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.

  A major strength of Bayesian inference is its ability to formally integrate previous research findings, expert judgment, or domain knowledge into the analysis through the use of prior distributions. This feature is especially valuable in contexts where data are limited, costly to obtain, or collected under restrictive conditions. By combining prior beliefs with observed data, the Bayesian approach produces estimates that reflect both historical evidence and newly acquired information.

- **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.

  Bayesian methods provide a coherent and comprehensive framework for statistical inference, prediction, and decision-making. All uncertainty about the model parameters is expressed through the posterior distribution, which synthesizes the prior information and the likelihood of the observed data. This posterior distribution contains the complete probabilistic description of the parameters, enabling analysts to make direct statements such as "there is a 90% probability that the parameter lies within this interval," which is not possible in classical (frequentist) inference.

  In Bayesian inference, one may state that "there is a 90% probability that the parameter lies within this interval" because the parameter is treated as a random quantity whose uncertainty is described by the posterior distribution. Thus, a Bayesian credible interval provides a direct probabilistic statement about the parameter.

  In contrast, classical (frequentist) interval estimation treats the parameter as a fixed but unknown constant. A 90% confidence interval therefore *does not* imply a 90% probability that the parameter lies within the computed interval. Instead, the correct interpretation is that if we were to repeatedly draw samples and construct confidence intervals in the same way, then 90% of those intervals would contain the true parameter value. The probability statement refers to the long–run performance of the procedure, not to the specific interval obtained from a single dataset.

- **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:

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

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: 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.

Suppose an email service wants to determine whether a new email is spam based on the presence of a particular word, say "prize". The following information is available from historical data:

- The proportion of emails that are spam is

  $$P(\theta = \text{spam}) = 0.30.$$

- The proportion of emails that are not spam is

  $$P(\theta = \text{not spam}) = 0.70.$$

- The probability that the word "prize" appears in a spam email:

$$P(D = \text{“prize”} \mid \theta = \text{spam}) = 0.40.$$

- The probability that the word "prize" appears in a legitimate email:

$$P(D = \text{“prize”} \mid \theta = \text{not spam}) = 0.05.$$

We observe a new email containing the word "prize". We want the posterior probability that this email is spam.

Step 1: Compute the marginal probability of observing the word

$$P(D = \text{“prize”}) = P(D \mid \text{spam})P(\text{spam}) + P(D \mid \text{not spam})P(\text{not spam}).$$

$$P(D) = (0.40)(0.30) + (0.05)(0.70) = 0.12 + 0.035 = 0.155.$$

Step 2: Apply Bayes' Theorem

$$P(\text{spam} \mid D) = \frac{P(D \mid \text{spam})\, P(\text{spam})}{P(D)}.$$

$$P(\text{spam} \mid D) = \frac{0.40 \times 0.30}{0.155} = \frac{0.12}{0.155} \approx 0.774.$$

Conclusion

After observing that the email contains the word "prize", the posterior probability that it is spam is approximately

$$P(\text{spam} \mid D) \approx 0.774.$$

Thus, there is a 77.4% chance that the email is spam, much higher than the original prior probability of 30%.

Example 2: 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%**.

Example 3: 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:

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

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

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

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

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

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

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.

1.5 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.

1.6 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.

1.7 Comparing Bayesian and Frequentist Approaches

1.7.1 Example 1: Coin Tossing

Scenario: A coin is flipped 10 times, resulting in 7 heads. We are interested in estimating the probability p that the coin will land heads in future flips.

Frequentist Approach

1. Basic Idea:
The frequentist perspective treats the parameter p as a fixed but unknown constant. In this view, probability represents the long-run frequency of events.

2. Estimation:
The natural estimator for p is the sample proportion,

$$\hat{p} = \frac{\text{Number of heads}}{\text{Number of flips}} = \frac{7}{10} = 0.7.$$

This value represents a *point estimate* of the true probability p.

3. Interpretation:
From a frequentist perspective, the best estimate of p given our data is $\hat{p} = 0.7$. This approach provides no direct probabilistic statement about the parameter p itself.

