STAT 431 — Applied Bayesian Analysis — Course Notes

The Two-Parameter Normal Model

Spring 2019

- Model:

$$Y_1, \ldots Y_n \mid \mu, \sigma^2 \quad \sim \quad \text{i.i.d. } \mathrm{N}(\mu,\, \sigma^2 = 1/\tau^2)$$

Let

$$\boldsymbol{y} \;=\; (y_1, \ldots y_n) \qquad \text{(observed version of } \boldsymbol{Y}\text{)}$$

$$\bar{y} \;=\; \frac{1}{n}\sum_i y_i \;=\; \text{usual est. of } \mu$$

$$s^2 \;=\; \frac{1}{n-1}\sum_i (y_i - \bar{y})^2$$

$$=\; \text{usual unbiased est. of } \sigma^2 \text{ (for } n > 1\text{)}$$

Both μ and σ^2 are unknown.

- Likelihood:

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

You can show that

$$
\sum_i (y_i - \mu)^2 = (n-1)s^2 + n(\bar{y} - \mu)^2
$$

so

$$
L(\mu, \sigma^2;\, \boldsymbol{y}) \propto \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{n-1}{2}\, s^2/\sigma^2} \cdot e^{-\frac{n}{2\sigma^2}(\mu-\bar{y})^2}
$$

Note: $(\bar{y}, s^2)$ is sufficient for (μ, σ^2) (why?)

Recall the MLEs (for $n > 1$):

$$\hat{\mu} \;=\; \bar{y} \qquad\qquad \hat{\sigma}^2 \;=\; \frac{n-1}{n}\, s^2$$

If there is a conjugate prior, it will be *joint* in μ and σ^2. We will need some distribution theory ...

Some Useful Distributions

X has a t-**distribution** with **location** μ, **scale** σ^2, and **degrees of freedom** $\nu > 0$ if it has (continuous) density

$$p(x) \;=\; \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi\sigma^2}} \left(1 \,+\, \frac{1}{\nu}\,\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

for $-\infty < x < \infty$.

(Note: $\mu = 0$ and $\sigma^2 = 1$ give the usual $t(\nu)$ distribution.)

We write

$$X \;\sim\; t(\mu, \sigma^2, \nu)$$

[Graph:]

Remarks:

- $\mathrm{E}(X) \;=\; \mu \quad$ if $\nu > 1$
- $\mathrm{Var}(X) \;=\; \frac{\nu}{\nu-2}\,\sigma^2 \quad$ if $\nu > 2$
- $X \;\Rightarrow\; \mathrm{N}(\mu, \sigma^2) \quad$ as $\nu \to \infty$
- $\frac{X-\mu}{\sigma} \;\sim\; t(\nu)$

(X, W) has a **normal-inverse gamma distribution** if

$$\begin{aligned} X \mid W = w &\sim \mathrm{N}(\mu_0, w/\kappa) \\ W &\sim \mathrm{IG}(\alpha, \beta) \end{aligned}$$

for some μ_0, $\kappa > 0$, $\alpha > 0$, $\beta > 0$.

The (joint) density is $\quad (w > 0)$

$$\begin{aligned} p(x, w) &= p(w)\, p(x \mid w) \\ &\propto \frac{1}{w^{\alpha+1}}\, e^{-\beta/w} \cdot \frac{1}{\sqrt{w}}\, e^{-\frac{\kappa(x-\mu_0)^2}{2w}} \\ &= \frac{1}{w^{\,\alpha+3/2}}\, e^{-\left(\beta + \frac{1}{2}\kappa(x-\mu_0)^2\right)\big/ w} \end{aligned}$$

Proposition

Let (X, W) *be normal-inverse gamma:*

$$\begin{aligned} X \mid W = w &\sim \mathrm{N}(\mu_0, w/\kappa) \\ W &\sim \mathrm{IG}(\alpha, \beta) \end{aligned}$$

Then the marginal distribution of X *is*

$$X \sim t\big(\mu_0,\, \beta/(\alpha\kappa),\, 2\alpha\big)$$

To show this ...

