Lecture Notes

BAYESIAN ANALYSIS

DSA 8505

Strathmore University

Lecturer: Dr. Jacob Ong'ala

Contents

1 Introduction to Bayesian Inference

Bayesian inference offers a probabilistic framework for incorporating prior knowledge and updating beliefs in light of new evidence. This approach fundamentally differs from the classical (frequentist) paradigm, which treats parameters as fixed and unknown quantities. Bayesian methods, in contrast, view parameters as random variables described by probability distributions. This distinction provides several advantages:

- **Incorporation of prior knowledge:** Bayesian inference allows analysts to incorporate existing knowledge or expert opinions into the analysis through prior distributions. This is particularly useful in situations where data are scarce or expensive to collect.
- **Unified framework:** Bayesian methods offer a cohesive approach to inference, prediction, and decision-making. The posterior distribution contains all relevant information about the parameters, facilitating direct probabilistic statements about their values.
- **Flexibility:** Bayesian models can easily accommodate complex structures, hierarchical relationships, and uncertainty quantification.

Connections with the classical approach include:

- **Likelihood functions:** Both Bayesian and frequentist methods rely heavily on the likelihood function to capture the relationship between the data and the model parameters. In Bayesian inference, the likelihood is combined with a prior to compute the posterior.
- **Large-sample behavior:** Under certain conditions, Bayesian posterior distributions converge to frequentist estimators as sample sizes increase. For example, the mean of the posterior distribution often approximates the maximum likelihood estimate (MLE), and credible intervals can align closely with confidence intervals.

1.1 Comparison of Bayesian and Classical Approaches

Parameter Treatment:

- Frequentist methods treat parameters as fixed and unknown constants, focusing on sampling variability of the data.
- Bayesian methods treat parameters as random variables with a prior distribution that reflects their uncertainty before observing data.

Uncertainty Quantification:

- Frequentist confidence intervals provide a range of values that, under repeated sampling, would contain the true parameter value with a certain frequency (e.g., 95%).
- Bayesian credible intervals directly express the probability that the parameter lies within a given range based on the observed data and the prior.

Incorporation of Prior Knowledge:

- Frequentist methods rely solely on the data and do not incorporate prior information.
- Bayesian methods combine prior distributions with the likelihood, enabling analysts to use external knowledge or past research.

Interpretation of Results:

- Frequentist p-values and confidence intervals can be challenging to interpret and may not provide direct probabilistic statements about parameters.
- Bayesian inference directly quantifies uncertainty, offering intuitive probabilistic interpretations of parameters and predictions.

Applications:

- Frequentist methods are well-suited for scenarios with large datasets and minimal prior information.
- Bayesian methods excel in fields such as medicine, engineering, and finance, where prior knowledge is critical and data may be limited.

This comparison highlights the complementary nature of Bayesian and frequentist approaches, with each offering distinct strengths depending on the context of the analysis.

1.2 Subjective Interpretation of Probability

Bayesian probability is interpreted subjectively as a degree of belief, representing an individual's uncertainty about a proposition or parameter. This contrasts with the frequentist interpretation, which defines probability as the long-run relative frequency of an event occurring in repeated experiments. The subjective interpretation offers several key advantages:

- **A coherent framework for updating beliefs:** By using Bayes' theorem, prior beliefs can be updated in light of new data to reflect current knowledge.
- **Quantifying uncertainty in unique events:** Unlike the frequentist view, Bayesian methods allow for probability statements about single events or unique situations (e.g., the likelihood of rain tomorrow).

Real-World Applications

The subjective interpretation of probability is particularly valuable in practical settings where uncertainty plays a significant role:

- **Risk assessment in financial markets:** Analysts can use Bayesian methods to estimate the probability of market crashes or evaluate investment risks based on historical data and expert opinions.
- **Medical decision-making under uncertainty:** Physicians can integrate prior clinical experience and trial data to determine the probability of treatment success for individual patients.
- **Engineering reliability analysis:** Bayesian methods help assess the probability of system failures or estimate the remaining lifespan of critical components, combining field data with expert knowledge.

Updating Beliefs Using Bayes' Theorem

Consider a practical scenario where subjective probabilities evolve with new information:

- **Example: Diagnosing a Disease**
  - **Prior belief:** Based on population data, a physician estimates that 5% of patients presenting with certain symptoms have Disease A.
  - **Likelihood:** A diagnostic test has a 95% sensitivity (true positive rate) and a 90% specificity (true negative rate).
  - **Data:** The patient tests positive for the disease.
  - **Posterior belief:** Using Bayes' theorem, the physician updates the probability of Disease A for the patient to approximately 34%.

