Normal Likelihood, Normal Prior

Posterior for the Mean (Known Variance) + Conjugacy + Interpretation

Bayesian Inference and Computation (Lecture)

Today

- Today is algebra-heavy: we will derive a posterior distribution by hand.
- We focus on a simplified but very common case:

$$Y_1, \ldots, Y_N \sim \mathcal{N}(\mu, \sigma^2), \quad \mu \textbf{ unknown, } \sigma^2 \textbf{ known.}$$

- Why simplify?
  - Normal model appears everywhere.
  - Two-parameter case (μ, σ^2) is doable but messier (worksheet / later).
  - This case isolates the key trick: **collect terms** + **complete the square**.

Setup and goal

Data model

$$Y_1, \ldots, Y_N \mid \mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2), \qquad \sigma > 0 \text{ known.}$$

Let $y = (y_1, \ldots, y_N)$ denote the observed data.

Goal

Use Bayes' theorem:

$$\pi(\mu \mid y) \propto \pi(y \mid \mu)\, \pi(\mu).$$

- We will work *up to proportionality* (ignore constants not involving μ).

Example 3.5: Normal prior on μ

Prior

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2).$$

- μ_0 = prior mean (best guess before seeing data).
- σ_0^2 = prior variance (how uncertain you are).
- Vague prior: choose σ_0^2 very large (e.g. 10^6).

Likelihood function (Example 3.5)

Because the observations are independent,

$$\pi(y \mid \mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}$$

$$= (2\pi\sigma^2)^{-N/2} \exp\left\{-\sum_{i=1}^{N} \frac{(y_i-\mu)^2}{2\sigma^2}\right\}.$$

Up to proportionality in μ

Drop the leading constant $(2\pi\sigma^2)^{-N/2}$:

$$\pi(y \mid \mu) \propto \exp\left\{-\sum_{i=1}^{N} \frac{(y_i-\mu)^2}{2\sigma^2}\right\}.$$

Prior density (Example 3.5)

$$\pi(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \propto \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}.$$

Bayes (up to proportionality)

$$\pi(\mu \mid y) \propto \exp\left\{-\sum_{i=1}^{N} \frac{(y_i-\mu)^2}{2\sigma^2}\right\} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}.$$

Posterior: expand the exponent and drop constants

Combine the exponents:

$$\pi(\mu \mid y) \;\propto\; \exp\left\{-\sum_{i=1}^{N}\frac{(y_i-\mu)^2}{2\sigma^2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}.$$

Expand terms (grouping by μ):

$$\begin{aligned}
&-\sum_{i=1}^{N}\frac{(y_i-\mu)^2}{2\sigma^2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\\
&=\underbrace{-\sum_{i=1}^{N}\frac{y_i^2}{2\sigma^2}}_{\text{no }\mu}+\mu\underbrace{\left(\frac{\sum_{i=1}^{N} y_i}{\sigma^2}\right)}_{\text{linear}}-\mu^2\underbrace{\left(\frac{N}{2\sigma^2}\right)}_{\text{quadratic}}-\mu^2\left(\frac{1}{2\sigma_0^2}\right)+\mu\left(\frac{\mu_0}{\sigma_0^2}\right)-\underbrace{\frac{\mu_0^2}{2\sigma_0^2}}_{\text{no }\mu}.
\end{aligned}$$

Drop constants

The first and last terms do not depend on μ, so they vanish into the proportionality constant.

Arranging

After dropping constants,

$$\pi(\mu \mid y) \;\propto\; \exp\left\{-\mu^2\left(\frac{N}{2\sigma^2}+\frac{1}{2\sigma_0^2}\right)+\mu\left(\frac{\sum_{i=1}^{N} y_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)\right\}.$$

Define (as in the notes)

$$a \;=\; \left(\frac{\sum_{i=1}^{N} y_i}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right), \qquad b^2 \;=\; \left(\frac{N}{\sigma^2}+\frac{1}{\sigma_0^2}\right)^{-1} \;=\; \frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}.$$

Then

$$\pi(\mu \mid y) \;\propto\; \exp\left\{-\frac{\mu^2}{2b^2}+\mu\, a\right\}.$$

Complete the square

Consider the exponent:

$$-\frac{\mu^2}{2b^2} + \mu a.$$

Complete the square:

$$-\frac{\mu^2}{2b^2} + \mu a = -\frac{1}{2b^2}\left(\mu - ab^2\right)^2 + \frac{a^2 b^2}{2}.$$

Drop the constant again

$$\pi(\mu \mid y) \propto \exp\left\{-\frac{1}{2b^2}\left(\mu - ab^2\right)^2\right\}.$$

This is the kernel of a Normal distribution.

