DSA 8505 Bayesian Statistics

CAT 1

Kevin Obote - 190696

SU

Field of study

Strathmore Institute of Mathematical Science
Strathmore University

January 23, 2026

Project Repository: `https://github.com/Kevinobote/Bayesian-Statistics/blob/main/CATS/CAT1/CAT1.ipynb`

Question 1

A researcher is studying the probability θ that a patient responds to a new treatment. Before collecting data, the researcher specifies the prior distribution:

$$\theta \sim \text{Beta}(\alpha, \beta), \quad \alpha = 3,\ \beta = 7.$$

A clinical trial is conducted on $n = 20$ patients and $k = 14$ respond successfully.
Required:

(a) Write the likelihood function $p(D \mid \theta)$ under the Binomial model.

(b) Using Bayes theorem, derive the posterior distribution $p(\theta \mid D)$ showing all steps.

(c) State the posterior distribution clearly in the form:

$$\theta \mid D \sim \text{Beta}(\alpha, \beta).$$

Solution

We are interested in modeling the probability $\theta \in (0, 1)$ that a patient responds successfully to a new treatment. Since θ represents a probability, it is naturally defined on the interval $(0, 1)$.
Before observing any data, the researcher specifies a prior belief about θ using a Beta distribution:

$$\theta \sim \text{Beta}(\alpha, \beta), \quad \alpha = 3,\ \beta = 7.$$

The Beta distribution is appropriate here because it is defined on $(0, 1)$ and is conjugate to the Binomial likelihood.
A clinical trial is then conducted on $n = 20$ patients, and $k = 14$ patients respond successfully.

(a) Likelihood Function

Let D denote the observed data, namely $k = 14$ successful responses out of $n = 20$ trials. Assuming each patient responds independently with probability θ, the number of successful responses follows a Binomial distribution:

$$D \mid \theta \sim \text{Binomial}(20, \theta).$$

The likelihood function, which gives the probability of observing the data given θ, is therefore:

$$p(D \mid \theta) = \binom{20}{14} \theta^{14} (1-\theta)^{20-14} = \binom{20}{14} \theta^{14} (1-\theta)^{6},$$

where $\theta \in (0, 1)$.

(b) Posterior Distribution via Bayes Theorem

Bayes theorem allows us to update our prior belief about θ using the observed data:

$$p(\theta \mid D) = \frac{p(D \mid \theta) p(\theta)}{p(D)},$$

where the marginal likelihood (or evidence) is given by:

$$p(D) = \int_0^1 p(D \mid \theta) p(\theta) \, d\theta.$$

Step 1: Prior Density

The probability density function of the Beta$(3, 7)$ prior is:

$$p(\theta) = \frac{1}{B(3,7)} \theta^{3-1} (1-\theta)^{7-1} = \frac{1}{B(3,7)} \theta^{2} (1-\theta)^{6},$$

where $B(\alpha, \beta)$ is the Beta function, which ensures the density integrates to one.

Step 2: Form the Unnormalized Posterior

Substituting the likelihood and the prior into Bayes theorem, we obtain:

$$p(\theta \mid D) \propto p(D \mid \theta) p(\theta).$$

Substituting explicitly:

$$p(\theta \mid D) \propto \left[\binom{20}{14} \theta^{14} (1-\theta)^{6} \right] \left[\theta^{2} (1-\theta)^{6} \right].$$

The binomial coefficient and the Beta normalization constant do not depend on θ, and hence can be absorbed into the proportionality constant.

Step 3: Simplify the Expression

Combining powers of θ and $(1-\theta)$:

$$p(\theta \mid D) \propto \theta^{14+2}(1-\theta)^{6+6} = \theta^{16}(1-\theta)^{12}.$$

