Conditional Probability

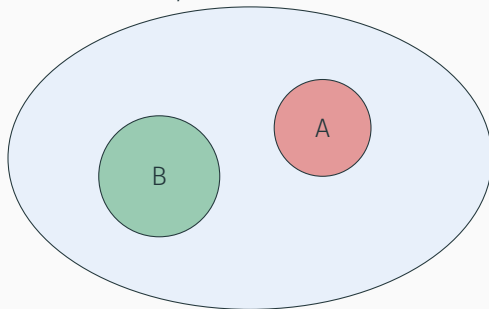## Theory and Biological Applications

DRME

2025

Foundations of Probability

The Logic of Uncertainty

- **Mathematics** is the logic of certainty
- **Probability** is the logic of uncertainty
- Provides framework for quantifying doubt and updating beliefs

Sample Space and Events

Ω: All possible outcomes

- Ω: Set of all possible outcomes
- Events: Subsets of Ω (e.g., "has disease allele")
- Pebble World: Each "pebble" = one possible outcome

Basic Probability Rules

Naive Definition (Equal Likelihood)
$$P(A) = \frac{|A|}{|\Omega|}$$

Caution: Requires equally likely outcomes

Counting Rules

- Multiplication Rule: $n_1 \times n_2 \times \cdots \times n_r$
- Combinations: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Conditional Probability: Definition and Intuition

Core Insight
All probabilities are conditional on available information

Biological Context: Genetic inheritance, disease risk, evolutionary relationships

Two Roles of Conditioning:
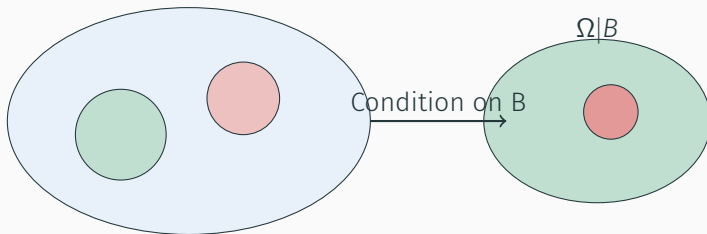
1. Updating beliefs with new evidence
2. Problem-solving strategy (break complex problems into simpler pieces)

Definition (Conditional Probability)
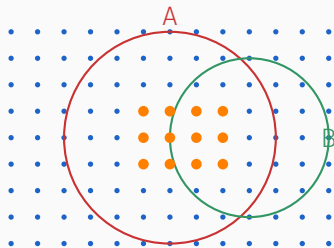If $P(B) > 0$, the conditional probability of A given B is:

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

$\Omega|B$

Condition on B

Three Step Process:

1. Start with all pebbles (mass = 1)
2. Discard pebbles not in B (mass now = $P(B)$)
3. Renormalize: divide all masses by $P(B)$ so total mass = 1

Example: Two Cards Problem

Problem: Draw two cards without replacement from a standard deck.

- A: First card is a heart
- B: Second card is red

Solution:

$$P(A \cap B) = \frac{13 \times 25}{52 \times 51} = \frac{25}{204}$$

$$P(A|B) = \frac{25/204}{1/2} = \frac{25}{102} \approx 0.2451$$

$$P(B|A) = \frac{25/204}{1/4} = \frac{25}{51} \approx 0.4902$$

Key Observation: $P(A|B) \neq P(B|A)$

The Two Children Problem

Case 1: Mr. Jones has two children. The **older** child is a girl. What is P(both girls)?

Case 2: Mr. Smith has two children. **At least one** is a boy. What is P(both boys)?

Solutions:

- Case 1: P(both girls|elder is girl) $= \frac{1/4}{1/2} = \frac{1}{2}$
- Case 2: P(both boys|at least one boy) $= \frac{1/4}{3/4} = \frac{1}{3}$

Why different? Different information restricts the sample space differently.

