Chapter 1: Probability and Counting

Introduction to Probability

DRME

December 25, 2025

Introduction

Why Study Probability?

- **Mathematics**: Logic of certainty
- **Probability**: Logic of uncertainty
- **Applications**:
  - Statistics (foundation for data analysis)
  - Physics (quantum mechanics, statistical mechanics)
  - Biology (genetics, evolutionary biology, ecology)
  - Computer science (randomized algorithms, machine learning)
  - Finance (stock modeling, risk assessment)
  - Medicine (clinical trials)
  - Everyday life (decision making, understanding ”coincidences”)

The Probabilistic Model

The Probabilistic Model

A probabilistic model is a mathematical description of an uncertain situation. It consists of:

- **Sample space** (Ω or S): set of all possible outcomes
- **Probability law**: assigns probabilities to events (subsets of outcomes)

Sample Space[1]

Definition (Sample Space)
The **sample space** S is the set of all possible outcomes of an experiment.

Requirements:

- Outcomes must be **distinct** and **mutually exclusive**
- The sample space must be **collectively exhaustive**
- When the experiment is carried out, exactly one outcome occurs

Examples:

- Coin flip: $S = \{\text{Heads}, \text{Tails}\}$
- Roll a die: $S = \{1, 2, 3, 4, 5, 6\}$
- Two dice: $S = \{(1, 1), (1, 2), \ldots, (6, 6)\}$

[1] https://en.wikipedia.org/wiki/Sample_space

Events[2]

Definition (Event)
An **event** is a subset of the sample space S. We say A occurred if the actual outcome is in A.

Special cases:

- Single outcome: $\{\text{Heads}\}$
- Multiple outcomes: $\{2, 4, 6\}$ (rolling an even number)
- Entire sample space: S (always occurs)
- Empty set: $\emptyset$ (never occurs)

[2] https://en.wikipedia.org/wiki/Event_(probability_theory)

Set Operations for Events[4]

For events $A, B \subseteq S$:

- **Union** $A \cup B$: *A or B* occurs (or both)
- **Intersection** $A \cap B$: Both *A and B* occur
- **Complement** A^c: A does *not* occur

De Morgan's Laws[3]:

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

[3] https://en.wikipedia.org/wiki/De_Morgan's_laws
[4] https://en.wikipedia.org/wiki/Algebra_of_sets

Probability Law

A function $P : \mathcal{F} \to [0, 1]$ that assigns a probability to each event.

Discrete Probability Law
For a finite or countably infinite sample space, assign a probability mass $p(\omega)$ to each outcome ω such that:

$$p(\omega) \geq 0, \qquad \sum_{\omega \in \Omega} p(\omega) = 1.$$

Then for any event A,

$$P(A) = \sum_{\omega \in A} p(\omega).$$

Discrete Uniform Probability Law
If all outcomes are equally likely, $p(\omega) = 1/|\Omega|$, and

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

This leads to the naive definition of probability.

Naive Definition of Probability

Naive Definition of Probability[5]

Definition (Naive Probability)
For an event A in a finite sample space S:

$$P_{\text{naive}}(A) = \frac{|A|}{|S|} = \frac{\text{\# outcomes favorable to } A}{\text{total \# outcomes}}$$

Assumptions:

1. Finite sample space
2. All outcomes equally likely

[5] https://en.wikipedia.org/wiki/Classical_definition_of_probability

When is the Naive Definition Applicable?

When applicable:

- Symmetry (fair coin, well-shuffled deck)
- Equally likely by design (simple random sample)
- As a null model

Pitfall: Misapplying without justification

Bad reasoning: "Either it will happen or it won't, so probability is 1/2"

Example: Life on Mars

- Wrong: $P(\text{life}) = 1/2$, $P(\text{intelligent life}) = 1/2$
- Why wrong? These events shouldn't have equal probability

Leibniz's Mistake[6]

Problem: Roll two fair dice. Which is more likely: sum of 11 or sum of 12?

Leibniz's incorrect reasoning:

- Sum 11: {5,6} (one way)
- Sum 12: {6,6} (one way)
- Conclusion: Both equally likely

Correct solution:

- Label dice A and B
- Sum 11: (5,6), (6,5) → $P = 2/36 = 1/18$
- Sum 12: (6,6) → $P = 1/36$

Lesson: Always label objects to avoid treating them as indistinguishable!

[6] https://en.wikipedia.org/wiki/Gottfried_Wilhelm_Leibniz

Counting Methods

Multiplication Rule[7]

Theorem (Multiplication Rule)
If Experiment A has a outcomes, and for each of these, Experiment B has b outcomes, then the compound experiment has $a \cdot b$ outcomes.

