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

This section introduces three definitions of probability:

- **Classical Definition**: Probability $= \frac{\text{number of ways event can occur}}{\text{number of outcomes in } S}$ (assuming equally likely outcomes)
- **Relative Frequency Definition**: Probability as limiting proportion in repeated experiments
- **Subjective Probability Definition**: Probability as measure of personal certainty

The mathematical approach uses:

- A sample space S of all possible outcomes
- A set of events (subsets of S)
- A mechanism for assigning probabilities 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.

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 the following two conditions hold:

1. $0 \leq P(a_i) \leq 1$
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.

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, or simply $P(A) = \sum_{a \in A} P(a)$.

Definition 1.2.5 (Odds)

The **odds in favour** of an event A is the probability the event occurs divided by the probability it does not occur, or simply $\frac{P(A)}{1 - P(A)}$. The **odds against** the event is the reciprocal of this quantity, or simply $\frac{1 - P(A)}{P(A)}$.

Example 1.2.1

Draw one card from a standard well-shuffled deck of cards, comprised of 13 cards (i.e., 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 $S = \{2\diamondsuit, 3\diamondsuit, \ldots, A\diamondsuit, 2\heartsuit, \ldots, A\clubsuit\}$. Each of the 52 cards has probability $\frac{1}{52}$. The event $A = \{2\clubsuit, 3\clubsuit, \ldots, A\clubsuit\}$ has 13 outcomes, so $P(A) = \frac{13}{52} = \frac{1}{4}$.

Example 1.2.2

Toss a coin twice. Find the probability of getting exactly one head.

Solution: Let $S = \{HH, HT, TH, TT\}$ and assume each simple event has probability $\frac{1}{4}$. Since one head occurs for simple events HT and TH, the event of interest is $A = \{HT, TH\}$ and $P(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Example 1.2.3

Roll a red die and a green die. Find the probability of the event $A =$ "the total number of dots showing on the upturned faces is 5".
Solution: Let (x, y) represent getting x on the red die and y on the green die. The sample space S has $6 \times 6 = 36$ equally probable outcomes. For the event of interest, $A = \{(1,4),(2,3),(3,2),(4,1)\}$ and therefore $P(A) = \frac{4}{36} = \frac{1}{9}$.

1.3 Counting in Uniform Probability Models

Key Concepts

Counting Rules:

- **Addition Rule:** If we can do job 1 in p ways and job 2 in q ways, we can do either job 1 OR job 2 (but not both) in $p + q$ ways.

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

Permutations and Combinations:

- $n! =$ number of arrangements of n distinct objects

- $n^{(k)} = n(n-1)\cdots(n-k+1) = \dfrac{n!}{(n-k)!} =$ permutations of k objects from n

- $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!} =$ combinations (subsets) of size k from n objects

Example 1.3.1

Consider an experiment in which we select two digits from the set $\{1, 2, 3, 4, 5\}$ with replacement. Find the probability that one digit is even.
Solution: The event can be re-worded as: "The first digit is even AND the second is odd (in 2×3 ways) OR the first digit is odd AND the second is even (in 3×2 ways)". There are $(2 \times 3) + (3 \times 2) = 12$ ways for this event to occur. Since S contains $5 \times 5 = 25$ outcomes, $P(\text{one digit is even}) = \frac{12}{25}$.

Example 1.3.2

Suppose the letters a, b, c, d, e, and f are arranged at random to form a six-letter word. Find the probability that the second letter is e or f.
Solution: The sample space has $6! = 720$ equally probable outcomes. For the event $A =$ "second letter is e or f": we can fill the second box in 2 ways, then the first box in 5 ways, then the remaining four boxes in $4! = 24$ ways. So $P(A) = \frac{2 \times 5 \times 24}{720} = \frac{240}{720} = \frac{1}{3}$.

