STA 360/602L: Module 3.5

The normal model: joint inference for mean and variance

Dr. Olanrewaju Michael Akande

Joint inference for mean and variance

- We have derived the posterior for the μ, conditional on $\sigma/\ \tau$ being known. What happens when $\sigma/\ \tau$ is unknown? We need a joint prior $\pi(\mu, \sigma^2)$ for μ and σ^2.

- Write the joint prior distribution for the mean and variance as the product of a conditional and a marginal distribution. That is,

$$\pi(\mu, \sigma^2) \;=\; \pi(\mu|\sigma^2)\pi(\sigma^2).$$

- From the previous module, we have seen that we can set the conditional prior $\pi(\mu|\sigma^2)$ to be a normal distribution.

- For $\pi(\sigma^2)$, we need a distribution with support on $(0, \infty)$. One such family is the gamma family, but this is NOT conjugate for the variance of a normal distribution.

- The gamma distribution is, however, conjugate for the precision τ, and in that case, we say that σ^2 has an inverse-gamma distribution.

Joint inference for mean and variance

- Recall that conjugacy means that for a prior $\pi(\theta)$ in a class of distributions $\mathcal{P}$, $\pi(\theta|Y)$ is also in class $\mathcal{P}$.
- However, when we have multiple parameters, the dependence structure in the prior must also be preserved in the posterior, for conjugacy to hold.
- So, if

$$\pi(\mu, \sigma^2) = \pi(\mu|\sigma^2)\pi(\sigma^2).$$

with $\pi(\mu|\sigma^2)$ a normal distribution, and $\pi(\sigma^2)$ an inverse-gamma distribution, we will have conjugacy if $\pi(\mu, \sigma^2|Y)$ can also be written as

$$\pi(\mu, \sigma^2|Y) = \pi(\mu|\sigma^2, Y)\pi(\sigma^2|Y),$$

where $\pi(\mu|\sigma^2, Y)$ is also a normal distribution, and $\pi(\sigma^2|Y)$ is an inverse-gamma distribution, just like the prior.

Inverse-gamma distribution

- As before, we will continue to work mostly in terms of the precision τ.
- That is, we will deal with the already familiar gamma distribution, instead of the inverse-gamma distribution.
- However, as a quick review, if $\theta \sim \mathcal{IG}(a,b)$, then the pdf is

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} e^{-\frac{b}{\theta}} \quad \text{for} \quad a, b > 0,$$

where

  - $\mathbb{E}[\theta] = \frac{b}{a-1}$;
  - $\mathbb{V}[\theta] = \frac{b^2}{(a-1)^2(a-2)} \;\; \text{for} \;\; a \geq 2.$;
  - $\text{Mode}[\theta] = \frac{b}{a+1}$.

Conjugate prior

- Once again, suppose $Y = (y_1, y_2, \ldots, y_n)$, where each

$$y_i \sim \mathcal{N}(\mu, \tau^{-1}).$$

- A conjugate joint prior is given by

$$\tau = \frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right)$$
$$\mu|\tau \sim \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0\tau}\right).$$

- This is often called a normal-gamma prior distribution.
- σ_0^2 is the prior guess for σ^2, while ν_0 is often referred to as the "prior degrees of freedom", our degree of confidence in σ_0^2.
- We do not have conjugacy if we replace $\frac{1}{\kappa_0\tau}$ in the normal prior with an arbitrary prior variance independent of τ/σ^2. To do inference in that scenario, we need Gibbs sampling (to come soon!).

Conjugate prior

- So, we have

$$\pi(\mu|\tau) = \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0\tau}\right) \propto \exp\left\{-\frac{1}{2}\kappa_0\tau(\mu-\mu_0)^2\right\}.$$

- and

$$\pi(\tau) = \mathrm{Ga}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right) \propto \tau^{\frac{\nu_0}{2}-1}\exp\left\{-\frac{\tau\nu_0\sigma_0^2}{2}\right\}.$$

- Thus, the kernel of the normal-gamma prior distribution is

$$\begin{aligned}\Rightarrow \pi(\mu,\tau) \;&=\; \pi(\mu|\tau)\cdot\pi(\tau) = \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0\tau}\right)\cdot \mathrm{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right)\\ &\propto \underbrace{\exp\left\{-\frac{1}{2}\kappa_0\tau(\mu-\mu_0)^2\right\}}_{\propto\,\pi(\mu|\tau)}\cdot\underbrace{\tau^{\frac{\nu_0}{2}-1}\exp\left\{-\frac{\tau\nu_0\sigma_0^2}{2}\right\}}_{\propto\,\pi(\tau)}.\end{aligned}$$

- Take note of this form. When we derive the posterior kernel, we will try to match it to this to recognize the parameters.

