Lecture Notes: Statistics

Part 1: Probability by example – A friendly introduction for computer scientists

Martin Simon

October 24, 2025

Contents

Acknowledgements

I would like to express my gratitude to Egbert Falkenberg for sharing his extensive teaching materials and experience from many years of teaching this course.

The main reference for these notes (particularly for the examples and exercises) is the excellent book *Probability and Statistics - The Science of Uncertainty* by Michael J. Evans and Jeffrey S. Rosenthal which is freely available online at https://utstat.utoronto.ca/mikevans/jeffrosenthal/

How to Use these Notes

These lecture notes come with a warning: they are full of examples, puzzles, and exercises. If you are looking for a text that you can skim once and then parrot back in an exam, you are in the wrong place. The goal here is not to turn you into a large language model that can autocomplete definitions, but into something far rarer: a human who thinks critically, solves problems creatively, and occasionally enjoys the process.

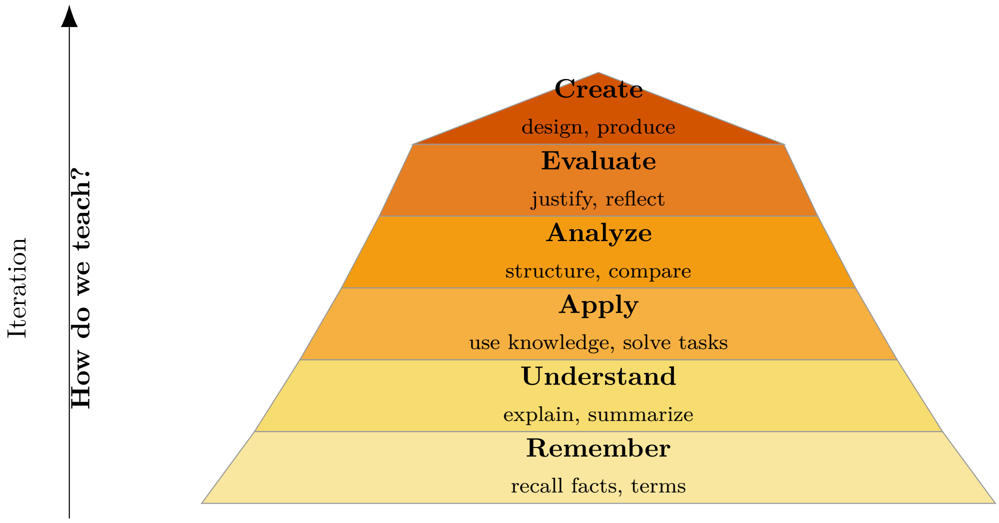
Tips for survival (and learning):

- Do the exercises. Reading solutions without trying them is like watching someone else go to the gym – you might learn the names of the machines, but your muscles won't grow.
- Talk to your peers. Explaining a tricky point to a classmate will show you whether you actually understand it, or whether you're just generating plausible-sounding sentences (again, very LLM-ish).
- Iterate. Struggle, make mistakes, revisit concepts. That's the cognitive ladder from "Remember" to "Create" in Bloom's Taxonomy below.
- Ask "why?" until you either get a satisfying answer or your neighbor starts avoiding eye contact.

Remember: an LLM can regurgitate what it has seen. My mission with these notes is to help you do what an LLM cannot – connect, analyze, evaluate, and create new ideas. If by the end of the semester you find yourself thinking more deeply than your favorite chatbot, then both of us have done our jobs.

From Reproduction to Creation

Bloom's Taxonomy describes cognitive learning objectives in six ascending stages. From bottom to top, cognitive complexity increases: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate*, and *Create*. The following figure visualizes this progression as a learning pyramid and emphasizes the iterative nature of effective teaching.



Higher levels build on the lower ones. Therefore, I will plan learning activities so that you gradually progress through the levels. We will use iteration and feedback to ensure sustainable competence development.

I hope these notes not only help you pass exams, but also spark curiosity and confidence that last well beyond the classroom. May you leave this course a little wiser, a little bolder, and definitely smarter than any large language model.

Chapter 1

Probability Models

1.1 Probability: A Measure of Uncertainty

In everyday life we routinely encounter the limits of our knowledge. Will traffic jam tonight? What weather arrives tomorrow? How will stocks move next week? Who wins an election? Where did we leave a hat? In many such cases we lack certainty and must instead rely on guesses, estimates, and hedges.

Probability is the mathematics of uncertainty. It supplies rigorous rules for reasoning with incomplete information. It will not reveal tomorrow's weather or next week's stock price, but it gives a structured framework to work with what we do and do not know so that our decisions are coherent. Saying "there's a 40% chance of rain tomorrow" does not determine the weather—it expresses our quantified uncertainty about it.

In these notes we refine the meaning of such statements. We will handle ideas like randomness, probability, expectation, prediction, and estimation in ways that are both intuitive and mathematically precise. Randomness can arise from many sources. Computers use *pseudorandom* numbers for games, simulation, and search; modern quantum theory suggests some phenomena are intrinsically random. The techniques we develop apply across these settings.

Another useful viewpoint is *relative frequency.* To say a fair coin has a 50% chance of heads can be read as: if flipped many times, about half the flips would be heads. This perspective has limitations (many events are not repeatable; "about" is vague) but remains a powerful source of intuition.

