STAT 431 — Applied Bayesian Analysis — Course Notes

The Two-Parameter Normal Model

→ Model

→ Likelihood

- sufficient stats
- MLE

→ 3 ways for joint prior

- conjugate priors
- jeffrey's "
- product jeffrey's

Fall 2022

→ [Useful distrⁿ

i) Student's t-distⁿ]

→ Normal-inverse gamma

- conjugate prior

- Model:

$$Y_1, \ldots, Y_n \mid \mu, \sigma^2 \quad \sim \quad iid \ \text{Normal}(\mu, \ \sigma^2 = 1/\tau^2)$$

Let

$$\begin{aligned}
\boldsymbol{y} &= (y_1, \ldots, y_n) && \text{(observed version of } \boldsymbol{Y}\text{)} \\
\bar{y} &= \frac{1}{n} \sum_i y_i \quad = \quad \text{usual est. of } \mu \\
s^2 &= \frac{1}{n-1} \sum_i (y_i - \bar{y})^2 \\
&= \text{usual unbiased est. of } \sigma^2 \text{ (for } n > 1\text{)}
\end{aligned}$$

Analytical sol^n instead of Gibbs sampling

Both μ and σ^2 are unknown.

- Likelihood:

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

You can show that

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

so

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

Note: $\boxed{(\bar{y}, s^2) \text{ is sufficient for } (\mu, \sigma^2)}$ (why?)

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

$$\hat{\mu} = \bar{y}$$

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

We will consider three ways to specify the joint prior $\pi(\mu, \sigma^2)$:

1 ▶ a conjugate prior

2 ▶ Jeffreys' prior

3 ▶ the "standard" noninformative ("product-Jeffreys'" or "reference") prior

We will need some distribution theory ...

Some Useful Distributions

X has a **(Student's)** t-**distribution** with **location** μ, **scale** $\sigma > 0$, and **degrees of freedom** $\nu > 0$ if it has PDF

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

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

(Note: $\mu = 0$ and $\sigma = 1$ give the standard t-distribution with ν degrees of freedom.)

We write

$$X \;\sim\; \mathrm{t}_\nu(\mu, \sigma^2)$$

[Graph PDF:]

Remarks:

- $\mathrm{E}(X) = \mu$ if $\nu > 1$
- $\mathrm{Var}(X) = \frac{\nu}{\nu-2}\,\sigma^2$ if $\nu > 2$
- $X \Rightarrow \mathrm{Normal}(\mu, \sigma^2)$ as $\nu \to \infty$
- $\frac{X-\mu}{\sigma} \sim \mathrm{t}_\nu(0,1)$

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

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

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

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

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

Proposition

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

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

Then the marginal distribution of X *is*

$$X \sim \mathrm{t}_{2\alpha}\big(\mu_0,\, \beta/(\alpha\kappa)\big)$$

To show this ...

$f(x, w) = f(x | w) \cdot f(w)$

Dependence b/w x & w

$$f(x) = \int f(x,w)\,dw$$

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

x, w are dependent but have cov of 0.

$$f(x) = \int f(x,w)\,dw$$

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

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

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

which is a kernel of $\mathrm{t}_{2\alpha}\big(\mu_0,\ \beta/(\alpha\kappa)\big)$ Marginal

Conjugate Prior

For $n \geq 2$ and

$$Y_1, \ldots, Y_n \mid \mu, \sigma^2 \quad \sim \quad iid \ \text{Normal}\big(\mu,\ \sigma^2 = 1/\tau^2\big)$$

consider the normal-inverse gamma prior

$$\begin{aligned} \mu \mid \sigma^2 \quad &\sim \quad \text{Normal}\big(\mu_0, \sigma^2/\kappa_0\big) \\ \sigma^2 \quad &\sim \quad \text{InvGamma}(\alpha_0, \beta_0) \end{aligned}$$

for some μ_0, $\kappa_0 > 0$, $\alpha_0 > 0$, $\beta_0 > 0$.

It can be shown that this prior is conjugate ...