Example 1.3.3

A password of length 4 is formed by randomly selecting with replacement 4 digits from $\{0, 1, \ldots, 9\}$. Find:

(a) $P(A)$ where $A =$ "password has only even digits": $P(A) = \dfrac{5^4}{10^4} = \dfrac{1}{16}$

(b) $P(B)$ where $B =$ "all digits unique": $P(B) = \dfrac{10^{(4)}}{10^4} = \dfrac{10 \times 9 \times 8 \times 7}{10^4} = \dfrac{63}{125}$

(c) $P(C)$ where $C =$ "password contains at least one 2": $P(C) = 1 - \dfrac{9^4}{10^4} = \dfrac{3439}{10000}$

Example 1.3.5

Consider a group of six third-year and seven fourth-year students. A committee of size five is randomly formed. Find:

(a) $P(A) =$ Roger (a third-year) is included: $P(A) = \dfrac{\binom{12}{4}}{\binom{13}{5}} = \dfrac{5}{13}$

(b) $P(B) =$ committee is all fourth-year: $P(B) = \dfrac{\binom{7}{5}}{\binom{13}{5}} = \dfrac{21}{1287} = \dfrac{7}{429}$

(c) $P(C) =$ at most four third-years: $P(C) = 1 - \dfrac{\binom{6}{5}}{\binom{13}{5}} = \dfrac{1281}{1287} = \dfrac{427}{429}$

Example 1.3.6

A box contains 3 red, 4 white, and 3 green balls. A sample of 4 balls is selected without replacement. Find:

(a) $P(A) =$ sample contains 2 red: $P(A) = \dfrac{\binom{3}{2}\binom{7}{2}}{\binom{10}{4}} = \dfrac{63}{210} = \dfrac{3}{10}$

(b) $P(B) =$ 2 red, 1 white, 1 green: $P(B) = \dfrac{\binom{3}{2}\binom{4}{1}\binom{3}{1}}{\binom{10}{4}} = \dfrac{36}{210} = \dfrac{6}{35}$

(c) $P(C) =$ 2 or more red: $P(C) = \dfrac{\binom{3}{2}\binom{7}{2} + \binom{3}{3}\binom{7}{1}}{\binom{10}{4}} = \dfrac{70}{210} = \dfrac{1}{3}$

Section 1.1 Problems

1. Try to think of examples of probabilities you have encountered which might have been obtained by each of the three "definitions".

2. Which definitions do you think could be used for obtaining the following probabilities?

   (a) You have a claim on your car insurance in the next year.

   (b) There is a meltdown at a nuclear power plant during the next 5 years.

(c) A person's birthday is in April.

Section 1.2 Problems

1. Students in a particular program have the same 4 math professors. Two students each independently ask one of their math professors for a letter of reference.
   (a) List a sample space for this "experiment".
   (b) Use this sample space to determine the odds in favour of both students asking the same professor.
2. Toss a fair coin 3 times.
   (a) List a sample space for this "experiment".
   (b) What are the odds against getting exactly 2 tails?

Section 1.3 Problems

1. A course has four sections with no limit on enrolment. Three students each pick a section at random.
   (i) Find the probability that all three students end up in the same section.
   (ii) Find the probability that all three students end up in different sections.
   (iii) Find the probability that no one picks section 1.
2. Canadian postal codes consist of 3 letters alternated with 3 digits. For a randomly chosen postal code, find the probability that:
   (a) all 3 letters are the same
   (b) the digits are all even or all odd
3. A binary sequence of length 10 is chosen uniformly at random. What is the probability it has exactly 5 zeros?
4. **The Birthday Problem:** Suppose there are r persons in a room. Find the probability that no two persons have the same birthday (ignoring Feb 29). Find the numerical value for $r = 20, 40, 60$.

Chapter 2

Probability Rules and Conditional Probability

2.1 Use of Sets

Key Concepts

Set Operations for Events:

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

De Morgan's Laws:

- $\overline{A \cup B} = \bar{A} \cap \bar{B}$
- $\overline{A \cap B} = \bar{A} \cup \bar{B}$

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 and B are two events with $A \subseteq B$, then $P(A) \leq P(B)$.

Rule 4a (Probability of the Union of Two Events)

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

Rule 4b (Probability of the Union of Three Events)

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

Definition 2.2.1 (Mutually Exclusive Events)

Events A and B are **mutually exclusive** if $AB = \emptyset$.

