Introduction to Bayesian Inference

DSA 8505: Bayesian Analysis

Dr. Jacob Ongála

Strathmore University

January 17, 2025

Outline

What is Bayesian Inference?

- A probabilistic framework for updating beliefs based on new evidence.
- Treats parameters as random variables with probability distributions.
- Key distinction from frequentist methods:
  - Bayesian: Parameters are random variables.
  - Frequentist: Parameters are fixed but unknown constants.

Advantages of Bayesian Inference

- **Incorporation of Prior Knowledge:** Use prior distributions to integrate existing knowledge.
- **Unified Framework:** Supports inference, prediction, and decision-making in one approach.
- **Flexibility:** Handles complex models and uncertainty quantification effectively.

Parameter Treatment

- **Frequentist:** Parameters are fixed constants, and variability comes from sampling.
- **Bayesian:** Parameters are random variables described by prior distributions.

Example:
Frequentist Method: Suppose we want to estimate the average height of adult women in a country. We take a sample of 100 women and calculate the sample mean height.

Parameter Treatment: Example

In the **Frequentist** approach, the population mean height is considered a fixed, unknown constant. Our goal is to estimate this constant from the sample.

- We might report a **confidence interval** for the population mean height, such as "the population mean height lies between 165 cm and 170 cm with 95% confidence." However, the population mean remains fixed, and our interval reflects the uncertainty in the estimate due to sampling variability.

Parameter Treatment: Example ...

In contrast, **Bayesian** methods treat parameters as random variables. Before observing any data, a **prior distribution** is assigned to the parameter based on prior knowledge or beliefs. Once data is observed, the prior is updated to form a **posterior distribution**, which represents the updated uncertainty about the parameter.

- For example, based on previous studies or expert knowledge, we might believe that the mean height of adult women is likely to be between 160 cm and 175 cm, with a higher probability around 170 cm. This belief is captured by a **prior distribution**.
- After collecting data from our sample of 100 women and calculating the sample mean, the Bayesian approach would update the prior distribution using the **likelihood** of the data (based on the observed sample mean and its variability). The result is a **posterior distribution**, which gives the updated belief about the population mean after incorporating both the prior knowledge and the observed data.

Incorporation of Prior Knowledge

Frequentist Method:

- Relies solely on the data, no incorporation of prior information.

Data (Sample)

Bayesian Method:

- Combines prior distributions with likelihood, using external knowledge or past research.

Prior Information

Data (Likelihood)

Posterior Distribution

Subjective Probability

- Bayesian probability is interpreted subjectively as a degree of belief, representing an individual's uncertainty about a proposition or parameter.
- Contrasts with frequentist interpretation (long-run frequency of an event).
- Enables:
  - Updating beliefs with new data.
  - Quantifying uncertainty for single events.

Real-World Applications

- **Risk Assessment:** Estimate probabilities in financial markets.
- **Medical Decision-Making:** Calculate treatment success probabilities for individual patients.
- **Engineering Reliability:** Assess system failure probabilities and component lifespans.

Introduction to Bayes' Theorem and Its Use in Updating Information

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

- $P(\theta|D)$: Posterior probability (updated belief).
- $P(D|\theta)$: Likelihood (data given the hypothesis).
- $P(\theta)$: Prior probability (initial belief).
- $P(D)$: Marginal likelihood (normalizing constant).

Conditional Probability and Bayes' Theorem

Bayes' Theorem from Conditional Probability:

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

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

Substituting $P(A \cap B)$ from the second equation into the first gives:

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

- Demonstrates how prior beliefs ($P(A)$) are updated with new evidence ($P(B|A)$).
- Forms the basis for $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$ in Bayesian inference.

Explanation of Terms

1. **Prior Probability** $P(\theta)$
   - Represents the initial belief about the hypothesis before observing data.
   - Can be subjective (expert knowledge) or derived from previous studies.
2. **Likelihood** $P(D|\theta)$
   - Probability of observing the data given the hypothesis.
   - Reflects how well the hypothesis explains the observed data.

Explanation of Terms (cont.)

