STAT 676 - Bayesian Statistics

Jorge Carlos Román
San Diego State University

Spring 2016

1-Sample Normal Model: Unknown SD

- Data: $Y_1, Y_2, \ldots, Y_n | \mu, \sigma \sim \mathrm{N}(\mu, \sigma^2)$,
- We now assume that both μ and σ are unknown parameters.
- It is convenient to use the precision parameter $\tau = 1/\sigma^2$, so we write $\mathrm{N}(\mu, 1/\tau)$ instead.
- The commonly used reference prior is the improper density

$$f(\mu, \tau) = \frac{1}{\tau}$$

- This prior denstity is *not* a probability density (i.e. it does not integrate to 1.)

1-Sample Normal Model: Unknown SD

- The joint posterior density of μ and τ is complicated.
- However, one can show that the marginal posterior density of τ is

$$\tau | Y \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

  where s is the usual sample standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

- The conditional posterior density of μ given τ is

$$\mu | \tau, Y \sim \text{N}\left(\bar{y}, \frac{1}{n\tau}\right) \quad ,$$

  where $\bar{y}$ is the sample mean and Y is the random data vector.
- The product of these two densities gives us the joint posterior density of (μ, τ).

1-Sample Normal Model: Unknown SD

- To obtain a samples from the posterior density of (μ, τ), we use the sequential method.
  1. Sample
  $$\tau|Y \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$
  2. Sample
  $$\mu|\tau, Y \sim \text{N}\left(\bar{y}, \frac{1}{n\tau}\right) \quad .$$

Probability Intervals for σ

- We showed in class that

$$(n-1)s^2\tau|Y = \frac{(n-1)s^2}{\sigma^2}|Y \sim \chi^2_{n-1} \ ,$$

- This leads to a $(1-\alpha)100\%$ PI for σ since

$$1-\alpha = P\left(\sqrt{\frac{(n-1)s^2}{u}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{l}}\middle| y\right) \ ,$$

where $l = \chi^2_{n-1}(\alpha/2)$ and $u = \chi^2_{n-1}(1-\alpha/2)$.

Probability Intervals for μ

- We showed in class that

$$\frac{\mu - \bar{Y}}{\frac{s}{\sqrt{n}}} \Big| Y \sim t_{n-1} \; .$$

- A $(1-\alpha)100\%$ PI for μ is

$$\bar{y} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \; .$$

- Note that these PIs look like frequentist confidence intervals; however, the interpretation is different.
- Bayesian analysis: $\bar{y}$ and s are treated as fixed numbers but μ and σ are random variables.
- Frequentist analysis: $\bar{y}$ and s are random variables but μ and σ are (unknown) fixed parameters.

PIs for a new observation Y_{n+1}

- We showed in class that

$$\left.\frac{Y_{n+1} - \bar{Y}}{s\sqrt{1+\frac{1}{n}}}\right| Y \sim t_{n-1} \ ,$$

- A $(1-\alpha)100\%$ PI for Y_{n+1} is

$$\bar{y} \pm t_{n-1,1-\frac{\alpha}{2}} s\sqrt{1+\frac{1}{n}} \ .$$

Practice Problems

- Read example 5.2.1 from the textbook.
- Derive the posterior distribution of the Normal model in Section 5.2.1. (This will also be done in class.)
- Derive the posterior distribution of the Normal model in Section 5.2.2. (Conjugate priors)
- Exercises: 5.19-5.26. Use R instead of WinBUGS.

1-Sample Normal Model: Unknown SD
Proper Independet Priors

- Data: $Y_1, Y_2, \ldots, Y_n | \mu, \sigma \sim \mathsf{N}(\mu, \sigma^2)$,
- We now assume that both μ and σ are unknown parameters.
- It is convenient to use the precision parameter $\tau = 1/\sigma^2$, so we write $\mathsf{N}(\mu, 1/\tau)$ instead.
- Let's now consider a proper prior distribution on μ and τ

$$\mu \sim \mathsf{N}(a, 1/b) \quad \perp \quad \tau \sim \mathsf{Gamma}(c, d) \ .$$

- The hyper-parameters a, b, c, and d can be chosen using the expert's knowledge (beliefs) about μ and τ.

- The joint posterior distribution cannot be obtained using calculus.
- We will use Markov chain Monte Carlo methods to obtain approximate samples from the intractable joint posterior density.
- STOP! - Before you continue you must read the notes on Markov chain Monte Carlo algorithms, in particular, the material on the Gibbs sampler.