Example: 10 runners for 1st, 2nd, 3rd place

$$10 \times 9 \times 8 = 720 \text{ possibilities}$$

Generalization: For k experiments with $n_1, n_2, \ldots, n_k$ outcomes:

$$n_1 \times n_2 \times \cdots \times n_k \text{ total outcomes}$$

[7] https://en.wikipedia.org/wiki/Rule_of_product

Permutations[8] (ordered arrangements):

- n distinct objects taken k at a time: $P(n,k) = n(n-1)\cdots(n-k+1)$
- All n objects: $n! = n(n-1)\cdots 2\cdot 1$

Combinations[9] (unordered selections):

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

Counts number of k-element subsets of an n-element set.

Example: $\binom{4}{2} = 6$ committees from 4 people

[8] https://en.wikipedia.org/wiki/Permutation
[9] https://en.wikipedia.org/wiki/Combination

Strategy: If we count each possibility exactly c times, divide by c

Example: Choose a 2-person committee from 4 people

- **Incorrect**: $4 \times 3 = 12$ (counts each committee twice: AB and BA)
- **Correct**: $(4 \times 3)/2 = 6$

Binomial Coefficient[10]:

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

Counts number of k-element subsets of an n-element set.

[10] `https://en.wikipedia.org/wiki/Binomial_coefficient`

Axiomatic Probability

Definition (Probability Space)
A probability space consists of:

- Sample space S
- Probability function P mapping events $A \subseteq S$ to $[0, 1]$

Axioms (Kolmogorov, 1933):

1. **Nonnegativity**: $P(A) \geq 0$ for all events A
2. **Normalization**: $P(S) = 1$
3. **Countable Additivity**: If $A_1, A_2, \dots$ are disjoint, then:

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

[11] https://en.wikipedia.org/wiki/Probability_axioms

Interpretations of Probability

- **Frequentist**[12]: long-run frequency
- **Bayesian**[13]: degree of belief
- **Pebble World**: mass distribution (total mass = 1)
- **Axiomatic**: mathematical structure satisfying the three axioms

All interpretations must satisfy the axioms to be mathematically consistent.

[12] https://en.wikipedia.org/wiki/Frequentist_probability
[13] https://en.wikipedia.org/wiki/Bayesian_probability

Theorem: For any events A and B:

1. $P(A^c) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. If $A \subseteq B$, then $P(A) \leq P(B)$ (monotonicity)
4. $P(A) \leq 1$ for all A
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Inclusion-Exclusion Principle[14]

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

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

For 3 events:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

[14] `https://en.wikipedia.org/wiki/Inclusion–exclusion_principle`

Story Proofs

What is a Story Proof?

A **story proof** proves a combinatorial identity by interpreting both sides as counting the same thing in different ways.

Advantages:

- More intuitive than algebra
- Provides insight into *why* identity holds
- Avoids messy calculations

Example 1: Symmetry of Binomial Coefficients

$$\binom{n}{k} = \binom{n}{n-k}$$

Story: Choosing k people for a committee = choosing $n-k$ people to *exclude*.

Both sides count the same thing: number of k-element subsets.

Vandermonde's Identity[15]

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j} \binom{n}{k-j}$$

Story: Choose k people from m juniors and n seniors.

Count directly: $\binom{m+n}{k}$

Count by cases: If we choose j juniors, we must choose $k-j$ seniors. Sum over all possible j:

$$\sum_{j=0}^{k} \binom{m}{j} \binom{n}{k-j}$$

Both count the same thing, so they must be equal.

[15] https://en.wikipedia.org/wiki/Vandermonde's_identity

Classic Probability Problems

Birthday Problem[16]

What's the probability that in a group of k people, at least two share a birthday?

Exact formula:

$$P(\text{no match}) = \frac{365 \times 364 \times \cdots \times (365 - k + 1)}{365^k}$$

Surprising results:

- With 23 people: $P(\text{match}) > 0.5$
- With 57 people: $P(\text{match}) > 0.99$

Intuition: Many possible pairs ($\binom{k}{2}$) increase collision probability.

[16] https://en.wikipedia.org/wiki/Birthday_problem

Birthday Problem: Step-by-Step Calculation

Problem: With $k = 23$ people, what's $P(\text{at least one birthday match})$?