Rule 5 (Probability of Union of Mutually Exclusive Events)

If $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then

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

Rule 6 (Probability of the Complement)

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

Example 2.2.1

In a standard deck of 52 cards, two suits are red and two are black. There are 3 face cards (J, Q, K) per suit. If one card is randomly drawn, find P(red card or face card).

Solution: Let A = red card, B = face card. Then $P(A) = \frac{26}{52} = \frac{1}{2}$, $P(B) = \frac{12}{52} = \frac{3}{13}$, and $P(AB) = \frac{6}{52} = \frac{3}{26}$.

Using Rule 4a: $P(A \cup B) = \frac{1}{2} + \frac{3}{13} - \frac{3}{26} = \frac{8}{13}$

Example 2.2.2

An elementary school offers Russian, French, and German classes. Among 100 students: 26 in Russian, 29 in French, 17 in German, 12 in Russian and French, 6 in Russian and German, 4 in French and German, 2 in all three. Find the probability a randomly chosen student takes no language classes.

Solution: Using Rule 4b: $P(R \cup F \cup G) = 0.26 + 0.29 + 0.17 - 0.12 - 0.06 - 0.04 + 0.02 = 0.52$

By De Morgan's Law: $P(\bar{R} \cap \bar{F} \cap \bar{G}) = 1 - P(R \cup F \cup G) = 1 - 0.52 = 0.48$

Example 2.2.3

Three fair dice are rolled. Calculate $P(\text{at least one 6})$.
Solution: Let $A_i =$ "6 on die i". Using the complement: $P(A_1 \cup A_2 \cup A_3) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216}$

2.3 Dependent and Independent Events

Definition 2.3.1 (Independent Events)

Events A and B are **independent events** if and only if $P(AB) = P(A)P(B)$. If the events are not independent, we refer to them as **dependent**.

Definition 2.3.2 (Mutual Independence)

The events $A_1, A_2, \ldots, A_n$, $n \geq 2$, are **mutually independent** if and only if

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

for all distinct subscripts $i_1, i_2, \ldots, i_k$ chosen from $\{1, 2, \ldots, n\}$.

Example 2.3.1

A fair coin is tossed twice. Define:

- $A =$ "head on 1st toss"
- $B =$ "head on both tosses"
- $C =$ "head on 2nd toss"

Using $S = \{HH, HT, TH, TT\}$: $P(A) = P(C) = \frac{1}{2}$, $P(B) = \frac{1}{4}$, $P(AB) = P(AC) = \frac{1}{4}$.
Since $P(A)P(B) = \frac{1}{8} \neq \frac{1}{4} = P(AB)$, A and B are **dependent**.
Since $P(A)P(C) = \frac{1}{4} = P(AC)$, A and C are **independent**.

Example 2.3.3

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

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

Example 2.3.4

A pseudo random number generator gives independent digits from $\{0, 1, \ldots, 9\}$, each with probability $\frac{1}{10}$.

(a) $P(\text{all 5 digits odd}) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$

(b) $P(\text{9 first occurs on trial 10}) = (0.9)^9(0.1)$

2.4 Conditional Probability and Product Rules for Intersections of Events

Definition 2.4.1 (Conditional Probability)

The **conditional probability** of event A, given event B, is

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

Rule 7a (Product Rule for Two Events)

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

Rule 7b (Product Rule for Three Events)

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

Rule 8 (Law of Total Probability)

Let $\{A_i\}_{i=1}^n$ be a partition of S (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^n A_i = S$) with $P(A_i) > 0$ for all i. For any event B:

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

Rule 9 (Bayes' Rule)

Let $\{A_i\}_{i=1}^n$ be a partition of S with $P(A_i) > 0$ for all i. For any event B with $P(B) > 0$:

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

From which follows

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

Example 2.4.1

A fair coin is tossed three times. Find the probability that if at least one head occurred, exactly two heads occurred.
Solution: Let A = "at least one head" and B = "exactly two heads".

