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n) \quad \text{for all } x_1, \ldots, x_n$$

4.2 Multinomial Distribution

Multinomial Distribution

n independent trials, each resulting in one of k outcomes with probabilities $p_1, \dots, p_k$ ($\sum p_i = 1$). If X_i = count of outcome i:

$$f(x_1, \dots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

for $x_i \geq 0$ with $\sum x_i = n$.
Marginals: $X_i \sim \text{Bin}(n, p_i)$
Special case: $k = 2$ gives the Binomial distribution.

4.3 Expectation, Covariance, and Correlation

Theorem 4.3.1

If X and Y are independent:

$$E(g_1(X) \cdot g_2(Y)) = E(g_1(X)) \cdot E(g_2(Y))$$

Special case: $E(XY) = E(X)E(Y)$ if independent.

Definition 4.3.1 (Covariance)

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

Interpreting Covariance

- $\text{Cov}(X, Y) > 0$: Large X tends to occur with large Y
- $\text{Cov}(X, Y) < 0$: Large X tends to occur with small Y
- $\text{Cov}(X, Y) = 0$: No linear relationship (but could still be dependent!)

Theorem 4.3.2

If X and Y are independent, then $\text{Cov}(X, Y) = 0$.
Warning: The converse is FALSE! Cov = 0 does not imply independence.

Definition 4.3.2 (Correlation Coefficient)

$$\rho = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Properties:

- $-1 \leq \rho \leq 1$
- $\rho = \pm 1$ iff $Y = aX + b$ for some constants (perfect linear relationship)
- $\rho = 0$ means X and Y are **uncorrelated**

Covariance vs. Correlation

Covariance depends on the scale of X and Y. Correlation is unitless and always between -1 and 1, making it easier to interpret.

4.4 Linear Combinations of Random Variables

[Key Results] For $X_1, \ldots, X_n$ and constants $a_1, \ldots, a_n$:

Mean:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

Variance:

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

If independent (or uncorrelated):

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i)$$

Important Special Case

For independent $X_1, \ldots, X_n$ with common mean μ and variance σ^2:
Sample mean $\bar{X} = \frac{1}{n}\sum X_i$ has:

- $E(\bar{X}) = \mu$
- $\mathrm{Var}(\bar{X}) = \sigma^2/n$ (decreases with n!)

4.5 Markov's Inequality, Chebyshev's Inequality, and the Law of Large Numbers

Theorem 4.5.1 (Markov's Inequality)

For any $\varepsilon > 0$:

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

Theorem 4.5.2 (Chebyshev's Inequality)

For any $\varepsilon > 0$:

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

Equivalently, for any $k > 0$:

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

Using Chebyshev

Chebyshev gives worst-case bounds without knowing the distribution shape. For specific distributions, actual probabilities are often much smaller.
Example: $P(|X - \mu| \geq 2\sigma) \leq 1/4 = 0.25$, but for Normal, it's about 0.046.

Theorem 4.5.3 (Weak Law of Large Numbers)

Let $X_1, X_2, \ldots$ be independent with common mean μ and variance σ^2. Then for any $\varepsilon > 0$:

$$\lim_{n \to \infty} P(|\bar{X} - \mu| > \varepsilon) = 0$$

The sample mean **converges in probability** to the true mean.

Significance of LLN

The Law of Large Numbers justifies:

1. The relative frequency interpretation of probability
2. Using sample means to estimate population means
3. Monte Carlo simulation methods

4.6 Conditional Probability Distributions

Definition 4.6.1 (Conditional PMF)

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

provided $f_Y(y) > 0$.

Definition 4.6.2 (Conditional Mean)

$$E(X|Y = y) = \sum_{\text{all } x} x \cdot f_{X|Y}(x|y)$$

Properties of Conditional Expectation

- Linearity holds: $E(aX + bZ|Y = y) = aE(X|Y = y) + bE(Z|Y = y)$
- If X, Y independent: $E(X|Y = y) = E(X)$

[Law of Total Expectation]

$$E(X) = \sum_{\text{all } y} E(X|Y = y) \cdot f_Y(y) = E[E(X|Y)]$$

Intuition: The overall mean is the weighted average of conditional means.

Chapter 5

Univariate Continuous Probability Distributions

5.1 Continuous Random Variables

Definition 5.1.1 (Probability Density Function)

The **pdf** of a continuous random variable X is a function $f(x)$ such that:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Properties:

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

PDF is NOT Probability!

