

Lecture 10: Reasoning Under Uncertainty

From Logic to Probability: The Paradigm Shift in AI

Professor Anis Koubaa

SE 444
Alfaisal University

December 1, 2025

Outline

- 1 Two Perspectives on Intelligence
- 2 Limits of Logic
- 3 The Uncertainty Paradigm
- 4 Probability Basics
- 5 Bayes' Rule
- 6 Independence & Distributions
- 7 Statistical Theorems & Bayesian Inference
- 8 Conclusion

Opening: The Twin Pillars of Intelligence

Philosophical Foundation

Humanity has long sought to formalize intelligence through two contrasting visions:

First Tradition

Certainty through **logic**

- Deduction from axioms
- Binary truth values
- Mathematical proof
- Complete knowledge

Second Tradition

Confidence through **probability**

- Induction from observations
- Degrees of belief $[0,1]$
- Statistical inference
- Partial knowledge

Key Question

Which vision better captures real-world intelligence?

The Knowledge vs. Faith Analogy

An Insightful Parallel

Human Inquiry \leftrightarrow AI

Human	AI Analogue
Knowledge Absolute truth	Logic-Based AI Certainty via symbolic reasoning, axioms, proof Examples: Theorem provers, expert systems
Faith Trust + evidence	Probabilistic AI Partial knowledge; rational action under uncertainty Examples: Bayesian networks, ML, robotics

Clarification

“Faith” = **rational acceptance of uncertainty**, not blind belief!

Two Schools of AI

School 1: Logician/Symbolic

"Intelligence is reasoning with certainty"

Philosophers:

- Aristotle (logic)
- Descartes (rationalism)
- Boole (Boolean algebra)
- Russell & Whitehead

AI Pioneers:

- John McCarthy
- Marvin Minsky
- Herbert Simon & Allen Newell

Methods:

- First-order logic
- Expert systems
- Theorem proving

School 2: Probabilistic

"Intelligence is reasoning under uncertainty"

Philosophers:

- Hume (induction)
- Bayes (probability)
- Laplace (probability theory)
- de Finetti

AI Pioneers:

- Judea Pearl
- Stuart Russell
- Sebastian Thrun
- Michael Jordan

Methods:

- Bayesian networks
- Machine learning

AI's Journey: From Certainty to Probability

Era	Development
1950s-60s	Logic Era: Logic Theorist, GPS, Lisp; Symbolic manipulation
1970s-80s	Expert Systems: MYCIN, DENDRAL; Struggled with uncertainty
1980s-90s	Probabilistic Revolution: Bayesian Networks (Pearl, 1988); Paradigm shift
2000s+	ML Era: Deep learning, probabilistic programming; Modern AI is probabilistic

The Trend

AI has moved from **deterministic logic** to **probabilistic reasoning**

Complementary, Not Competing

*“Symbolic AI pursues truth. Probabilistic AI pursues confidence.
Modern AI seeks to combine both.”*

When Logic Excels:

- Complete information
- Deterministic environments
- Formal verification needed
- Guaranteed correctness required

Example: Software verification

When Probability Excels:

- Incomplete information
- Noisy sensors
- Stochastic environments
- Learning from data

Example: Robotics, vision, NLP

Logic assumes a world it does not inhabit

Logic's Requirements

- **Complete** knowledge
- **Perfect** information about current state
- **Deterministic** rules with no exceptions
- **Exact** truth values (true/false only)

Reality Check

The real world is:

- **Partial** – never know everything
- **Uncertain** – noisy sensors, incomplete observations
- **Stochastic** – probabilistic outcomes
- **Continuous** – degrees of truth

The Qualification Problem

Definition

The impossibility of listing all preconditions for an action to succeed.

Example: “If I go outside, I will get wet if raining”

Logical Rule: $\text{Outside} \wedge \text{Raining} \rightarrow \text{Wet}$

But what if...