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

Example 2.4.4 (Insurance Risk Classes)

In an insurance portfolio: 10% in class 1 (high risk) with claim probability 0.4, 40% in class 2 with claim probability 0.2, 50% in class 3 (low risk) with claim probability 0.05. Find $P(\text{class 1}|\text{claim})$.
Solution: Using the Law of Total Probability: $P(\text{claim}) = (0.10)(0.4) + (0.40)(0.2) + (0.50)(0.05) = 0.145$
Using Bayes' Rule: $P(\text{class 1}|\text{claim}) = \dfrac{(0.10)(0.4)}{0.145} = \dfrac{0.04}{0.145} \approx 0.276$

Example 2.4.5 (Diagnostic Testing)

A medical test has 99% sensitivity (true positive rate) and 98% specificity (true negative rate). If 0.5% of the population has the disease, find $P(\text{disease}|\text{positive test})$.
Solution: Let D = disease, T^+ = positive test.

$$\begin{aligned} P(T^+) &= P(D)P(T^+|D) + P(\bar{D})P(T^+|\bar{D}) \\ &= (0.005)(0.99) + (0.995)(0.02) = 0.02485 \end{aligned}$$

$$P(D|T^+) = \frac{(0.005)(0.99)}{0.02485} \approx 0.199$$

Only about 20% of positive tests are true positives!

Section 2.2 Problems

1. According to a survey, 55% of voters are female, 55% are politically right, and 15% are male and politically left. What percentage are female and politically right?

2. If A and B are mutually exclusive with $P(A) = 0.25$ and $P(B) = 0.4$, find $P(\bar{A})$, $P(A \cup B)$, $P(A \cap B)$, $P(\bar{A} \cup \bar{B})$, $P(\bar{A} \cap \bar{B})$.

Section 2.3 Problems

1. If events A and B are independent with $P(A) = 0.3$ and $P(B) = 0.2$, determine $P(A \cup B)$.

2. Three digits are chosen at random with replacement from $\{0, 1, \ldots, 9\}$. Find $P(\text{all identical})$, $P(\text{all exceed 4})$, $P(\text{all different})$.

Section 2.4 Problems

1. Consider three identically-shaped dice: one fair, one with faces $\{1, 2, 3, 4, 5, 5\}$, and one with all 6's. A die is chosen at random and rolled twice, showing 6 both times. Find $P(\text{die was the all-6 die})$.

2. A drawer contains 3 red socks and 4 blue socks. If 2 socks are drawn at random without replacement, find $P(\text{both same color})$.

Chapter 3

Univariate Discrete Probability Distributions

3.1 Discrete Random Variables

Definition 3.1.1 (Random Variable)

A **random variable** X is a function (i.e., $X : S \mapsto \mathbb{R}$) 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 random variable X is the function

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

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

Definition 3.1.3 (Cumulative Distribution Function)

The **cumulative distribution function (cdf)** of a random variable X is the function

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

Properties: $0 \leq F(x) \leq 1$, F is non-decreasing, $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

Example 3.1.1

Two fair dice are thrown. Let X be the sum of the values on the upturned faces. The range of X is $A = \{2, 3, \ldots, 12\}$. The pmf is:

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

3.2 Functions of Random Variables

Key Concepts

If X is a random variable and $Y = g(X)$ for some function g, then Y 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)

The **expected value** (also called the **mean** or **expectation**) 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.

Theorem 3.3.1 (Expectation of a Function)

Suppose that X is a discrete random variable with range A and pmf $f(x)$. Then, the expected value of any real-valued function $g(X)$ is

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

Theorem 3.3.2 (Linearity of Expectation)

For real constants $c_1, c_2, \ldots, c_n$ and real-valued functions $g_1, g_2, \ldots, g_n$,

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

Corollary 3.3.1

For real constants a and b:

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

Definition 3.3.2 (Variance)

The **variance** of a random variable X is given by

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

Definition 3.3.3 (Standard Deviation)

The **standard deviation** of a random variable X is given by

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

Theorem 3.3.3 (Variance Formulas)

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

Theorem 3.3.4 (Variance of Linear Function)

For real constants a and b:

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

Example 3.3.1

A lottery sells 1000 tickets numbered 000–999 for \$10 each. A 3-digit winning number is drawn. Prizes: \$500 for matching all 3 digits, \$100 for matching last 2 digits only, \$10 for matching last digit only. Find $E(\text{net winnings})$.

