

STA 360/602L: MODULE 3.4

THE NORMAL MODEL: CONDITIONAL INFERENCE FOR THE MEAN

DR. OLANREWaju MICHAEL AKANDE

NORMAL MODEL

- Suppose we have independent observations $Y = (y_1, y_2, \dots, y_n)$, where each $y_i \sim \mathcal{N}(\mu, \sigma^2)$ or $y_i \sim \mathcal{N}(\mu, \tau^{-1})$, with unknown parameters μ and σ^2 (or τ).
- Then, the likelihood is

$$\begin{aligned} P(Y|\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \tau^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \tau (y_i - \mu)^2 \right\} \\ &\propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \tau \sum_{i=1}^n (y_i - \mu)^2 \right\} \\ &\propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \tau \sum_{i=1}^n [(y_i - \bar{y}) - (\mu - \bar{y})]^2 \right\} \\ &\propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \tau \left[\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (\mu - \bar{y})^2 \right] \right\} \\ &\propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \tau \left[\sum_{i=1}^n (y_i - \bar{y})^2 - n(\mu - \bar{y})^2 \right] \right\} \\ &\propto \tau^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \tau s^2 (n-1) \right\} \exp \left\{ -\frac{1}{2} \tau n (\mu - \bar{y})^2 \right\}. \end{aligned}$$

LIKELIHOOD FOR NORMAL MODEL

- Likelihood:

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

where

- $\bar{y} = \sum_{i=1}^n y_i$ is the sample mean; and
 - $s^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / (n-1)$ is the sample variance.
- Sufficient statistics:
 - Sample mean \bar{y} ; and
 - Sample sum of squares $SS = s^2(n-1) = \sum_{i=1}^n (y_i - \bar{y})^2$.
 - MLEs:
 - $\hat{\mu} = \bar{y}$.
 - $\hat{\tau} = n/SS$, and $\hat{\sigma}^2 = SS/n$.

INFERENCE FOR MEAN, CONDITIONAL ON VARIANCE

- We can break down inference problem for this two-parameter model into two one-parameter problems.
- First start by developing inference on μ when σ^2 is known. Turns out we can use a conjugate prior for $\pi(\mu|\sigma^2)$. We will get to unknown σ^2 in the next module.
- For σ^2 known, the normal likelihood further simplifies to

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

leaving out everything else that does not depend on μ .

- For $\pi(\mu|\sigma^2)$, we consider $\mathcal{N}(\mu_0, \sigma_0^2)$, i.e., $\mathcal{N}(\mu_0, \tau_0^{-1})$, where $\tau_0^{-1} = \sigma_0^2$.
- Let's derive the posterior $\pi(\mu|Y, \sigma^2)$.

INFERENCE FOR MEAN, CONDITIONAL ON VARIANCE

- First, the prior $\pi(\mu|\sigma^2) = \mathcal{N}(\mu_0, \tau_0^{-1})$ can be written as

$$\begin{aligned}\Rightarrow \pi(\mu|\sigma^2) &= \frac{1}{\sqrt{2\pi}} \tau_0^{\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} \tau_0 (\mu - \mu_0)^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \tau_0 (\mu^2 - 2\mu\mu_0 + \mu_0^2) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \tau_0 (\mu^2 - 2\mu\mu_0) \right\}.\end{aligned}$$

- When the normal density is written in this form, note the following details in the exponent.
 - First, we must have $\mu^2 - 2\mu$, and whatever term we see multiplying 2μ must be the mean, in this case, μ_0 .
 - Second, the precision τ_0 is outside the parenthesis.

INFERENCE FOR MEAN, CONDITIONAL ON VARIANCE

- Now to the posterior:

$$\pi(\mu|Y, \sigma^2) \propto \pi(\mu|\sigma^2)P(Y|\mu, \sigma^2) \propto \exp\left\{-\frac{1}{2}\tau_0(\mu - \mu_0)^2\right\} \exp\left\{-\frac{1}{2}\tau n(\mu - \bar{y})^2\right\}$$

- Expanding out squared terms

$$\Rightarrow \pi(\mu|Y, \sigma^2) \propto \exp\left\{-\frac{1}{2}\tau_0(\mu^2 - 2\mu\mu_0 + \mu_0^2)\right\} \exp\left\{-\frac{1}{2}\tau n(\mu^2 - 2\mu\bar{y} + \bar{y}^2)\right\}$$