- Umbrella? Covered walkway? Rain stops? Raincoat? Car? Too light?

The Problem

Exhaustively listing all conditions is impossible! Logic breaks down.

Types of Uncertainty in AI

- 1 **Laziness** – Too many conditions to list explicitly
 - Getting wet: umbrella? raincoat? covered? etc.
- 2 **Theoretical Ignorance** – No complete theory exists
 - Medical diagnosis: diseases with unknown mechanisms
- 3 **Practical Ignorance** – Information not accessible
 - Traffic conditions ahead, stock market tomorrow
- 4 **Sensor Noise** – Observations are imperfect
 - Robot sensors, medical tests, image recognition
- 5 **Adversarial** – Opponents' actions are unpredictable
 - Games, security, competitive markets

Solution

Move from **true/false** to **degrees of belief**: **Probability**

The Paradigm Shift

From Knowing to Believing

The Old Paradigm: Logic

- Facts and rules
- Deduce with certainty
- Works: complete info, deterministic world

The New Paradigm: Probability

- **Degrees of belief**
- Update as **evidence accumulates**
- Works: incomplete info, uncertain world (**reality**)

Core Principle

Rational agents maximize expected utility under uncertainty

Subjective vs. Objective Probability

Frequentist (Objective)

Probability = long-run frequency

$$P(\text{Heads}) = \lim_{n \rightarrow \infty} \frac{\text{count}(\text{Heads})}{n}$$

"If we flip this coin infinitely many times, half will be heads"

Problems:

- What about one-time events?
- "Probability AI will be invented in 2050"?

Bayesian (Subjective)

Probability = degree of belief

Based on current knowledge and evidence

"Based on what I know, I believe there's a 70% chance of rain tomorrow"

Advantages:

- Applies to any proposition
- Updates with new evidence (Bayes' Rule)
- Rational framework for decisions

AI Uses Bayesian Interpretation

Quantifies uncertainty as rational degrees of belief, updateable with evidence

Probability Fundamentals

Sample Space (Ω)

Set of all possible outcomes

Example: $\Omega = \{H, T\}$ for a coin flip

Event (A)

Subset of the sample space

Example: $A = \{H\}$ (“coin lands heads”)

Probability Function P

Maps events to real numbers $[0, 1]$

The Three Axioms of Probability

Kolmogorov's Axioms (1933)

Axiom 1: Non-negativity

$$P(A) \geq 0 \quad \text{for all events } A$$

Probabilities cannot be negative

Axiom 2: Normalization

$$P(\Omega) = 1$$

Something must happen

Axiom 3: Additivity

$$P(A \cup B) = P(A) + P(B) \quad \text{if } A \cap B = \emptyset$$

Probabilities of disjoint events add up

Joint, Conditional, and Marginal Probabilities

Joint Probability

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

Probability that both A and B occur

Conditional Probability

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

Probability of A given that B occurred

Marginal Probability

$$P(A) = \sum_b P(A, B = b)$$

Probability of A regardless of B

The Product Rule (Chain Rule)

Derivation from Conditional Probability

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

Rearranging:

$$P(A, B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

General Chain Rule

$$P(X_1, X_2, \dots, X_n) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_3|X_1, X_2) \cdots P(X_n|X_1, \dots, X_{n-1})$$

Intuition: Joint probability = product of conditional probabilities

Example: Medical Diagnosis

Scenario: Patient has disease D and symptom S

Given Information:

- $P(D) = 0.01$ (1% of people have the disease)
- $P(S|D) = 0.9$ (90% of sick people show the symptom)
- $P(S|\neg D) = 0.2$ (20% of healthy people show the symptom)

Question: If someone has the symptom, what's $P(D|S)$?

Need Bayes' Rule!

