Homework 5: STA 721 Fall19

Your Name

October 11, 2019

1. Suppose $\mathbf{Y} \sim N(\mathbf{1}_n\beta_0 + \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n/\phi)$ where $\mathbf{X}^T\mathbf{1} = 0$ i.e. the columns of $\mathbf{X}$ have been centered to have mean 0, and you are using the g-prior for $\boldsymbol{\beta}$ specified as follows:

$$\boldsymbol{\beta} \mid \beta_0, \phi, g \sim N(0, \frac{g}{\phi}(\mathbf{X}^T\mathbf{X})^{-1}$$

where $\mathbf{X}$ is full column rank and independent Jeffreys priors for β_0 and ϕ,

$$p(\beta_0, \phi) = p(\beta_0)p(\phi) \propto 1 \cdot 1/\phi$$

   (a) Show that the likelihood function for $\beta_0, \boldsymbol{\beta}, \phi$ can be expressed as

$$\mathcal{L}(\beta_0, \boldsymbol{\beta}, \phi) \propto \frac{n}{2}\log(\phi) - \frac{\phi}{2}\left(n(\beta_0 - \bar{y})^2 + (\mathbf{Y}_c - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y}_c - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right)$$

   where $\mathbf{Y}_c = \mathbf{Y} - \mathbf{1}_n\bar{y}$. (suggestion: work by Monday)

   (b) Find the posterior distribution of $\beta_0, \boldsymbol{\beta} \mid \phi, g$ and $\phi \mid g$. Simplify so that results are functions of sufficient statistics $\bar{\mathbf{Y}}$, $\hat{\boldsymbol{\beta}}$ and SSEas in notes. Are β_0 and $\boldsymbol{\beta}$ independent given ϕ and $\mathbf{Y}$? You should state the distributions and their hyperparameters. (suggestion: work by Monday)

   (c) Find $\tilde{\beta} = E_{\beta \mid Y, g}[\beta \mid Y, g]$, the posterior mean under the Zellner g-prior from above. (suggestion: work by Monday)

   (d) Find the sampling distribution of $\tilde{\beta}$. (i.e treating $\tilde{\beta}$ as a function of $\mathbf{Y}$ given the true parameter, say $\boldsymbol{\beta}_T, \phi, \beta_0$ what is the distribution of $\tilde{\boldsymbol{\beta}}$? (attempt by Monday)

   (e) Is the posterior mean under the g-prior used here an unbiased estimator of $\boldsymbol{\beta}_T$. If not, what is the bias?

   (f) Using the sampling distribution of the data, calculate the expected loss for using the posterior mean to estimate $\boldsymbol{\beta}$ under squared error lossL

$$\mathsf{E}_{\mathbf{Y} \mid \beta_0, \boldsymbol{\beta}, \phi, g}[\|\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|\|^2]$$

   Hint: recall theorem regarding expectations of quadratic forms

   (g) The Gauss-Markov Theorem showed that out of the class of unbiased linear estimators that the MLE (OLS) estimator had the smallest variance. If we use the posterior mean, $\tilde{\boldsymbol{\beta}}$ as an estimator for $\boldsymbol{\beta}$, can the posterior mean have a smaller expected loss than the MLE for estimating $\boldsymbol{\beta}$? Can it be worse? Make a plot to illustrate with $g/(1+g)$ on the x-axis and MSE (expected loss) on the y-axis for the Bayes estimator (posterior mean). You may need to assume or fix values for some quantities that go into the loss, if so how sensitive are the plots/conclusions to those assumptions?)

   (h) Find a value of g that minimizes the expected MSE of the Bayes estimator. Add this point to the graph above. With this value for g will the expected MSE with the Bayes estimator always be smaller that the expected MSE under the MLE/OLS estimator (Bayes with independent Jeffreys)? If this depends on unknown parameters, describe how you might estimate g.

(i) Repeat 1f, but now consider estimation of $\boldsymbol{\mu} = \mathbf{1}_n \beta_0 + \mathbf{X}\boldsymbol{\beta}$.

(j) Repeat 1g, but now consider estimation of $\boldsymbol{\mu} = \mathbf{1}_n \beta_0 + \mathbf{X}\boldsymbol{\beta}$.

(k) Repeat 1h, but now consider estimation of $\boldsymbol{\mu} = \mathbf{1}_n \beta_0 + \mathbf{X}\boldsymbol{\beta}$.

(l) Does the optimal choice for g depend on whether you are estimating $\boldsymbol{\beta}$ or $\boldsymbol{\mu}$?

2. If $\tau = 1/g \sim G(1/2, n/2)$, find the marginal prior for $\boldsymbol{\beta} \mid \phi, \beta_0$.

3. Derive the conditional posterior distribution of $\tau \mid \beta_0, \beta, \phi, \mathbf{Y}$. *Use the likelihood decomposition above*

4. Refer to the Prostate data from HW4 and the model `lcavol` $\sim$ `1 + factor(Gleason)`,

   (a) Obtain 95% credible intervals analytically (i.e. using the appropriate t-distributions) for the 4 group means with the g-prior with $g = n$ for $\boldsymbol{\beta} \mid \phi, \beta_0$ (with centered $\mathbf{X}$) and calculate for the data.

   (b) Use JAGS and an approximation to the independent Jeffreys prior $p(\beta_0, \phi) \propto 1/\phi$, to obtain 95% credible intervals for the 4 group means using MCMC based on your distributions above under the g-prior with $g = n$ for $\boldsymbol{\beta} \mid \phi, \beta_0$ (with centered $\mathbf{X}$). Use the multivariate normal distribution in JAGS and remember jags specifies normal distributions using precisions rather than variances. (Try before Wednesday). Your results should be close to the non-simulation based solutions above.

   (c) Extend the JAGS code to use a mixture of the g-prior by adding that $\tau = 1/g \sim G(1/2, n/2)$.

   (d) Find the Student-t distribution for $\boldsymbol{\beta} \mid \phi$ by integrating out τ. Use the (`dmt` in the JAGS code instead of the multivariate normal g-prior. Are you intervals comparable to those above? (if code is correct they should agree)

   (e) Will your results above be the same if you change the baseline category? Explain.

5. (Optional Challenge) If W_1 and W_2 have a joint multivariate normal distribution with

$$\begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} \\ \Sigma_{21} \\ \Sigma_{21} \\ \Sigma_{22} \end{bmatrix} \right)$$

then the conditional distribution of W_1 given W_2 is

$$W_1 \mid W_2 = w_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-}(w_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21})$$

where Σ_{22}^{-} is a generalized inverse of Σ_{22}. Suppose that for the model $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathsf{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ we use a g-prior for $\boldsymbol{\beta}$ with a generalized inverse

$$\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{0}, g/\phi(\mathbf{X}^T\mathbf{X})^{-}).$$

   (a) Find the joint (normal) distribution of $\mathbf{Y}$ and $\boldsymbol{\beta}$ given ϕ.

   (b) Use the result about conditional normals above to find the posterior distribution of $\boldsymbol{\beta}$ given ϕ. If $\mathbf{X}$ is full rank do you obtain the same results based on using densities to derive the posterior distribution? Rearrange until you do! Does this suggest an alternative representation in the case of the non-full rank case?

   (c) Find the posterior distribution of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ given ϕ. Does the latter depend on the choice of generalized inverse? (Express the result as a function of $\mathbf{P_X}$)