Although uncertainty is ancient, the mathematical theory of probability crystallized in the 17th century. In 1654, the gambler Le Chevalier de Méré posed betting questions to Blaise Pascal, prompting a correspondence with Pierre de Fermat and leading to developments on binomial coefficients (Pascal's triangle) and the binomial distribution. In the early 20th century, Markov, Kolmogorov, and Chebyshev in Russia and Norbert Wiener in the U.S. formalized foundations; later, William Feller and Joe Doob popularized the subject and its applications across physics, chemistry, computer science, economics, and finance.

1.1.1 Why Do We Need Probability Theory?

Because it appears everywhere.

Lotteries. In "Lotto 6/49" you pick six distinct numbers from 1–49; the lottery draws six distinct numbers. The chance of an exact match is 1/13,983,816—roughly one in 14 million. Not good odds.

Expected value and lotteries. If a ticket costs \$1, you shouldn't buy unless the jackpot exceeds about \$14 million; even then, many players will join, increasing the chance of splitting the pot—typically still unfavorable.

Three-card bet. A hat contains three cards: RR, BB, RB. One is chosen and placed flat; you see a *red* face. A friend bets \$4 against your \$3 that the hidden face is red. Intuition might say 50–50; conditional probability shows the chance is actually 2/3 that the other side is red, so the bet is unfavorable.

Coin-flip trap. A friend proposes 1,000 fair flips: if heads ≥ 600, you get \$100; otherwise you pay \$1. The law of large numbers implies the proportion of heads concentrates near 1/2; getting 600+ heads is astronomically unlikely (less than 10^{-9}). Decline.
These examples show that probabilistic thinking leads to better judgments – and can both save and earn money.

Probability also underlies safety-critical engineering. A reactor cannot be made absolutely risk-free: components may fail and extreme natural events can occur. Probability lets us quantify uncertainties and design systems so catastrophic release is *extremely* improbable (e.g., on the order of once per billion years). In short, we all navigate *risk* (e.g., driving, flying, walking). We accept these activities because the probabilities of harm are very low. As problems grow complex, intuition must be refined into precise tools—hence probability theory. Insurance, for instance, prices the transfer of risk using probability. Finally, and most importantly, probability theory is fundamental in computer science because it provides the mathematical framework for dealing with uncertainty, randomness, and incomplete information. Many core areas—such as randomized algorithms, machine learning, cryptography, and distributed systems—rely on probabilistic reasoning to ensure efficiency and reliability. In modern artificial intelligence, probability underpins methods for learning from data, modeling uncertainty, and making predictions. Even large language models (LLMs) are ultimately powered by probabilistic principles: they assign likelihoods to sequences of words in order to generate coherent text. In case, you are still not convinced that probability theory is indeed at the very heart of modern computer science, I would like to conclude this introduction with another motivating example:

Battling fake accounts on social media Twitter, Instagram, and Facebook fight a constant war against bots. How do they detect them? Not with a single trick, but with statistics: If someone posts 500 times a day at exactly the same interval, what's the probability it's human? How do you cluster suspiciously similar accounts? What's the cost of a false alarm (banning a real person) versus a missed detection (letting bots spread disinformation)?

At the end of the semester you will have the tools to answer (approximately) at least some of these questions.

Discussion Topics

1. Are phenomena like tomorrow's weather or next week's stock prices "truly" random, or is randomness a modeling convenience?
2. Can probabilities depend on the observer or on when they are assessed?
3. Is it surprising that probability wasn't formalized until the 17th century? Why or why not?
4. Where does probability matter in physics, computer science, and finance?
5. List situations from your own life where thinking probabilistically saved (or could have saved) money or led to better choices.
6. List portrayals of probability in movies. Were those portrayals reasonable?

1.2 Probability Models

A probability model begins with a *sample space* Ω: the set of all possible *outcomes* of the situation under study. Examples:

- Weather tomorrow: $\Omega = \{\text{rain}, \text{snow}, \text{clear}\}$.
- Next week's stock price: Ω could be the positive real numbers.

We call $\omega \in \Omega$ an *elementary event.* A probability model also specifies a family of *events*—subsets of Ω—to which we assign probabilities. For the weather example, $\{\text{rain}\}$, $\{\text{snow}\}$, $\{\text{rain, snow}\}$, Ω, and $\emptyset$ are all events. Formally, the collection of all possible events is called the *σ-algebra* $\mathcal{F}$, which is a collection of subsets of Ω satisfying the following properties:

1. $\Omega \in \mathcal{F}$.
2. If $A \in \mathcal{F}$, then its complement $A^c := \Omega \setminus A$ is also in $\mathcal{F}$.
3. If $(A_i)_{i \in \mathbb{N}}$ is a countable sequence of sets in $\mathcal{F}$, then the union $\bigcup_{i=1}^{\infty} A_i$ is in $\mathcal{F}$.

In finite settings, we typically treat all subsets as events (this is the power set you already know from your algebra lectures).