(c) Posterior Distribution

The expression $\theta^{16}(1-\theta)^{12}$ is the kernel of a Beta distribution. Comparing with the general Beta density:

$$\theta^{\alpha^*-1}(1-\theta)^{\beta^*-1},$$

we identify the updated parameters as:

$$\alpha^* = 17, \qquad \beta^* = 13.$$

Equivalently, the posterior parameters can be obtained using the conjugate updating rules:

$$\alpha^* = \alpha + k = 3 + 14 = 17,$$

$$\beta^* = \beta + (n-k) = 7 + 6 = 13.$$

Final Posterior Distribution

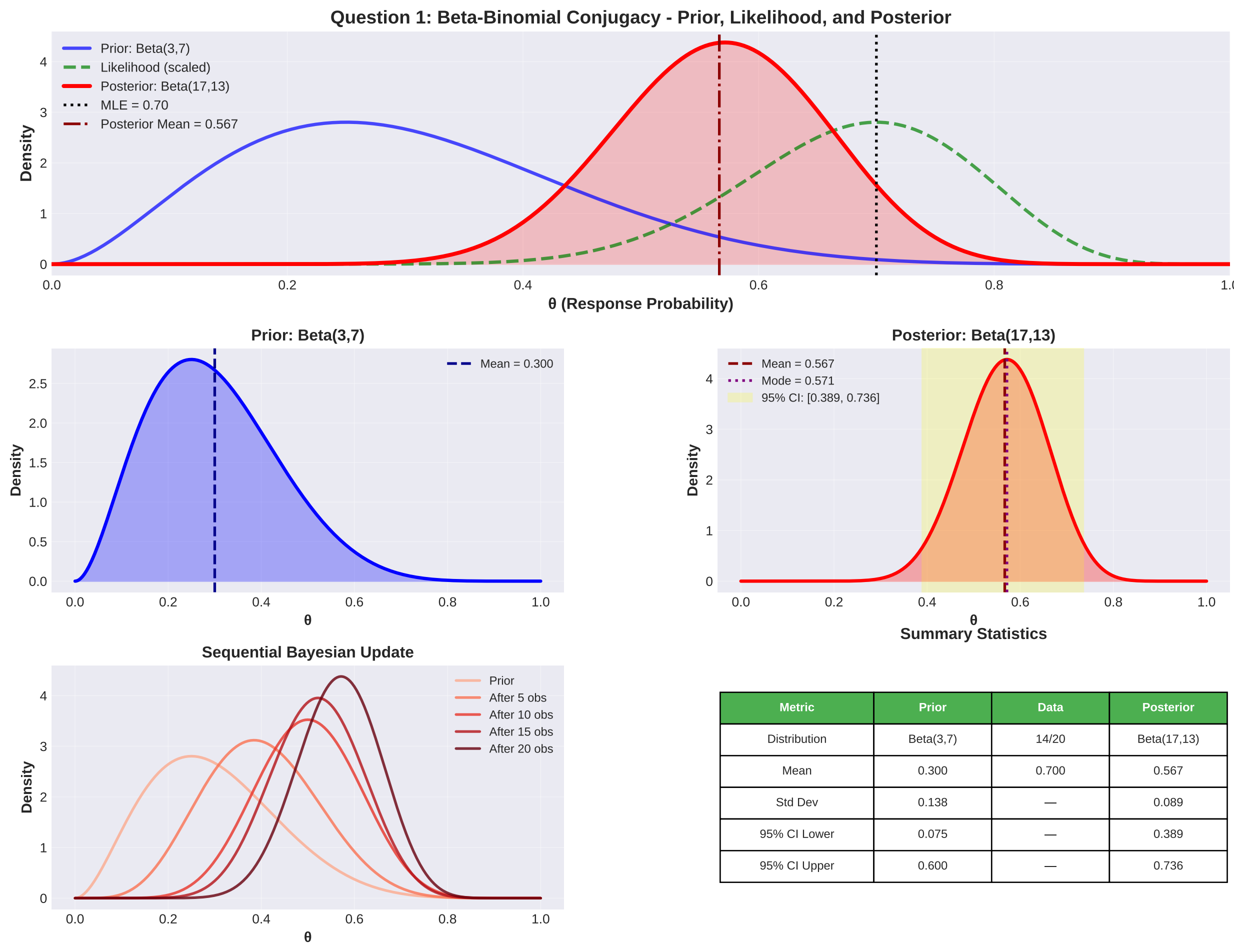
Therefore, the posterior distribution of θ given the observed data is:

$$\boxed{\theta \mid D \sim \text{Beta}(17,\, 13)}$$

This result demonstrates the conjugacy between the Beta prior and the Binomial likelihood, whereby the posterior distribution remains in the Beta family with updated parameters reflecting both prior information and observed data.