For continuous X:

- $P(X = x) = 0$ for any single value x
- $f(x)$ can be greater than 1 (it's a density, not a probability)
- Only integrals of f give probabilities

CDF and PDF Relationship

- $F(x) = \int_{-\infty}^{x} f(t)\,dt$ (CDF from PDF)
- $f(x) = F'(x)$ where F is differentiable (PDF from CDF)

Definition 5.1.2 (Quantile)

The p**th quantile** $q(p)$ satisfies $F(q(p)) = p$.
Median $= q(0.5)$ (50th percentile)

5.2 Functions of Random Variables

Theorem 5.2.1 (Transformation Method)

If X has pdf $f_X(x)$ and g is strictly monotonic and differentiable, then $Y = g(X)$ has pdf:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Corollary 5.2.1 (Linear Transformation)

If $Y = aX + b$ where $a \neq 0$:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

5.3 Expectation

For continuous random variables:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

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

5.4 Special Continuous Distributions

5.4.1 Continuous Uniform Distribution

Uniform Distribution $X \sim U(a, b)$

PDF: $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$
CDF: $F(x) = \frac{x-a}{b-a}$ for $a \leq x \leq b$
Mean: $E(X) = \frac{a+b}{2}$
Variance: $\text{Var}(X) = \frac{(b-a)^2}{12}$

5.4.2 Exponential Distribution

Exponential Distribution $X \sim \textbf{Exp}(\theta)$

PDF: $f(x) = \frac{1}{\theta} e^{-x/\theta}$ for $x > 0$
CDF: $F(x) = 1 - e^{-x/\theta}$ for $x > 0$
Mean: $E(X) = \theta$
Variance: $\text{Var}(X) = \theta^2$

Memoryless Property

$$P(X > s + t | X > s) = P(X > t)$$

The exponential is the ONLY continuous distribution with this property. Useful for modeling "waiting times" where age doesn't matter.

5.4.3 Normal (Gaussian) Distribution

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

PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for $x \in \mathbb{R}$
Mean: $E(X) = \mu$
Variance: $\text{Var}(X) = \sigma^2$
MGF: $M(t) = e^{\mu t + \sigma^2 t^2/2}$

Theorem 5.4.2 (Standardization)

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

68-95-99.7 Rule

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

- $P(|X - \mu| < \sigma) \approx 0.68$
- $P(|X - \mu| < 2\sigma) \approx 0.95$
- $P(|X - \mu| < 3\sigma) \approx 0.997$

Theorem 5.4.1 (Probability Integral Transform)

If X has continuous cdf F_X that is strictly increasing, then $Y = F_X(X) \sim U(0, 1)$.
Inverse: If $U \sim U(0, 1)$, then $X = F^{-1}(U)$ has cdf F.

5.5 The Central Limit Theorem

Theorem 5.5.1 (Central Limit Theorem)

Let $X_1, X_2, \ldots, X_n$ be independent with common mean μ and variance σ^2. As $n \to \infty$:

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

Equivalently:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

Practical Application

For large n (typically $n \geq 30$):

- $S_n = \sum X_i \approx N(n\mu, n\sigma^2)$
- $\bar{X} \approx N(\mu, \sigma^2/n)$

The CLT works regardless of the original distribution of X_i!

Normal Approximation to Binomial

If $X \sim \text{Bin}(n,p)$ with $np \geq 5$ and $n(1-p) \geq 5$:

$$X \approx N(np, np(1-p))$$

Continuity correction: For integer k, use $P(X \leq k) \approx P(Z \leq \frac{k + 0.5 - np}{\sqrt{np(1-p)}})$

CLT Example: Quality Control

100 items, each defective with probability 0.1 independently. Find $P(X \geq 15)$.
$X \sim \text{Bin}(100, 0.1)$, so $\mu = 10$, $\sigma = 3$.
$P(X \geq 15) \approx P\left(Z \geq \frac{15-10}{3}\right) = P(Z \geq 1.67) \approx 0.048$

Appendix A

Summary of Distributions

Distribution	PMF/PDF	Mean	Variance	MGF
$\mathrm{DU}(a,b)$	$\dfrac{1}{b-a+1}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)(b-a+2)}{12}$	–
$\mathrm{Bin}(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$(pe^t+1-p)^n$
$\mathrm{Geo}(p)$	$(1-p)^x p$	$\dfrac{1-p}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{p}{1-(1-p)e^t}$
$\mathrm{NB}(k,p)$	$\binom{x+k-1}{x}p^k(1-p)^x$	$\dfrac{k(1-p)}{p}$	$\dfrac{k(1-p)}{p^2}$	$\left(\dfrac{p}{1-(1-p)e^t}\right)^k$
$\mathrm{Poi}(\mu)$	$\dfrac{\mu^x e^{-\mu}}{x!}$	μ	μ	$e^{\mu(e^t-1)}$
$U(a,b)$	$\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{tb}-e^{ta}}{t(b-a)}$
$\mathrm{Exp}(\theta)$	$\dfrac{1}{\theta}e^{-x/\theta}$	θ	θ^2	$(1-\theta t)^{-1}$
$N(\mu,\sigma^2)$	$\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu t+\sigma^2 t^2/2}$