Finally, a *probability measure* P assigns a number $P(A)$ to each event $A \in \mathcal{F}$, satisfying:

1. $0 \le P(A) \le 1$.
2. $P(\emptyset) = 0$.
3. $P(\Omega) = 1$.
4. Countable additivity: If $A_1, A_2, \ldots$ are disjoint, then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \;=\; \sum_{i=1}^{\infty} P(A_i). \tag{1.2.1}$$

Definition 1.2.1 (Probability model)**.** A probability model consists of a nonempty sample space Ω, a σ-algebra $\mathcal{F}$ of possible events and a probability measure P on $\mathcal{F}$.

Recap: Venn Diagrams and Set Operations

Venn diagrams visualize Ω and its subsets. If $A \subseteq \Omega$, the *complement* is $A^c = \{\omega \in \Omega : \omega \notin A\}$. For two subsets $A, B \subseteq \Omega$:

$$\begin{aligned} A \cap B &= \{\omega :\ \omega \in A \text{ and } \omega \in B\}, \\ A \cup B &= \{\omega :\ \omega \in A \text{ or } \omega \in B\}, \\ A \cap B^c &= \{\omega :\ \omega \in A \text{ and } \omega \notin B\}. \end{aligned}$$

Useful identities: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$ (De Morgan). If A and B are disjoint, they do not overlap.

Examples

Example 1.2.2 (Cholesterol)**.** A medical doctor is interested in modeling the cholesterol levels of her patients probabilistically. Every time a patient visits her, she tests their cholesterol level. Here the *experiment* is the cholesterol test, the *outcome* is the measured cholesterol level, and the *sample space* Ω is the positive real line. The doctor is mainly interested in whether the patients have low, borderline-high, or high cholesterol. The event L (*low cholesterol*) contains all outcomes below 200 mg/dL, the event B (*borderline-high cholesterol*) contains all outcomes between 200 and 240 mg/dL (including the boundaries), and the event H (*high cholesterol*) contains all outcomes above 240 mg/dL. The σ-algebra $\mathcal{F}$ of possible events therefore equals

$$\mathcal{F} := \{L \cup B \cup H,\ L \cup B,\ L \cup H,\ B \cup H,\ L,\ B,\ H,\ \emptyset\}. \tag{1.3}$$

Example 1.2.3 (Weather)**.** $\Omega = \{\text{rain}, \text{snow}, \text{clear}\}$ with $P(\text{rain}) = 0.40$, $P(\text{snow}) = 0.15$, $P(\text{clear}) = 0.45$. Then $P(\text{rain or snow}) = 0.55$; $P(\emptyset) = 0$; $P(\Omega) = 1$.

Example 1.2.4 (Election)**.** If your candidate wins with probability 60%, set $\Omega = \{\text{win}, \text{lose}\}$, $P(\text{win}) = 0.6$, $P(\text{lose}) = 0.4$.

Example 1.2.5 (One fair coin)**.** $\Omega = \{H, T\}$, $P(H) = P(T) = 0.5$.

Example 1.2.6 (Three fair coins: sequences)**.**

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

each with $P = 1/8$. Then $P(\text{first } H \text{ and second } T) = P(HTH) + P(HTT) = 1/4$.

Example 1.2.7 (Three fair coins: number of heads)**.** If we only track the number of heads, $\Omega = \{0, 1, 2, 3\}$, but outcomes are not equally likely. Later we will see $P(0) = P(3) = 1/8$ and $P(1) = P(2) = 3/8$.

Example 1.2.8 (Countably infinite sample space)**.** Consider the experiment of repeatedly tossing a fair coin until the first head appears. The sample space is

$$\Omega = \{H, TH, TTH, TTTH, \dots\},$$

that is, the set of all finite sequences of tosses ending in the first occurrence of a head. This sample space is *countably infinite*, since its elements can be put in one-to-one correspondence with the natural numbers:

$$H \leftrightarrow 1, \quad TH \leftrightarrow 2, \quad TTH \leftrightarrow 3, \quad \dots$$

If the coin is fair, the probability of obtaining the first head on the n-th toss is

$$P(\text{first head on toss } n) = \left(\frac{1}{2}\right)^n.$$

The total probability is

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = 1,$$

so this defines a valid probability measure on a countably infinite sample space.

Example 1.2.9 (Uncountable sample space)**.** Let $\Omega = [0, 1]$. Define P so that for any interval $[a, b] \subseteq [0, 1]$,

$$P([a, b]) = b - a.$$

This is the *uniform distribution* on $[0, 1]$.

Summary of Section 1.2

- A probability model = sample space Ω + probability measure P on events $\mathcal{F}$.
- Sample spaces can be finite, countable, or uncountable.
- Venn diagrams aid set reasoning.

Exercises for Section 1.2

Exercise 1.2.10. $\Omega = \{1, 2, 3\}$ with $P(\{1\}) = \frac{1}{2}$, $P(\{2\}) = \frac{1}{3}$, $P(\{3\}) = \frac{1}{6}$.
(a) $\mathcal{F}$? (b) $P(\{1, 2\})$? (c) $P(\{1, 2, 3\})$? (d) List all events $A \in \mathcal{F}$ with $P(A) = \frac{1}{2}$.