- To implement the two-component Gibbs sampler we need to sample from following two conditional distributions.
- The conditional posterior density of τ given μ is

$$\tau|\mu, Y \sim \text{Gamma}\left(c + \frac{n}{2}, d + \frac{\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]}{2}\right)$$

  where s is the usual sample standard deviation.
- The conditional posterior density of μ given τ is

$$\mu|\tau, Y \sim \text{N}\left(\hat{\mu}, \frac{1}{n\tau + b}\right) \ ,$$

  where

$$\hat{\mu} = \hat{\mu}(\tau) = \left(\frac{n\tau}{n\tau + b}\right)\bar{y} + \left(\frac{b}{n\tau + b}\right)a$$

- To obtain a samples from the posterior density of (μ, τ), we use the Gibbs sampler Markov chain $\{(\mu_n, \tau_n)\}_{n=0}^{\infty}$ as follows.
  1. Select a starting pair (μ_0, τ_0).
  2. If the current state of the Markov chain is (μ_n, τ_n), we obtain the next state (μ_{n+1}, τ_{n+1}) in two steps:
     a. Generate τ_{n+1} from the

$$\mathrm{Gamma}\left(c + \frac{n}{2}, d + \frac{\left[(n-1)s^2 + n(\bar{y} - \mu_n)^2\right]}{2}\right)$$

distribution.

     b. Generate μ_{n+1} from the

$$\mathrm{N}\left(\hat{\mu}_n, \frac{1}{n\tau_{n+1} + b}\right)$$

distribution, where

$$\hat{\mu}_n = \left(\frac{n\tau_{n+1}}{n\tau_{n+1} + b}\right)\bar{y} + \left(\frac{b}{n\tau_{n+1} + b}\right)a \ .$$

Practice Problems

- Derive the conditional distributions used in the Gibbs sampler for the posterior distribution of μ and τ. See Example 6.3.1.
- Implement the Gibbs sampler in R. Write code that runs the Gibbs sampler. Hint: Follow the code for the toy example given earlier.

Choosing the normal prior for μ

- To find a $N(a, 1/b)$ prior density for μ, we can ask an expert the following questions:
  1. What is your best guess for μ?
  2. What is the largest reasonable value for μ? More precisely, a value with only a 5% chance that μ would exceed it.
- One can set the prior mean to the experts' best guess.
- The second question gives us the 95th percentile of the prior distribution; that is, a number c such that

$$P_{a,b}(\mu < c) = a + 1.645\sqrt{\frac{1}{b}} = 0.95 \ .$$

- To obtain the pair (a, b) that satisfies these constraints, we solve a simple 2×2 system of equations.

Choosing the gamma prior for τ

- Recall that $\tau \sim$ Gamma(c, d). We need to find appropriate values for c and d.
- One can ask an expert about σ or about a *percentile* of the data and then translate the information into statements about τ.
- One can also work directly with the prior distribution of σ^2.
- Since the data are assumed to follow a $N(\mu, \sigma^2)$ distribution, we know that the (100α)th percentile is

$$\text{Percentile}_\alpha = \mu + z_\alpha \sigma \ .$$

- In a Bayesian analysis, this percentile is assumed to be a random number and it has a prior distribution.

Choosing the gamma prior for τ

- To find the values of c and d, we can ask an expert the following questions about Percentile$_\alpha$:
  1. What is your best guess for Percentile$_\alpha$?
  2. What is the largest reasonable value for Percentile$_\alpha$? This would be the 95th percentile of (the random) Percentile$_\alpha$. (A percentile of a percentile.)
- When talking to the expert, you should pick a value of α, say $\alpha = 0.90$, and ask about the 90th percentile of the *data*.

Choosing the gamma prior for τ

- Assume that the expert has already given you a guess for the prior mean of μ. This is the a in $N(a, 1/b)$.
- Given $\mu = a$, information about Percentile$_\alpha$ can be translated into information about σ (or σ^2) since

$$\sigma = \frac{\text{Percentile}_\alpha - a}{z_\alpha} .$$

- Let $p_{0.95}$ be the 95th percentile of Percentile$_\alpha$, then

$$0.95 = P(\text{Percentile}_\alpha < p_{0.95} | \mu = a) = P\left(\sigma < \frac{p_{0.95} - a}{z_\alpha}\right) .$$

- We now work with the prior distribution of $\sigma^2 = 1/\tau$.

Definition

A random variable θ has an Inverted Gamma distribution with parameters $a, b > 0$ if it has p.d.f

$$f_\theta(\theta; a, b) = \begin{cases} \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} e^{-\frac{b}{\theta}} & \text{for } \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

where Γ is the gamma function. Notation: $\theta \sim \text{IG}(a, b)$.

Theorem

If $\theta \sim$ IG(a, b), then

$$\textit{mode}(\theta) = \frac{b}{a+1} \,.$$

Also, $\theta \sim$ IG(a, b) if and only if $1/\theta \sim$ Gamma(a, b). The mean of θ exists if $a > 1$ and the variance exists if $a > 2$.