4. Uncertainty:
Uncertainty in the estimate can be quantified through a confidence interval. For example, an approximate 95% confidence interval for p is given by

$$0.7 \pm 1.96\sqrt{\frac{0.7(1-0.7)}{10}} = (0.42, 0.98).$$

Importantly, the interval is interpreted as: if we repeated the experiment many times, 95% of such intervals would contain the true parameter p.

Bayesian Approach

1. Basic Idea:
The Bayesian perspective treats p as a random variable. Prior beliefs about p are updated using observed data.

2. Prior Distribution:
Assume a prior distribution for p of the form

$$p \sim \text{Beta}(1, 1),$$

which is equivalent to a uniform distribution over $[0, 1]$, representing no strong prior beliefs.

3. Observed Data:
We observe $x = 7$ heads and $n = 10$ total flips.

4. Posterior Distribution:
Given a $\text{Beta}(\alpha, \beta)$ prior and binomial data with x heads in n flips, the posterior distribution is

$$p \mid \text{data} \sim \text{Beta}(\alpha + x, \beta + n - x).$$

Substituting the prior parameters $\alpha = \beta = 1$ gives

$$p \mid \text{data} \sim \text{Beta}(1 + 7, 1 + 3) = \text{Beta}(8, 4).$$

5. Posterior Mean:
The posterior mean is given by

$$\mathbb{E}[p \mid \text{data}] = \frac{\alpha + x}{(\alpha + x) + (\beta + n - x)} = \frac{8}{12} = 0.667.$$

This estimate is slightly lower than the sample proportion 0.7 due to the influence of the prior distribution.

6. Posterior Credible Interval:
A 95% credible interval for p can be computed from the $\text{Beta}(8, 4)$ distribution. For illustration,

$$p \in (0.41, 0.87).$$

This is interpreted as: with 95% probability, the true value of p lies within this interval.

7. Prediction for Future Flips:
The posterior predictive probability that the next flip results in a head is

$$P(\text{next head} \mid \text{data}) = \frac{8}{12} = 0.667.$$

Key Differences Between the Approaches

- Frequentist: p is a fixed but unknown constant. Estimate p using the sample proportion; provides a point estimate and confidence interval about the estimator.
- Bayesian: p is a random variable. Prior beliefs plus observed data yield a posterior distribution; provides probability statements about the parameter and predictions.

Conclusion:
The frequentist estimate of p is $\hat{p} = 0.7$, based solely on observed data. The Bayesian approach yields a posterior distribution $\text{Beta}(8, 4)$ with posterior mean 0.667 and credible interval $(0.41, 0.87)$, which reflects both the data and prior beliefs.

Example 2: Drug Effectiveness

Scenario: A new drug is tested on 100 patients suffering from a certain condition. After treatment, 60 patients show a positive improvement. We are interested in estimating the probability p that the drug is effective for a randomly selected future patient with the same condition.

Frequentist Approach
1. Basic Idea:
From the frequentist point of view, the parameter p (the probability that the drug is effective) is a fixed but unknown constant. Probability is interpreted as the long-run relative frequency of success if the treatment were repeated on many similar patients.

2. Estimation:
The natural estimator for p is the sample proportion of successes:

$$\hat{p} = \frac{\text{Number of patients who improved}}{\text{Total number of patients}} = \frac{60}{100} = 0.6.$$

This is a *point estimate* of the true effectiveness probability p.

3. Interpretation:
Based solely on the data, the frequentist estimate of the probability that the drug is effective for a future patient is $\hat{p} = 0.6$.

4. Quantifying Uncertainty (Confidence Interval):
To capture uncertainty about this estimate, the frequentist approach often uses a confidence interval. An approximate 95% confidence interval for p is

$$\hat{p} \pm z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.6 \pm 1.96\sqrt{\frac{0.6 \cdot 0.4}{100}}.$$

Compute the standard error:

$$\sqrt{\frac{0.6 \cdot 0.4}{100}} = \sqrt{\frac{0.24}{100}} = \sqrt{0.0024} \approx 0.049.$$

Thus,

$$0.6 \pm 1.96 \times 0.049 \approx 0.6 \pm 0.096 \approx (0.504,\ 0.696).$$

Interpretation: if we were to repeat many such clinical trials, approximately 95% of the confidence intervals constructed in this way would contain the true parameter p.