Indeed, the posterior turns out to be

$$\begin{aligned} \mu \mid \sigma^2, \boldsymbol{y} &\sim \mathrm{Normal}\big(\mu_1,\, \sigma^2/\kappa_1\big) \\ \sigma^2 \mid \boldsymbol{y} &\sim \mathrm{InvGamma}(\alpha_1, \beta_1) \end{aligned}$$

where

$$\mu_1 = \frac{n\bar{y} + \kappa_0\mu_0}{n + \kappa_0} \qquad \kappa_1 = n + \kappa_0 \qquad \alpha_1 = \frac{n}{2} + \alpha_0$$

$$\beta_1 = \frac{1}{2}\frac{n\kappa_0}{n + \kappa_0}(\bar{y} - \mu_0)^2 + \frac{(n-1)s^2}{2} + \beta_0$$

Notice:

$$\mu_1 \;=\; \frac{n}{n+\kappa_0}\,\bar{y} + \frac{\kappa_0}{n+\kappa_0}\,\mu_0$$

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

Also note that κ_0 acts like a "prior sample size" for μ in this weighting: The prior is like adding κ_0 new observations that have average μ_0.

Q: What kind of κ_0 values make the prior less informative?

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

$$\mu \mid \boldsymbol{y} \sim \mathrm{t}_{2\alpha_1}\left(\mu_1, \frac{1}{\tau_1^2 \kappa_1}\right)$$

where

$$\tau_1^2 = \frac{\alpha_1}{\beta_1}$$

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

It follows that

Classical frequentist approach

$$
\begin{aligned}
E(\mu \mid \boldsymbol{y}) &= \mu_1 \\
\mathrm{Var}(\mu \mid \boldsymbol{y}) &= \frac{2\alpha_1}{2\alpha_1 - 2} \cdot \frac{1}{\tau_1^2 \kappa_1} \\
E(\sigma^2 \mid \boldsymbol{y}) &= \frac{1}{\tau_1^2 \left(1 - \frac{1}{\alpha_1}\right)}
\end{aligned}
$$

$\approx \bar{y}$ for large n

$\approx \frac{s^2}{n}$ for " "

$\approx s^2$ " " "

(see formulas in BSM, Appendix A.1)

For n large, you can show

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

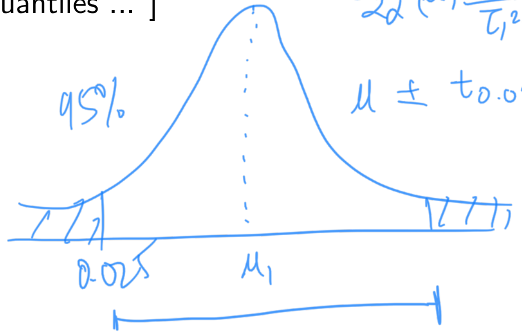
That is, asymptotically (for n large), the Bayesian will agree with the frequentist.

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

[Illustrate t quantiles ...]

$t_{2\alpha}\left(\mu_1, \frac{1}{\tau_1^2 \kappa_1}\right)$

$\mu \pm t_{0.025, 2\alpha_1} \frac{1/\sqrt{\tau_1^2}}{\sqrt{\kappa_1}}$



Recall: μ can be "Studentized" to the usual kind of t-distribution by subtracting its posterior location and dividing by its posterior scale.

Recalling that

$$\mu \mid \boldsymbol{y} \quad \sim \quad \mathrm{t}_{2\alpha_1}\left(\mu_1,\ \frac{1}{\tau_1^2 \kappa_1}\right)$$

we find the following 95% credible interval for μ:

$$\mu_1 \quad \pm \quad t_{0.025, 2\alpha_1} \cdot \frac{1/\sqrt{\tau_1^2}}{\sqrt{\kappa_1}}$$

where $t_{0.025,2\alpha_1}$ is the usual 0.025 upper quantile of $\mathrm{t}_{2\alpha_1}(0,1)$