Solution: Let X = winnings.

$$\begin{aligned} E(X) &= 500 \cdot \frac{1}{1000} + 100 \cdot \frac{9}{1000} + 10 \cdot \frac{90}{1000} + 0 \cdot \frac{900}{1000} \\ &= 0.50 + 0.90 + 0.90 = 2.30 \end{aligned}$$

So $E(\text{net winnings}) = E(X) - 10 = -\$7.70$

Example 3.3.3

For the pmf: $x \in \{1, 2, \ldots, 9\}$ with $f(x) = \{0.07, 0.1, 0.12, 0.13, 0.16, 0.13, 0.12, 0.1, 0.07\}$
Since the distribution is symmetric about $x = 5$: $\mu_X = 5$
$E(X^2) = \sum x^2 f(x) = 30.26$
$\sigma_X^2 = E(X^2) - \mu^2 = 30.26 - 25 = 5.26$, so $\sigma_X \approx 2.29$

3.4 Moment Generating Functions

Definition 3.4.1 (Moment Generating Function)

Consider a discrete random variable X with range A and pmf $f(x)$. The **moment generating function (mgf)** of X is defined as

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

provided the mgf is finite for t in an open neighborhood of 0.

Definition 3.4.2 (Moments)

For a random variable X, the quantity $E(X^n)$, $n = 1, 2, 3, \ldots$, is called the n**th moment** of the probability distribution.

Theorem 3.4.1 (Uniqueness Theorem)

Suppose that random variables X and Y have moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. If $M_X(t) = M_Y(t)$ for all $t \in (-a, a)$ for some $a > 0$, then X and Y have the same probability distribution.

Theorem 3.4.2 (Moments from MGF)

If X is a discrete random variable with mgf $M(t)$, then

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

Theorem 3.4.3 (MGF of Linear Transformation)

Let X be a random variable with mgf $M_X(t)$. 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 $X \sim \mathbf{DU}(a, b)$

X takes values in $\{a, a+1, \ldots, b\}$ with equal probability.

PMF: $f(x) = \dfrac{1}{b-a+1}$ for $x = a, a+1, \ldots, b$

Mean: $E(X) = \dfrac{a+b}{2}$

Variance: $\mathrm{Var}(X) = \dfrac{(b-a)(b-a+2)}{12}$

3.5.2 Binomial Distribution

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

X = number of successes in n independent Bernoulli trials, each with success probability p.

PMF: $f(x) = \dbinom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \ldots, n$

Mean: $E(X) = np$

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

MGF: $M(t) = (pe^t + 1 - p)^n$

3.5.3 Hypergeometric Distribution

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

From a population of N objects (r successes, $N - r$ failures), n are selected without replacement. X = number of successes.

PMF: $f(x) = \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ for $x = \max(0, n-N+r), \ldots, \min(r, n)$

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

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

3.5.4 Geometric Distribution

Geometric Distribution $X \sim$ **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 $t < \ln(1-p)^{-1}$

3.5.5 Negative Binomial Distribution

Negative Binomial Distribution $X \sim$ **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: $\mathrm{Var}(X) = \dfrac{k(1-p)}{p^2}$
MGF: $M(t) = \left(\dfrac{p}{1-(1-p)e^t}\right)^k$ for $t < \ln(1-p)^{-1}$

3.5.6 Poisson Distribution

Poisson Distribution $X \sim$ **Poi**(μ)

Used to model counts of rare events; approximates $\mathrm{Bin}(n, p)$ when n large, p small, $\mu = np$.
PMF: $f(x) = \dfrac{\mu^x e^{-\mu}}{x!}$ for $x = 0, 1, 2, \ldots$
Mean: $E(X) = \mu$
Variance: $\mathrm{Var}(X) = \mu$
MGF: $M(t) = e^{\mu(e^t - 1)}$

Example 3.5.1 (Hypergeometric vs Binomial Approximation)

15 cans of soup: 6 tomato, 9 mushroom. Select 8 at random. Find $P(3 \text{ tomato})$.
Exact (Hypergeometric): $P(X = 3) = \dfrac{\binom{6}{3}\binom{9}{5}}{\binom{15}{8}} = 0.3916$
Binomial Approximation: $P(X = 3) = \binom{8}{3}\left(\dfrac{6}{15}\right)^3\left(\dfrac{9}{15}\right)^5 = 0.2787$
The approximation is poor because we're sampling over half the population. With 1500 cans (600 tomato), the hypergeometric gives 0.2794 vs binomial 0.2787 – much closer!