Exercise 1.2.11. $\Omega = \{1, \ldots, 8\}$ with $P(\{\omega\}) = \frac{1}{8}$ for all $\omega \in \Omega$.
(a) $P(\{1, 2\})$? (b) $P(\{1, 2, 3\})$? (c) How many $A \in \mathcal{F}$ satisfy $P(A) = \frac{1}{2}$?

Exercise 1.2.12. $\Omega = \{1, 2, 3\}$ with $P(\{1\}) = \frac{1}{2}$ and $P(\{1, 2\}) = \frac{2}{3}$. Find $P(\{2\})$.

Exercise 1.2.13. We attempt to define P on $\Omega = \{1, 2, 3\}$ by
$P(\{1, 2, 3\}) = 1$, $P(\{1, 2\}) = 0.7$, $P(\{1, 3\}) = 0.5$, $P(\{2, 3\}) = 0.7$, $P(\{1\}) = 0.2$, $P(\{2\}) = 0.5$, $P(\{3\}) = 0.3$. Is this valid? Why/why not?

Exercise 1.2.14. Given a uniform distribution on $[0, 1]$. For a single point $\omega \in [0, 1]$, what is $P(\{\omega\})$? Surprising?

Exercise 1.2.15. Label the subregions in a three-set Venn diagram (A, B, C) using intersections and complements.

Exercise 1.2.16. Depict elements in A or B but not both. Express the set using unions, intersections, and complements.

Exercise 1.2.17. $\Omega = \{1, 2, 3\}$ with $P(\{1, 2\}) = \frac{1}{3}$ and $P(\{2, 3\}) = \frac{2}{3}$. Compute $P(\{1\}), P(\{2\}), P(\{3\})$.

Exercise 1.2.18. $\Omega = \{1, 2, 3, 4\}$ with $P(\{1\}) = \frac{1}{12}$, $P(\{1, 2\}) = \frac{1}{6}$, $P(\{1, 2, 3\}) = \frac{1}{3}$. Compute $P(\{1\}), P(\{2\}), P(\{3\}), P(\{4\})$.

Exercise 1.2.19. $\Omega = \{1, 2, 3\}$ with $P(\{1\}) = P(\{3\}) = 2P(\{2\})$. Compute $P(\{1\}), P(\{2\}), P(\{3\})$.

Exercise 1.2.20. $\Omega = \{1, 2, 3\}$ with $P(\{1\}) + P(\{2\}) = \frac{1}{6}$ and $P(\{3\}) = 2P(\{2\})$. Compute $P(\{1\}), P(\{2\}), P(\{3\})$.

Exercise 1.2.21. $\Omega = \{1, 2, 3, 4\}$ with $P(\{1\}) = \frac{1}{8}$, $P(\{2\}) = 3P(\{3\}) = 4P(\{4\})$. Determine all four probabilities.

More challenging Problems for Section 1.2

Problem 1.2.22. Assume a uniform distribution on $[0, 1]$. Is $P([0, 1]) \stackrel{?}{=} \sum_{\omega \in [0,1]} P(\{\omega\})$? Relate to countable additivity.

Problem 1.2.23. If Ω is finite or countable, can $P(\{\omega\}) = 0$ for every $\omega \in \Omega$? Explain.

Problem 1.2.24. If Ω is uncountable, can $P(\{\omega\}) = 0$ for every $\omega \in \Omega$? Explain.

Discussion for Section 1.2

Discussion 1.2.25. Does additivity feel intuitive? Why or why not?

Discussion 1.2.26. Why is $P(\Omega) = 1$ essential? How would the theory change otherwise?

1.3 Properties of Probability Models

From additivity follow several universal facts. For any event A, its complement is A^c ("A does not occur"). Because A and A^c are disjoint and $A \cup A^c = \Omega$, additivity yields

$$P(A^c) = 1 - P(A). \tag{1.3.1}$$

If $A_1, A_2, \ldots$ are disjoint and cover the sample space ($\bigcup_i A_i = \Omega$), they form a *partition*. For any event B, we have the

Law of total probability (unconditioned): For each partition of Ω we have

$$P(B) = \sum_i P(A_i \cap B).$$

Theorem 1.3.1. *If $B \subseteq A$, then*

(i) $P(B) \leq P(A)$ (Monotonicity)

(ii) $P(A \cap B^c) = P(A) - P(B)$.

Two-event inclusion–exclusion:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.3.4}$$

Subadditivity: For any (finite or countable) collection $A_1, A_2, \ldots$ (not necessarily disjoint),

$$P\Big(\bigcup_i A_i\Big) \leq \sum_i P(A_i).$$

Summary of Section 1.3

- Complement rule (1.3.1).
- Monotonicity, subadditivity, and total probability hold in all models.
- Inclusion–exclusion (1.3.4) computes $P(A \cup B)$ from simpler pieces.

Exercises for Section 1.3

Exercise 1.3.2. If you pick a random integer between 1 and 100 uniformly, what is the probability that it is divisible by 3 or by 5?

Exercise 1.3.3. $\Omega = \{1, 2, \ldots, 100\}$ with $P(\{1\}) = 0.1$.
(a) $P(\{2, 3, \ldots, 100\})$? (b) Smallest possible $P(\{1, 2, 3\})$?

