Stat 111 Spring 2011 Week 13 - More Bayesian Inference

0. **Warm-up problem** (for everyone): The puzzler on *Car Talk* a few weeks ago was a simple version of an old problem often referred to as "The secretary problem" (check Wikipedia for a nice description).

   `http://www.cartalk.com/content/puzzler/transcripts/201107/index.html`

```
RAY: This is from Norm Leyden from Franktown, Colorado. The date on it is 1974--I'm
a little behind.

Three different numbers are chosen at random, and one is written on each of three
slips of paper. The slips are then placed face down on the table. The objective is
to choose the slip upon which is written the largest number.

Here are the rules: You can turn over any slip of paper and look at the amount
written on it. If for any reason you think this is the largest, you're done; you
keep it. Otherwise you discard it and turn over a second slip.

Again, if you think this is the one with the biggest number, you keep that one and
the game is over. If you don't, you discard that one too.

TOM: And you're stuck with the third. I get it.

RAY: The chance of getting the highest number is one in three. Or is it? Is there a
strategy by which you can improve the odds?
```

   Think about how this problem relates to Bayesian hierarchical models, and "shrinkage" estimates like the Stein estimate.

1. **Bayesian vs. Frequentist thinking: The Two Envelopes Problem**
   Suppose you and a friend are recruited by a wealthy-looking visitor to give a campus tour. Afterwards the visitor hands you two envelopes and tells you that one contains twice as much money as the other. You choose an envelope at random and see that it contains \$100. Your friend's envelope either contains \$50 or \$200. Having chosen one at random, you figure the expected amount in the other envelope is $(.5)(50) + (.5)(200) = \$125 > \$100$. Expecting to gain money, you offer to switch envelopes with your friend. Meanwhile, your friend figures that if her envelope contains x dollars, the expected amount in your envelope is $(.5)(x/2) + (.5)(2x) = 1.25x > x$. So she wants to switch too. Notice that your friend's argument works even if both envelopes were left unopened. You and your friend might decide to switch envelopes, then want to switch back, and so on. There couldn't possibly be any expected gain in doing that, so something about the logic must be faulty.

   a) Explain the flaw in the above reasoning. Define θ to be the larger amount, and X to be the amount in your envelope. So $P(X = \theta \,|\, \theta) = P(X = \theta/2 \,|\, \theta) = 0.5$. If you knew θ you would know whether or not you had the larger amount.

   b) Outline a Bayesian procedure for making inference about θ. What prior density for θ would lead to the paradox described?

2. The following table gives the means and standard deviations (y_i and s_i) of the points scored for each of the $k = 13$ WNBA teams in their first five games of 2004, and the averages for the remaining $n = 29$ games (which we hope are close to the true means θ_i, $i = 1, \ldots, k$). The last column gives a covariate x_i, which is team i's average points per game in the 2003 season.

Team	n_i	y_i	s_i	(θ_i)	x_i
Charlotte	5	55.4	5.5	61.4	65.2
Connecticut	5	68.2	10.6	67.5	70.1
Detroit	5	67.8	5.8	69.2	75.1
Houston	5	62.8	7.5	63.8	66.0
Indiana	5	69.6	5.9	63.2	68.7
Los Angeles	5	69.8	8.2	71.8	73.5
Minnesota	5	63.6	11.2	63.0	70.0
New York	5	62.2	7.8	65.7	66.0
Phoenix	5	69.8	12.3	67.1	61.7
Sacramento	5	68.4	6.6	67.3	67.6
San Antonio	5	61.0	6.4	64.0	65.1
Seattle	5	75.6	14.8	70.7	70.2
Washington	5	66.4	7.6	68.1	68.5
Overall	65	66.2	8.9	66.4	68.3

a) Consider only the first three columns. Compute the Mean Square Model (MSM) and Mean Square Error (MSE) from these summaries of the first $n = 5$ games. Carry out an ANOVA F test to see if there is evidence that the θ_i's (underlying team means in 2004) differ. State the hypotheses for the F test in terms of the parameter $B = V/(V + A)$, where $V = \sigma^2/n$ is the variance of the y_i's, and A is the variance of the θ_i's.

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

$$B\left(\frac{\text{MSM}}{\text{MSE}}\right) \mid B \;\sim\; F_{(k-1,N-k)}, \qquad N \;=\; nk.$$

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 θ_i's.

c) Estimate V using the MSE and treat it as known. Simulate values from $\theta_i \mid y$ by first drawing from $f_{B|y}(B \mid y)$ and then from $f_{\theta_i|y,B}(\theta_i \mid y, B)$, $i = 1, \ldots, k$. Plot the means of the simulated θ_i's (Bayesian estimates of each θ_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: $\hat{V} = \text{MSE}/n$. 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 presentation 2 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).

c) Generalize the model from presentation 2 to allow the level-2 mean to be $\mu_i = X_i'\beta$.

Problems to turn in:

1. Consider data values $X_1, \ldots, X_n \sim$ Binomial(1, θ). Show that Jeffreys' prior for the success probability θ is a proper Beta(1/2, 1/2) density. Work out the implied posterior density and posterior mean value. Notice that prior specification is like assuming an additional trial that was half success and half failure.

2. Consider the usual linear regression model assuming a flat (improper) prior density for the log of the residual variance σ^2 and the $q \times 1$ regression coefficient β:

$$Y \mid \beta, \sigma \sim N_n(\mathbf{X}\beta, \sigma^2\mathbf{I}); \qquad p(\beta, \sigma^2) \propto 1/\sigma^2.$$

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

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

c) (optional) For the height, shoe length and gender data, generate 1000 draws from the joint posterior density for $\beta, \sigma^2 \mid y$ 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 you found in week 8.

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

```
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 $\text{chol}(\sigma^2(\mathbf{X}'\mathbf{X}^{-1}) = \sigma\, \text{chol}((\mathbf{X}'\mathbf{X})^{-1})$, so you only need to evaluate this once.