Part 2: Visual Explanation of the Solution

To complement the analytical derivation in Question 1, a graphical illustration of the Bayesian updating process is provided in Figure 1.



Metric	Prior	Data	Posterior
Distribution	Beta(3,7)	14/20	Beta(17,13)
Mean	0.300	0.700	0.567
Std Dev	0.138	—	0.089
95% CI Lower	0.075	—	0.389
95% CI Upper	0.600	—	0.736

Figure 1: Bayesian updating for Question 1 showing the prior, likelihood, posterior, sequential updates, and credible interval.

Figure 1 visually demonstrates how prior beliefs are updated using observed data through Bayes theorem.

1. Components of the Bayesian Update

Prior Distribution (Blue Curve): The blue curve represents the prior distribution Beta$(3, 7)$, which encodes the researchers initial belief about the treatment response probability before observing any data. Its mean is:

$$E(\theta) = \frac{3}{3+7} = 0.30,$$

indicating an initial expectation of relatively low treatment effectiveness.

Likelihood Function (Green Dashed Curve): The green dashed curve corresponds to the likelihood induced by the observed data of 14 successes out of 20 trials. It peaks at the Maximum Likelihood Estimate (MLE):

$$\hat{\theta}_{\text{MLE}} = \frac{14}{20} = 0.70,$$

representing the value of θ that best explains the observed data alone.

Posterior Distribution (Red Curve): The red curve shows the posterior distribution $\text{Beta}(17, 13)$ obtained analytically in Question 1. It represents a compromise between the prior belief and the observed data, and its mean is:

$$E(\theta \mid D) = \frac{17}{17+13} \approx 0.567.$$

As seen in the figure, the posterior lies between the prior and the likelihood, reflecting the combined influence of both sources of information.

2. Key Insights from the Visuals

Posterior Mean vs. MLE: While the sample data alone suggest a success probability of 70%, Bayesian inference moderates this estimate by incorporating prior information. As a result, the posterior mean of approximately 0.567 is pulled toward the prior mean of 0.30, illustrating the regularizing effect of the prior.

Sequential Bayesian Updating: The bottom-left panel of Figure 1 shows posterior distributions after observing increasing sample sizes (from 5 to 20 patients). As more data are observed, the posterior distribution becomes increasingly concentrated, indicating reduced uncertainty about θ. This demonstrates a key property of Bayesian inference: uncertainty decreases as evidence accumulates.

Credible Interval (Yellow Shaded Region): The posterior distribution includes a 95% credible interval given by:

$$[0.389,\ 0.736].$$

This interval has a direct probabilistic interpretation: given the data and the prior, there is a 95% probability that the true treatment success probability θ lies within this range.

3. Consistency with the Analytical Results

The summary statistics table in Figure 1 confirms the analytical posterior derived earlier:

$$\alpha^* = \alpha + k = 3 + 14 = 17, \quad \beta^* = \beta + (n - k) = 7 + 6 = 13.$$

Thus, both the mathematical derivation and the graphical evidence consistently lead to the posterior distribution:

$$\theta \mid D \sim \text{Beta}(17, 13).$$

The visual analysis therefore reinforces and validates the analytical solution obtained in Question 1.

Question 2

In a hospital, the number of emergency cases arriving per night is assumed to follow a Poisson distribution:

$$Y \mid \lambda \sim \text{Poisson}(\lambda).$$

Assume a Gamma prior for λ:

$$\lambda \sim \Gamma(\alpha, \beta), \quad \alpha = 4,\ \beta = 2.$$

In one night, $y = 7$ emergency cases are observed.
Required:

(a) Derive the posterior distribution for λ.

(b) Find the posterior mean $E(\lambda \mid y)$.

Solution

Step 1: Define the likelihood

Let Y be the number of emergency cases per night, modeled as Poisson with rate λ:

$$Y \mid \lambda \sim \text{Poisson}(\lambda), \quad \lambda > 0.$$

The likelihood function for observing $Y = y$ is:

$$\underbrace{L(\lambda; y)}_{\text{likelihood}} = f(y \mid \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}.$$

This represents the probability of observing y events given the rate λ.