Exercise 1.3.4. Al watches the 6pm news 2/3 of days, the 11pm news 1/2 of days, and both 1/3 of days.
(a) Probability he watches only the 6pm news? (b) Probability he watches neither?

Exercise 1.3.5. An employee arrives late 10%, leaves early 20%, both 5%. Probability of "late or early (or both)"?

Exercise 1.3.6. A patient has right knee sore 15% of the time, left knee 10%. Largest and smallest possible chance at least one knee is sore?

Exercise 1.3.7. Five fair coin tosses. (a) $P(\text{all heads})$? (b) $P(\text{at least one tail})$?

Exercise 1.3.8. Random card from 52-card deck. (a) $P(\text{jack})$? (b) $P(\text{club})$? (c) $P(\text{jack of clubs})$? (d) $P(\text{jack or club})$?

Exercise 1.3.9. Team has 30% chance to win and 40% chance to win or tie. Probability of tie?

Exercise 1.3.10. 55% of students are female; of these, 4/5 (i.e., 44% overall) have long hair. 45% are male; of these, 1/3 (15% overall) have long hair. $P(\text{female or long hair (or both)})$?

More challenging Problem for Section 1.3

Problem 1.3.11. We choose a positive integer with unknown distribution. Given $P(\{1, 2, 3, 4, 5\}) = 0.3$, $P(\{4, 5, 6\}) = 0.4$, and $P(\{1\}) = 0.1$. Find the largest and smallest possible values of $P(\{2\})$.

Challenge for Section 1.3

Challenge 1.3.12 (Inclusion–exclusion generalization)**.** (a) For three events A, B, C, show $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$. (b) Extend to n events $A_1, \ldots, A_n$ (hint: induction).

Discussion for Section 1.3

Discussion 1.3.13. Which results here are most/least important in your view? Why?

1.4 Uniform Probability on a Finite Sample Space

When Ω is finite and all outcomes are equally likely, the *uniform* measure assigns $P(\{\omega\}) = 1/|\Omega|$ for all elementary events $\omega \in \Omega$. Thus for any event $A \in \mathcal{F}$,

$$P(A) = \frac{|A|}{|\Omega|}. \tag{1.4.1}$$

Examples

Example 1.4.1 (Die roll). $\Omega = \{1, \dots, 6\}$. For any $A \subseteq \Omega$, $P(A) = |A|/6$.

Example 1.4.2 (Fair coin). $\Omega = \{H, T\}$, $P(H) = P(T) = 1/2$.

Example 1.4.3 (Three fair coins (ordered)). $|\Omega| = 8$ sequences; each has probability $1/8$.

Example 1.4.4 (Die + coin). $\Omega = \{1H, \dots, 6H, 1T, \dots, 6T\}$ with $|\Omega| = 12$.

Uniform problems reduce to counting $|A|$ and $|\Omega|$. Key tools:

Multiplication Principle (Counting Sequences)

If a process has n_1 choices for step 1, n_2 for step 2, etc., then total outcomes $= n_1 n_2 \cdots$.

Example 1.4.5. Three fair coins and two fair dice $\Rightarrow 2^3 \cdot 6^2 = 288$ outcomes. Probability of HHH and both dice showing 6 is $1/288$. Equivalently: *Independence* gives $(1/2)^3(1/6)^2$.

Example 1.4.6. Two dice: 36 outcomes; we are interested in the event that three ordered outcomes sum to 10 (4,6), (5,5), (6,4). So $P(\text{sum} = 10) = 3/36 = 1/12$.

Permutations

Number of ordered length-k selections without replacement from n distinct items: $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$. For $k = n$: $n!$.

Example 1.4.7 (Coat check). Four coats randomly returned to four people: total matchings $4! = 24$. Exactly one perfect assignment $\Rightarrow P = 1/24$.

Subsets (Binomial Coefficients)

Number of k-element subsets from n items:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}. \tag{1.4.2}$$

Example 1.4.8 (Ten coins: exactly 7 heads). Probability $= \binom{10}{7}/2^{10} = 120/1024 \approx 0.117$. For biased coins with $P(H) = p$, probability of exactly k heads in n flips is $\binom{n}{k}p^k(1-p)^{n-k}$.

Partitions (Multinomial Coefficients)

Number of ways to split n labeled items into l ordered groups of sizes $k_1, \dots, k_l$ (disjoint, covering all n):

$$\frac{n!}{k_1!\,k_2!\cdots k_l!}. \tag{1.4.4}$$

Example 1.4.9 (Bridge deals). Divide 52 cards into four labeled hands of 13 each: $\dfrac{52!}{13!\,13!\,13!\,13!}$ (enormous).

Summary of Section 1.4

- In finite uniform spaces, $P(A) = |A|/|\Omega|$ for all $A \in \mathcal{F}$.
- Counting uses multiplication, factorials, and binomial/multinomial coefficients.

Exercises for Section 1.4

Exercise 1.4.10. Eight fair dice. (a) Probability all eight are 6? (b) Probability all eight show the same number? (c) Probability sum equals 9?

Exercise 1.4.11. Ten fair dice. Probability exactly two show "2".

Exercise 1.4.12. 100 fair independent coins. Probability at least three heads? (Hint: use (1.3.1).)