We'll answer this in the next section

Bayes' Theorem

The Most Important Formula in AI

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

Where:

- $P(A|B)$ = **Posterior** (what we want to know)
- $P(B|A)$ = **Likelihood** (how likely is evidence given hypothesis)
- $P(A)$ = **Prior** (what we believed before seeing evidence)
- $P(B)$ = **Evidence/Normalization constant**

Intuitive Form

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Derivation: Directly from the product rule!

Solving the Medical Example

Given:

- $P(D) = 0.01$
- $P(S|D) = 0.9$
- $P(S|\neg D) = 0.2$

Step 1: Compute $P(S)$ using Law of Total Probability

$$\begin{aligned}P(S) &= P(S|D) \cdot P(D) + P(S|\neg D) \cdot P(\neg D) \\ &= 0.9 \times 0.01 + 0.2 \times 0.99 \\ &= 0.009 + 0.198 = 0.207\end{aligned}$$

Step 2: Apply Bayes' Rule

$$\begin{aligned}P(D|S) &= \frac{P(S|D) \cdot P(D)}{P(S)} \\ &= \frac{0.9 \times 0.01}{0.207}\end{aligned}$$

0.009

The Base Rate Fallacy

Surprising Result

Even with a 90% accurate test, if you test positive, there's only a **4.35% chance** you have the disease!

Why?

- The disease is **rare** (base rate = 1%)
- The test has **false positives** (20% of healthy people test positive)
- Most positive results come from the much larger healthy population

Lesson: Base Rates Matter!

People (and naive AI systems) often ignore base rates and focus only on the test accuracy. Bayes' Rule forces us to account for **prior probabilities**.

Diagnostic vs. Causal Reasoning

Causal Reasoning

From cause to effect

$$P(\text{Effect}|\text{Cause})$$

“If it rains, what’s the probability the grass is wet?”

Usually **easy to estimate**
(we understand causation)

Diagnostic Reasoning

From effect to cause

$$P(\text{Cause}|\text{Effect})$$

“If the grass is wet, what’s the probability it rained?”

Usually **hard to estimate directly**
(reverse inference is difficult)

Bayes' Rule to the Rescue

Bayes' Rule lets us go from **easy causal probabilities** to **hard diagnostic probabilities**:

$$P(\text{Cause}|\text{Effect}) = \frac{P(\text{Effect}|\text{Cause}) \cdot P(\text{Cause})}{P(\text{Effect})}$$

Independence

Definition

Two events A and B are **independent** if:

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

Equivalently: $P(A|B) = P(A)$ (knowing B doesn't change belief about A)

Example:

- Coin flip 1 and Coin flip 2 are independent
- $P(H_1, H_2) = P(H_1) \cdot P(H_2) = 0.5 \times 0.5 = 0.25$

Conditional Independence

A and B are **conditionally independent** given C if:

$$P(A, B|C) = P(A|C) \cdot P(B|C)$$

Common Probability Distributions

Bernoulli Distribution

Single binary outcome:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

Example: Single coin flip

Binomial Distribution

Number of successes in n trials:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Example: Number of heads in 10 coin flips

Normal (Gaussian) Distribution

Continuous distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Central Limit Theorem

One of the Most Important Theorems in Statistics

The sum (or average) of many independent random variables tends toward a normal distribution, **regardless of the original distributions**.

Formal Statement:

If X_1, X_2, \dots, X_n are i.i.d. with mean μ and variance σ^2 :

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1) \quad \text{where } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Why it matters:

- Explains why normal distributions appear everywhere
- Foundation for statistical inference
- Justifies many ML assumptions

Law of Large Numbers

Strong Law of Large Numbers

As the number of trials $n \rightarrow \infty$, the sample average converges to the expected value:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} \mathbb{E}[X] = \mu$$

Intuition:

- Flip a coin many times: proportion of heads $\rightarrow 0.5$
- Roll a die many times: average value $\rightarrow 3.5$

Connects Probability to Reality

This theorem justifies using frequencies to estimate probabilities!