Step 1: Calculate $P(\text{no match})$

$$P(\text{no match}) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{365-22}{365}$$

Step 2: Use product notation

$$P(\text{no match}) = \prod_{i=0}^{22} \frac{365-i}{365} = \frac{365 \times 364 \times \cdots \times 343}{365^{23}}$$

Step 3: Take complement

$$P(\text{match}) = 1 - P(\text{no match}) = 1 - \frac{365 \times 364 \times \cdots \times 343}{365^{23}}$$

Step 4: Numerical result

$$P(\text{match}) \approx 0.5073 \text{ (just over 50\%)}$$

Birthday Problem: Approximation Derivation

Approximation using Taylor series:

Step 1: Take natural log

$$\ln P(\text{all different}) = \sum_{i=0}^{k-1} \ln\left(1 - \frac{i}{365}\right)$$

Step 2: Use approximation $\ln(1-x) \approx -x$ for small x

$$\ln P(\text{all different}) \approx -\sum_{i=0}^{k-1} \frac{i}{365} = -\frac{1}{365} \cdot \frac{k(k-1)}{2}$$

Step 3: Exponentiate

$$P(\text{all different}) \approx e^{-\frac{k(k-1)}{2\times 365}}$$

Step 4: Complement

$$P(\text{match}) \approx 1 - e^{-\frac{k(k-1)}{730}}$$

For $k = 23$, $P(\text{match}) \approx 1 - e^{-0.693} \approx 1 - 0.500 = 0.500$

de Montmort's Matching Problem[17]

Problem: Shuffle n cards numbered 1 to n. Flip cards one by one, saying "1, 2, ..., n". Win if any card matches its number.

Solution: Let A_i = "card i is in position i"

Using inclusion-exclusion:

$$P(\text{win}) = P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}\frac{1}{n!}$$

For large n:

$$P(\text{win}) \approx 1 - \frac{1}{e} \approx 0.632$$

[17] https://en.wikipedia.org/wiki/Matching_problem

de Montmort's Problem: Step-by-Step for n=4

Problem: 4 cards numbered 1-4, shuffled. Win if any card matches its position.

Step 1: Define events Let $A_i =$ "card i is in position i" for $i = 1, 2, 3, 4$

Step 2: Use inclusion-exclusion

$$P(\text{win}) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - P(A_1 \cap A_2 \cap A_3 \cap A_4)$$

Step 3: Calculate and count terms

$$P(A_i) = \frac{1}{4} \quad \text{(4 terms)}$$

$$P(A_i \cap A_j) = \frac{1}{4 \times 3} = \frac{1}{12} \quad (\binom{4}{2} = 6 \text{ terms})$$

$$P(A_i \cap A_j \cap A_k) = \frac{1}{24} \quad \text{(4 terms)}$$

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{1}{24} \quad \text{(1 term)}$$

Step 4: Calculate

$$P(\text{win}) = 4 \times \frac{1}{4} - 6 \times \frac{1}{12} + 4 \times \frac{1}{24} - \frac{1}{24} = \frac{5}{8} = 0.625$$

Newton-Pepys Problem[18]

Which is most likely?

- A: At least one 6 in 6 dice
- B: At least two 6's in 12 dice
- C: At least three 6's in 18 dice

Isaac Newton correctly solved this problem in correspondence with Samuel Pepys (1693).

[18] https://en.wikipedia.org/wiki/Newton-Pepys_problem

Newton-Pepys Solution

Event A: At least one 6 in 6 dice

$$P(A) = 1 - \left(\frac{5}{6}\right)^6 \approx 0.6651$$

Event B: At least two 6's in 12 dice

$$P(B) = 1 - \left[\left(\frac{5}{6}\right)^{12} + \binom{12}{1}\left(\frac{5}{6}\right)^{11}\left(\frac{1}{6}\right)\right] \approx 0.6187$$

Event C: At least three 6's in 18 dice

$$P(C) = 1 - \left[\left(\frac{5}{6}\right)^{18} + \binom{18}{1}\left(\frac{5}{6}\right)^{17}\left(\frac{1}{6}\right) + \binom{18}{2}\left(\frac{5}{6}\right)^{16}\left(\frac{1}{6}\right)^2\right] \approx 0.5973$$

A has the highest probability.

Newton's insight: While expected number of 6's increases (1, 2, 3), the probability of meeting or exceeding these thresholds decreases.

