STATISTICS 225 – Fall 2018

Homework 2 - solutions

1. **Improper hyperpriors in hierarchical models, problem 5.10 in BDA3:** We need to show that $p(\theta, \mu, \tau|y)$ has finite integral. In this case we can do some of the required integration analytically

$$\int\int\int p(\theta,\mu,\tau|y)d\theta\ d\mu\ d\tau = \int\int\int p(\theta|\mu,\tau,y)d\theta\ p(\mu|\tau,y)d\mu\ p(\tau|y)d\tau = \int p(\tau|y)d\tau$$

so that $p(\theta, \mu, \tau|y)$ is proper if and only if $p(\tau|y)$ is a proper univariate distribution. The display above uses the fact that the conditional distributions for θ and μ are known normal distributions that must integrate to 1. As for the rest of the problem:

- For τ near zero, the posterior distribution is a constant multiple of the prior. So we need only check integrability of the prior distribution. An important point to note is that $p(\tau)$ being infinite at $\tau = 0$ is not the problem; if $p(\tau) = 1/\sqrt{\tau}$ then the prior distribution tends to infinity as τ approaches zero but this prior distribution has finite integral near zero.
- The argument from (a) show that if $p(\tau) = 1$ then the posterior is integrable near zero. Now we need to check the behavior as τ gets large. All we need is an upper bound for each term. The exponential term is clearly less than one. We can rewrite the remaining terms as $(\sum_{j=1}^{J}[\prod_{k\neq j}(\sigma_k^2+\tau^2)])^{-1/2}$. For τ larger than one we make this quantity bigger by dropping all of the σ^2 to yield $(J\tau^{2(J-1)})^{-1/2}$. An upper bound on $p(\tau|y)$ for τ large is $J^{-1/2}/\tau^{J-1}$. This upper bound is integrable on the interval (M, ∞) for any large M if $J > 2$ so that $p(\tau|y)$ is integrable if $J > 2$. (For $J = 2$ we would actually need to find a lower bound and show that it is not integrable but the argument is similar).
- For part (c) several people suggested a proper prior distribution for τ. Others noted that $p(\tau) \propto \tau^{-1/2}$ would work fine in the arguments of parts (a) and (b). That would work. A key point is that the data provides very little information about τ if $J = 2$ (only a poor point estimate is possible for τ). I would probably do a two-sample analysis and not worry about the hierarchical structure.

2. **Exponential/gamma hierarchical model - theory:**

(a) The unnormalized joint posterior density is

$$p(\theta,\alpha,\beta|y) \propto p(\alpha,\beta)p(\theta|\alpha,\beta)p(y|\theta) \propto \prod_{i=1}^{9}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\theta_i^{\alpha-1}e^{-\beta\theta_i}\right)\prod_{i=1}^{9}\prod_{j=1}^{2}\left(\theta_i e^{-\theta_i y_{ij}}\right)$$

(b) To find the conditional posterior distribution of θ given the hyperparameters, consider the terms that involve θ in the above:

$$p(\theta|\alpha,\beta,y) \propto \prod_{i=1}^{9}\left(\theta_i^{\alpha+2-1}e^{-(\beta+y_{i1}+y_{i2})\theta_i}\right)$$

Thus the θ_i's are independent in this conditional posterior distribution with $\theta_i \sim \text{Gamma}(\alpha+2, \beta+y_{i1}+y_{i2})$.

(c) There are two equivalent methods for finding $p(\alpha, \beta|y)$. We can take the expression in (a) and integrate over θ (using what we noted in (b)). Or we can take the expression in (a) and divide by the product of independent gamma densities (with their normalizing constants) implied by (b). In either case:

$$p(\alpha,\beta|y) \propto \prod_{i=1}^{9}\left(\frac{\beta^\alpha\Gamma(\alpha+2)}{(\beta+y_{i1}+y_{i2})^{\alpha+2}\Gamma(\alpha)}\right)$$

(d) Take $\alpha = c\beta$ where c is a constant. This yields

$$\begin{aligned} p(\alpha,\beta|y) &\propto \prod_{i=1}^{9}\left(\frac{c\beta(c\beta+1)}{(\beta+y_{i1}+y_{i2})^2(1+\frac{y_{i1}+y_{i2}}{\beta})^{c\beta}}\right) \\ &= \prod_{i=1}^{9} c^2\frac{1+\frac{1}{c\beta}}{(1+\frac{y_{i1}+y_{i2}}{\beta})^2}(1+\frac{y_{i1}+y_{i2}}{\beta})^{-c\beta} \\ \lim_{\beta\to\infty} p(\alpha,\beta|y) &\propto \prod_{i=1}^{9} c^2 e^{-c(y_{i1}+y_{i2})} \end{aligned}$$

Like the examples we saw in class the posterior distribution tends to a positive constant along rays extending out from the origin so that the posterior distribution is not a proper distribution.