Posterior distribution (Example 3.5)

Therefore,

$$\mu \mid Y \sim \mathcal{N}(\mu_{\text{post}},\ b^2), \qquad \text{where} \quad \mu_{\text{post}} := ab^2.$$

Posterior mean (expanded)

$$\mu_{\text{post}} = ab^2 = \frac{\sigma_0^2 \sum_{i=1}^N y_i + \mu_0 \sigma^2}{N\sigma_0^2 + \sigma^2} = \underbrace{\frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}}_{\text{weight on } \bar{y}} \bar{y} + \underbrace{\frac{\sigma^2}{N\sigma_0^2 + \sigma^2}}_{\text{weight on } \mu_0} \mu_0, \qquad \bar{y} = \frac{1}{N}\sum_{i=1}^N y_i.$$

Posterior variance

$$b^2 = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}.$$

Interpretation: precision-weighted averaging

Rewrite μ_{post} using precisions:

$$\mu_{\text{post}} = \frac{\frac{N}{\sigma^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}}\,\bar{y} + \frac{\frac{1}{\sigma_0^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}}\,\mu_0.$$

Precision = inverse variance

- Prior precision: $1/\sigma_0^2$.
- Sample mean (MLE): $\hat{\mu}_{\text{MLE}} = \bar{y}$ and

$$\bar{y} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right) \quad \Rightarrow \quad \text{precision} = \frac{N}{\sigma^2}.$$

Memory aid

$$\frac{1}{b^2} = \frac{N}{\sigma^2} + \frac{1}{\sigma_0^2},$$

Sanity checks: limiting cases

- **Vague prior** ($\sigma_0^2 \to \infty$):

$$\mu_{\text{post}} \to \bar{y}, \qquad b^2 \to \frac{\sigma^2}{N}.$$

Sanity checks: limiting cases

- **Vague prior** ($\sigma_0^2 \to \infty$):

$$\mu_{\text{post}} \to \bar{y}, \qquad b^2 \to \frac{\sigma^2}{N}.$$

- **Very strong prior** ($\sigma_0^2 \to 0$):

$$\mu_{\text{post}} \to \mu_0, \qquad b^2 \to 0.$$

Sanity checks: limiting cases

- **Vague prior** ($\sigma_0^2 \to \infty$):

$$\mu_{\text{post}} \to \bar{y}, \qquad b^2 \to \frac{\sigma^2}{N}.$$

- **Very strong prior** ($\sigma_0^2 \to 0$):

$$\mu_{\text{post}} \to \mu_0, \qquad b^2 \to 0.$$

- **More data** ($N \uparrow$):

$$b^2 = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1} \downarrow,$$

  so uncertainty about μ shrinks as we observe more data.

Credible intervals (equal-tailed, level $1-\alpha$)

Let $\alpha < 1/2$. We want $l, u \in \mathbb{R}$ such that

$$\pi(\mu < l \mid y) = \pi(\mu > u \mid y) = \alpha/2, \quad \Rightarrow \quad \pi(\mu \in [l, u] \mid y) = 1 - \alpha.$$

Since $\mu \mid y \sim \mathcal{N}(\mu_{\text{post}}, b^2)$,

$$Z := \frac{\mu - \mu_{\text{post}}}{b} \sim \mathcal{N}(0, 1).$$

Thus

$$u = \mu_{\text{post}} + b\,\Phi^{-1}(1 - \alpha/2), \qquad l = \mu_{\text{post}} - b\,\Phi^{-1}(1 - \alpha/2).$$

Credible interval

$$\boxed{\mu_{\text{post}} \pm b\,\Phi^{-1}(1 - \alpha/2)}$$

Large N link to confidence intervals

If N is large (or prior is vague), then $\mu_{\text{post}} \approx \bar{y}$ and $b^2 \approx \sigma^2/N$, so the Bayesian credible interval is approximately

$$\bar{y} \pm \frac{\sigma}{\sqrt{N}} \Phi^{-1}(1 - \alpha/2),$$

which matches the usual Normal-theory confidence interval (known σ).