Full House in Poker: A Counting Example

Problem: Probability of a full house (3 of one rank, 2 of another) in 5-card poker[19]

Solution:

1. Total hands: $\binom{52}{5} = 2,598,960$
2. Choose rank for triple: 13 choices
3. Choose 3 suits for triple: $\binom{4}{3} = 4$ choices
4. Choose different rank for pair: 12 choices
5. Choose 2 suits for pair: $\binom{4}{2} = 6$ choices

$$P(\text{full house}) = \frac{13 \times 4 \times 12 \times 6}{\binom{52}{5}} = \frac{3,744}{2,598,960} \approx 0.00144$$

[19] https://en.wikipedia.org/wiki/Poker_probability

Full House Probability: Step-by-Step Calculation

Step 1: Total number of 5-card hands

$$\binom{52}{5} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2,598,960$$

Step 2: Choose rank for triple (13), suits for triple ($\binom{4}{3} = 4$)

$$13 \times 4 = 52$$

Step 3: Choose different rank for pair (12), suits for pair ($\binom{4}{2} = 6$)

$$12 \times 6 = 72$$

Step 4: Apply multiplication rule

$$52 \times 72 = 3,744 \text{ possible full house hands}$$

Step 5: Calculate probability

$$P(\text{full house}) = \frac{3,744}{2,598,960} \approx 0.0014406$$

Step 6: Interpretation

$$\approx 0.144\% \text{ chance, or about 1 in 694 hands}$$

Poker Hand Probabilities: Comparison Table

Calculate probabilities for various poker hands:

Hand	Probability
Royal Flush	$4/2,598,960 \approx 0.00000154$
Straight Flush	$36/2,598,960 \approx 0.0000139$
Four of a Kind	$624/2,598,960 \approx 0.000240$
Full House	$3,744/2,598,960 \approx 0.00144$
Flush	$5,108/2,598,960 \approx 0.00197$
Straight	$10,200/2,598,960 \approx 0.00392$
Three of a Kind	$54,912/2,598,960 \approx 0.0211$
Two Pair	$123,552/2,598,960 \approx 0.0475$
One Pair	$1,098,240/2,598,960 \approx 0.423$
High Card	$1,302,540/2,598,960 \approx 0.501$

Poker Hand Calculation: Three of a Kind

Step-by-step calculation for **Three of a Kind**:

Step 1: Choose rank for the triple: 13 choices

Step 2: Choose 3 suits for the triple: $\binom{4}{3} = 4$ choices

Step 3: Choose 2 other ranks (different from triple rank and from each other)

$$\binom{12}{2} = \frac{12 \times 11}{2} = 66 \text{ choices}$$

Step 4: Choose suits for the 2 other cards

$$4 \times 4 = 16 \text{ choices (each can be any of 4 suits)}$$

Step 5: Apply multiplication rule

$$13 \times 4 \times 66 \times 16 = 54,912$$

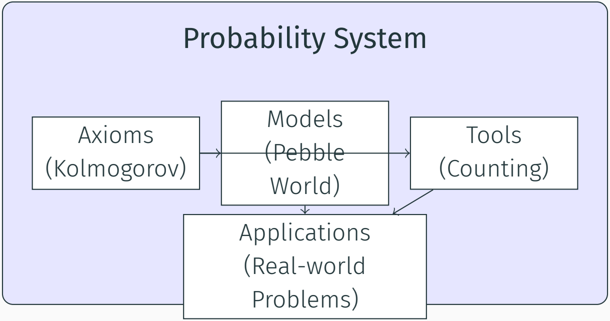
Step 6: Calculate probability

$$\frac{54,912}{2,598,960} \approx 0.02112845 \text{ (about 1 in 47.3 hands)}$$

Recap and Analogy

1. **Sample space** S: Set of all possible outcomes
2. **Event**: Subset of S
3. **Naive probability**: $P(A) = |A|/|S|$ (requires equally likely outcomes)
4. **Counting tools**:
   - Multiplication rule
   - Permutations: $n!$, $n(n-1)\cdots(n-k+1)$
   - Combinations: $\binom{n}{k}$
   - Adjusting for overcounting
5. **Story proofs**: Interpret both sides of identity
6. **Axiomatic probability**: Nonnegativity, Normalization, Additivity
7. **Properties**: Complements, monotonicity, inclusion-exclusion

Probability as a Coherent System



Foundation: Axioms provide rigorous mathematical basis for all probability theory.

References

References and Further Reading

- Blitzstein, J. K., & Hwang, J. (2019). *Introduction to Probability* (2nd ed.). CRC Press.
- Wikipedia: Probability
- Wikipedia: Combinatorics
- Wikipedia: Andrey Kolmogorov
- Stanford Encyclopedia of Philosophy: Interpretations of Probability
- **Probability axioms**: `https://en.wikipedia.org/wiki/Probability_axioms`
- **Naive definition**: Chapter 1 of "Introduction to Probability" by Blitzstein and Hwang
- **Birthday problem**: `https://en.wikipedia.org/wiki/Birthday_problem`
- **de Montmort's problem**: Classic probability texts
- **Newton–Pepys problem**: Historical correspondence
- **Poker probabilities**: Standard combinatorics sources