Step 2: Specify the prior

Assume a Gamma prior on λ:

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0,$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the rate parameter. This prior encodes prior belief about the typical number of emergency cases per night.

Step 3: Formulate the unnormalized posterior

By Bayes' theorem, the posterior density is proportional to the product of the likelihood and the prior:

$$\pi(\lambda \mid y) \propto f(y \mid \lambda)\pi(\lambda).$$

Substitute the explicit forms:

$$\pi(\lambda \mid y) \propto \underbrace{\lambda^y e^{-\lambda}}_{\text{likelihood}} \underbrace{\lambda^{\alpha-1} e^{-\beta\lambda}}_{\text{prior}} = \lambda^{\alpha+y-1} e^{-(\beta+1)\lambda}.$$

Step 4: Recognize the posterior distribution

The unnormalized posterior has the form:

$$\pi(\lambda \mid y) \propto \lambda^{\alpha'-1} e^{-\beta'\lambda}, \quad \lambda > 0,$$

which is the kernel of a Gamma distribution. Hence:

$$\lambda \mid y \sim \Gamma(\alpha', \beta'), \quad \text{where} \quad \alpha' = \alpha + y, \quad \beta' = \beta + 1.$$

Step 5: Update parameters with observed data

Given $\alpha = 4$, $\beta = 2$, and observation $y = 7$:

$$\alpha' = 4 + 7 = 11, \quad \beta' = 2 + 1 = 3.$$

Thus, the posterior distribution is:

$$\boxed{\lambda \mid y \sim \Gamma(11, 3)}$$

Step 6: Compute the posterior mean

For a Gamma distribution with shape α' and rate β', the mean is:

$$E[\lambda \mid y] = \frac{\alpha'}{\beta'}.$$

Substitute the posterior parameters:

$$E[\lambda \mid y] = \frac{11}{3} \approx 3.667.$$

The variance of a Gamma distribution is:

$$\mathrm{Var}(\lambda \mid y) = \frac{\alpha'}{\beta'^2} = \frac{11}{9} \approx 1.222.$$