Posterior for the mean given variance, under normal-gamma prior

- Based on the normal-gamma prior, we need $\pi(\mu|Y, \tau)$ and $\pi(\tau|Y)$.
- For $\pi(\mu|Y, \tau)$, we already know from the previous module that it will be a normal distribution.
- However, some algebra is required to get $\pi(\tau|Y)$.
- Infact, we need to write the full joint posterior and go from there, because we will need to keep some of the terms we discarded in the derivation in the last module.
- First, recall that the likelihood is

$$P(Y|\mu, \tau) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\tau s^2(n-1)\right\} \exp\left\{-\frac{1}{2}\tau n(\mu - \bar{y})^2\right\}.$$

Posterior derivation

Then, $\pi(\mu,\tau|Y) \propto \pi(\mu|\tau) \times \pi(\tau) \times P(Y|\mu,\tau)$

$$\propto \underbrace{\exp\left\{-\frac{1}{2}\kappa_0\tau(\mu-\mu_0)^2\right\}}_{\propto\, \pi(\mu|\sigma^2)} \times \underbrace{\tau^{\frac{\nu_0}{2}-1}\exp\left\{-\frac{\tau\nu_0\sigma_0^2}{2}\right\}}_{\propto\, \pi(\tau)}$$

$$\times \underbrace{\tau^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\tau s^2(n-1)\right\}\exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}}_{\propto\, P(Y|\mu,\tau)}$$

$$= \underbrace{\exp\left\{-\frac{1}{2}\kappa_0\tau(\mu-\mu_0)^2\right\}\exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}}_{\text{Terms involving } \mu}$$

$$\times \underbrace{\tau^{\frac{\nu_0}{2}-1}\exp\left\{-\frac{\tau\nu_0\sigma_0^2}{2}\right\}\tau^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\tau s^2(n-1)\right\}}_{\text{Terms involving } \tau \text{ but NOT } \mu}$$

Posterior derivation

$$\pi(\mu,\tau|Y) \propto \underbrace{\exp\left\{-\frac{1}{2}\kappa_0\tau(\mu^2 - 2\mu\mu_0 + \mu_0^2)\right\} \exp\left\{-\frac{1}{2}\tau n(\mu^2 - 2\mu\bar{y} + \bar{y}^2)\right\}}_{\text{Terms involving } \mu}$$

$$\times \underbrace{\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2 + s^2(n-1)\right]}{2}\right\}}_{\text{Terms involving } \tau \text{ but NOT } \mu}$$

$$= \underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau(\mu^2 - 2\mu\mu_0) + \tau n(\mu^2 - 2\mu\bar{y})\right]\right\}}_{\text{Terms involving } \mu}$$

$$\times \underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2 + \tau n\bar{y}^2\right]\right\} \cdot \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2 + s^2(n-1)\right]}{2}\right\}}_{\text{Terms involving } \tau \text{ but NOT } \mu}$$

$$= \underbrace{\exp\left\{-\frac{1}{2}\left[\mu^2(n\tau + \kappa_0\tau) - 2\mu(n\tau\bar{y} + \kappa_0\tau\mu_0)\right]\right\}}_{\text{Terms involving } \mu}$$

$$\times \underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2 + \tau n\bar{y}^2\right]\right\} \cdot \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2 + s^2(n-1)\right]}{2}\right\}}_{\text{Terms involving } \tau \text{ but NOT } \mu}$$

Posterior derivation

- To match the terms for the terms involving μ to the normal kernel in the prior, we need to complete the square so that we have something that looks like the $(\mu - \mu_0)^2$ term in our prior.

- Recall how to complete the square. Specifically, we can write

$$a\mu^2 + b\mu$$

as

$$a(\mu + d)^2 + e,$$

where

  - $d = \dfrac{b}{2a}$, and

  - $e = -\dfrac{b^2}{4a}$.

Posterior derivation

- First, write out the posterior again:

$$\pi(\mu,\tau|Y) = \underbrace{\exp\left\{-\frac{1}{2}\left[(n\tau+\kappa_0\tau)\mu^2 - 2\mu(n\tau\bar{y}+\kappa_0\tau\mu_0)\right]\right\}}_{\text{Terms involving } \mu}$$

$$\times \underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2 + \tau n\bar{y}^2\right]\right\} \cdot \tau^{\frac{\nu_0+n}{2}-1} \exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2 + s^2(n-1)\right]}{2}\right\}}_{\text{Terms involving } \tau \text{ but NOT } \mu}$$

- Set $a^\star = (n\tau+\kappa_0\tau)$ and $b^\star = (n\tau\bar{y}+\kappa_0\tau\mu_0)$, then complete the square for the first part.

$$\Rightarrow \pi(\mu,\tau|Y) \propto \underbrace{\exp\left\{-\frac{1}{2}\left[a^\star\mu^2 - 2b^\star\mu\right]\right\}}_{\text{Terms involving } \mu}$$

$$\times \underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2 + \tau n\bar{y}^2\right]\right\} \cdot \tau^{\frac{\nu_0+n}{2}-1} \exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2 + s^2(n-1)\right]}{2}\right\}}_{\text{Terms involving } \tau \text{ but NOT } \mu}$$