Bayes' Rule and Law of Total Probability

The Multiplication Rule

Joint Probability Factorization
$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$
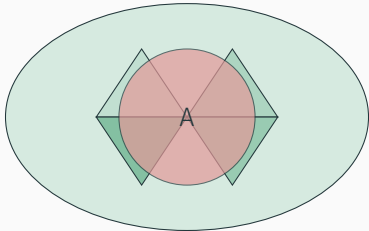
Genetic Application: Compound Heterozygotes
- Event A: Inherit mutant allele from mother
- Event B: Inherit mutant allele from father
- $P(\text{disease}) = P(A \cap B) = P(A|B)P(B)$
- For recessive disorders: $P(A|B) = P(A)$ if parents unrelated

Theorem (Law of Total Probability)
If $B_1, B_2, \ldots, B_n$ partition Ω, then:

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

Bayes' Rule: Derivation and Statement

Theorem (Bayes' Rule)
For events A and B with $P(B) > 0$:

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

Derivation:

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \quad \text{(definition)} \\
&= \frac{P(B|A)P(A)}{P(B)} \quad \text{(multiplication rule)}
\end{aligned}
$$

Terminology:

- $P(A)$: Prior probability (before evidence)
- $P(A|B)$: Posterior probability (after evidence)
- $P(B|A)$: Likelihood of evidence if hypothesis true

Example: Rare Disease Testing

Setup:

- Disease prevalence:
  $P(D) = 0.001$
- Test sensitivity: $P(+|D) = 0.99$
- Test specificity:
  $P(-|\neg D) = 0.99$
- Question: $P(D|+)$ = ?

Solution:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)}$$
$$= \frac{0.99 \times 0.001}{0.0095 + 0.0495}$$
$$\approx 0.161$$

Population

Diseased (0.001)   Healthy (0.999)

+ (0.99)   - (0.01)   + (0.01)   - (0.99)

Key Insight

Only 16% chance of having disease despite positive test! False positives outnumber true positives.

Example: Random Coin Problem

Problem: You have a fair coin ($P(H) = 1/2$) and a biased coin ($P(H) = 3/4$). Pick one at random (50-50), flip it 3 times, get HHH. What is $P(\text{fair}|HHH)$?

Solution:

$$\begin{aligned} P(F|HHH) &= \frac{P(HHH|F)P(F)}{P(HHH|F)P(F) + P(HHH|F^c)P(F^c)} \\ &= \frac{(1/8)(1/2)}{(1/8)(1/2) + (27/64)(1/2)} \\ &= \frac{8}{35} \approx 0.2286 \end{aligned}$$

Interpretation: After seeing HHH, probability it's the fair coin drops from 0.5 to about 0.23.

Extended Forms

Bayes' Rule with Extra Conditioning
$$P(A|B, E) = \frac{P(B|A, E)P(A|E)}{P(B|E)}$$

LOTP with Extra Conditioning
$$P(B|E) = \sum_{i=1}^{n} P(B|A_i, E)P(A_i|E)$$

for partition $A_1, \ldots, A_n$.

Application: Sequential updating of beliefs

- Today's posterior becomes tomorrow's prior
- Coherency: Sequential updating gives same result as updating with all evidence at once

Conditional Probability as Probability

Conditional Probability Satisfies Probability Axioms

Theorem: For fixed E with $P(E) > 0$, $\tilde{P}(A) = P(A|E)$ satisfies:

1. **Non-negativity**: $\tilde{P}(A) \geq 0$ for all A
2. **Normalization**: $\tilde{P}(\Omega) = 1$
3. **Additivity**: For disjoint $A_1, A_2, \ldots$:

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

Proof of additivity:

$$\tilde{P}\left(\bigcup_i A_i\right) = \frac{P\left((\bigcup_i A_i) \cap E\right)}{P(E)}$$

$$= \frac{P\left(\bigcup_i (A_i \cap E)\right)}{P(E)}$$

$$= \frac{\sum_i P(A_i \cap E)}{P(E)} = \sum_i \tilde{P}(A_i)$$

All Probabilities are Conditional

Fundamental Principle
There are no unconditional probabilities. All probabilities depend on background information *K*.

Example: *P*(Rain today) implicitly conditions on:

- Location, season, time of day
- Current weather conditions
- Meteorological models

Mathematical notation:

- $P(A)$ is shorthand for $P(A|K)$ where K is background knowledge
- The vertical bar is always there, even if not written

Biological Context:

- Mutation rates: Conditional on sequence context
- Disease risk: Conditional on genetic background, environment
- Evolutionary probabilities: Conditional on population size, selection

Independence

Definition of Independence

Definition (Independence)
Events *A* and *B* are **independent** if:

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

If $P(A) > 0$ and $P(B) > 0$, equivalent to:

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B)$$

Interpretation: *A* provides no information about *B*, and vice versa.