$$
\begin{aligned}
p(x) &= \int p(x,w)\,dw \\
&\underset{\text{in } x}{\propto} \int_0^\infty \underbrace{\frac{1}{w^{\,\alpha+3/2}}\, e^{-\left(\beta+\frac{1}{2}\kappa(x-\mu_0)^2\right)/w}} \; dw
\end{aligned}
$$

$$
\begin{aligned}
p(x) &= \int p(x,w)\,dw \\
&\underset{\text{in } x}{\propto} \int_0^\infty \underbrace{\frac{1}{w^{\alpha+3/2}}\, e^{-\left(\beta + \frac{1}{2}\kappa(x-\mu_0)^2\right)/w}}_{\text{kernel of } \mathrm{IG}\left(\alpha+1/2,\ \beta\ +\ \frac{1}{2}\kappa(x-\mu_0)^2\right)}\ dw
\end{aligned}
$$

$$
\begin{aligned}
p(x) &= \int p(x,w)\,dw \\
&\underset{\text{in } x}{\propto} \int_0^\infty \underbrace{\frac{1}{w^{\alpha+3/2}}\, e^{-\left(\beta + \frac{1}{2}\kappa(x-\mu_0)^2\right)/w}}_{\text{kernel of } \mathrm{IG}\left(\alpha+1/2,\ \beta\ +\ \frac{1}{2}\kappa(x-\mu_0)^2\right)} \, dw \\
&= \frac{\Gamma(\alpha+1/2)}{\left(\beta\ +\ \frac{1}{2}\kappa(x-\mu_0)^2\right)^{\alpha+1/2}}
\end{aligned}
$$

$$
\begin{aligned}
p(x) &= \int p(x,w)\,dw \\
&\underset{\text{in } x}{\propto} \int_0^\infty \underbrace{\frac{1}{w^{\,\alpha+3/2}}\, e^{-\left(\beta + \frac{1}{2}\kappa(x-\mu_0)^2\right)/w}}_{\text{kernel of } \mathrm{IG}\left(\alpha+1/2,\ \beta\ +\ \frac{1}{2}\kappa(x-\mu_0)^2\right)}\ dw \\
&= \frac{\Gamma(\alpha+1/2)}{\left(\beta\ +\ \frac{1}{2}\kappa(x-\mu_0)^2\right)^{\alpha+1/2}} \\
&\underset{\text{in } x}{\propto} \left(1\ +\ \frac{1}{2\alpha}\,\frac{(x-\mu_0)^2}{\beta/(\alpha\kappa)}\right)^{-\frac{2\alpha+1}{2}}
\end{aligned}
$$

which is a kernel of $t\big(\mu_0,\ \beta/(\alpha\kappa),\ 2\alpha\big)$

Conjugate Prior

A normal-inverse gamma prior is conjugate for the two-parameter normal model:

Let

$$\begin{aligned} \mu \mid \sigma^2 &\sim \mathrm{N}(\mu_0, \sigma^2/\kappa) \\ \sigma^2 &\sim \mathrm{IG}(\alpha, \beta) \end{aligned}$$

for some μ_0, $\kappa > 0$, $\alpha > 0$, $\beta > 0$.

Then the posterior is $(\sigma^2 > 0)$

$$\begin{aligned} p(\mu, \sigma^2 \mid \boldsymbol{y}) &\propto p(\sigma^2)\, p(\mu \mid \sigma^2)\, \underbrace{p(\boldsymbol{y} \mid \mu, \sigma^2)}_{L(\mu, \sigma^2; \boldsymbol{y})} \\ &\propto \cdots \end{aligned}$$

$$
\begin{aligned}
\cdots \quad \propto \quad & \frac{1}{(\sigma^2)^{\alpha+1}}\, e^{-\beta/\sigma^2} \cdot \frac{1}{\sqrt{\sigma^2}}\, e^{-\frac{\kappa}{2\sigma^2}(\mu-\mu_0)^2} \\
& \cdot \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{n-1}{2}\, s^2/\sigma^2 - \frac{n}{2\sigma^2}(\mu-\bar{y})^2} \\
\\
= \quad & \frac{1}{(\sigma^2)^{\alpha + n/2 + 3/2}} \\
& \cdot \exp\left(-\frac{\beta}{\sigma^2} \; - \; \frac{n-1}{2}\frac{s^2}{\sigma^2} \; - \; \frac{\kappa(\mu-\mu_0)^2 + n(\mu-\bar{y})^2}{2\sigma^2} \right)
\end{aligned}
$$

