STAT 431 — Applied Bayesian Analysis — Course Notes

# Conjugate Priors for One-Parameter Normal Models

Fall 2022

- Model:

$$Y_1, \ldots Y_n \mid \mu, \sigma^2 \quad \sim \quad iid \ \text{Normal}(\mu, \sigma^2)$$

Let

$$\boldsymbol{Y} \quad = \quad (Y_1, \ldots Y_n)$$

$$\boldsymbol{y} \quad = \quad (y_1, \ldots y_n) \qquad \text{(observation of } \boldsymbol{Y}\text{)}$$

$$\bar{y} \quad = \quad \frac{1}{n} \sum_i y_i \quad = \quad \text{usual estimate of } \mu$$

The **precision** is defined as

$$\tau^2 \;=\; 1/\sigma^2$$

which

- measures *concentration*, not spread
- can lead to less complicated derivations (later)
- is used in an alternative parameterization, especially in some Bayesian software

(Note: BSM denotes precision as "$\tau$" rather than $\tau^2$.)

# Known Variance

Assume $\sigma^2$ (or $\tau^2$) is known.

- Likelihood

  Joint PDF of $\boldsymbol{Y}$:

$$
\begin{aligned}
f(\boldsymbol{y} \mid \mu) &= \prod_i f(y_i \mid \mu) \\
&= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\
&\propto e^{-\frac{1}{2\sigma^2}\sum_i (y_i-\mu)^2}
\end{aligned}
$$

  (where the proportionality is in $\mu$)

Can show

$$\sum_i (y_i - \mu)^2 \quad = \quad \sum_i (y_i - \bar{y})^2 \; + \; n(\bar{y} - \mu)^2$$

so

$$\begin{aligned}
\text{likelihood} \quad &\propto \quad e^{-\frac{1}{2\sigma^2}\left(\sum_i (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right)} \\
&= \quad e^{-\frac{1}{2\sigma^2}\sum_i (y_i - \bar{y})^2} \;\cdot\; e^{-\frac{1}{2\sigma^2} n(\bar{y} - \mu)^2} \\
&\propto \quad e^{-\frac{1}{2\sigma^2} n(\mu - \bar{y})^2}
\end{aligned}$$

(where the proportionality is in $\mu$)

Note: $\bar{y}$ is a **sufficient statistic**. (How can you tell?)

[ Draw likelihood ... ]

- Conjugate Prior

$$\mu \quad \sim \quad \mathrm{Normal}(\mu_0, \sigma_0^2) \quad = \quad \mathrm{Normal}\big(\mu_0, 1/\tau_0^2\big)$$

Why is this conjugate? Let's derive the posterior ...

$$\begin{aligned} p(\mu \mid \boldsymbol{y}) \quad &\propto \quad f(\boldsymbol{y} \mid \mu)\, \pi(\mu) \\ &\propto \quad e^{-\frac{1}{2\sigma^2} n(\mu - \bar{y})^2} \ \cdot \ e^{-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2} \\ &= \quad e^{-\frac{1}{2}\left(n\tau^2(\mu - \bar{y})^2 \,+\, \tau_0^2(\mu - \mu_0)^2\right)} \end{aligned}$$

The exponent is a concave quadratic function of $\mu$, and thus the expression is the kernel of a normal distribution.

Next we identify the posterior mean and variance ...

$$
\begin{aligned}
n\tau^2(\mu-\bar{y})^2 + \tau_0^2(\mu-\mu_0)^2 \\
&= \; (n\tau^2+\tau_0^2)\mu^2 \; - \; 2(n\tau^2\bar{y}+\tau_0^2\mu_0)\mu \\
&\quad + \; \text{constant (without } \mu\text{)}
\end{aligned}
$$

$$
\begin{aligned}
n\tau^2(\mu-\bar{y})^2 + \tau_0^2(\mu-\mu_0)^2 & \\
&= (n\tau^2+\tau_0^2)\mu^2 \;-\; 2(n\tau^2\bar{y}+\tau_0^2\mu_0)\mu \\
&\quad + \text{ constant (without } \mu) \\
&= \cdots \;\; \text{(complete the square)} \;\; \cdots \\
&= (n\tau^2+\tau_0^2)\left(\mu \;-\; \frac{n\tau^2\bar{y}+\tau_0^2\mu_0}{n\tau^2+\tau_0^2}\right)^2 \\
&\quad + \text{ constant (without } \mu)
\end{aligned}
$$