Bayesian Approach

1. Basic Idea:
In the Bayesian framework, the parameter p is treated as a random variable. We start with a *prior* distribution that encodes our beliefs about p before seeing the data, and then update this prior using the observed data to obtain a *posterior* distribution.

2. Choosing a Prior:
Suppose we begin with a non-informative prior, expressing initial uncertainty about the drug's effectiveness:

$$p \sim \text{Beta}(1, 1),$$

which is equivalent to a uniform distribution on $[0, 1]$ (all values of p between 0 and 1 are equally likely a priori).

Alternatively, if we had prior belief that the drug is somewhat likely to be effective (e.g., based on earlier small studies), we might use a more informative prior such as

$$p \sim \text{Beta}(3, 2),$$

whose mean is $\frac{3}{3+2} = 0.6$. For this example, we first use the non-informative prior $\text{Beta}(1, 1)$ for simplicity.

3. Observed Data:
We observe

$$x = 60 \quad \text{successes (improved patients)}, \quad n = 100 \quad \text{total patients}.$$

4. Posterior Distribution (Conjugacy):
For a binomial likelihood and a Beta prior, the posterior distribution is also Beta. If

$$p \sim \text{Beta}(\alpha, \beta)$$

and we observe x successes out of n trials, then

$$p \mid \text{data} \sim \text{Beta}(\alpha + x,\ \beta + n - x).$$

Using the non-informative prior Beta(1, 1):

$$p \mid \text{data} \sim \text{Beta}(1 + 60,\ 1 + 40) = \text{Beta}(61, 41).$$

5. Posterior Mean:
The posterior mean is

$$\mathbb{E}[p \mid \text{data}] = \frac{\alpha + x}{(\alpha + x) + (\beta + n - x)} = \frac{61}{61 + 41} = \frac{61}{102} \approx 0.598.$$

This value is very close to the frequentist estimate 0.6, reflecting that the data are relatively informative (sample size is large), and the prior was non-informative.

6. Posterior Credible Interval:
A 95% credible interval for p is obtained from the Beta(61, 41) distribution. Symbolically, we can write

$$P\big(p \in (a, b) \,\big|\, \text{data}\big) = 0.95,$$

for some values a and b determined from the quantiles of the Beta(61, 41) distribution. The key interpretation is:

There is a 95% probability that p lies between a and b.

7. Prediction for a Future Patient:
The posterior predictive probability that the drug will be effective for a new patient is the posterior mean:

$$P(\text{drug effective for next patient} \mid \text{data}) = \mathbb{E}[p \mid \text{data}] \approx 0.598.$$

Comparison of Frequentist and Bayesian Results

- *Frequentist:*
  - Treats p as a fixed but unknown constant.
  - Estimate: $\hat{p} = 0.6$.
  - 95% confidence interval: approximately (0.504, 0.696).
  - The interval has a long-run frequency interpretation.
- *Bayesian:*
  - Treats p as a random variable with a prior distribution.
  - Prior: $p \sim \text{Beta}(1, 1)$.
  - Posterior: $p \mid \text{data} \sim \text{Beta}(61, 41)$.
  - Posterior mean: ≈ 0.598 (very close to 0.6).
  - 95% credible interval: (a, b) such that $P(a < p < b \mid \text{data}) = 0.95$.

  - The interval has a direct probability interpretation about p.

Conclusion:
In this drug effectiveness example, both approaches produce similar point estimates because the sample size is large and the prior is weak. The main difference lies in the interpretation: the frequentist method provides a confidence interval based on repeated sampling, while the Bayesian approach yields a posterior distribution and credible interval that directly reflect uncertainty about the true effectiveness probability p.