Stat 111 Spring 2010 Week 12 Problems - Bayesian Inference

1. Consider the 2-level Normal hierarchical model with constant and known level-1 variance $V$:

$$ Y_i | \theta_i \sim^{\text{indep}} N(\theta_i, V), \quad i = 1, \ldots, k. $$

$$ \theta_i | \mu, A \sim^{\text{i.i.d.}} N(\mu, A). $$

For example, each $Y_i$ might represent the average point total for a particular team in their first $n$ games. Assuming the same standard deviation for all teams, we have $V = \sigma/n$, with $\sigma$ estimated based on $k(n - 1)$ degrees of freedom.

a) It is well-established that a Uniform density on $(-\infty, \infty)$ is a reasonable non-informative prior density for the mean $\mu$ of a Normal random variable. Treating $\mu$ as known (for now) derive the Jeffreys prior density for $B = V/(V + A)$:

$$ p(B) \propto \left(-E \left( \frac{\partial^2}{\partial^2 B} \log(L(B)) \right) \right)^{1/2}, \quad 0 < B \leq 1. $$

b) If an independent Uniform prior distribution is assumed for $\mu$, then your answer to (a) is the implicit joint-prior density for $\mu$ and $B$. Making this prior specification, find the marginal posterior density $f(B|Y)$ and show that the posterior mode for $B$ (the value of $B$ that maximizes its marginal posterior density) is the Stein estimate: $\hat{B} = \min\{1, (k - 3)V/\sum((y_i - \bar{y})^2)\}$.

c) Find the conditional posterior density for $\mu | y, B$ and for $\theta | y, B$. Use these to find the modified posterior standard deviation estimates for the $\theta_i$'s described in the article Stein’s Paradox Revisited.

2. The following table gives the means and standard deviations of the points scored for each of the $k = 13$ WNBA teams in their first five games of 2004 ($y_i$) and the averages for the remaining $n = 29$ games (which we hope is close to $\theta_i$).

<table>
<thead>
<tr>
<th>Team</th>
<th>$n_i$</th>
<th>Average ($y_i$)</th>
<th>SD ($s_i$)</th>
<th>($\theta_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charlotte</td>
<td>5</td>
<td>55.4</td>
<td>5.5</td>
<td>61.4</td>
</tr>
<tr>
<td>Connecticut</td>
<td>5</td>
<td>68.2</td>
<td>10.6</td>
<td>67.5</td>
</tr>
<tr>
<td>Detroit</td>
<td>5</td>
<td>67.8</td>
<td>5.8</td>
<td>69.2</td>
</tr>
<tr>
<td>Houston</td>
<td>5</td>
<td>62.8</td>
<td>7.5</td>
<td>63.8</td>
</tr>
<tr>
<td>Indiana</td>
<td>5</td>
<td>69.6</td>
<td>5.9</td>
<td>63.2</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>5</td>
<td>69.8</td>
<td>8.2</td>
<td>71.8</td>
</tr>
<tr>
<td>Minnesota</td>
<td>5</td>
<td>63.6</td>
<td>11.2</td>
<td>63.0</td>
</tr>
<tr>
<td>New York</td>
<td>5</td>
<td>62.2</td>
<td>7.8</td>
<td>65.7</td>
</tr>
<tr>
<td>Phoenix</td>
<td>5</td>
<td>69.8</td>
<td>12.3</td>
<td>67.1</td>
</tr>
<tr>
<td>Sacramento</td>
<td>5</td>
<td>68.4</td>
<td>6.6</td>
<td>67.3</td>
</tr>
<tr>
<td>San Antonio</td>
<td>5</td>
<td>61.0</td>
<td>6.4</td>
<td>64.0</td>
</tr>
<tr>
<td>Seattle</td>
<td>5</td>
<td>75.6</td>
<td>14.8</td>
<td>70.7</td>
</tr>
<tr>
<td>Washington</td>
<td>5</td>
<td>66.4</td>
<td>7.6</td>
<td>68.1</td>
</tr>
<tr>
<td>Overall</td>
<td>65</td>
<td>66.2</td>
<td>8.9</td>
<td>66.4</td>
</tr>
</tbody>
</table>
a) Compute the Mean Square Model (MSM) and Mean Square Error (MSE) from these summaries and carry out an ANOVA F test to see if there is evidence, based on the first five games, that the \( \theta_i \)'s (underlying team means in 2004) differ. State the hypotheses for the F test in terms of the parameter \( B = V/(V + A) \).