Genetic Independence

- **Unlinked loci**: Inheritance of alleles at different chromosomes
- **Hardy-Weinberg equilibrium**: Random mating assumption
- **Linkage equilibrium**: No association between alleles at different loci

Properties of Independence

Theorem: If *A* and *B* are independent, then:

1. *A* and B^c are independent
2. A^c and *B* are independent
3. A^c and B^c are independent

Caution: Independence $\neq$ Disjointness!

- Disjoint: $P(A \cap B) = 0$
- Independent: $P(A \cap B) = P(A)P(B)$
- Can only be both if $P(A) = 0$ or $P(B) = 0$

Independence of Three Events

Definition (Independence of Three Events)
Events A, B, C are **independent** if:

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

Important: Need *all four* conditions!

Pairwise Independence $\neq$ Independence:

- Pairwise: first three conditions hold
- Independence: all four conditions hold

Example: Pairwise Independence ≠ Independence

Setup: Two fair, independent coin tosses

- A: First toss is H
- B: Second toss is H
- C: Both tosses have same result

Check pairwise independence:

$$P(A) = P(B) = P(C) = 1/2$$
$$P(A \cap B) = P(HH) = 1/4 = P(A)P(B)$$
$$P(A \cap C) = P(HH) = 1/4 = P(A)P(C)$$
$$P(B \cap C) = P(HH) = 1/4 = P(B)P(C)$$

So A, B, C are pairwise independent.

Check triple independence:

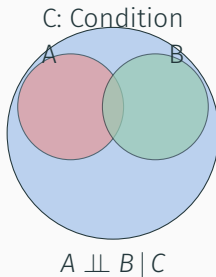$$P(A \cap B \cap C) = P(HH) = 1/4 \neq P(A)P(B)P(C) = 1/8$$

So NOT independent!

Conditional Independence

Definition (Conditional Independence)
Events *A* and *B* are **conditionally independent given** *E* if:

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

C: Condition

$A \perp\!\!\!\perp B \mid C$

Caution: Independence $\neq$ Conditional Independence!

- *A*, *B* can be independent but not conditionally independent given *E*

Problem-Solving Strategies

Strategy: Condition on What You Wish You Knew

Approach:

1. Identify the unknown quantity
2. Determine what information would make the problem trivial
3. Condition on that information
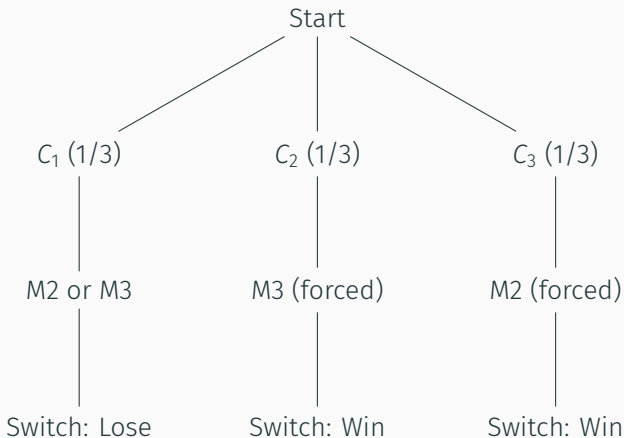4. Use Law of Total Probability to combine cases

Monty Hall Problem

- 3 doors: 1 car, 2 goats
- You pick door 1
- Monty opens a door with a goat
- Should you switch?

