Homework 5: STA 721 Fall24

Your Name

September 26, 2024; Due in one week (see Gradescope)

You are encouraged to write up solutions in LaTeX.

1. Let $\mathbf{Y} \mid \boldsymbol{\beta} \sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, (\phi\boldsymbol{\Phi})^{-1})$. Find the conditional distribution of $\boldsymbol{\beta}$ given $\mathbf{Y}$, $\boldsymbol{\Phi} > 0$ and ϕ under the prior distribution $\boldsymbol{\beta} \sim \mathsf{N}(\mathbf{b}_0, \boldsymbol{\Phi}_0^{-1})$. In the case that $\mathbf{b}_0 = \mathbf{0}$, compare the posterior expectation of $\boldsymbol{\beta}$ to the GLS estimator where $\sigma^2\boldsymbol{\Sigma} = (\phi\boldsymbol{\Phi})^{-1}$. (Try to express the posterior expectation in terms of the GLS estimator.)

2. Consider the Bayes estimator $\hat{\boldsymbol{\beta}}_g = \frac{g}{1+g}\hat{\boldsymbol{\beta}}$ under the g-prior where $\hat{\boldsymbol{\beta}}$ is the OLS estimator under the usual linear model $\mathbf{Y} \mid \boldsymbol{\beta} \sim \mathsf{N}_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}/\phi)$.

   (a) Compute the bias, covariance and MSE under squared error loss of this estimator as a function of g and $\boldsymbol{\beta}$ (and ϕ).

   (b) Plot the MSE as a function of g, for fixed $\boldsymbol{\beta}$ and $\phi = 1$. Overlay lines that show the squared bias term and the trace of the covariance matrix, along with a line for the MSE of the OLS estimator. For what values of g is the MSE of $\hat{\boldsymbol{\beta}}_g$ lower than that of $\hat{\boldsymbol{\beta}}$?

   (c) Plot the MSE as a function of $\boldsymbol{\beta}$, for fixed g. for what values of $\boldsymbol{\beta}$ is the MSE of $\hat{\boldsymbol{\beta}}_g$ lower than that of $\hat{\boldsymbol{\beta}}$?

   (d) Find the value of g that minimizes the MSE of $\hat{\boldsymbol{\beta}}_g$. Does this value guarantee that the MSE of $\hat{\boldsymbol{\beta}}_g$ is lower than the MSE of $\hat{\boldsymbol{\beta}}$? Is it possible to use this prior in practice?

3. Suppose researcher 1 will estimate $\boldsymbol{\beta}$ from $\mathbf{Y}$ and $\mathbf{X}$ in the linear model $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n/\phi)$, whereas researcher 2 will estimate $\boldsymbol{\alpha}$ in the linear model $\mathbf{Y} \sim \mathsf{N}(\mathbf{W}\boldsymbol{\alpha}, \mathbf{I}_n/\phi)$, where $\mathbf{W} = \mathbf{X}\mathbf{A}$ for some non-singular $p \times p$ matrix $\mathbf{A}$. Note that if $\mathbf{a} \in \mathbb{R}^p$ is the true value of $\boldsymbol{\alpha}$ in the second model, then $\mathbf{A}\mathbf{a}$ is the true value of $\boldsymbol{\beta}$ in the first model.

   (a) Find the OLS estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ based on $(\mathbf{Y}, \mathbf{X})$, the OLS estimator $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ based on $(\mathbf{Y}, \mathbf{W})$ and describe the relationship between the estimators. Is $\hat{\boldsymbol{\beta}} = \mathbf{A}\hat{\boldsymbol{\alpha}}$?

   (b) Now both researchers use $g-$priors with the same value of g to obtain their estimators $\hat{\boldsymbol{\beta}}_g$ and $\hat{\boldsymbol{\alpha}}_g$. Describe the relationship between the estimators. Is $\hat{\boldsymbol{\beta}}_g = \mathbf{A}\hat{\boldsymbol{\alpha}}_g$?

   (c) Now both researchers use ridge priors with the same value of κ to obtain their estimators $\hat{\boldsymbol{\beta}}_\kappa$ and $\hat{\boldsymbol{\alpha}}_\kappa$. Describe the relationship between the estimators. Under what conditions on $\mathbf{A}$ is $\hat{\boldsymbol{\beta}}_\kappa = \mathbf{A}\hat{\boldsymbol{\alpha}}_\kappa$?