1.3 Introduction to Bayes' Theorem and Its Use in Updating Information

Bayes' Theorem is a fundamental concept in probability theory and statistics. It provides a way to update the probability of a hypothesis (θ) given new data (D). In Bayesian inference, this theorem allows us to revise our beliefs about a parameter θ after observing new evidence D. This theorem is crucial for decision-making in various fields, such as medicine, machine learning, and scientific research.

1.3.1 Bayes' Theorem Formula and Conditional Distributions

Bayes' Theorem is rooted in the concept of conditional probability. It describes how to update the probability of a hypothesis θ based on new data D. This is done by utilizing prior knowledge and the likelihood of the observed data.

The general form of Bayes' Theorem is given as:

$$P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{P(D)} \tag{1.1}$$

Where:

- $P(\theta|D)$ is the **posterior probability**: the probability of the hypothesis θ being true after observing the data D. It represents the updated belief about the hypothesis after accounting for the data.
- $P(D|\theta)$ is the **likelihood**: the probability of observing the data D, given that the hypothesis θ is true. It reflects how well the hypothesis explains the observed data.
- $P(\theta)$ is the **prior probability**: the probability assigned to the hypothesis θ before observing any data. This prior encapsulates the initial belief or background knowledge about θ.
- $P(D)$ is the **marginal likelihood** or **evidence**: the total probability of observing the data D under all possible hypotheses. It normalizes the posterior probability to ensure it is a valid probability distribution.

1.3.2 Conditional Probability and Bayes' Theorem

Bayes' Theorem is a direct result of the definition of conditional probability. The conditional probability of an event A given another event B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.2}$$

Similarly, the conditional probability of B given A is:

$$\begin{gathered} P(B|A) = \frac{P(A \cap B)}{P(A)} \\ P(B|A) \cdot P(A) = P(A \cap B) \end{gathered} \tag{1.3}$$

substituting $P(A \cap B)$ of Equation 1.3 in Equation 1.2, we derive the following Bayes' Theorem:

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

In the context of Bayes' Theorem applied to statistical inference, we substitute the hypothesis θ for event A and the data D for event B, yielding the formula:

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

1.3.3 Explanation of Terms

1. Prior Probability $P(\theta)$

The prior probability represents the initial belief about the hypothesis θ before any data is observed. It can be subjective, based on expert knowledge or past experience, or it can be derived from prior data. In the Bayesian framework, this is where any existing information about θ is incorporated.

For example, in a medical context, the prior probability $P(\theta)$ could represent the prevalence of a disease in the population before any diagnostic tests are performed. If a certain disease affects 2% of the population, the prior probability that a randomly chosen person has the disease is 0.02.

2. Likelihood $P(D|\theta)$

The likelihood is the probability of observing the data D, assuming that the hypothesis θ is true. The likelihood function plays a crucial role in Bayesian inference as it measures how well the hypothesis θ explains the data.

In many practical cases, the likelihood is determined from a statistical model. For example, in a clinical trial, the likelihood $P(D|\theta)$ might be based on a binomial or normal distribution, depending on the type of data collected (e.g., success/failure outcomes or continuous measurements).

3. Marginal Likelihood $P(D)$

The marginal likelihood (also called the evidence) is the total probability of observing the data D, regardless of which hypothesis θ is true. It is computed by summing (or integrating) over all possible hypotheses:

$$P(D) = \sum_{\theta} P(D|\theta) \cdot P(\theta) \tag{1.4}$$

in discrete case in $\theta = \{\theta_1 \texttt{ and } \theta_2\}$ then

$$P(D) = P(D|\theta_1) \cdot P(\theta_1) + P(D|\theta_2) \cdot P(\theta_2) \tag{1.5}$$

For continuous distributions, this becomes an integral:

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

The marginal likelihood ensures that the posterior probability $P(\theta|D)$ is properly normalized. It also plays a critical role in model comparison in Bayesian inference, where models are compared based on their ability to explain the observed data.

4. Posterior Probability $P(\theta|D)$

The posterior probability is the main quantity of interest in Bayesian inference. It represents the updated probability of the hypothesis θ after considering the observed data D. The posterior combines both the prior belief $P(\theta)$ and the likelihood $P(D|\theta)$, providing a new probability distribution over the hypotheses.In decision-making contexts, the posterior probability helps in making informed decisions based on updated beliefs. For instance, after observing a positive test result, the posterior probability tells us the likelihood that a patient has a disease.