$$n\tau^2(\mu - \bar{y})^2 + \tau_0^2(\mu - \mu_0)^2$$

$$= (n\tau^2 + \tau_0^2)\mu^2 - 2(n\tau^2\bar{y} + \tau_0^2\mu_0)\mu + \text{constant (without } \mu)$$

$$= \cdots \text{ (complete the square) } \cdots$$

$$= (n\tau^2 + \tau_0^2)\left(\mu - \frac{n\tau^2\bar{y} + \tau_0^2\mu_0}{n\tau^2 + \tau_0^2}\right)^2 + \text{constant (without } \mu)$$

$$= \tau_1^2(\mu - \mu_1)^2 \quad + \text{constant (without } \mu)$$

where

$$\tau_1^2 = n\tau^2 + \tau_0^2 \qquad\qquad \mu_1 = \frac{n\tau^2\bar{y} + \tau_0^2\mu_0}{n\tau^2 + \tau_0^2}$$

So we find

$$p(\mu \mid \boldsymbol{y}) \quad \propto \quad e^{-\frac{1}{2}\tau_1^2(\mu-\mu_1)^2}$$

which we recognize as the kernel of a $\mathrm{Normal}(\mu_1, 1/\tau_1^2)$:

$$\mu \mid \boldsymbol{y} \quad \sim \quad \mathrm{Normal}(\mu_1, 1/\tau_1^2)$$

So

$$\mathrm{E}(\mu \mid \boldsymbol{y}) \quad = \quad \mu_1 \qquad\qquad \mathrm{Var}(\mu \mid \boldsymbol{y}) \quad = \quad 1/\tau_1^2 \quad \equiv \quad \sigma_1^2$$

The posterior mean estimate of $\mu$ is thus $\mu_1$, with a posterior standard deviation of $\sigma_1$.

(Notice: The posterior depends on the data values only through the sufficient statistic $\bar{y}$.)

Notice that $\mu_1$ is a weighted average of the sample average $\bar{y}$ and prior mean $\mu_0$:

$$
\begin{aligned}
\mu_1 &= \frac{n\tau^2}{n\tau^2 + \tau_0^2}\,\bar{y} \;+\; \frac{\tau_0^2}{n\tau^2 + \tau_0^2}\,\mu_0 \\
&= w\,\bar{y} \;+\; (1-w)\,\mu_0
\end{aligned}
$$

(What happens as $\tau_0^2 \to 0$? As $n \to \infty$?)

Notice that $\mu_1$ is a weighted average of the sample average $\bar{y}$ and prior mean $\mu_0$:

$$\begin{aligned} \mu_1 &= \frac{n\tau^2}{n\tau^2 + \tau_0^2}\,\bar{y} \;+\; \frac{\tau_0^2}{n\tau^2 + \tau_0^2}\,\mu_0 \\ &= w\,\bar{y} \;+\; (1-w)\,\mu_0 \end{aligned}$$

(What happens as $\tau_0^2 \to 0$? As $n \to \infty$?)

Note: $\mu_1$ is generally biased as an estimator of $\mu$. (Why?)

Letting $m = \tau_0^2/\tau^2$,

$$\mu_1 \quad = \quad \frac{n}{n+m}\,\bar{y} \;+\; \frac{m}{n+m}\,\mu_0$$

Interpret:

$$\begin{aligned} m &= \text{“prior sample size”} \\ \mu_0 &= \text{“prior average”} \end{aligned}$$

Also, $\tau_1^2 = (n+m)\,\tau^2$. (What if $n \to \infty$?)