$$
\begin{aligned}
\cdots \quad \propto \quad & \frac{1}{(\sigma^2)^{\alpha+1}}\, e^{-\beta/\sigma^2} \cdot \frac{1}{\sqrt{\sigma^2}}\, e^{-\frac{\kappa}{2\sigma^2}(\mu-\mu_0)^2} \\
& \cdot \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{n-1}{2}\, s^2/\sigma^2 \,-\, \frac{n}{2\sigma^2}(\mu-\bar{y})^2} \\
= \quad & \frac{1}{(\sigma^2)^{\alpha + n/2 + 3/2}} \\
& \cdot \exp\left(-\frac{\beta}{\sigma^2} \;-\; \frac{n-1}{2}\frac{s^2}{\sigma^2} \;-\; \frac{\kappa(\mu-\mu_0)^2 + n(\mu-\bar{y})^2}{2\sigma^2} \right)
\end{aligned}
$$

You can show (Cowles, Sec. 7.3)

$$
\kappa(\mu-\mu_0)^2 + n(\mu-\bar{y})^2 \quad = \quad (\kappa+n)(\mu-\mu_1)^2 + \frac{\kappa n}{\kappa+n}(\mu_0-\bar{y})^2
$$

where

$$
\mu_1 \quad = \quad \frac{\kappa\mu_0 + n\bar{y}}{\kappa+n}
$$

So

$$
\begin{aligned}
p(\mu, \sigma^2 \mid \boldsymbol{y}) \quad &\propto \quad \frac{1}{(\sigma^2)^{\alpha + n/2 + 1}} \, e^{-\left(\beta + \frac{(n-1)s^2}{2} + \frac{1}{2}\frac{\kappa n}{\kappa + n}(\mu_0 - \bar{y})^2\right) \Big/ \sigma^2} \\
&\qquad \cdot \frac{1}{\sqrt{\sigma^2}} \, e^{-\frac{\kappa + n}{2\sigma^2}(\mu - \mu_1)^2} \\
&\propto \quad p(\sigma^2 \mid \boldsymbol{y}) \, p(\mu \mid \sigma^2, \boldsymbol{y})
\end{aligned}
$$

So

$$
\begin{aligned}
p(\mu, \sigma^2 \mid \boldsymbol{y}) \;\; &\propto \;\; \frac{1}{(\sigma^2)^{\alpha + n/2 + 1}} \, e^{-\left(\beta + \frac{(n-1)s^2}{2} + \frac{1}{2}\frac{\kappa n}{\kappa + n}(\mu_0 - \bar{y})^2\right) \Big/ \sigma^2} \\
&\qquad \cdot \frac{1}{\sqrt{\sigma^2}} \, e^{-\frac{\kappa + n}{2\sigma^2}(\mu - \mu_1)^2} \\
&\propto \;\; p(\sigma^2 \mid \boldsymbol{y}) \, p(\mu \mid \sigma^2, \boldsymbol{y})
\end{aligned}
$$

We recognize this as a normal-inverse gamma distribution:

$$
\begin{aligned}
\mu \mid \sigma^2, \boldsymbol{y} \;\; &\sim \;\; \mathrm{N}\big(\mu_1, \, \sigma^2/(\kappa + n)\big) \\
\sigma^2 \mid \boldsymbol{y} \;\; &\sim \;\; \mathrm{IG}\left(\alpha + \frac{n}{2}, \, \beta + \frac{(n-1)s^2}{2} + \frac{1}{2}\frac{\kappa n}{\kappa + n}(\mu_0 - \bar{y})^2\right)
\end{aligned}
$$

where

$$
\mu_1 \;\; = \;\; \frac{\kappa \mu_0 + n \bar{y}}{\kappa + n}
$$

Notice:

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

is a weighted average of the prior location parameter μ_0 and the sample mean $\bar{y}$.

Also note that κ acts like a “prior sample size” for μ in this weighting: The prior is like adding κ new observations that have average μ_0.

Q: What kind of κ values make the prior less informative?