Formally, if we treat θ as a random variable with prior distribution $P(\theta)$, and D as the observed data, the posterior distribution $P(\theta|D)$ is updated via Bayes' Theorem as:

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

1.4 Examples and Applications of Bayes' Theorem

Example 1: Medical Diagnosis

Consider a scenario where we want to 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 in the population is 1%, meaning $P(\theta) = 0.01$.
- **Likelihood ($P(D|\theta)$):** The probability of testing positive, given that 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 $P(D|\neg\theta) = 0.05$.
- **Marginal Probability ($P(D)$):** The total probability of a positive test result, regardless of whether the patient has the disease or not.

We now apply Bayes' Theorem to calculate the **posterior probability** $P(\theta|D)$, the probability that the patient has the disease given a positive test result:

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

First, we compute $P(D)$, the marginal probability:

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

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

Now we can compute the posterior:

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

So, 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 making decisions.

Example 2: Spam Filtering

Bayes' Theorem is widely used in email spam filters. The goal is to determine whether an email is spam (θ) given the data (D), such as the words contained in the email.

- **Prior Probability** $P(\theta)$**:** This could represent the proportion of spam emails in your inbox, based on past observations.
- **Likelihood** $P(D|\theta)$**:** This is the probability of observing certain words or phrases in spam emails, such as "prize," "win," or "money."
- **Posterior Probability** $P(\theta|D)$**:** After observing the words in a new email, the posterior probability gives the updated belief that the message is spam.

1.5 Concept of Belief Updating

Belief updating is a fundamental concept in probability and statistics, often used in Bayesian inference. It refers to the process of adjusting or revising one's beliefs (probability estimates) in light of new evidence.

In statistical terms:

- **Prior belief**: The initial probability or belief before observing any data.
- **Likelihood**: The probability of observing the new data given the current state of belief.
- **Posterior belief**: The updated belief after incorporating the new evidence.

The process of belief updating helps us make more accurate predictions and decisions by continuously refining our understanding based on observed data.

1.5.1 Bayes' Theorem

The mathematical foundation for belief updating is based on **Bayes' Theorem**:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

Where:

- $P(H|E)$ is the **posterior probability** (updated belief) of hypothesis H given evidence E.
- $P(E|H)$ is the **likelihood**, the probability of observing the evidence E given that the hypothesis H is true.
- $P(H)$ is the **prior probability** of hypothesis H before observing the evidence.
- $P(E)$ is the **marginal likelihood**, the overall probability of observing the evidence E, which can be computed by summing over all possible hypotheses.

1.6 Step-by-Step Inference Examples

1.6.1 Example 1: Medical Diagnosis

Imagine a scenario where a doctor wants to update the belief that a patient has a certain disease based on a test result. Let's walk through the steps:

1. **Define Prior Belief**
   The doctor knows from past experience that 1% of patients have the disease. This is the prior belief:
   $$P(\text{Disease}) = 0.01$$

2. **Define the Likelihood**
   Suppose a test for the disease is 90% accurate. This means that if a person has the disease, the test will show positive 90% of the time:

$$P(\text{Positive Test}|\text{Disease}) = 0.90$$

   However, the test can also give false positives. Let's assume 5% of healthy patients test positive (false positive rate):

$$P(\text{Positive Test}|\text{No Disease}) = 0.05$$

3. **Calculate the Marginal Likelihood**
   We now need to compute the overall probability of a positive test result, $P(\text{Positive Test})$:

$$P(\text{Positive Test}) = P(\text{Positive Test}|\text{Disease}) \cdot P(\text{Disease}) + P(\text{Positive Test}|\text{No Disease}) \cdot P(\text{No Dise}$$

$$= (0.90 \cdot 0.01) + (0.05 \cdot 0.99) = 0.009 + 0.0495 = 0.0585$$

4. **Apply Bayes' Theorem**
   Now, we update the belief using Bayes' Theorem:

$$P(\text{Disease}|\text{Positive Test}) = \frac{P(\text{Positive Test}|\text{Disease}) \cdot P(\text{Disease})}{P(\text{Positive Test})}$$

$$= \frac{0.90 \cdot 0.01}{0.0585} = \frac{0.009}{0.0585} \approx 0.154$$

   Therefore, the updated probability that the patient has the disease after a positive test is approximately **15.4%**.

1.6.2 Example 2: Coin Tossing

Consider a scenario where you are unsure whether a coin is fair or biased toward heads. Initially, you believe the coin is fair, but after tossing the coin several times and observing the outcomes, you update your belief.