Maximum Likelihood Estimation (MLE)

The Idea

Choose the parameter θ that makes the observed data most likely

$$\hat{\theta}_{MLE} = \arg \max_{\theta} P(\text{Data}|\theta)$$

Example: Coin Bias

- Flip a coin 100 times: 60 heads, 40 tails
- What's the bias p (probability of heads)?

$$P(\text{Data}|p) = \binom{100}{60} p^{60} (1-p)^{40}$$
$$\frac{d}{dp} \log P(\text{Data}|p) = \frac{60}{p} - \frac{40}{1-p} = 0$$

Bayesian Parameter Estimation

The Bayesian Approach

Don't just estimate a single “best” parameter. Compute the full **distribution** over parameters:

$$P(\theta|Data) = \frac{P(Data|\theta) \cdot P(\theta)}{P(Data)}$$

Components:

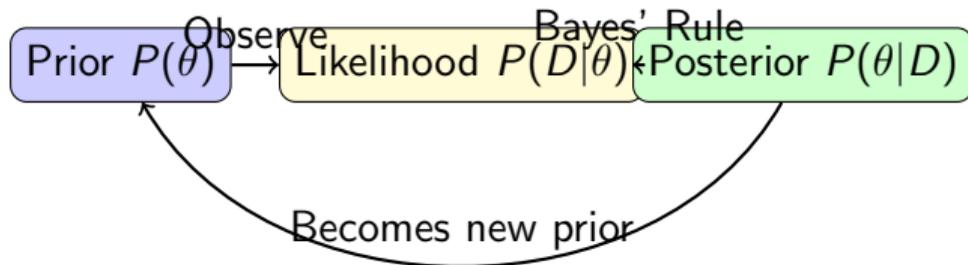
- $P(\theta)$ = **Prior**: belief before seeing data
- $P(Data|\theta)$ = **Likelihood**: how well θ explains data
- $P(\theta|Data)$ = **Posterior**: updated belief after seeing data

Advantage Over MLE

- Quantifies uncertainty about parameters
- Incorporates prior knowledge

Prior to Posterior: Belief Updating

The Bayesian Learning Cycle



Key Insight: Today's posterior is tomorrow's prior!

Sequential Belief Updating

$$P(\theta|D_1, D_2) = \frac{P(D_2|\theta) \cdot P(\theta|D_1)}{P(D_2)}$$

where $P(\theta|D_1)$ is the posterior after seeing D_1

Conjugate Priors

Definition

A prior is **conjugate** to a likelihood if the posterior has the same functional form as the prior.

Example: Beta-Binomial Conjugacy

- Prior: $p \sim \text{Beta}(\alpha, \beta)$
- Likelihood: $\text{Data} | p \sim \text{Binomial}(n, p)$
- Posterior: $p | \text{Data} \sim \text{Beta}(\alpha + k, \beta + n - k)$

where k = number of successes in n trials

Why Useful?

- Closed-form posterior (no numerical integration!)
- Easy to update sequentially
- Intuitive interpretation

Summary: The Probabilistic Revolution

1 Two Visions of Intelligence

- Logic: certainty, deduction, complete knowledge
- Probability: confidence, induction, partial knowledge

2 Logic's Limits

- Qualification problem, uncertainty types
- Real world is inherently uncertain

3 Probability Theory

- Three axioms, conditional probability, Bayes' Rule
- Rational framework for uncertainty

4 Bayesian Inference

- Belief updating, prior to posterior
- Principled learning from data

What You Should Remember

- 1 Modern AI is **fundamentally probabilistic**
- 2 Bayes' Rule: $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$ – most important formula
- 3 Base rates matter! (avoid base rate fallacy)
- 4 Rational agents maximize expected utility under uncertainty
- 5 Probability is degrees of belief, not just frequencies
- 6 Bayesian inference: learning = belief updating with evidence

Questions?

References & Further Reading

- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Online Resources:

- Course materials: SE444 Lecture 10
- Interactive demos available on course website