3. **Exponential/Gamma hierarchical model - analysis:**

   (a) This is just a two-dimensional transformation of variables. We are given that $p_{\alpha,\beta}(\alpha,\beta) \propto \beta^{-5/2}$. We would like to derive $p_{\phi_1,\phi_2}(\phi_1,\phi_2)$ where $\phi_1 = log(\alpha/\beta)$ and $\phi_2 = log(\beta)$. Note that the inverse transformations are $\beta = e^{\phi_2}$ and $\alpha = e^{\phi_1+\phi_2}$. Then $p_{\phi_1,\phi_2}(\phi_1,\phi_2) = p_{\alpha,\beta}(e^{\phi_1+\phi_2}, e^{\phi_2})|J|$ where J is the Jacobian (matrix of derivatives $d(\alpha,\beta)/d(\phi_1,\phi_2)$. Here the first term is $e^{-5\phi_2/2}$ and $|J| = \begin{vmatrix} e^{\phi_1+\phi_2} & e^{\phi_1+\phi_2} \\ 0 & e^{\phi_2} \end{vmatrix} = e^{\phi_1+2\phi_2}$. Thus $p_{\phi_1,\phi_2}(\phi_1,\phi_2) = e^{\phi_1-0.5\phi_2}$.

   (b) As we discussed in class on Tuesday October 30, the posterior distribution here seems poorly behaved and I don't understand why. Those who tried to run Stan (including me) had trouble getting the algorithm to provide reliable results. I've attached R code and output that uses a grid approximation to the marginal posterior distribution (of transformation of alpha, beta). These "seem" to work on the face of it. However, if you rerun with a larger grid, then you will find that inferences for α and β are not stable. Fortunately, inferences for the θ's which are what is needed below are stable.

   (c) The main point of this question is to show that the posterior distribution $p(\theta,\alpha,\beta|y)$ can be used to provide the posterior probability that manufacturer 3's bulbs have greater mean lifetime than manufacturer 7's bulbs, i.e., $\Pr(1/\theta_3 > 1/\theta_7) = \Pr(\theta_3 < \theta_7|y)$. With our posterior simulations, we merely count the number of times (out of 1000) that θ_3 was less than θ_7 to yield the estimate 0.956.

   (d) To study the predictive distribution for new bulbs from manufacturer one we take each simulated θ_1 and generate a single random exponential lifetime corresponding to each plausible value for the manufacturer parameter. The resulting simulations are plausible lightbulb lifetimes for a new bulb from this manufacturer. The process yields an estimated posterior median of 1.6 and a 95% prediction interval of (0.04, 14.2).

   (e) To obtain a predictive distribution for a new manufacturer we must first simulate a θ parameter for the new manufacturer from the gamma population distribution. To do so, for each simulated (α,β) pair we generate a new θ from the gamma distribution (not updated with any data). Then for each simulated θ we generate a random exponential lifetime. This process yields an estimated posterior median of 2.4 and a 95% prediction interval of (0.06, 90.4). The wider interval relative to part (c) reflects the uncertainty in the θ for this new manufacturer. Another interesting thing to note is that the posterior mean of the predictive distribution is impacted by the small chance of an extremely good manufacturer (value of θ near zero). One might argue that this suggests the model is not reasonable.

   Note an alternative way to do this would be to generate θ for a new manufacturer (as above) and then generate many random exponential lifetimes to form a predictive distribution for that manufacturer. This would not however accommodate our uncertainty about the quality of the new manufacturer. Thus we would have to repeat the above for many θ's and combine the results. Our approach (one simulated lifetime per θ) achieves the same result.

4. **Normal approximation** – The first step here is to create the posterior (or logposterior) distribution for β.

$$\begin{aligned} p(\beta|y) &\propto p(y|\beta)p(\beta) = \prod_{i=1}^{n} \left(\frac{e^{\beta x}}{1+e^{\beta x}}\right)^{y_i} \left(\frac{1}{1+e^{\beta x}}\right)^{1-y_i} \\ \log p(\beta|y) &= \sum_{i=1}^{n} (\beta x_i y_i - \log(1+e^{\beta x})) + \text{constant} \end{aligned}$$

   (a) R code and output in the accompanying R Markdown document.

   (b) You can just use the grid to identify the mode. The attached R code uses Newton's method. Noting that the first derivative is $der1 = \sum_{i=1}^{n} x_i y_i - x_i \frac{e^{\beta x_i}}{1+e^{\beta x_i}}$ and $der2 = -\sum_i x_i^2 \frac{e^{\beta x_i}}{1+e^{\beta x_i}} \frac{1}{1+e^{\beta x_i}}$. The mode turns out to be $\hat{\beta} = .7075$ and the negative 2nd derivative is 5.3897.

   (c) Then the normal approximation is $N(.7075, 1/5.3897)$ (note the variance is 1/5.3897 not the s.d.). The normal approximation fits reasonably well but does not capture the asymmetry in the posterior distribution. The left tail (negative β values) is much less likely than predicted by the normal approximation and the right tail (positive β values) is much more likely.