Step 8: Interpretation and insight

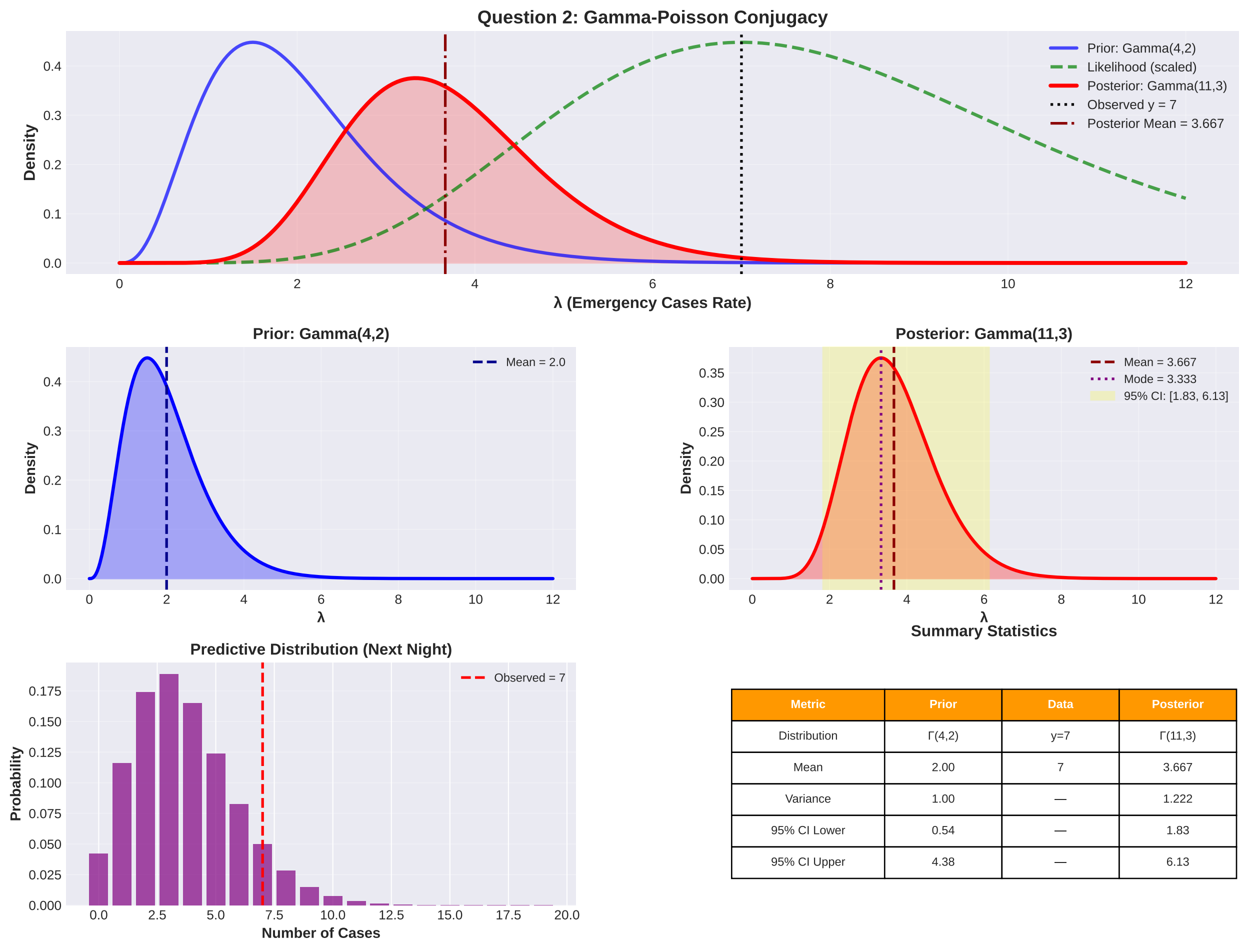
- **Conjugacy:** The Gamma prior is conjugate to the Poisson likelihood, which guarantees that the posterior is also a Gamma distribution.
- **Bayesian updating:** The prior mean was $\alpha/\beta = 2$. Observing $y = 7$ updates the expected rate to $E[\lambda \mid y] = 3.667$, showing the influence of both prior and data.
- **Uncertainty:** The posterior variance ≈ 1.222 quantifies remaining uncertainty about λ after observing the data.

Step 9: Results

$$\begin{aligned}
&\text{Posterior distribution: } \lambda \mid y \sim \Gamma(11, 3),\\
&\text{Posterior mean: } E[\lambda \mid y] \approx 3.667,\\
&\text{Posterior variance: } \mathrm{Var}(\lambda \mid y) \approx 1.222.
\end{aligned}$$

Part 2: Visual Explanation of the Solution

To complement the analytical solution of Question 2, a graphical illustration of the Bayesian updating process for the Poisson Gamma model is provided in Figure 2.



Metric	Prior	Data	Posterior
Distribution	Γ(4,2)	y=7	Γ(11,3)
Mean	2.00	7	3.667
Variance	1.00	—	1.222
95% CI Lower	0.54	—	1.83
95% CI Upper	4.38	—	6.13

Figure 2: Bayesian updating for Question 2 showing the prior, likelihood, posterior, credible interval, and posterior predictive distribution.

Figure 2 illustrates how prior beliefs about the emergency arrival rate are updated after observing new data using Bayes theorem.

1. Distribution Comparison

The main panel of Figure 2 compares the prior distribution, the likelihood induced by the data, and the resulting posterior distribution.

Prior Distribution (Blue Curve): The blue curve represents the prior distribution $\Gamma(4, 2)$, which encodes the initial belief about the nightly emergency arrival rate λ before observing data. This prior has mean:

$$E(\lambda) = \frac{4}{2} = 2.0,$$

and is relatively broad, reflecting substantial uncertainty about the true value of λ.

Likelihood Function (Green Dashed Curve): The green dashed curve represents the likelihood derived from the observed count $y = 7$. As a function of λ, the Poisson likelihood is proportional to:

$$\lambda^7 e^{-\lambda},$$

and it peaks at $\lambda = 7.0$, which corresponds to the Maximum Likelihood Estimate. This curve reflects what the data alone suggest about the arrival rate.