- Ignoring terms not containing μ

$$\begin{aligned}\Rightarrow \pi(\mu|Y, \sigma^2) &\propto \exp\left\{-\frac{1}{2}\tau_0(\mu^2 - 2\mu\mu_0)\right\} \exp\left\{-\frac{1}{2}\tau n(\mu^2 - 2\mu\bar{y})\right\} \\ &= \exp\left\{-\frac{1}{2}[\tau_0(\mu^2 - 2\mu\mu_0) + \tau n(\mu^2 - 2\mu\bar{y})]\right\} \\ &= \exp\left\{-\frac{1}{2}[\mu^2(\tau n + \tau_0) - 2\mu(\tau n\bar{y} + \tau_0\mu_0)]\right\}.\end{aligned}$$

INFERENCE FOR MEAN, CONDITIONAL ON VARIANCE

- This sort of looks like a normal kernel but we need to do a bit more work to get there.
- Particularly, we need to have it be of the form $b(\mu^2 - 2\mu a)$, so that we have a as the mean and b as the precision.
- We have

$$\begin{aligned}\pi(\mu|Y, \sigma^2) &\propto \exp \left\{ -\frac{1}{2} [\mu^2(\tau n + \tau_0) - 2\mu(\tau n \bar{y} + \tau_0 \mu_0)] \right\} \\ &= \exp \left\{ -\frac{1}{2} \cdot (\tau n + \tau_0) \left[\mu^2 - 2\mu \left(\frac{\tau n \bar{y} + \tau_0 \mu_0}{\tau n + \tau_0} \right) \right] \right\}.\end{aligned}$$

which now looks like the kernel of a normal distribution.

POSTERIOR WITH PRECISION TERMS

- Again, the posterior is

$$\pi(\mu|Y, \sigma^2) \propto \exp \left\{ -\frac{1}{2} \cdot (\tau n + \tau_0) \left[\mu^2 - 2\mu \left(\frac{\tau n \bar{y} + \tau_0 \mu_0}{\tau n + \tau_0} \right) \right] \right\}.$$

- So, in terms of precision, we have

$$\mu|Y, \sigma^2 \sim \mathcal{N}(\mu_n, \tau_n^{-1})$$

where

$$\mu_n = \frac{\tau n \bar{y} + \tau_0 \mu_0}{\tau n + \tau_0}$$

and

$$\tau_n = \tau n + \tau_0.$$

POSTERIOR WITH PRECISION TERMS

- As mentioned before, Bayesians often prefer to talk about precision instead of variance.
- We have
 - τ as the sampling precision (how close the y_i 's are to μ).
 - τ_0 as the prior precision (our prior belief about the uncertainty about μ around our prior guess μ_0).
 - τ_n as the posterior precision
- From the posterior, we can see that, *the posterior precision equals the prior precision plus the data precision.*
- That is, once again, the posterior information is a combination of the prior information and the information from the data.

POSTERIOR WITH PRECISION TERMS: COMBINING INFORMATION

- Posterior mean is weighted sum of prior information plus data information:

$$\begin{aligned}\mu_n &= \frac{n\tau\bar{y} + \tau_0\mu_0}{\tau n + \tau_0} \\ &= \frac{\tau_0}{\tau_0 + \tau n} \mu_0 + \frac{n\tau}{\tau_0 + \tau n} \bar{y}\end{aligned}$$

- Recall that σ^2 (and thus τ) is known for now.
- If we think of the prior mean as being based on κ_0 prior observations from a similar population as y_1, y_2, \dots, y_n , then we might set $\sigma_0^2 = \frac{\sigma^2}{\kappa_0}$, which implies $\tau_0 = \kappa_0\tau$, and then the posterior mean is given by

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}.$$

POSTERIOR WITH VARIANCE TERMS

- In terms of variances, we have

$$\mu|Y, \sigma^2 \sim \mathcal{N}(\mu_n, \sigma_n^2)$$

where

$$\mu_n = \frac{\frac{n}{\sigma^2} \bar{y} + \frac{1}{\sigma_0^2} \mu_0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

and

$$\sigma_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}.$$

- It is still easy to see that we can re-express the posterior information as a sum of the prior information and the information from the data.

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!