Example: Jevons's Coin Data

- coins (gold sovereigns) collected in England ca. 1870
- legal standard weight: 7.9876 g
- min. legal weight: 7.9379 g

For $n = 24$ coins minted before 1830,

$$\begin{aligned} \bar{y} &= \text{avg. wt.} = 7.8730 \text{ g} \\ s &= \text{sample std. dev.} = 0.05353 \text{ g} \end{aligned}$$

For illustration, let's *assume*

$$\sigma^2 = s^2 = (0.05353)^2$$

Let's take a normal prior with

$$\begin{aligned} \mu_0 &= \text{standard weight} = 7.9876 \\ \sigma_0^2 &= (0.025)^2 \end{aligned}$$

(so that $\sigma_0$ is about half the difference between the standard and minimum legal weights)

How informative is this prior?

$$m = \frac{\tau_0^2}{\tau^2} = \frac{1/(0.025)^2}{1/(0.05353)^2} \approx 4.6$$

so equivalent to 4 or 5 "prior observations."

Posterior is normal with

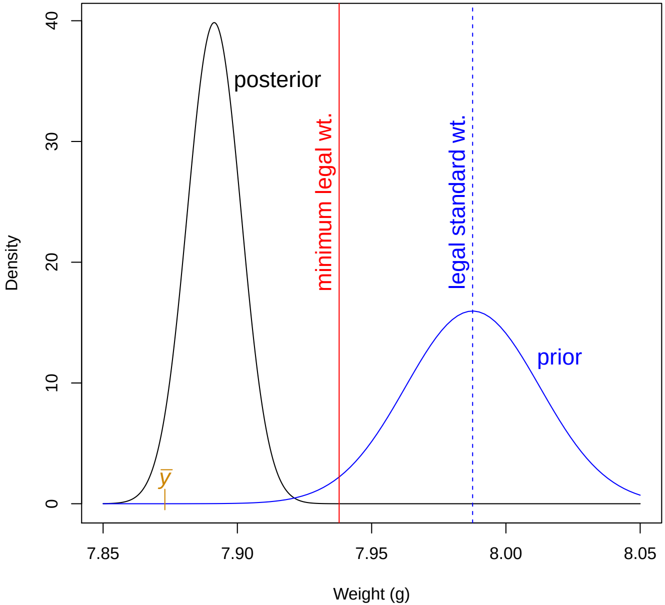
$$\mu_1 = \frac{n\tau^2 \bar{y} + \tau_0^2 \mu_0}{n\tau^2 + \tau_0^2} \approx 7.891381$$

$$\sigma_1^2 = \frac{1}{n\tau^2 + \tau_0^2} \approx 0.0001002444$$

$$(\sigma_1 \approx 0.01001221)$$

So $\bar{y} = 7.8730$ is barely within 2 posterior standard deviations of the posterior mean.

Perhaps our prior is a bit too informative (too much bias)?


posterior
minimum legal wt.
legal standard wt.
prior
$\bar{y}$
Density
Weight (g)
0
10
20
30
40
7.85
7.90
7.95
8.00
8.05

- Posterior Predictive Distribution

  Let $Y^*$ be a hypothetical new observation sampled independently of the data (conditional on $\mu$).

  Then

$$Y^* \mid \mu \;\;=\;\; Y^* \mid \mu, \boldsymbol{y} \;\;\sim\;\; \mathrm{Normal}(\mu, \sigma^2)$$

  and we can write

$$Y^* \;\;=\;\; \mu \,+\, \varepsilon^* \qquad\qquad \varepsilon^* \;\sim\; \mathrm{Normal}(0, \sigma^2)$$

  where $\varepsilon^*$ is independent of $\mu$, $\boldsymbol{Y}$. (Why?)

So

$$\mu \mid \boldsymbol{y} \;\sim\; \text{Normal}(\mu_1, \sigma_1^2)$$
$$\varepsilon^* \mid \boldsymbol{y} \;\sim\; \text{Normal}(0, \sigma^2)$$

and $\mu$ and $\varepsilon^*$ are conditionally independent given $\boldsymbol{Y}$.

This makes it easy to find the posterior predictive distribution:

$$Y^* \mid \boldsymbol{y} \;=\; \mu + \varepsilon^* \mid \boldsymbol{y} \;\sim\; \text{Normal}(\mu_1, \sigma_1^2 + \sigma^2)$$

(Why?)

Note: This distribution always has variance at least $\sigma^2$, no matter how small $\sigma_1^2$ is.