Posterior Distribution (Red Curve): The red curve shows the posterior distribution $\Gamma(11, 3)$ derived analytically in Question 2. It lies between the prior and the likelihood, representing a compromise between prior belief and observed evidence. Compared to the prior, the posterior is taller and narrower, indicating reduced uncertainty after incorporating the data.

2. Uncertainty Quantification

The posterior-specific plot in Figure 2 highlights the 95% credible interval for λ.

The yellow shaded region represents the interval:

$$[1.83,\ 6.13].$$

In Bayesian terms, this interval has a direct probabilistic interpretation: given the observed data and the prior distribution, there is a 95% probability that the true nightly emergency arrival rate λ lies within this range.

This contrasts with frequentist confidence intervals, whose interpretation does not assign probability directly to the parameter.

3. Posterior Predictive Distribution

The bottom-left panel of Figure 2 displays the posterior predictive distribution for the number of emergency cases on the next night.

This distribution integrates over the posterior uncertainty in λ:

$$p(y_{\text{new}} \mid y) = \int p(y_{\text{new}} \mid \lambda)\, p(\lambda \mid y)\, d\lambda.$$

Unlike a single point prediction based on the posterior mean, this predictive distribution accounts for:

- the inherent randomness of Poisson arrivals, and
- uncertainty in the estimated rate parameter λ.

As a result, it provides a more realistic and informative description of future emergency arrival counts.

4. Consistency with the Analytical Results

The visual summary in Figure 2 confirms the analytical posterior obtained earlier:

$$\lambda \mid y \sim \Gamma(11, 3), \quad E(\lambda \mid y) = \frac{11}{3}.$$

Thus, the graphical analysis fully supports the mathematical derivation and illustrates how Bayesian inference combines prior knowledge with observed data to produce both parameter estimates and predictive insights.

Question 3

Suppose a single observation $x = 42$ is recorded from the model:

$$x \mid \theta \sim N(\theta, \sigma^2), \quad \sigma^2 = 9.$$

Assume the prior distribution:

$$\theta \sim \text{Uniform}(0, 100).$$

Required:

(a) Write Bayes theorem for $p(\theta \mid x)$.

(b) Show that the posterior has a Normal kernel.

(c) State the posterior distribution form and explain why it is a truncated Normal distribution.

Solution

We are given a single observation $x = 42$ from the model:

$$x \mid \theta \sim N(\theta, \sigma^2), \quad \sigma^2 = 9,$$

and a prior distribution:

$$\theta \sim \text{Uniform}(0, 100).$$

We are asked to derive and interpret the posterior.

Step 1: Write the likelihood function

For a Normal model with known variance σ^2, the likelihood of θ given the observation x is

$$L(\theta; x) = f(x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}.$$

Here, the likelihood quantifies the plausibility of different values of θ given the observed data.

Step 2: Write Bayes theorem

By Bayes theorem, the posterior distribution is

$$p(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{\int_0^{100} f(x \mid \theta)\pi(\theta)\, d\theta}.$$

Since the prior is Uniform$(0, 100)$:

$$\pi(\theta) = \frac{1}{100}, \quad 0 \leq \theta \leq 100.$$

Substituting the prior:

$$p(\theta \mid x) \propto f(x \mid \theta)\pi(\theta) = \frac{1}{100} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}, \quad 0 \leq \theta \leq 100.$$

The normalization constant ensures that the posterior integrates to 1 over the interval $[0, 100]$.

Step 3: Identify the Normal kernel

Ignoring constants independent of θ, the posterior density has the form

$$p(\theta \mid x) \propto \exp\left\{-\frac{(\theta - x)^2}{2\sigma^2}\right\}, \quad 0 \leq \theta \leq 100.$$

Notice that this is exactly the **kernel of a Normal distribution** with mean x and variance σ^2:

$$\exp\left\{-\frac{(\theta - \mu)^2}{2\tau^2}\right\} \quad \text{is the kernel of } N(\mu, \tau^2).$$