Since the posterior for (μ, σ^2) is normal-inverse gamma, the marginal posterior for μ is

$$\mu \mid \boldsymbol{y} \;\; \sim \;\; t\left(\mu_1,\ \frac{1}{\tau_1^2(\kappa+n)},\ 2\alpha+n\right)$$

where

$$\tau_1^2 \;\; = \;\; \frac{\alpha \ + \ n/2}{\beta \ + \ \frac{(n-1)s^2}{2} \ + \ \frac{1}{2}\frac{\kappa n}{\kappa+n}(\mu_0 - \bar{y})^2}$$

(τ_1^2 turns out to be the posterior mean of $\tau^2 = 1/\sigma^2$.)

It follows that (provided $2\alpha + n > 2$)

$$\begin{aligned}
\mathrm{E}(\mu \mid \boldsymbol{y}) &= \mu_1 \\
\mathrm{Var}(\mu \mid \boldsymbol{y}) &= \frac{2\alpha + n}{2\alpha + n - 2} \cdot \frac{1}{\tau_1^2(\kappa + n)} \\
\mathrm{E}(\sigma^2 \mid \boldsymbol{y}) &= \frac{1}{\tau_1^2 \left(1 - \frac{1}{\alpha + n/2}\right)}
\end{aligned}$$

(see formulas in Cowles, Table A.2)

For n large, you can show

$$\begin{aligned} \mathrm{E}(\mu \mid \boldsymbol{y}) &\approx \bar{y} \\ \mathrm{Var}(\mu \mid \boldsymbol{y}) &\approx \frac{s^2}{n} \\ \mathrm{E}(\sigma^2 \mid \boldsymbol{y}) &\approx s^2 \end{aligned}$$

That is, asymptotically (for large n), the Bayesian will agree with the frequentist.

Since the t-distribution is symmetric, it is easy to find credible intervals for μ that are both equal-tailed and HPD ...

[Illustrate t distribution ...]

Recall: μ can be "Studentized" to the usual kind of t distribution by subtracting its posterior location and dividing by (the square root of) its posterior scale parameter.

Recalling that

$$\mu \mid \boldsymbol{y} \;\; \sim \;\; t\left(\mu_1,\ \frac{1}{\tau_1^2(\kappa+n)},\ 2\alpha+n\right)$$

we find ...

A 95% Credible Interval for μ:

$$\mu_1 \;\; \pm \;\; t_{0.025,2\alpha+n} \cdot \frac{1\big/\sqrt{\tau_1^2}}{\sqrt{\kappa+n}}$$

where $t_{0.025,2\alpha+n}$ is the usual 0.025 upper quantile of $t(2\alpha+n)$

To test

$$H_0 : \mu \geq \mu_* \qquad\qquad H_1 : \mu < \mu_*$$

we can compute the posterior probability

$$\begin{aligned}
\mathrm{P}(H_0 \mid \boldsymbol{y}) &= \mathrm{P}(\mu \geq \mu_* \mid \boldsymbol{y}) \\
&= \mathrm{P}\left(\frac{\mu - \mu_1}{1/\sqrt{\tau_1^2(\kappa + n)}} \geq \frac{\mu_* - \mu_1}{1/\sqrt{\tau_1^2(\kappa + n)}} \;\Big|\; \boldsymbol{y} \right) \\
&= 1 - F_t\left(\frac{\mu_* - \mu_1}{1/\sqrt{\tau_1^2(\kappa + n)}} \right)
\end{aligned}$$

where F_t is the (cumulative) distribution function of $t(2\alpha + n)$.

To find a credible interval for σ^2, we start by considering

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

for which the posterior is (why?)

$$\tau^2 \mid \boldsymbol{y} \;\sim\; \mathrm{gamma}(\alpha^*, \beta^*)$$

where

$$\alpha^* \;=\; \alpha + \frac{n}{2}$$

$$\beta^* \;=\; \beta + \frac{(n-1)s^2}{2} + \frac{1}{2}\,\frac{\kappa n}{\kappa + n}(\mu_0 - \bar{y})^2$$

Then, with posterior probability 0.95,

$$g_{0.975,\,\alpha^*,\,\beta^*} \quad < \quad \tau^2 \quad < \quad g_{0.025,\,\alpha^*,\,\beta^*}$$

where the limits are upper $\mathrm{gamma}(\alpha^*, \beta^*)$ quantiles.