b) Under the Null hypothesis, the F-ratio follows an F distribution. Describe how the F ratio relates to the Stein estimate. Show that

\[
B \left( \frac{\text{MSM}}{\text{MSE}} \right) \mid B \sim F_{(k-1,N-k)}, \quad N = nk.
\]

Use this fact to find a 95% confidence interval for \( B \) for the WNBA data. Check that the Stein estimate falls in the interval. Also find the implied interval for \( \sqrt{A} \), the standard deviation of the \( \theta_i \)'s.

c) Estimate \( V \) using the MSE and treat it as known. Simulate values from \( \theta_i \mid y \) by first drawing from \( f_{B\mid y}(B \mid y) \) and then from \( f_{\theta_i\mid y,B}(\theta_i \mid y,B), i = 1, \ldots, k \) (using the results from part c of presentation 1). Plot the means of the simulated \( \theta_i \)'s (Bayesian estimates of each \( \theta_i \)) against the observed \( y_i \)'s to see the anticipated regression to the mean. Add in points for the actual averages for each team’s remaining 29 games. Draw in posterior interval estimates for predicting the remaining averages.

3. **Unknown** \( V \). If we assume equal variances \( V = \sigma^2/n \) for each \( Y_i \mid \theta_i \) distribution, we can estimate \( V \) using the mean squared error: \( V = \text{MSE}/n \). This gives a 2-part level-1 model:

\[
Y_i \mid \theta_i, V \sim N(\theta_i, V), \quad i = 1, \ldots, k
\]

\[
\hat{V} \mid V \sim \text{Gamma} \left( \frac{N - k}{2}, \frac{N - k}{2V} \right).
\]

a) Assuming the level-2 distribution from problem 1 and prior density \( f(\mu, B, V) \propto (V^{-1})B^{-1} \), work out the marginal posterior density for \( B = V/(V + A) \) and point out its connection to the F density. Also identify the conditional posterior distribution for \( V \mid y, \hat{V}, B \).

b) Simulate values from \( f_{V,B}(V,B \mid y) \) and then from \( f(\theta_i \mid y,B,V), i = 1, \ldots, k \). Compare the results to simulations assuming \( V = \text{MSE}/n \) is known (from presentation 2).
Problem to turn in:
Consider the usual linear regression model assuming a flat (improper) prior density for the log of the residual variance $\sigma^2$ and the $q \times 1$ regression coefficient $\beta$:

$$Y | \beta, \sigma \sim N_n(X\beta, \sigma^2 I); \quad p(\beta, \sigma^2) \propto 1/\sigma^2.$$

a) Find the conditional posterior density for $\beta | y, \sigma$.

b) Find the marginal posterior density for $\sigma^2 | y$ (it’s inverse Gamma). Identify the posterior mean and posterior mode for $\sigma^2$.

c) For the height, shoe length and gender data, generate 1000 draws from the joint posterior density for $\beta, \sigma^2 | y$ by first drawing from $\sigma^2 | y$ and then from $\beta | y, \sigma^2$. Assuming the model with shoe length and gender (no interaction) is correct, find a 95% posterior interval estimate for the difference in the mean height of men and women with the same shoe length. Compare this to the usual confidence interval.

The following commands will produce a draw $\beta \sim N_q(\mu, Vbeta)$:

```r
rtVbeta = t(chol(Vbeta))
beta = mu + rtVbeta %*% rnorm(q)
```

The function chol returns the Choleski decomposition of a square matrix. It is an upper triangle square root matrix such that $t(chol(Vbeta))%*%chol(Vbeta) = Vbeta$. Note that $chol(\sigma^2 (X'X)^{-1}) = \sigma chol((X'X)^{-1})$, so you only need to evaluate this once.