But...

For small N or strong prior belief, credible and confidence intervals can differ substantially.

Conjugate priors: the "cheat code" (with a warning)

Definition (Conjugate prior)

A prior $\pi(\theta)$ is **conjugate** for a likelihood $\pi(y \mid \theta)$ if the posterior $\pi(\theta \mid y)$ has the **same distributional form** as the prior.

Here

Normal likelihood (known σ^2) $+$ Normal prior on $\mu \implies$ Normal posterior for μ.

Important warning

Conjugacy gives algebraic convenience, not correctness. Choose priors because they are sensible for the application—not just because they make the maths tidy.

Compute it in R: simulate + posterior parameters (Example 3.5)

```r
set.seed(1)

N <- 30
sigma2 <- 1            # known variance
mu_true <- 5

# simulate data
y <- rnorm(N, mean = mu_true, sd = sqrt(sigma2))

# prior on mu: Normal(mu0, sigma0^2)
mu0 <- 0
sigma0 <- 1000
sigma0_2 <- sigma0^2

# Example 3.5 definitions
a <- sum(y)/sigma2 + mu0/sigma0_2
b2 <- 1 / (N/sigma2 + 1/sigma0_2)
```

Posterior distribution plot (conceptual code)

```
# grid for plotting posterior of mu
grid <- seq(4, 6, length.out = 400)

# posterior density: Normal(mu_post, b2)
post <- dnorm(grid, mean = mu_post, sd = sqrt(b2))

plot(grid, post, type="l",
     main="Posterior for mu",
     xlab="mu", ylab="density")
abline(v = mu_post, lty = 2)
```

Interpretation

This curve is your **belief distribution** for μ after seeing the data: not just a point estimate, but a full distribution you can query.

Credible interval (95%) in R

Equal-tailed credible interval

For $\alpha = 0.05$,

$$[l, u] = \mu_{\text{post}} \pm b\,\Phi^{-1}(1 - \alpha/2), \qquad b = \sqrt{b^2}.$$

```
alpha <- 0.05
b <- sqrt(b2)

l <- mu_post - b * qnorm(1 - alpha/2)
u <- mu_post + b * qnorm(1 - alpha/2)

c(l, u)
```

- You can also compute probabilities like $\Pr(\mu > 5 \mid y)$ via `pnorm`.

History break: David Blackwell

Who?

David Harold Blackwell (1919–2010):

- major contributions to Bayesian statistics, information theory, and game theory,
- known for work on decision theory and games with partial information.

Why mention him here?

Bayesian inference is not just "a method":

- it links **belief updating** to **decision making** under uncertainty.

Blackwell: the decision-theory connection

Bayes gives a posterior; decisions need a loss

- Posterior $\pi(\theta \mid y)$ tells you uncertainty about θ.
- To choose an action a, introduce a loss $L(a, \theta)$.
- Bayesian decision rule: choose a minimizing posterior expected loss

$$a^\star \in \arg\min_a \mathbb{E}[L(a, \theta) \mid y] \, .$$

- This is the bridge between inference (what you believe) and action (what you do).

Wrap-up

Main result (Example 3.5)

If $Y_i \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$ with known σ^2 and $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, then

$$\mu \mid y \sim \mathcal{N}(\mu_{\text{post}}, b^2), \quad b^2 = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}, \quad \mu_{\text{post}} = \frac{\frac{N}{\sigma^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}}\bar{y} + \frac{\frac{1}{\sigma_0^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}}\mu_0.$$

What to focus on

- Posterior mean is a precision-weighted average of sample mean and prior mean.
- Posterior variance shrinks as information increases.
- Conjugacy is convenient, but prior choice should be principled.

Questions?

Thanks everyone!