Exercise 1.4.13. Five-card hand from a 52-card deck. Probability of: (a) four aces plus the king of spades; (b) all spades; (c) no pairs (all values distinct); (d) full house.

Exercise 1.4.14. Bridge (four 13-card hands). Probability that (a) all 13 spades go to one hand; (b) all four aces go to the same hand.

Exercise 1.4.15. Two random cards. Values: J/Q/K count 10, Ace counts 1. Probability sum ≥ 4?

Exercise 1.4.16. Deal cards until the first jack appears. Probability at least 10 cards precede it?

Exercise 1.4.17. In a random shuffle, probability ace of spades and ace of clubs are adjacent?

Exercise 1.4.18. Repeatedly roll two dice; consider sums. Probability first 7 occurs on the third roll?

Exercise 1.4.19. Three dice. Probability exactly two show the same value and the third is different?

Exercise 1.4.20. Urn #1: 5 red, 7 blue. Urn #2: 6 red, 12 blue. Draw 3 from each urn uniformly. Probability all six drawn are the same color?

Exercise 1.4.21. Roll one die and flip three fair coins. Probability "#heads equals the die value"?

Exercise 1.4.22. Flip 2 pennies, 3 nickels, 4 dimes. Probability the *value* showing heads totals \$0.31?

More challenging Problems for Section 1.4

Problem 1.4.23. Show (1.4.1) implies additivity (1.2.1).

Problem 1.4.24. Eight dice: compute $P(\text{sum} = 9)$, $P(\text{sum} = 10)$, $P(\text{sum} = 11)$.

Problem 1.4.25. Roll one die; flip six coins. Probability #heads equals die outcome?

Problem 1.4.26. Ten dice. Probability exactly two 2's and exactly three 3's?

Problem 1.4.27. Bridge hands. Probability North and East have exactly the same number of spades?

Problem 1.4.28. Pick a card and flip 10 coins. Probability #heads equals card value (J/Q/K=10, Ace=1)?

Challenges for Section 1.4

Challenge 1.4.29. Roll two dice and flip 12 coins. Probability heads count equals the sum on the dice?

Challenge 1.4.30 (Birthday problem)**.** C people with birthdays equally likely among 365 days (non-leap year).
(a) $C = 2$: probability same birthday?
(b) General C: probability *all* share the same birthday?
(c) General C: probability *some pair* matches?
(d) Smallest C where (c) exceeds 0.5? Surprising?

1.5 Conditional Probability and Independence

Consider three fair coin flips with equally likely sequences. Unconditionally, $P(\text{first } H) = 1/2$. If we learn that *exactly two* heads occurred, the possible sequences are HHT, HTH, THH (three equally likely), and in two of them the first flip is H. Thus

$$P(\text{first } H \mid \text{two heads}) = 2/3. \tag{1.5.0}$$

Conditional Probability

For events A, B with $P(B) > 0$,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{1.5.1}$$

From (1.5.1):

$$\textbf{Multiplication rule:} \quad P(A \cap B) = P(A)\, P(B \mid A). \tag{1.5.2}$$

Law of total probability (conditioned). If $A_1, A_2, \dots$ is a partition with $P(A_i) > 0$,

$$P(B) = \sum_i P(A_i)\, P(B \mid A_i).$$

Bayes' Theorem. For $P(A), P(B) > 0$,

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

Example 1.5.1 (Two-stage urn)**.** Urn #1: 3R,2B; Urn #2: 4R,7B. Choose urn uniformly; then draw one ball uniformly from that urn.
$P(\text{urn 2 and blue}) = \frac{1}{2} \cdot \frac{7}{11}$;
$P(\text{blue}) = \frac{1}{2} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{7}{11}$;
$P(\text{urn 2} \mid \text{blue}) = \dfrac{\frac{1}{2} \cdot \frac{7}{11}}{P(\text{blue})} \approx 0.614.$

Independence

Two events A, B are *independent* if

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

equivalently (when $P(A), P(B) > 0$) $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

Example 1.5.2 (Die and coin)**.** Let $A = \{\text{die} = 5\}$, $B = \{\text{coin} = T\}$. Then $P(A) = 1/6$, $P(B) = 1/2$, and $P(A \cap B) = 1/12 = (1/6)(1/2)$, so A, B are independent.

For multiple events $A_1, \dots, A_n$, independence means every finite subcollection factors: $P(\bigcap_j A_{i_j}) = \prod_j P(A_{i_j})$. Pairwise independence alone is not sufficient (counterexamples exist).

Example 1.5.3 (Monty Hall Problem)**.** Consider a game show with three closed doors. Behind one door is a car (the prize), and behind the other two are goats. A contestant chooses one door, say Door 1. Monty, the host, who knows what is behind each door, then opens one of the remaining doors, always revealing a goat (for example, Door 3). The contestant is now offered a choice:

- Stay: Keep the original choice (Door 1), or
- Switch: Change to the other unopened door (Door 2).

Question: Should the contestant stay or switch?

Solution: Will be presented in the lecture!

Summary of Section 1.5

- Conditional probability (1.5.1); multiplication rule (1.5.2); Bayes' Theorem.
- Conditioned total probability decomposes $P(B)$ across a partition.
- Independence means events do not affect each other's probabilities; for multiple events, require all finite intersections to factor.