Posterior derivation

$$
\begin{aligned}
\Rightarrow \pi(\mu,\tau|Y) \propto\ & \exp\left\{-\frac{1}{2}a^{\star}\left[\mu-\frac{b^{\star}}{a^{\star}}\right]^2+\frac{(b^{\star})^2}{2a^{\star}}\right\} \cdot \exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2\right]\right\} \\
& \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\} \\
=\ & \exp\left\{-\frac{1}{2}a^{\star}\left[\mu-\frac{b^{\star}}{a^{\star}}\right]^2\right\} \underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2-\frac{(b^{\star})^2}{a^{\star}}\right]\right\}}_{\text{Next, substitute the values for } a^{\star} \text{ and } b^{\star} \text{ back}} \\
& \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\} \\
=\ & \exp\left\{-\frac{1}{2}a^{\star}\left[\mu-\frac{b^{\star}}{a^{\star}}\right]^2\right\} \underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2-\frac{(n\tau\bar{y}+\kappa_0\tau\mu_0)^2}{(n\tau+\kappa_0\tau)}\right]\right\}}_{\text{Next, expand terms and recombine}} \\
& \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}
\end{aligned}
$$

Posterior derivation

$$
\begin{aligned}
\Rightarrow \pi(\mu,\tau|Y) \;\propto\; & \exp\left\{-\frac{1}{2}a^\star\left[\mu-\frac{b^\star}{a^\star}\right]^2\right\}\exp\left\{-\frac{1}{2}\left[\frac{n\kappa_0\tau^2(\mu_0^2-2\mu_0\bar{y}+\bar{y}^2)}{\tau(\kappa_0+n)}\right]\right\} \\
& \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\} \\
= \;& \exp\left\{-\frac{1}{2}a^\star\left[\mu-\frac{b^\star}{a^\star}\right]^2\right\}\exp\left\{-\frac{\tau}{2}\left[\frac{n\kappa_0(\bar{y}-\mu_0)^2}{(\kappa_0+n)}\right]\right\} \\
& \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\} \\
= \;& \exp\underbrace{\left\{-\frac{1}{2}a^\star\left[\mu-\frac{b^\star}{a^\star}\right]^2\right\}}_{\text{Substitute the values for } a^\star \text{ and } b^\star \text{ back}} \\
& \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\exp\left\{-\frac{\tau}{2}\left[\frac{n\kappa_0(\bar{y}-\mu_0)^2}{(\kappa_0+n)}\right]\right\}
\end{aligned}
$$

Posterior derivation

$$\Rightarrow \pi(\mu,\tau|Y) \propto \underbrace{\exp\left\{-\frac{1}{2}(n\tau+\kappa_0\tau)\left[\mu-\frac{(n\tau\bar{y}+\kappa_0\tau\mu_0)}{(n\tau+\kappa_0\tau)}\right]^2\right\}}_{\text{Normal Kernel}}$$

$$\times \underbrace{\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau}{2}\left[\nu_0\sigma_0^2+s^2(n-1)+\frac{n\kappa_0}{(\kappa_0+n)}(\bar{y}-\mu_0)^2\right]\right\}}_{\text{Gamma Kernel}}$$

$$= \underbrace{\exp\left\{-\frac{1}{2}\tau(\kappa_0+n)\left[\mu-\frac{(\kappa_0\mu_0+n\bar{y})}{(\kappa_0+n)}\right]^2\right\}}_{\text{Normal Kernel}}$$

$$\times \underbrace{\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau}{2}\left[\nu_0\sigma_0^2+s^2(n-1)+\frac{n\kappa_0}{(\kappa_0+n)}(\bar{y}-\mu_0)^2\right]\right\}}_{\text{Gamma Kernel}}$$

Posterior derivation

$$\Rightarrow \pi(\mu,\tau|Y) \ = \ \mathcal{N}\left(\mu_n, \frac{1}{\kappa_n \tau}\right) \times \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}\right) = \pi(\mu|Y,\tau)\pi(\tau|Y),$$

where

$$\begin{aligned}
\kappa_n &= \kappa_0 + n \\
\mu_n &= \frac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_n} = \frac{\kappa_0}{\kappa_n}\mu_0 + \frac{n}{\kappa_n}\bar{y} \\
\nu_n &= \nu_0 + n \\
\sigma_n^2 &= \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + s^2(n-1) + \frac{n\kappa_0}{\kappa_n}(\bar{y}-\mu_0)^2\right] = \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{n\kappa_0}{\kappa_n}(\bar{y}-\mu_0)^2\right]
\end{aligned}$$

- Turns out that the marginal posterior of μ, that is, $\pi(\mu|Y) = \int_0^\infty \pi(\mu,\tau|Y)d\tau$ is a **t-distribution**.
- You can derive that distribution if you are interested, we won't spend time on it in class. We will be able to sample from it through Monte Carlo anyway.

What's next?

Move on to the readings for the next module!