1. **Define the Prior**
   You start with the prior belief that the coin is fair:

$$P(\text{Fair}) = 0.50$$

   and that it is biased:

$$P(\text{Biased}) = 0.50$$

2. **Define the Likelihood**
   Suppose you toss the coin 10 times, and you observe 7 heads. If the coin is fair, the probability of observing 7 heads is:

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

   If the coin is biased, suppose it lands heads 70% of the time, so:

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

3. **Calculate the Marginal Likelihood**
The overall probability of observing 7 heads is:

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

$$= (0.117 \cdot 0.50) + (0.267 \cdot 0.50) = 0.0585 + 0.1335 = 0.192$$

4. **Apply Bayes' Theorem**
Now, update the belief:

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

$$= \frac{0.117 \cdot 0.50}{0.192} = 0.305$$

Similarly, update for the biased hypothesis:

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

$$= \frac{0.267 \cdot 0.50}{0.192} = 0.695$$

Thus, after observing 7 heads, the updated belief is that there's approximately a **30.5%** chance that the coin is fair and a **69.5%** chance that it is biased toward heads.

1.7 Applications of Bayes' Theorem

Bayes' Theorem finds application in a wide range of areas:

- **Medical Diagnosis**: Helps in updating probabilities of diseases after observing test results.
- **Spam Filtering**: Used by email services to filter spam based on word frequencies.
- **Forensic Science**: Applied in drug testing or DNA matching to calculate the likelihood of guilt or innocence.
- **Machine Learning and AI**: Forms the foundation for Bayesian classifiers, like the Naive Bayes classifier.
- **Predictive Analytics**: Used to update predictions in various domains, including finance, weather forecasting, and market analysis.

Practical Exercise

Question 2.1: In a clinical study, 5% of the population is known to carry a particular virus. A new test is developed with the following characteristics:

- If a person has the virus, the test is positive 92% of the time.

- If a person does not have the virus, the test is positive 8% of the time.

Suppose a random person from the population tests positive. Calculate:

(i) The probability that the person actually has the virus.

(ii) The probability of getting a positive test result for this population.

Question 2.2: Using the following study data of a population of 10,000 individuals where a disease test is conducted:

- Total population (N): 10,000 individuals
- Number of people with the disease (D^+): 500 individuals
- Number of people without the disease (D^-): 9,500 individuals

The test characteristics are as follows:

- True Positive (TP): 450 individuals
- False Negative (FN): 50 individuals
- True Negative (TN): 8,550 individuals
- False Positive (FP): 950 individuals

Calculate the following:

(a) The marginal probability of testing negative, $P(\neg D)$.

(b) The probability that a patient has the disease given that they tested negative, $P(\theta|\neg D)$.

Bayes' theorem is the foundation of Bayesian inference. It is expressed as:

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

where:

- $P(\theta|D)$ is the posterior probability of the parameter θ given data D.
- $P(D|\theta)$ is the likelihood of the data given the parameter θ.
- $P(\theta)$ is the prior probability of the parameter θ.
- $P(D)$ is the marginal likelihood or evidence.

Bayes' theorem enables the updating of prior beliefs $P(\theta)$ using observed data D to obtain the posterior distribution $P(\theta|D)$. This process is iterative and allows for:

- Dynamic incorporation of new evidence.
- Improved decision-making as more data becomes available.

Practical Examples Comparing Bayesian and Frequentist Approaches

Example 1: Coin Tossing

Scenario: A coin is flipped 10 times, resulting in 7 heads. What is the probability of the coin landing heads in future flips?

- **Frequentist approach:** Estimate the probability using the sample proportion: $\hat{p} = 7/10 = 0.7$. This provides a point estimate but no direct measure of uncertainty.
- **Bayesian approach:** Assume a prior distribution (e.g., Beta(1, 1)) for the probability of heads. Using Bayes' theorem, update the prior with the observed data to obtain a posterior distribution (e.g., Beta(8, 4)), which allows for interval estimates and predictions.

Example 2: Drug Effectiveness

Scenario: A new drug is tested on a small sample, showing promising results. How confident can we be about its effectiveness?

- **Frequentist approach:** Perform a hypothesis test (e.g., t-test) and calculate a p-value to assess significance.
- **Bayesian approach:** Use prior knowledge about similar drugs to construct a prior distribution. Update this with the observed data to obtain a posterior probability distribution, providing a more intuitive measure of confidence.

Key Insights from Comparisons

- Bayesian methods provide a richer output (e.g., full posterior distributions) compared to frequentist point estimates or p-values.
- Bayesian approaches are more flexible in incorporating prior information and adapting to small sample sizes.
- Frequentist methods are simpler to implement in standard cases and do not rely on subjective priors.