3. **Marginal Likelihood** $P(D)$
   - Total probability of observing the data across all possible hypotheses:

$$P(D) = \sum_{\theta} P(D|\theta)P(\theta)$$

   if $\theta = \theta_1 \& \theta_2\}$ then $P(D) = P(D|\theta_1) \cdot P(\theta_1) + P(D|\theta_2) \cdot P(\theta_2)$
   - For continuous cases:

$$P(D) = \int P(D|\theta)P(\theta)\, d\theta$$

   - Used to normalize the posterior distribution.
4. **Posterior Probability** $P(\theta|D)$
   - Updated belief about the hypothesis after observing data:

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

   - Combines prior knowledge and likelihood of the data.

Example: Coin Tossing

Scenario: Toss a coin 10 times, observe 7 heads.

- **Frequentist:** Estimate $\hat{p} = 7/10 = 0.7$.
- **Bayesian:** Use a Beta(1, 1) prior.
  - Posterior: Beta(8, 4), allowing interval estimates.

Example 1: Medical Diagnosis

Scenario: Calculate the probability that a patient has a certain disease after receiving a positive test result.

- **Prior Probability ($P(\theta)$):** The prevalence of the disease is 1%, meaning $P(\theta) = 0.01$.
- **Likelihood ($P(D|\theta)$):** The probability of testing positive if the patient has the disease is 90%, meaning $P(D|\theta) = 0.9$.
- **False Positive Rate ($P(D|\neg\theta)$):** The probability of testing positive when the patient does not have the disease is 5%, meaning $P(D|\neg\theta) = 0.05$.

Example 1: Medical Diagnosis ...

- **Marginal Probability ($P(D)$):** The total probability of a positive test result:

$$P(D) = P(D|\theta)P(\theta) + P(D|\neg\theta)P(\neg\theta)$$

$$P(D) = (0.9 \times 0.01) + (0.05 \times 0.99) = 0.009 + 0.0495 = 0.0585$$

- **Posterior Probability ($P(\theta|D)$):**

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

$$P(\theta|D) = \frac{0.9 \times 0.01}{0.0585} = \frac{0.009}{0.0585} \approx 0.154$$

Conclusion: Even after a positive test result, the probability that the patient actually has the disease is about 15.4%, highlighting the importance of prior probabilities in decision-making.

Example 2: Coin Tossing

Scenario: You are unsure whether a coin is fair or biased toward heads. Initially, you believe the coin is fair, but after tossing it 10 times and observing 7 heads, you update your belief.

- **Define the Prior:**

$$P(\text{Fair}) = 0.50, \quad P(\text{Biased}) = 0.50$$

- **Define the Likelihood:**

$$P(7 \text{ heads}|\text{Fair}) = \binom{10}{7}(0.5)^7(0.5)^3 = 0.117$$

$$P(7 \text{ heads}|\text{Biased}) = \binom{10}{7}(0.7)^7(0.3)^3 = 0.267$$

- **Calculate the Marginal Likelihood:**

$$P(7 \text{ heads}) = P(7 \text{ heads}|\text{Fair})P(\text{Fair}) + P(7 \text{ heads}|\text{Biased})P(\text{Biased})$$

$$P(7 \text{ heads}) = (0.117 \times 0.50) + (0.267 \times 0.50) = 0.0585 + 0.1335 = 0.192$$

Example 2: Coin Tossing ...

- **Apply Bayes' Theorem:**

$$P(\text{Fair}|7\text{ heads}) = \frac{P(7\text{ heads}|\text{Fair})P(\text{Fair})}{P(7\text{ heads})}$$

$$P(\text{Fair}|7\text{ heads}) = \frac{0.117 \times 0.50}{0.192} = 0.305$$

Similarly, for the biased hypothesis:

$$P(\text{Biased}|7\text{ heads}) = \frac{P(7\text{ heads}|\text{Biased})P(\text{Biased})}{P(7\text{ heads})}$$

$$P(\text{Biased}|7\text{ heads}) = \frac{0.267 \times 0.50}{0.192} = 0.695$$

Conclusion: After observing 7 heads, there is approximately a 30.5% chance the coin is fair and a 69.5% chance it is biased toward heads.