Hence, the posterior **ignoring truncation** is:

$$\theta \mid x \sim N(\mu = x, \tau^2 = \sigma^2) = N(42, 9).$$

Step 4: Truncated Normal form

Because the prior restricts θ to the interval $[0, 100]$, the posterior is **not a standard Normal**; it is **truncated** to this interval.
The posterior density is therefore:

$$p(\theta \mid x) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\theta-x)^2}{2\sigma^2}\right\}}{\int_0^{100} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\theta-x)^2}{2\sigma^2}\right\} d\theta}, \quad 0 \leq \theta \leq 100.$$

Equivalently:

$$\boxed{\theta \mid x \sim \text{Truncated Normal}\left(\mu = 42, \sigma^2 = 9, \text{lower} = 0, \text{upper} = 100\right)}.$$

Step 5: Explanation and interpretation

1. **Normal kernel:** The likelihood for a Normal model produces an exponential quadratic in θ, which is the defining kernel of a Normal distribution. This shows that the posterior is Gaussian-shaped.

2. **Effect of the prior:** The uniform prior acts as a flat prior over $[0, 100]$. It does not affect the shape of the posterior within the interval but truncates it outside. This is why the posterior is a **truncated Normal** rather than an unconstrained Normal.

3. **Bayesian updating:** The posterior mean is at $x = 42$, which coincides with the maximum likelihood estimate (MLE) in the case of a flat prior. The variance $\sigma^2 = 9$ is unchanged because we have only one observation and a uniform prior; the data dominate the posterior.

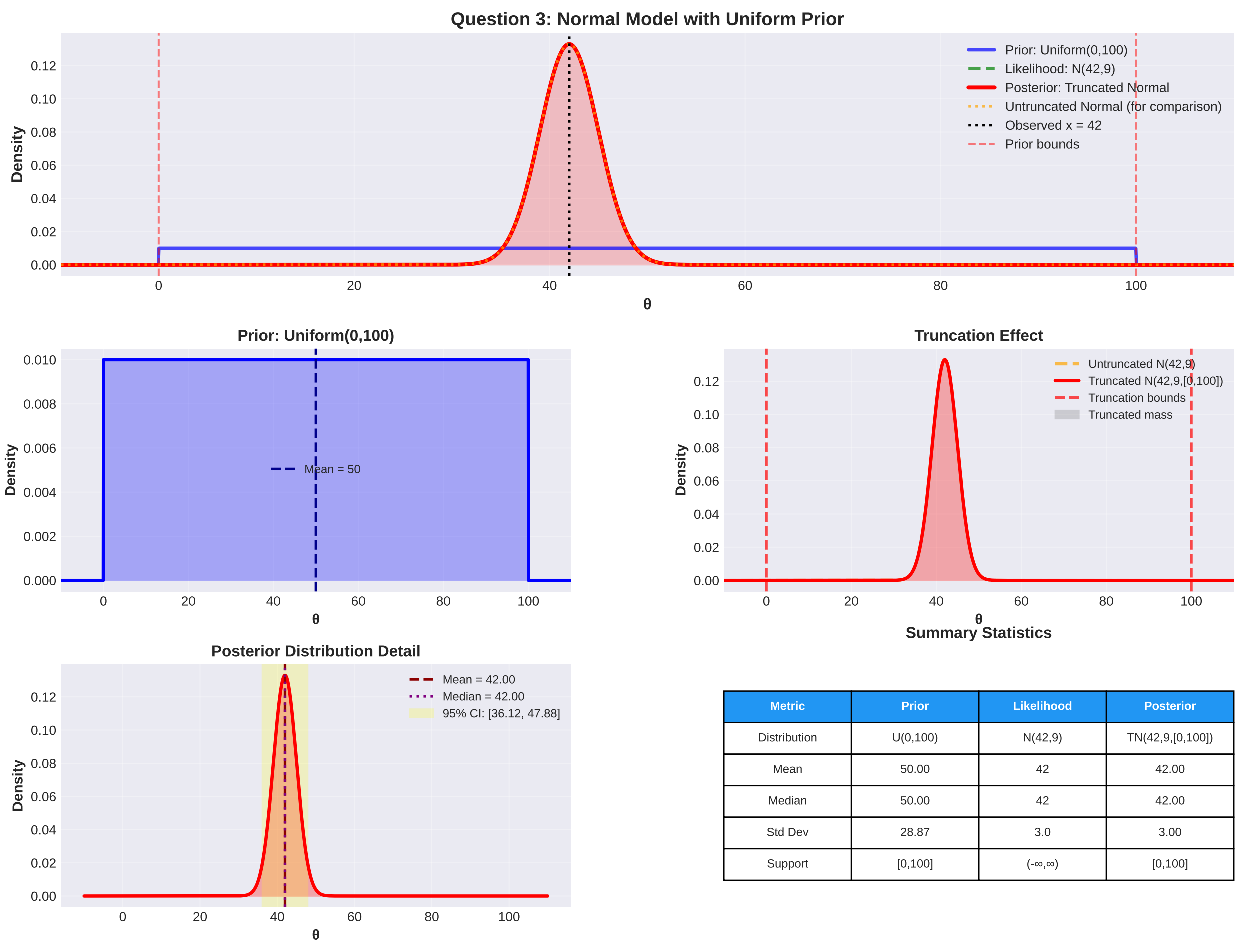
4. **Truncation effects:** If x were close to the bounds (0 or 100), the posterior would be skewed due to truncation. In this case, $x = 42$ is well within $(0, 100)$, so truncation has minimal effect on the posterior mean and variance.