Example 3.5.2 (Poisson Approximation)

200 wedding guests. Find $P(\text{exactly 2 born on January 1})$.

Exact (Binomial): $P(X=2) = \binom{200}{2}\left(\frac{1}{365}\right)^2\left(\frac{364}{365}\right)^{198} = 0.08677$

Poisson Approximation: With $\mu = \frac{200}{365}$: $P(X=2) = \frac{(200/365)^2 e^{-200/365}}{2!} = 0.08679$

Excellent approximation!

Section 3.5 Problems

1. A random four-digit number is created from $\{1, \ldots, 9\}$ with replacement. Find the pmf of $X =$ smallest digit.

2. The fraction of a population with blood type O+ is 0.38. Find $P($more than 10 people tested before finding 5 v

3. An airline sells 122 tickets for a 120-seat flight. If each passenger shows up independently with probability 0.97, find the probability of overbooking. Compare with Poisson approximation.

Chapter 4

Multivariate Discrete Probability Distributions

4.1 Basic Terminology and Techniques

Key Concepts

Joint PMF: For discrete random variables X and Y:

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

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

Marginal PMFs:

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

Definition 4.1.1 (Independent Random Variables)

X and Y are **independent random variables** if and only if

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

Definition 4.1.2 (Mutual Independence)

$X_1, X_2, \ldots, X_n$ are **independent random variables** if and only if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdots f_n(x_n) \quad \text{for all } x_1, \ldots, x_n$$

4.2 Multinomial Distribution

Multinomial Distribution

Consider n independent trials where each trial results in one of k outcomes with probabilities $p_1, p_2, \ldots, p_k$ (where $\sum_{i=1}^{k} p_i = 1$). If X_i = number of outcomes of type i, then $(X_1, \ldots, X_k)$ has a multinomial distribution.

Joint PMF:

$$f(x_1, \ldots, x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

for $x_i \geq 0$ integers with $\sum_{i=1}^{k} x_i = n$.

Marginals: $X_i \sim \text{Bin}(n, p_i)$

4.3 Expectation, Covariance, and Correlation

Theorem 4.3.1

Suppose that X and Y are independent random variables. If g_1 and g_2 are two real-valued functions, then

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

Definition 4.3.1 (Covariance)

The **covariance** of random variables X and Y is given by

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

Theorem 4.3.2

If X and Y are independent random variables, then $\text{Cov}(X, Y) = 0$.
Note: The converse is NOT generally true!

Definition 4.3.2 (Correlation Coefficient)

The **correlation coefficient** of random variables X and Y is

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

Properties: $-1 \leq \rho \leq 1$

4.4 Linear Combinations of Random Variables

Key Results

For random variables $X_1, \ldots, X_n$ and constants $a_1, \ldots, a_n$:

Mean of Linear Combination:

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

Variance of Linear Combination:

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

If $X_1, \ldots, X_n$ are independent:

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

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

Theorem 4.5.1 (Markov's Inequality)

If X is a random variable with $E[|X|] < \infty$, then for any $\varepsilon > 0$:

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

Theorem 4.5.2 (Chebyshev's Inequality)

If X is a random variable with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, then for any $k > 0$:

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

Equivalently, for any $\varepsilon > 0$:

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

Theorem 4.5.3 (Weak Law of Large Numbers)

Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with common mean μ and common variance σ^2. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then for any $\varepsilon > 0$:

$$\lim_{n\to\infty} P(|\bar{X} - \mu| < \varepsilon) = 1$$

4.6 Conditional Probability Distributions

Definition 4.6.1 (Conditional PMF)

Suppose that X and Y are discrete random variables with joint pmf $f(x, y)$. The **conditional pmf** of X given $Y = y$ is

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

provided $f_Y(y) > 0$.