Exercises for Section 1.5

Exercise 1.5.4. If you flip three fair coins, what is the probability that the first coin shows heads given that exactly two of the three coins are heads?

Exercise 1.5.5. Four dice. (a) $P(\text{first die} = 2 \mid \text{exactly three dice show } 2)$?
(b) $P(\text{first die} = 2 \mid \text{at least three dice show } 2)$?

Exercise 1.5.6. Two fair coins + one fair die. (a) $P(\#H = \text{die})$?
(b) $P(\#H = \text{die} \mid \text{die} = 1)$?
(c) Is (b) larger or smaller than (a)? Explain.

Exercise 1.5.7. Three fair coins. (a) $P(HHH)$? (b) $P(HHH \mid \text{odd \# heads})$? (c) $P(HHH \mid \text{even \# heads})$?

Exercise 1.5.8. Five-card hand. $P(\text{all spades} \mid \text{at least four spades})$?

Exercise 1.5.9. Five-card hand. $P(\text{all four aces} \mid \text{at least four aces in hand})$?

Exercise 1.5.10. Five-card hand. $P(\text{no pairs} \mid \text{no spades})$?

Exercise 1.5.11. Baseball: Pitcher throws fastballs 80%, curveballs 20%. Batter homerun (HR) rates: 8% on fastballs, 5% on curves. $P(\text{HR on next pitch})$?

Exercise 1.5.12. $P(\text{snow}) = 0.2$, $P(\text{accident}) = 0.1$, and $P(\text{accident} \mid \text{snow}) = 0.4$. Compute $P(\text{snow} \mid \text{accident})$.

Exercise 1.5.13. Two dice: red and blue. Events A: equal numbers; B: sum $= 12$; C: red die $= 4$; D: blue die $= 4$.
Check independence for (a) A, B; (b) A, C; (c) A, D; (d) C, D; (e) A, C, D jointly.

Exercise 1.5.14. Roll a die, then flip #coins equal to the die. (a) $P(\#H = 3)$?
(b) Given $\#H = 3$, what is $P(\text{die} = 5)$?

Exercise 1.5.15. Roll a die, then draw as many cards as the die shows. (a) $P(\text{\# jacks } = 2)$?
(b) Given # jacks $= 2$, what is $P(\text{die} = 3)$?

More challenging Problems for Section 1.5

Problem 1.5.16 (Three cards puzzle)**.** 2-sided colored cards: RR, BB, RB. Choose one at random, place flat, view one side.
(a) P(visible side is red)? (b) Given red is visible, P(card is RR)?
(c) Describe how to verify (b) experimentally.

Problem 1.5.17. Prove A and B are independent iff A^c and B are independent.

Problem 1.5.18. For events with positive probability, prove $P(A \mid B) = P(A)$ iff $P(B \mid A) = P(B)$.

Challenges for Section 1.5

Challenge 1.5.19. An urn contains 8 red and 12 blue balls (20 total). You draw 5 balls without replacement. What is the probability of drawing exactly k red balls, for $k = 0, 1, 2, 3, 4, 5$?

Challenge 1.5.20. Three dice. Compute $P(\text{first die} = 4 \mid \text{sum} = 12)$.

Challenge 1.5.21. First roll: win on 7 or 11; lose on 2,3,12; otherwise the sum becomes the *point*. Then roll until either the point reappears (win) or 7 occurs (lose).
(a) If point is 4, compute $P(\text{win} \mid \text{point} = 4)$.
(b) Compute $p_i = P(\text{win} \mid \text{point} = i)$ for $i = 2, \ldots, 12$.
(c) Overall win probability (use (b) and total probability).

Discussion for Section 1.5

Discussion 1.5.22. If you can repeat an experiment often, how would you test whether two events are independent?

Discussion 1.5.23 (Monty Hall)**.** Reflect on why the problem generated controversy (and how host rules change the analysis).

1.6 Advanced: Continuity of P

We often want to pass to limits of events. Define:

- $A_n \uparrow A$ ("increasing to A"): $A_1 \subseteq A_2 \subseteq \cdots$ and $\bigcup_{n=1}^{\infty} A_n = A$.
- $A_n \downarrow A$ ("decreasing to A"): $A_1 \supseteq A_2 \supseteq \cdots$ and $\bigcap_{n=1}^{\infty} A_n = A$.

Theorem 1.6.1 (Continuity of probability)**.** *If $A_n \uparrow A$ or $A_n \downarrow A$, then $\lim_{n\to\infty} P(A_n) = P(A)$.*

Example 1.6.2. Let $\Omega = \{1, 2, 3, \dots\}$ with $P(\{\omega\}) = 2^{-\omega}$. For $A = \{5, 6, 7, \dots\}$, take $A_n = \{5, 6, \dots, n\} \uparrow A$. Then P is indeed a probability measure, since

$$P(\Omega) = \sum_{\omega=1}^{\infty} 2^{-\omega} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