Eg: Jevons's Coin Data

Consider randomly selecting another coin of the same kind (minted before 1830). Its (random) weight will be $Y^*$.

Under the posterior obtained previously,

$$Y^* \mid \boldsymbol{y} \;\; \sim \;\; \mathrm{Normal}\big(7.891381,\, 0.0001002444 + (0.05353)^2\big)$$

The posterior predictive standard deviation works out to be about $0.05446$.

For example, the posterior predictive prob. that *this coin* is of legal weight:

$$\begin{aligned} \mathrm{Prob}(Y^* \geq 7.9379 \mid \boldsymbol{y}) \;\; &\approx \;\; 1 \,-\, \Phi\left(\frac{7.9379 - 7.891381}{0.05446}\right) \\ &\approx \;\; 0.1965 \end{aligned}$$

Remark:

For a posterior predictive check, we might compare the posterior predictive PDF with a histogram of the original data.

Any serious mismatch might indicate a problem with the prior density we are using.

(Unfortunately, Jevons's original data values are not available.)

# Known Mean

Now assume $\mu$ is known, but not $\sigma^2$.

- Likelihood

$$\begin{aligned} f(\boldsymbol{y} \mid \sigma^2) &= \prod_i f(y_i \mid \sigma^2) \\ &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\, e^{-\frac{1}{2\sigma^2}\sum_i (y_i-\mu)^2} \quad \propto \quad \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{SSE}{2\sigma^2}} \end{aligned}$$

in terms of sufficient statistic

$$SSE = \sum_i (y_i - \mu)^2$$

So

$$\text{likelihood} \quad \propto \quad \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{SSE}{2\sigma^2}} \qquad\qquad \sigma^2 > 0$$

[ Draw likelihood ... ]

(Can show $SSE/n$ is the MLE.)

- Conjugate Prior

  We say $X$ has an **inverse gamma distribution** with parameters $\alpha > 0$ and $\beta > 0$ if it has (continuous) density

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{1}{x^{\alpha+1}} e^{-\beta/x} \qquad x > 0$$

  and write

$$X \sim \text{InvGamma}(\alpha, \beta)$$

If

$$X \sim \mathrm{InvGamma}(\alpha, \beta)$$

it can be shown that

- $$1/X \sim \mathrm{Gamma}(\alpha, \beta)$$

  (in the parameterization of BSM, Appendix A.1)

- if $\alpha > 1$,

  $$\mathrm{E}(X) = \frac{\beta}{\alpha - 1}$$

- if $\alpha > 2$,

  $$\mathrm{Var}(X) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}$$

The inverse gamma distribution is a conjugate prior for $\sigma^2$:

Suppose

$$\sigma^2 \quad \sim \quad \mathrm{InvGamma}(\alpha, \beta)$$

Then (for $\sigma^2 > 0$)

$$\begin{aligned} p(\sigma^2 \mid \boldsymbol{y}) \quad &\propto \quad \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{SSE}{2\sigma^2}} \;\cdot\; \frac{1}{(\sigma^2)^{\alpha+1}}\, e^{-\beta/\sigma^2} \\ &= \quad \frac{1}{(\sigma^2)^{n/2\,+\,\alpha\,+\,1}}\, e^{-(SSE/2\,+\,\beta)/\sigma^2} \end{aligned}$$

which is the kernel of $\mathrm{InvGamma}(n/2\,+\,\alpha,\, SSE/2\,+\,\beta)$:

$$\sigma^2 \mid \boldsymbol{y} \quad \sim \quad \mathrm{InvGamma}(n/2\,+\,\alpha,\, SSE/2\,+\,\beta)$$

We could alternatively consider the *reparameterization*

$$\tau^2 \;\; = \;\; 1/\sigma^2$$

It follows that the prior

$$\tau^2 \;\; \sim \;\; \mathrm{Gamma}(\alpha, \beta)$$

produces posterior

$$\tau^2 \mid \boldsymbol{y} \;\; \sim \;\; \mathrm{Gamma}(n/2 + \alpha,\, SSE/2 + \beta)$$

so the gamma distribution is conjugate for this situation.