[Illustrate gamma quantiles ...]

So a 95% *equal-tailed* credible interval for σ^2 is given by

$$\frac{1}{g_{0.025,\,\alpha^*,\,\beta^*}} \;<\; \sigma^2 \;<\; \frac{1}{g_{0.975,\,\alpha^*,\,\beta^*}}$$

(Note: We use gamma rather than inverse gamma quantiles because R functions for the gamma distribution are more readily available.)

R Example 7.1:

Two-Parameter Normal Inference — Conjugate Prior

Noninformative Priors

There are various possibilities:

- Jeffreys:

  Turns out to be the improper prior

  $$p(\mu, \sigma^2) \quad \propto \quad \frac{1}{(\sigma^2)^{3/2}} \qquad (\sigma^2 > 0)$$

  (Not often used in practice.)

- Limiting ("Vague") Conjugate:

  Using the kernel of the conjugate prior, formally letting

  $$\kappa \to 0 \qquad \alpha \to 0 \qquad \beta \to 0$$

  produces

  $$p(\mu, \sigma^2) \;\; \propto \;\; \frac{1}{(\sigma^2)^{3/2}} \qquad (\sigma^2 > 0)$$

  (same as Jeffreys)

- "Standard" (Product-Jeffreys)

$$p(\mu, \sigma^2) \quad \propto \quad \frac{1}{\sigma^2} \qquad\qquad (\sigma^2 > 0)$$

(Note: improper)

Equals the product of the Jeffreys priors for μ alone (σ^2 known) and for σ^2 alone (μ known):

$$\mu, \sigma^2 \quad \sim \quad 1\, d\mu \;\cdot\; \frac{1}{\sigma^2}\, d\sigma^2$$

We will use this "standard" noninformative prior.

Since it is improper, can we be sure the “standard” noninformative prior will give a proper posterior?

Yes, provided there are at least two *distinct* observed y values.

The posterior is $\quad (\sigma^2 > 0)$

$$
\begin{aligned}
p(\mu, \sigma^2 \mid \boldsymbol{y}) &\propto p(\mu, \sigma^2)\, p(\boldsymbol{y} \mid \mu, \sigma^2) \\
&\propto \frac{1}{\sigma^2} \cdot \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{n-1}{2} s^2/\sigma^2}\; e^{-\frac{n}{2\sigma^2}(\mu-\bar{y})^2} \\
&= \frac{1}{(\sigma^2)^{\frac{n+1}{2}}}\, e^{-\frac{n-1}{2} s^2/\sigma^2} \cdot \frac{1}{\sqrt{\sigma^2}}\, e^{-\frac{n}{2\sigma^2}(\mu-\bar{y})^2} \\
&\propto p(\sigma^2 \mid \boldsymbol{y})\, p(\mu \mid \sigma^2, \boldsymbol{y})
\end{aligned}
$$

The posterior is $\quad (\sigma^2 > 0)$

$$\begin{aligned}
p(\mu, \sigma^2 \mid \boldsymbol{y}) &\propto p(\mu, \sigma^2)\, p(\boldsymbol{y} \mid \mu, \sigma^2) \\
&\propto \frac{1}{\sigma^2} \cdot \frac{1}{(\sigma^2)^{n/2}}\, e^{-\frac{n-1}{2} s^2/\sigma^2}\, e^{-\frac{n}{2\sigma^2}(\mu - \bar{y})^2} \\
&= \frac{1}{(\sigma^2)^{\frac{n+1}{2}}}\, e^{-\frac{n-1}{2} s^2/\sigma^2} \cdot \frac{1}{\sqrt{\sigma^2}}\, e^{-\frac{n}{2\sigma^2}(\mu - \bar{y})^2} \\
&\propto p(\sigma^2 \mid \boldsymbol{y})\, p(\mu \mid \sigma^2, \boldsymbol{y})
\end{aligned}$$