Let $A = \{5, 6, 7, \dots\}$ and $A_n = \{5, 6, \dots, n\}$. Clearly, $A_n \uparrow A$, i.e.,

$$A_1 \subseteq A_2 \subseteq \cdots \quad \text{and} \quad \bigcup_n A_n = A.$$

By continuity from below of probability measures,

$$P(A) = \lim_{n\to\infty} P(A_n).$$

Now,

$$P(A_n) = \sum_{\omega=5}^{n} 2^{-\omega},$$

hence

$$P(A) = \lim_{n\to\infty} \sum_{\omega=5}^{n} 2^{-\omega} = \sum_{\omega=5}^{\infty} 2^{-\omega}.$$

The series is geometric with first term $2^{-5} = \frac{1}{32}$ and ratio $\frac{1}{2}$, so

$$\sum_{\omega=5}^{\infty} 2^{-\omega} = \frac{2^{-5}}{1 - \frac{1}{2}} = \frac{\frac{1}{32}}{\frac{1}{2}} = \frac{1}{16}.$$

Alternatively,

$$P(A) = 1 - \sum_{\omega=1}^{4} 2^{-\omega} = 1 - \left(\tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16}\right) = 1 - \tfrac{15}{16} = \tfrac{1}{16}.$$

Thus,

$$P(A) = \lim_{n\to\infty} \sum_{\omega=5}^{n} 2^{-\omega} = \sum_{\omega=5}^{\infty} 2^{-\omega} = \frac{1}{16}.$$

Summary of Section 1.6

- If events increase to A or decrease to A, their probabilities converge to $P(A)$.
- This enables limit calculations and bounds that would otherwise be awkward.

Chapter 2

Random Variables and Distributions

Rather than specifying a probability model entirely by a measure on a sample space, we often work with *random variables* (functions on the sample space), and with their *distributions*: probability mass functions (pmfs), density functions (pdfs), and cumulative distribution functions (cdfs). This chapter develops these objects as the main computational interface to a probability model. We also study conditional distributions, independence, change of variables, joint distributions, and basic simulation methods.

2.1 Random Variables

A random variable assigns a numerical value to each outcome.

Definition 2.1.1 (Random variable)**.** Given a probability model with sample space Ω, a *random variable* is any function $X : \Omega \to \mathbb{R}$.

Example 2.1.2. Weather outcomes $\Omega = \{\text{rain}, \text{snow}, \text{clear}\}$; define $X(\text{rain}) = 3$, $X(\text{snow}) = 6$, $X(\text{clear}) = 2.7$. The random variable X may, e.g., model the heating cost that results from a given weather outcome. That is, if we have a probability model for the weather, introducing X enables designing a probabilistic model connecting weather and heating costs.

Example 2.1.3 (Indicator)**.** For any event $A \subseteq \Omega$, the *indicator* $\mathbf{1}_A(\omega) = 1$ if $\omega \in A$ and 0 otherwise is a random variable.

Arithmetic on random variables is pointwise; e.g. $(X + Y)(\omega) = X(\omega) + Y(\omega)$. We write $X \leq Y$ if $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$. Constants are random variables ($X \equiv c$). Caution: Unbounded random variables are possible if Ω is infinite.

Summary. A random variable is any real-valued function on the sample space Ω. Indicators and algebraic combinations of random variables are again random variables.

Exercises for Section 2.1

Exercise 2.1.4. Let $\Omega = \{1, 2, 3\}$, $X(\omega) = \omega^2$, $Y(\omega) = 1/\omega$. Determine $\min X$, $\max X$, $\min Y$, $\max Y$.

Exercise 2.1.5. Let $\Omega = \{\text{high}, \text{middle}, \text{low}\}$. Define $X(\text{high}) = 12$, $X(\text{middle}) = 2$, $X(\text{low}) = -3$; $Y(\text{high}) = 0$, $Y(\text{middle}) = 0$, $Y(\text{low}) = 1$; $Z(\text{high}) = 6$, $Z(\text{middle}) = 0$, $Z(\text{low}) = -4$. Decide each relation: $X \leq Y$; $X \geq Y$; $Y \leq Z$; $Y \geq Z$; $XY \leq Z$; $XY \geq Z$.

2.2 Distributions of Random Variables

For $B \subseteq \mathbb{R}$, the probability that X lands in B is $P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})$.

Definition 2.2.1 (Distribution of a random variable)**.** The *distribution* of X is the set function $B \mapsto P(X \in B)$ (for Borel $B \subseteq \mathbb{R}$).

Example 2.2.2. If $S = \{\text{rain}, \text{snow}, \text{clear}\}$ with $P(\text{rain}) = 0.40$, $P(\text{snow}) = 0.15$, $P(\text{clear}) = 0.45$ and $X(\text{rain}) = 3$, $X(\text{snow}) = 6$, $X(\text{clear}) = 2.7$, then $P(X \in B) = 0.40 \cdot \mathbf{1}_B(3) + 0.15 \cdot \mathbf{1}_B(6) + 0.45 \cdot \mathbf{1}_B(2.7)$.

Summary. The distribution of X is fully specified by the probabilities $P(X \in B)$.

Exercises for Section 2.2

Exercise 2.2.3. Flip two fair coins. Let X be the number of heads. Compute $P(X = x)$ for all x.

Exercise 2.2.4. Roll two fair dice. Let Y be their sum. Compute $P(Y = y)$ and $P(Y \in B)$ for general B.