To test

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

we can compute the posterior probability

$$\begin{aligned}
\mathrm{Prob}(H_0 \mid \boldsymbol{y}) &= \mathrm{Prob}(\mu \geq \mu_* \mid \boldsymbol{y}) \\
&= \mathrm{Prob}\left(\frac{\mu - \mu_1}{1/\sqrt{\tau_1^2 \kappa_1}} \geq \frac{\mu_* - \mu_1}{1/\sqrt{\tau_1^2 \kappa_1}} \,\middle|\, \boldsymbol{y} \right) \\
&= 1 - F_{\mathrm{t}}\left(\frac{\mu_* - \mu_1}{1/\sqrt{\tau_1^2 \kappa_1}} \right)
\end{aligned}$$

upper tailed posterior

where F_{t} is the (cumulative) distribution function of $\mathrm{t}_{2\alpha_1}(0, 1)$.

Jeffreys' Prior

Individual jeffrey: $\pi(\mu) \propto 1$

$\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ $(\sigma^2 > 0)$

Turns out to be the improper prior

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

(See BSM, Appendix A.3.)

Note:

- Not the product of the individual Jeffreys' priors
- Can be obtained from the conjugate prior by formally letting

$$\kappa_0 \to 0 \qquad \alpha_0 \to 0 \qquad \beta_0 \to 0$$

Less informative

var is high $\to N(\mu_0, \sigma^2/\kappa_0^2)$

Since it is improper, can we be sure this prior will give a proper posterior?

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

If there are atleast 2 distinct values among $y_1, \ldots, y_n \Rightarrow s_n^2 > 0$

Under those conditions, the posterior turns out to be proper, and is actually normal-inverse gamma:

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

It follows that the marginal posterior for μ is

$$\mu \mid \boldsymbol{y} \sim \mathrm{t}_n\big(\bar{y},\, \hat{\sigma}^2/n\big)$$

which gives the following 95% credible interval for μ:

(1) n-1 degrees usually

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

$$\bar{y} \pm t_{0.025,n} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

(2) usual has s

How different from usual 95% confidence interval? (2)

$\hat{\sigma}^2 = \left(\frac{n-1}{n}\right)s^2$, $\hat{\sigma}^2 < s$ | Quantiles size: small df, heavier the tail

$\therefore$ This credible interval is smaller

Inferences based on this Jeffreys' prior don't quite match the usual frequentist inferences.

Underestimate the variability

Is there a prior that provides a better match to frequentist inference?

Yes ...

The “Standard” Noninformative Prior

The “standard” noninformative (or “product-Jeffreys”', or “reference”) prior is the improper prior

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

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

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

Provided there are at least two *distinct* observed y values, the posterior turns out to be proper.

In fact, it is normal-inverse gamma:

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

It follows that the marginal posterior for μ is

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

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,

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

implies

$$\left.\frac{\mu - \bar{y}}{s/\sqrt{n}} \;\right|\; \boldsymbol{y} \quad \sim \quad \mathrm{t}_{n-1}(0,1)$$

Compare with the usual frequentist result:

$$\left.\frac{\overline{Y} - \mu}{S/\sqrt{n}} \;\right|\; \mu, \sigma^2 \quad \sim \quad \mathrm{t}_{n-1}(0,1)$$

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

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

where F_{t} is the (cumulative) distribution function of $\mathrm{t}_{n-1}(0,1)$.

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

Posterior Predictive Distribution

Consider predicting the "new" Y value

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

where

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

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

Under the "standard" noninformative prior,

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

so we get

$$Y^* \;=\; \mu + \varepsilon^* \;\Big|\; \sigma^2, \boldsymbol{y} \quad\sim\quad \mathrm{Normal}\left(\bar{y},\, \sigma^2\Big(1 + \frac{1}{n}\Big)\right)$$

Since also

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

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

Therefore,

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

that is,

$$\frac{Y^* - \bar{y}}{s\sqrt{1+\frac{1}{n}}} \;\Bigg|\; \boldsymbol{y} \quad \sim \quad \mathrm{t}_{n-1}(0,1)$$

Compare with the frequentist result

$$\frac{Y^* - \overline{Y}}{S\sqrt{1+\frac{1}{n}}} \;\Bigg|\; \mu, \sigma^2 \quad \sim \quad \mathrm{t}_{n-1}(0,1)$$

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.