   (d) Since $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} = \mathbf{W}\boldsymbol{\alpha}$ does not depend on the model parameterization, we may wish that the posterior distribution of $\mathbf{X}\boldsymbol{\beta}$ equals the posterior distribution of $\mathbf{W}\boldsymbol{\alpha}$ for all linear reparameterizations of the model of the form above. Restricting attention to normal prior distributions with mean zero and precision $\phi\boldsymbol{\Phi}_0$, show that this posterior distribution of $\mathbf{X}\boldsymbol{\beta}$ is the same as $\mathbf{W}\boldsymbol{\alpha}$ if and only if $\boldsymbol{\Phi}_0 = \mathbf{C}^T\mathbf{C}$ where $\boldsymbol{\mu} \in C(\mathbf{C})$.

4. Group priors: Consider Bayesian estimation of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in the linear model $\mathbf{Y} \sim \mathsf{N}(\mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n/\phi)$.

(a) Show that this model may always be reparameterized so that that the blocks in the design matrix are orthogonal, i.e., $\mathbf{W}^T\tilde{\mathbf{X}} = \mathbf{0}$. Hint: let $\mathbf{P_W} = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$ denote the orthogonal projection onto the column space of $\mathbf{W}$ and write $\mathbf{X} = \mathbf{P_W}\mathbf{X} + (\mathbf{I}_n - \mathbf{P_W})\mathbf{X}$. Do the coefficients of $\mathbf{W}$ change in the reparameterization? Do the coefficients of $\mathbf{X}$ change?

(b) Assume now that $\mathbf{W}^T\mathbf{X} = \mathbf{0}$ in the model $\mathbf{Y} \sim \mathsf{N}(\mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n/\phi)$. Under the prior distributions $\boldsymbol{\alpha} \sim \mathsf{N}(\mathbf{0}, g_\alpha(\mathbf{W}^T\mathbf{W})^{-1}/\phi)$ and $\boldsymbol{\beta} \sim \mathsf{N}(\mathbf{0}, g_\beta(\mathbf{X}^T\mathbf{X})^{-1}/\phi)$, with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being a priori independent. Find the posterior distribution of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ conditional on ϕ using results on partitioned matrices (express the posterior mean and covariance in more detail than in terms of inverses of large matrices, i.e., your answer should be more detailed than including something like $([W\,X][W\,X]^T)^{-1}$).

(c) Are $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ given $\mathbf{Y}$ and ϕ independent? Prove.

5. Consider the orthogonal regression setting, $\mathbf{Y} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$ where $\mathbf{X}^T\mathbf{X} = \mathbf{I}_p$, with prior distribution $\boldsymbol{\beta} \sim \mathsf{N}(\mathbf{0}, {\sigma^2}_0\mathbf{I}_p)$. Let $\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$ be the OLS estimator of $\boldsymbol{\beta}$.

(a) Find the distribution of $\hat{\boldsymbol{\beta}}$ conditional on $\boldsymbol{\beta}$ and σ^2, and unconditionally given σ^2 and σ_0^2.

(b) Find the posterior distribution of $\boldsymbol{\beta}$ given $\hat{\boldsymbol{\beta}}$, σ^2 and σ^2 and show that the posterior mean is $(1-\omega)\hat{\boldsymbol{\beta}}$ where $\omega = \frac{\sigma^2}{\sigma^2 + {\sigma^2}_0}$.

(c) Find the distribution of

$$\frac{\|\hat{\boldsymbol{\beta}}\|^2}{\sigma^2 + \sigma_0^2}$$

unconditionally given σ^2 and σ_0^2. Without doing a change of variables, find the

$$\mathsf{E}\left[\frac{\sigma^2 + \sigma_0^2}{\|\hat{\boldsymbol{\beta}}\|^2}\right]$$

and show that $\frac{p-2}{\|\hat{\boldsymbol{\beta}}\|^2}$ is an unbiased estimator of $\frac{1}{\sigma^2+\sigma_0^2}$.

(d) With the estimate $\hat{\omega} \equiv (p-2)\sigma^2/\|\hat{\boldsymbol{\beta}}\|^2$ as an Empirical Bayes estimate of ω, solve for the estimate of σ_0^2 from $\hat{\omega} = \frac{\sigma^2}{\sigma^2+\hat{\sigma}_0^2}$.

(e) Allowing for data dependent priors, will this Empirical Bayes estimate of σ_0^2 lead to a valid "prior" distribution for $\boldsymbol{\beta}$? Why or why not?