Recognize as a normal-inverse gamma:

$$\begin{aligned}
\mu \mid \sigma^2, \boldsymbol{y} &\sim \mathrm{N}(\bar{y},\, \sigma^2/n) \\
\sigma^2 \mid \boldsymbol{y} &\sim \mathrm{IG}\left(\frac{n-1}{2},\ \frac{n-1}{2}\, s^2\right)
\end{aligned}$$

It follows that the posterior marginal for μ is

$$\mu \mid \boldsymbol{y} \;\sim\; t(\bar{y},\, s^2/n,\, n-1)$$

This implies

$$\begin{aligned} \mathrm{E}(\mu \mid \boldsymbol{y}) &= \bar{y} && (n > 2) \\ \mathrm{Var}(\mu \mid \boldsymbol{y}) &= \frac{n-1}{n-3} \cdot \frac{s^2}{n} && (n > 3) \end{aligned}$$

(So the posterior standard deviation is a bit larger than the usual standard error.)

Also,

$$\mu \mid \boldsymbol{y} \;\sim\; t(\bar{y},\, s^2/n,\, n-1)$$

implies

$$\left.\frac{\mu - \bar{y}}{s/\sqrt{n}}\;\right|\; \boldsymbol{y} \quad\sim\quad t(n-1)$$

Compare with the usual frequentist result:

$$\left.\frac{\overline{Y} - \mu}{S/\sqrt{n}}\;\right|\; \mu, \sigma^2 \quad\sim\quad t(n-1)$$

It follows that posterior inference is much like the usual frequentist inference:

95% credible interval (equal-tailed and HPD) for μ:

$$\bar{y} \quad \pm \quad t_{0.025,\, n-1} \cdot \frac{s}{\sqrt{n}}$$

(why?)

Consider testing

$$H_0 : \ \mu \geq \mu_* \qquad\qquad H_1 : \ \mu < \mu_*$$

Then the posterior probability of H_0 is

$$\begin{aligned} \mathrm{P}(\mu \geq \mu_* \mid \boldsymbol{y}) &= \mathrm{P}\left(\frac{\mu - \bar{y}}{s/\sqrt{n}} \ \geq \ \frac{\mu_* - \bar{y}}{s/\sqrt{n}} \ \middle| \ \boldsymbol{y} \right) \\ &= 1 \ - \ F_t\left(\frac{\mu_* - \bar{y}}{s/\sqrt{n}} \right) \end{aligned}$$

where F_t is the (cumulative) distribution function of $t(n-1)$.

Notice: This also happens to be the (one-sided) p-value.

Since the posterior marginal of σ^2 is

$$\sigma^2 \mid \boldsymbol{y} \;\sim\; \mathrm{IG}\left(\frac{n-1}{2},\ \frac{n-1}{2}\, s^2\right)$$

(why?) we have $(n > 3)$

$$\begin{aligned} \mathrm{E}(\sigma^2 \mid \boldsymbol{y}) &= \frac{\frac{n-1}{2}\, s^2}{\frac{n-1}{2} - 1} = \frac{n-1}{n-3}\, s^2 \\ &= \frac{1}{n-3} \sum_i (y_i - \bar{y})^2 \end{aligned}$$

(Compare with the usual estimate s^2. Bias?)

Also, for $\tau^2 = 1/\sigma^2$,

$$\tau^2 \mid \boldsymbol{y} \;\sim\; \text{gamma}\left(\frac{n-1}{2},\, \frac{n-1}{2}\, s^2\right)$$

(why?) so

$$(n-1)s^2\tau^2 \mid \boldsymbol{y} \;\sim\; \text{gamma}\left(\frac{n-1}{2},\, \frac{1}{2}\right) \;=\; \chi^2(n-1)$$

that is,

$$(n-1)\,\frac{s^2}{\sigma^2} \;\Big|\; \boldsymbol{y} \quad\sim\quad \chi^2(n-1)$$

Compare with the usual frequentist result

$$(n-1)\,\frac{S^2}{\sigma^2} \;\Big|\; \mu, \sigma^2 \quad\sim\quad \chi^2(n-1)$$

So a 95% equal-tailed credible interval for σ^2 can be derived as

$$\chi^2_{0.975,n-1} \ < \ (n-1)\,\frac{s^2}{\sigma^2} \ < \ \chi^2_{0.025,n-1}$$

$$\Rightarrow \qquad \frac{(n-1)\,s^2}{\chi^2_{0.025,n-1}} \ < \ \sigma^2 \ < \ \frac{(n-1)\,s^2}{\chi^2_{0.975,n-1}}$$

and this coincides with the usual 95% confidence interval.