Step 6: Results

$$\text{Posterior kernel: } p(\theta \mid x) \propto \exp\left\{-\frac{(\theta-42)^2}{18}\right\}, \quad 0 \le \theta \le 100,$$
$$\text{Posterior distribution: } \theta \mid x \sim \text{Truncated Normal}(42, 9, 0, 100),$$
$$\text{Reason for truncation: prior restriction } \theta \in [0, 100].$$

Part 2: Visual Explanation of the Solution

To complement the analytical solution of Question 3, Figure 3 provides a graphical illustration of Bayesian updating when a Uniform prior is combined with a Normal likelihood.



Metric	Prior	Likelihood	Posterior
Distribution	U(0,100)	N(42,9)	TN(42,9,[0,100])
Mean	50.00	42	42.00
Median	50.00	42	42.00
Std Dev	28.87	3.0	3.00
Support	[0,100]	(-∞,∞)	[0,100]

Figure 3: Bayesian updating for Question 3 showing the prior distribution, posterior distribution, truncation effect, and credible interval.

Figure 3 illustrates how a single observation refines a broad prior belief through the likelihood function.

1. Prior and Posterior Comparison

Prior Distribution (Uniform$(0, 100)$): The prior distribution is uniform over the interval $[0, 100]$, resulting in a constant density of:

$$p(\theta) = \frac{1}{100} = 0.010.$$

This flat line, shown in Figure 3, represents complete prior ignorance regarding the true value of θ within the specified bounds.

Posterior Distribution: After observing a single data point $x = 42$, the posterior distribution becomes sharply concentrated around this value. The likelihood function has a Normal kernel centered at 42, and hence the posterior distribution is centered at the observed value, reflecting the strong influence of the data when the prior is non-informative.

2. The Truncation Effect

The truncation effect is clearly illustrated in the truncation plot in Figure 3. The dashed red vertical lines at $\theta = 0$ and $\theta = 100$ indicate the boundaries imposed by the Uniform prior. Although the underlying likelihood corresponds to an unbounded Normal distribution, the posterior is restricted to the interval $[0, 100]$. As a result, the posterior distribution is a *truncated Normal distribution*, with all probability mass lying strictly within the prior bounds.

3. Summary Statistics and Credible Interval

The summary statistics shown in Figure 3 confirm the analytical results derived earlier.
The posterior mean and median both shift from the prior value of 50.00 to the observed value:

$$E(\theta \mid x) = 42.00,$$

indicating that the data completely dominates the prior belief.
The 95% credible interval for θ is:

$$[36.12,\ 47.88].$$

In Bayesian terms, this interval has a direct probabilistic interpretation: given the observed data and the prior distribution, there is a 95% probability that the true value of θ lies within this interval.

4. Interpretation

The visual analysis in Figure 3 demonstrates how Bayesian inference transforms a non-informative prior into a concentrated posterior using even a single observation. The truncation imposed by the prior ensures parameter validity, while the likelihood determines the shape and location of the posterior.