Wish: I wish I knew where the car was!
Solution: $P(\text{win by switching}) = 2/3$

Monty Hall Problem: Tree Diagram

Probability: $\frac{2}{3}$ chance of winning by switching

Strategy: First-Step Analysis

Method:

1. Consider the first event/step in a process
2. Condition on possible outcomes of first step
3. Set up recursive equations
4. Solve the system

Amoeba Problem

- Single amoeba
- After 1 minute: die (1/3), stay same (1/3), split into two (1/3)
- All future amoebas behave independently
- $p = P$(population eventually dies out)

First-step analysis: $p = \frac{1}{3} + \frac{1}{3}p + \frac{1}{3}p^2$ **Solution**: $p = 1$ (certain extinction)

Gambler's Ruin: First-Step Analysis

Problem: Two gamblers A and B

- A starts with \$i, B with \$(N-i)
- Each bet: A wins with probability p, loses with probability $q = 1 - p$
- Game ends when someone has all \$N
- $p_i = P(\text{A wins}|\text{starts with } i)$

First-step analysis:

$$p_i = p \cdot p_{i+1} + q \cdot p_{i-1}$$

with boundary conditions $p_0 = 0$, $p_N = 1$.

Solution:

$$p_i = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)^N} & \text{if } p \neq 1/2 \\ \frac{i}{N} & \text{if } p = 1/2 \end{cases}$$

Genetics Applications

Example: Two Cards Problem (Genetics Version)

Genetic Counseling Scenario:

- Two alleles: N (normal), M (mutant)
- Parent genotype: NM (carrier)
- First child tested: Has mutation (M)
- Question: Probability second child has mutation?

First child

	M	
M	NM MM	MM
N	NN	NM

Solution:

$$P(\text{2nd M|1st M}) = \frac{P(\text{both M})}{P(\text{1st M})}$$
$$= \frac{(1/2)^2}{1/2} = \frac{1}{2}$$

Key Insight
Knowing first child's status doesn't change probability for second child (Mendelian inheritance)

Strategy: Condition on What You Wish You Knew (Genetics)

Evolutionary Example: Coalescent Theory

- **Unknown**: Time to most recent common ancestor (TMRCA)
- **Wish I knew**: Number of generations since divergence
- **Condition on**: Population size, mutation rate
- **Result**: $E[T] = \sum_t P(T > t)$ conditioned on demographic history

Population Genetics Example

- Gene frequency in next generation
- Condition on current frequency
- Wright-Fisher model: $X_{t+1}|X_t \sim$ Binomial

Pitfalls and Paradoxes

The Prosecutor's Fallacy

Real Case: Sally Clark (1998) - two infant deaths

- Expert: $P(\text{SIDS}) = 1/8500$
- So $P(\text{two SIDS}) = (1/8500)^2 \approx 1/73$ million
- Conclusion: Probability of innocence is 1 in 73 million

Two Major Errors:

1. Assumed independence (SIDS might run in families)
2. Confused $P(\text{evidence}|\text{innocent})$ with $P(\text{innocent}|\text{evidence})$

Correct by Bayes':

$$P(\text{innocent}|\text{evidence}) = \frac{P(\text{evidence}|\text{innocent})P(\text{innocent})}{P(\text{evidence})}$$

$P(\text{innocent})$ is very high (double infanticide is also rare!)

The Defense Attorney's Fallacy

Scenario: Murder case, husband accused

- Evidence: Husband abused wife
- Defense: Only 1 in 10,000 abusive husbands murder their wives
- So $P(\text{guilty}|\text{abuse}) = 0.0001$, evidence irrelevant

Error: Fails to condition on all evidence!

Correct: Should compute $P(\text{guilty}|\text{abuse}, \text{wife murdered})$

Given:

$$P(\text{abuse}) = 0.1$$
$$P(\text{guilty}|\text{murder}) = 0.2$$
$$P(\text{abuse}|\text{guilty}, \text{murder}) = 0.5$$
$$P(\text{abuse}|\text{innocent}, \text{murder}) = 0.1$$

Then:

$$P(\text{guilty}|\text{abuse}, \text{murder}) = \frac{0.5 \times 0.2}{0.5 \times 0.2 + 0.1 \times 0.8} = \frac{5}{9} \approx 0.56$$

Simpson's Paradox

Subpopulation	Drug A Success	Total	Drug B Success	Total
Severe cases	30/100 (30%)	100	210/700 (30%)	700
Mild cases	90/100 (90%)	100	10/100 (10%)	100
Combined	**120/200 (60%)**	200	**220/800 (27.5%)**	800

Paradox:

- Drug A better for both severe and mild cases
- But Drug B has higher overall success rate

Why? Drug A used more on severe cases (harder to treat)

Understanding Simpson's Paradox Mathematically

Let:

- A: Successful treatment
- B: Drug B is used
- C: Severe case

We have:

$$P(A|C, B) < P(A|C, B^c) \quad (30\% < 30\%)$$
$$P(A|C^c, B) < P(A|C^c, B^c) \quad (10\% < 90\%)$$

But:

$$P(A|B) > P(A|B^c) \quad (27.5\% > 60\%)$$

Understanding Simpson's Paradox Mathematically

Reason: Different weights:

$$P(A|B) = P(A|C, B)P(C|B) + P(A|C^c, B)P(C^c|B)$$
$$= 0.3 \times 0.875 + 0.1 \times 0.125 = 0.275$$
$$P(A|B^c) = P(A|C, B^c)P(C|B^c) + P(A|C^c, B^c)P(C^c|B^c)$$
$$= 0.3 \times 0.5 + 0.9 \times 0.5 = 0.6$$

Drug B's average weights easy cases more heavily!

Summary and Key Formulas

Chapter 2: Key Formulas

1. **Definition**: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(B) > 0$
2. **Bayes' Rule**: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
3. **LOTP**: $P(B) = \sum_i P(B|A_i)P(A_i)$ for partition $\{A_i\}$
4. **Extended forms**:
   - $P(A|B, E) = \frac{P(B|A,E)P(A|E)}{P(B|E)}$
   - $P(B|E) = \sum_i P(B|A_i, E)P(A_i|E)$
5. **Independence**: $P(A \cap B) = P(A)P(B)$
6. **Conditional Independence**: $P(A \cap B|E) = P(A|E)P(B|E)$
7. **Odds form of Bayes'**:

$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(B|A)}{P(B|A^c)} \cdot \frac{P(A)}{P(A^c)}$$

Problem-Solving Strategies

1. **Condition on what you wish you knew**
   - Monty Hall: Wish you knew where car is
   - Disease testing: Wish you knew true disease status

2. **First-step analysis**
   - Amoeba population: Condition on first minute
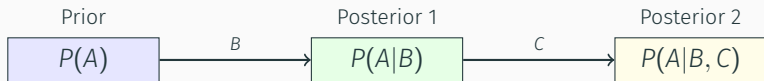   - Gambler's ruin: Condition on first bet

3. **Use Bayes' rule when $P(B|A)$ easier than $P(A|B)$**

4. **Use LOTP to break complex $P(B)$ into simpler conditional pieces**

Common Pitfalls to Avoid:

1. Prosecutor's fallacy: Confusing $P(A|B)$ with $P(B|A)$
2. Defense attorney's fallacy: Not conditioning on all evidence
3. Assuming independence without justification
4. Misinterpreting Simpson's paradox

Prior		Posterior 1		Posterior 2
$P(A)$	$\xrightarrow{B}$	$P(A\|B)$	$\xrightarrow{C}$	$P(A\|B, C)$

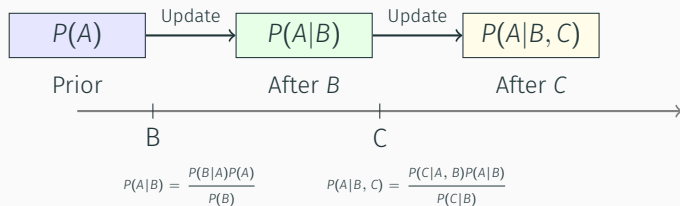**Step 1: Update with B** $\qquad \Rightarrow \qquad$ **Step 2: Update with C**

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

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

Key Insight: Each posterior becomes the prior for the next update

$P(A)$ →Update $P(A|B)$ →Update $P(A|B, C)$

Prior   After B After C

B   C

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad P(A|B, C) = \frac{P(C|A, B)P(A|B)}{P(C|B)}$$

Sequential Updating of Beliefs

Key Insight: Today's posterior becomes tomorrow's prior.

Coherency: Sequential updating gives same result as updating with all evidence at once:

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

where $P(A|B)$ is the "new prior" after observing B.

Conditioning is not just a technique

It's the fundamental way we incorporate information

Before Conditioning
Uncertain world

Conditioning Process
Zoom in on evidence

After Conditioning
Refined understanding

- Blitzstein & Hwang, *Introduction to Probability*