SDS 383C: Statistical Modeling I, Fall 2021
Homework 3, Due Oct 28, 12:00 Noon

Instructor: Abhra Sarkar (abhra.sarkar@utexas.edu)
Teaching Assistant: Angela Ting (angelating@utexas.edu)
Department of Statistics and Data Sciences
The University of Texas at Austin
2317 Speedway D9800, Austin, TX 78712-1823, USA

All homework must be submitted typed-in as a single pdf file. Name the file "firstname-lastname-SDS383C-HW-3.pdf" Submit this file without compression such as zip or rar. Figures accompanying the solutions must be presented close to the actual solution. Computer codes will be rarely evaluated but must still be submitted separately from the main file. Codes must be commented properly and should run easily on other machines. Precise, concise, clear, innovative solutions may be rewarded with bonus points. Explain your answer with logic reasoning and/or mathematical proofs. Organize your solutions in the same order as they were presented. If you can solve a problem using multiple techniques, present only your best solution.

1. (15 points) Let $\mathbf{y}_1, \ldots, \mathbf{y}_n \stackrel{iid}{\sim} p(\mathbf{y} \mid \boldsymbol{\theta}) = h(\mathbf{y}) \exp\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{y} - \psi(\boldsymbol{\theta})\}$ for $\mathbf{y} \in \mathcal{Y}$ with $\mathcal{Y}$ not depending on $\boldsymbol{\theta}$.

   (a) Show that $\mathbb{E}(\mathbf{y} \mid \boldsymbol{\theta}) = \boldsymbol{\xi}(\boldsymbol{\theta}) = \dfrac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$.

   Also, let $\boldsymbol{\theta}$ have prior $p(\boldsymbol{\theta} \mid \mathbf{y}_0, \lambda) = h(\mathbf{y}_0, \lambda) \exp\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{y}_0 - \lambda\psi(\boldsymbol{\theta})\}$ with $\mathbf{y}_0 \in \mathcal{Y}$.

   (b) Show that

   $$\mathbb{E}\{\boldsymbol{\xi}(\boldsymbol{\theta})\} = \frac{\mathbf{y}_0}{\lambda} + \text{constant} \qquad \text{and} \qquad \mathbb{E}\{\boldsymbol{\xi}(\boldsymbol{\theta}) \mid \mathbf{y}_{1:n}\} = \frac{\mathbf{y}_0 + n\overline{\mathbf{y}}}{\lambda + n} + \text{constant}.$$

   [Hints: For part (a), take the derivative of $\int p(\mathbf{y} \mid \boldsymbol{\theta}) d\mathbf{y} = 1$ w.r.t. $\boldsymbol{\theta}$ and apply Leibniz rule. For part (b), first prove the prior expectation result. Note that the integration is now w.r.t. $\boldsymbol{\theta}$, not $\mathbf{y}$.

   $$\text{Write} \quad \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\lambda}\left\{\mathbf{y}_0 - \left(\mathbf{y}_0 - \lambda\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right\}$$

   $$\text{and use} \quad \frac{\partial}{\partial \boldsymbol{\theta}} \exp\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{y}_0 - \lambda\psi(\boldsymbol{\theta})\} = \left(\mathbf{y}_0 - \lambda\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \exp\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{y}_0 - \lambda\psi(\boldsymbol{\theta})\}.$$

   For the posterior expectation result, simply appeal to conjugacy.]

2. (10 points) Show that binomial and negative binomial distributions belong to exponential families.

3. (15 points) Let $y \sim \text{Bin}(10, \theta)$. Also, let the observed value of $y = 3$. Let the prior on θ be a mixture of two Beta's as

$$p(\theta) = \frac{0.5}{\text{Beta}(10,20)}\theta^{10-1}(1-\theta)^{20-1} + \frac{0.5}{\text{Beta}(20,10)}\theta^{20-1}(1-\theta)^{10-1}.$$

(a) Show that the posterior has the form

$$p(\theta \mid y) = \frac{\frac{\text{Beta}(13,27)}{\text{Beta}(10,20)}\frac{1}{\text{Beta}(13,27)}\theta^{12}(1-\theta)^{26} + \frac{\text{Beta}(23,17)}{\text{Beta}(20,10)}\frac{1}{\text{Beta}(23,17)}\theta^{22}(1-\theta)^{16}}{\frac{\text{Beta}(13,27)}{\text{Beta}(10,20)} + \frac{\text{Beta}(23,17)}{\text{Beta}(20,10)}}.$$

(b) Plot the posterior superimposed on the prior.

(c) Compute a 90% posterior credible interval for θ.

4. (10 points) Prove that Jeffreys' priors satisfy the invariance principle: starting with $p(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2}$, show that the induced prior on $\boldsymbol{\psi} = g(\boldsymbol{\theta})$, where g is one-one, is $p(\boldsymbol{\psi}) \propto |\mathbf{I}(\boldsymbol{\psi})|^{1/2}$.

5. (5 points) For the Poisson likelihood model $y_1, \ldots, y_n \stackrel{iid}{\sim} \text{Poisson}(\lambda)$, Jeffrey's (improper) prior was derived in class. Compute the corresponding posterior. Is it proper?

6. (25 points) Consider the likelihood model $y_1, \ldots, y_n \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2)$, μ known.
(a) Compute the Jeffreys' prior for σ^2.
(b) Compute also the corresponding posterior.
(c) Draw a random sample of size 20 from a Normal$(0, 1)$ distribution. Using these sampled values as data points and assuming the variance to now be unknown, plot the posterior superimposed on the general shape of the prior.
(d) Compute a 90% centered quantile based credible interval for σ^2.
(e) Compute also a 90% HPD interval for σ^2.