To test, for example,

$$H_0:\ \sigma^2 \geq \sigma_*^2 \qquad\qquad H_1:\ \sigma^2 < \sigma_*^2$$

the posterior probability of H_0 is

$$\begin{aligned} \mathrm{P}(\sigma^2 \geq \sigma_*^2 \mid \boldsymbol{y}) &= \mathrm{P}\left((n-1)\,\frac{s^2}{\sigma^2} \ \leq\ (n-1)\,\frac{s^2}{\sigma_*^2} \;\middle|\; \boldsymbol{y} \right) \\ &= F_{\chi^2}\left((n-1)\,\frac{s^2}{\sigma_*^2} \right) \end{aligned}$$

where F_{χ^2} is the (cumulative) distribution function of $\chi^2(n-1)$.

(This also happens to be the usual one-sided p-value.)

R Example 7.2:

Two-Parameter Normal Inference — Noninformative Prior

Posterior Predictive Distribution

Consider predicting the "new" Y value

$$Y^* = \mu + \varepsilon^*$$

where

$$\varepsilon^* \mid \mu, \sigma^2, \boldsymbol{y} \sim \mathrm{N}(0, \sigma^2)$$

So, conditional on σ^2 and $\boldsymbol{Y}$, ε^* is independent of μ. (Why?)

Under the standard noninformative prior,

$$\mu \mid \sigma^2, \boldsymbol{y} \;\; \sim \;\; \mathrm{N}(\bar{y},\, \sigma^2/n)$$

so we get

$$Y^* \;=\; \mu \,+\, \varepsilon^* \;\Bigm|\; \sigma^2, \boldsymbol{y} \quad \sim \quad \mathrm{N}\biggl(\bar{y},\, \sigma^2\Bigl(1 + \frac{1}{n}\Bigr)\biggr)$$

Since also

$$\sigma^2 \mid \boldsymbol{y} \;\; \sim \;\; \mathrm{IG}\biggl(\frac{n-1}{2},\; \frac{n-1}{2}\, s^2\biggr)$$

we find that the posterior (predictive) distribution of (Y^*, σ^2) is normal-inverse gamma.

Therefore,

$$Y^* \mid \boldsymbol{y} \quad \sim \quad t\left(\bar{y},\, s^2\left(1 + \frac{1}{n}\right),\, n - 1\right)$$

that is,

$$\left.\frac{Y^* - \bar{y}}{s\sqrt{1 + \frac{1}{n}}}\ \right|\ \boldsymbol{y} \quad \sim \quad t(n-1)$$

Compare with the frequentist result

$$\left.\frac{Y^* - \overline{Y}}{S\sqrt{1 + \frac{1}{n}}}\ \right|\ \mu, \sigma^2 \quad \sim \quad t(n-1)$$

The Bayesian result implies the 95% posterior predictive interval for Y^* given by

$$\bar{y} \quad \pm \quad t_{0.025,\, n-1} \cdot s \sqrt{1 + \frac{1}{n}}$$

(Note: Also happens to be a frequentist prediction interval.)

Similarly, you can compute posterior predictive probabilities.

R Example 7.3:

Two-Parameter Normal Prediction — Noninformative Prior

Semi-Conjugate Prior

Recall:

- For fixed σ^2, the normal distribution is conjugate for μ.
- For fixed μ, the inverse gamma distribution is conjugate for σ^2.

In the two-parameter model, we call each of these a **semi-conjugate** distribution.

A possible (semi-conjugate) prior specification:

$$\left.\begin{aligned} \mu &\sim \mathrm{N}(\mu_0, \sigma_0^2) \\ \sigma^2 &\sim \mathrm{IG}(\alpha, \beta) \end{aligned}\right\} \text{independent}$$

You can show that this prior is not (fully) conjugate. Indeed, the posterior marginals are not even from common parametric families.

Note: Letting $\sigma_0^2 \to \infty,\ \alpha \to 0,\ \beta \to 0$ produces the "standard" noninformative prior.