- Note that if $\tau \sim$ Gamma(c, d) then $1/\sigma^2 \sim$ IG(c, d).
- To find the hyperparameters c and d, we set the prior mode of σ^2 to the prior guess of σ^2. This gives us the first equation of the desired 2×2 system:

$$\frac{d}{c+1} = \left(\frac{\text{Percentile}_\alpha - a}{z_\alpha}\right)^2 .$$

- The second equation comes from the guess of the 95th percentile of Percentile$_\alpha$:

$$0.95 = P_{c,d}\left(\sigma^2 < \left(\frac{p_{0.95} - a}{z_\alpha}\right)^2\right)$$

  and is equivalent to

$$0.05 = P_{c,d}\left(\tau < \left(\frac{z_\alpha}{p_{0.95} - a}\right)^2\right) .$$

- This 2×2 system can be solved numerically. See the R-code in the `Bayesfunctions.r` file.
- An alternate approach would be to obtain guesses for τ (and not σ^2). In this case, the guess for τ is set to the prior mode of a Gamma(c, d) density. This avoids using the mode of an inverted gamma density but leads to different c and d values.
- The approach discussed here is different from the one given in the textbook (Chapter 5, p. 117-119).

2-Sample Normal Model: Unknown SDs

Sample 1: $Y_{11}, Y_{12}, \ldots, Y_{1n_2} | \mu_1, \sigma_2 \sim \mathrm{N}(\mu_1, \sigma_1^2)$

Sample 2: $Y_{21}, Y_{22}, \ldots, Y_{2n_2} | \mu_2, \sigma_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$

- The samples are assumed to be independent.
- The parameters are $\mu = (\mu_1, \mu_2)$ and $\sigma = (\sigma_1, \sigma_2)$.
- The commonly used reference prior is the improper density

$$f(\mu, \tau) = \frac{1}{\tau_1} \times \frac{1}{\tau_2} \ ,$$

  where $\tau_i = 1/\sigma_i^2$ and $\tau = (\tau_1, \tau_2)$.
- This prior denstity is *not* a probability density (i.e. it does not integrate to 1.)

2-Sample Normal Model: Unknown SDs

- The posterior joint distribution of (μ, τ) (or (μ, σ)) can be obtained in closed form.
- Since we have two independent samples, the derivation of the posterior is almost identical that in the 1-sample case.
- The predictive densities (of new observations) can also be obtained in closed form.
- When the values of σ_1 and σ_2 are close, one can look at the posterior distribution of $\mu_1 - \mu_2$ to assess group differences.
- When the values of σ_1 and σ_2 *not* close, it is recommended to look at the predictive distribution of new observations in each group.

2-Sample Model w/ Proper Priors

$$\text{Sample 1:} \quad Y_{11}, Y_{12}, \ldots, Y_{1n_2} | \mu_1, \sigma_2 \sim \mathsf{N}(\mu_1, \sigma_1^2)$$
$$\text{Sample 2:} \quad Y_{21}, Y_{22}, \ldots, Y_{2n_2} | \mu_2, \sigma_2 \sim \mathsf{N}(\mu_2, \sigma_2^2)$$

- Consider now the same data model, but use the following proper priors

$$\mu_i \sim \mathsf{N}(a_i, 1/b_i) \quad \perp \quad \tau_i \sim \mathsf{Gamma}(c_i, d_i) \quad i = 1, 2\ .$$

- The hyper-parameters (a_i, b_i, c_i, and d_i) can be obtained using the 1-sample techniques discussed earlier.

2-Sample Model w/ Proper Priors

- The posterior joint distribution of (μ, τ) (or (μ, σ)) cannot be obtained in closed form.
- One can use a Gibbs sampler to obtain approximate posterior draws.
- Approximate draws from the predictive densities can be obtained using the Gibbs sampler and sequential sampling. (See the R-code for details.)

Gibbs Sampler

- Recall that $\mu = (\mu_1, \mu_2)$ and $\tau = (\tau_1, \tau_2)$.
- Simulate the GS Markov chain $\{(\mu^{(n)}, \tau^{(n)})\}_{n=0}^{\infty}$ as follows:
  1. Select a starting point $(\mu^{(0)}, \tau^{(0)}) \in \mathbb{R}^2 \times \mathbb{R}_+^2$.
  2. If the current state of the Markov chain is $(\mu^{(n)}, \tau^{(n)})$, we obtain the next state $(\mu^{(n+1)}, \tau^{(n+1)})$ in two steps:
     a. For $j = 1, 2$, generate

$$\tau_j^{(n+1)} \sim \text{Gamma}\left(c_j + \frac{n_j}{2}, d_j + \frac{\left[(n_j - 1)s_j^2 + n_j\left(\bar{y}_j - \mu_j^{(n)}\right)^2\right]}{2}\right)$$

     b. For $j = 1, 2$, generate

$$\mu_j^{(n+1)} \sim \text{N}\left(\hat{\mu}_j^{(n)}, \frac{1}{n_j\tau_j^{(n+1)} + b_j}\right) ,$$

where

$$\hat{\mu}_j^{(n)} = \left(\frac{n_j\tau_j^{(n+1)}}{n_j\tau_j^{(n+1)} + b_j}\right)\bar{y}_j + \left(\frac{b_j}{n_j\tau_j^{(n+1)} + b_j}\right)a_j .$$

Practice Problems

- Read Example 5.2.2 from the textbook.
- Solve exercises 5.28 ans 5.29.
- Perform a sensitivity analysis of the model in Example 5.2.2 using the common reference prior.
- Use R instead of WinBUGS.