Day 2: Conditional Probability and Independence

Updating Beliefs and Simplifying Reasoning in AI Systems

1. The Fundamental Question

How do we update our beliefs when new information arrives, and when can we simplify complex probabilistic relationships?

Conditional probability provides the mathematical framework for updating beliefs, while independence allows us to reason efficiently about complex systems by identifying when events don't influence each other.

2. Mathematical Foundations

2.1. Conditional Probability Definition

For events A and B with $P(B) > 0$:

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

2.2. Independence Definition

Events A and B are independent if and only if:

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

Equivalently: $P(A|B) = P(A)$ and $P(B|A) = P(B)$

Key Properties and Derived Rules

- **Multiplication Rule:** $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$
- **Chain Rule:** $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots$
- **Conditional Independence:** A and B are conditionally independent given C if $P(A \cap B|C) = P(A|C)P(B|C)$
- **Law of Total Probability:** $P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$ for partition $\{B_1, \ldots, B_n\}$

3. Numerical Example: Dice Roll Analysis

Problem Setup

Fair six-sided die: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $P(\{i\}) = \frac{1}{6}$
Let:

- $A = \{\text{even outcome}\} = \{2, 4, 6\}$
- $B = \{\text{outcome ¿ 3}\} = \{4, 5, 6\}$

3.1. Basic Probability Calculations

$$P(A) = \frac{3}{6} = 0.5, \quad P(B) = \frac{3}{6} = 0.5, \quad P(A \cap B) = P(\{4, 6\}) = \frac{2}{6} \approx 0.333$$

3.2. Conditional Probability Analysis

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/6}{3/6} = \frac{2}{3} \approx 0.667$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{2/6}{3/6} = \frac{2}{3} \approx 0.667$$

3.3. Independence Testing

$$P(A)P(B) = 0.5 \times 0.5 = 0.25$$
$$P(A \cap B) = 0.333 \neq 0.25$$

Conclusion: A and B are NOT independent.

Statistical Interpretation

- Knowing the outcome is ¿3 increases probability of even number from 50% to 66.7%
- The 16.7% increase indicates positive correlation between events
- In AI terms: Evidence B provides information about A

4. Why This Matters in AI Systems

Critical Applications in Machine Learning

- **Naive Bayes Classifiers:** Rely on conditional independence assumptions
- **Bayesian Networks:** Represent complex conditional dependencies
- **Hidden Markov Models:** Use conditional probabilities for state transitions **Reinforcement Learning:** Policy evaluation uses conditional state transitions
- **Computer Vision:** Conditional probabilities for object recognition given features

5. Real-World AI Examples

5.1. Spam Detection with Naive Bayes

- Assume word occurrences are conditionally independent given spam/not-spam
- $P(\text{spam}|w_1, \ldots, w_n) \propto P(\text{spam}) \prod_{i=1}^{n} P(w_i|\text{spam})$
- Achieves 95%+ accuracy despite independence assumption

5.2. Medical Diagnosis Systems

- $P(\text{disease}|\text{symptoms}) = \frac{P(\text{symptoms}|\text{disease})P(\text{disease})}{P(\text{symptoms})}$
- Updates diagnosis probability as new test results arrive
- Handles conditional dependencies between symptoms

5.3. Autonomous Vehicles

- $P(\text{collision}|\text{sensor readings})$ updates in real-time
- Conditional independence between different sensor modalities
- Enables rapid risk assessment and decision making

6. Independence Assumptions: When and Why They Work

Practical Guidelines for Independence

- **Use independence:** When features are approximately uncorrelated
- **Check independence:** Calculate correlations or mutual information
- **Conditional independence:** Often more realistic than full independence
- **Performance impact:** False independence assumptions can reduce accuracy by 5-15%

7. Implementation in Code

7.1. Python Example: Conditional Probability Calculation

```
def conditional_probability(p_a_and_b, p_b):
    """Calculate P(A|B) = P(AB)/P(B)"""
    if p_b == 0:
        return 0  # Handle division by zero
    return p_a_and_b / p_b

def check_independence(p_a, p_b, p_a_and_b, tolerance=1e-6):
    """Check if P(AB)  P(A)P(B)"""
    return abs(p_a_and_b - p_a * p_b) < tolerance
```


7.2. Empirical Verification

- Test independence with statistical tests (chi-square, p-values)
- Use cross-validation to validate conditional independence assumptions
- Monitor model performance when relaxing independence assumptions

8. Advanced Topics and Extensions

8.1. Conditional Independence in Graphical Models

$$X \perp Y | Z \iff P(X, Y | Z) = P(X|Z)P(Y|Z)$$

Essential for efficient inference in Bayesian networks.

8.2. d-separation

Graphical criterion for determining conditional independence in directed graphs.

8.3. Markov Blankets

The minimal set of nodes that renders a node conditionally independent of all others.

9. Practical Exercises for Mastery

Hands-On Practice Problems

1. A deck of cards: $P(\text{heart}|\text{face card}) = ?$
2. Two coin tosses: Are "first heads" and "second heads" independent?
3. Given $P(A) = 0.4, P(B) = 0.5, P(A \cap B) = 0.2$:
   - Compute $P(A|B)$ and $P(B|A)$
   - Test for independence
   - Calculate the degree of dependence
4. Implement a simple Naive Bayes classifier using conditional probabilities
5. Analyze a real dataset: Test independence assumptions between features

10. Common Pitfalls and Debugging Strategies

Troubleshooting Conditional Probability Models

- **Zero Probability:** Handle $P(B) = 0$ with Laplace smoothing **False Independence:** Validate assumptions with correlation analysis
- **Numerical Stability:** Use log probabilities for small values
- **Conditioning on Colliders:** Avoid introducing spurious dependencies
- **Sample Size:** Ensure sufficient data for reliable conditional probability estimates

11. Performance Metrics and Validation

11.1. Model Quality Assessment

- Log-likelihood on validation data
- Calibration curves for probability estimates
- Conditional independence tests (conditional mutual information)
- Cross-validation performance with/without independence assumptions

11.2. Empirical Results

- Naive Bayes: 85-95% accuracy despite strong independence assumptions
- Relaxing independence: Typically improves accuracy by 2-8%
- Computational cost: Full dependency models can be 10-100x slower

12. Key Insight: Intelligent Simplification

Conditional probability and independence are not just mathematical concepts—they are engineering tools that enable us to build scalable, efficient AI systems. The art lies in knowing when to assume independence for computational efficiency versus when to model dependencies for accuracy.

- **Trade-off:** Independence assumptions vs. model accuracy
- **Practical wisdom:** Start with independence, then relax if needed
- **Validation:** Always test assumptions with real data
- **Interpretability:** Conditional probabilities provide transparent reasoning

Next: Bayes' Theorem and Bayesian Inference

Tomorrow we'll explore how to combine prior knowledge with new evidence—the mathematical foundation of learning from data.