Definition 4.6.2 (Conditional Mean)

The **conditional mean** of X given $Y = y$ is

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

Law of Total Expectation

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

Chapter 5

Univariate Continuous Probability Distributions

5.1 Continuous Random Variables

Definition 5.1.1 (Probability Density Function)

The **probability density function (pdf)** of a continuous random variable X is the 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, and $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Note: For continuous random variables, $P(X = x) = 0$ for any specific value x.

CDF: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$

Relationship: $f(x) = F'(x)$ wherever F is differentiable.

Definition 5.1.2 (Quantile)

For $p \in (0, 1)$, the **pth quantile** (or **100pth percentile**) of X is the value $q(p)$ such that:

$$F(q(p)) = P(X \leq q(p)) = p$$

The **median** is $q(0.5)$.

5.2 Functions of Random Variables

Theorem 5.2.1 (Transformation Method)

Let X be a continuous random variable with pdf $f_X(x)$. Suppose that g is a strictly monotonic (increasing or decreasing) differentiable function. Then $Y = g(X)$ has pdf:

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

Corollary 5.2.1 (Linear Transformation)

If X has pdf $f_X(x)$ and $Y = aX + b$ where $a \neq 0$, then:

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

5.3 Expectation of a Random Variable

Key Formulas for Continuous Random Variables

Expected Value:

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

Expected Value of a Function:

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

Variance:

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

5.4 Special Continuous Probability Distributions

5.4.1 Continuous Uniform Distribution

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

PDF: $f(x) = \dfrac{1}{b-a}$ for $a \leq x \leq b$

CDF: $F(x) = \dfrac{x-a}{b-a}$ for $a \leq x \leq b$

Mean: $E(X) = \dfrac{a+b}{2}$

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

MGF: $M(t) = \dfrac{e^{tb} - e^{ta}}{t(b-a)}$ for $t \neq 0$

5.4.2 Exponential Distribution

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

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: $\text{Var}(X) = \theta^2$
MGF: $M(t) = (1 - \theta t)^{-1}$ for $t < 1/\theta$
Memoryless Property: $P(X > s + t | X > s) = P(X > t)$

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

Theorem 5.4.2 (Standardization)

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

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

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$$

Properties: $\Phi(-z) = 1 - \Phi(z)$, $P(Z > z) = 1 - \Phi(z)$

Theorem 5.4.1 (Probability Integral Transform)

Suppose that X is a continuous random variable with cdf $F_X(x)$ that is strictly increasing on the support of X. Then $Y = F_X(X) \sim U(0, 1)$.
Corollary: If $U \sim U(0, 1)$ and F is a continuous cdf with inverse F^{-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 random variables from a distribution with mean μ and variance σ^2. Let $S_n = \sum_{i=1}^{n} X_i$. Then as $n \to \infty$:

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

Equivalently, for the sample mean $\bar{X} = S_n/n$:

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

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

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

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

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

Example: Normal Approximation to Binomial

100 tomato seeds germinate independently with probability 0.8 each. Find $P(X \geq 75)$.
Solution: $X \sim \text{Bin}(100, 0.8)$, so $\mu = 80$, $\sigma^2 = 16$.
Using CLT: $P(X \geq 75) \approx P\left(Z \geq \frac{75-80}{4}\right) = P(Z \geq -1.25) = \Phi(1.25) \approx 0.894$

Example: Sum of Random Variables

80 independent metal parts have costs C_i with $E(C_i) = 13$ and $\text{Var}(C_i) = 861$. Find $P(\text{total cost} > 1200)$.

Solution: Let $C = \sum_{i=1}^{80} C_i$. Then $E(C) = 1040$, $\text{Var}(C) = 68880$.

$P(C > 1200) \approx P\left(Z > \frac{1200 - 1040}{\sqrt{68880}}\right) = P(Z > 0.61) \approx 0.271$

Section 5.5 Problems

1. Let $X_1, \ldots, X_{100}$ be independent from a distribution with $\mu = 0.5$ and $\sigma^2 = 1/24$. Find $P(49 < \sum X_i < 50.5)$.
2. A game has three dice. Let $Y =$ (number of hearts) -1. Find $P(\text{profit} > 0)$ after $n = 10, 25, 50$ plays.
3. How many voters should be surveyed so that $P\left(\left|\frac{X}{n} - 0.16\right| \leq 0.03\right) = 0.95$?