**PROBLEM 2: Exact Sequential Posterior Simulation**

In some very simple situations, your posterior distribution has a CDF and inverse CDF you can compute, in which case you can find any posterior probabilities explicitly. More often, Bayesian inference relies on simulation. Consider the Normal hierarchical model with equal and known level-1 variances $V$.

$$\text{level-1:} \quad Y_i \mid \theta_i, \sigma^2 \overset{\text{indep}}{\sim} N(\theta_i, V), \quad i = 1, \dots, k$$
$$\text{level-2:} \quad \theta_i \mid \mu, A \overset{\text{i.i.d.}}{\sim} N(\mu, A)$$

a) The marginal distribution of the $Y_i$'s is $N(\mu, V + A)$. Write out the log-likelihood function for $\mu$ and the shrinkage factor $B = \dfrac{V}{V+A}$. Show that Jeffrey's prior for $\mu$ is constant, and for $B$ it is $1/B$.

**Solution:**

We have derived earlier that the marginal distribution of the $Y_i$'s given $\mu, A$, and $V$ is:

$$Y_i \mid \mu, A, V \sim N(\mu, V + A)$$

The shrinkage factor is $B = \dfrac{V}{V+A} \implies V + A = \dfrac{V}{B}$. So, given $\mu, B$, and $V$,

$$Y_i \mid \mu, B, V \sim N(\mu, V/B)$$

The likelihood function for $\mu$ and $B$ is,

$$L(\mu, B) \propto B^{k/2} e^{-\frac{B}{2V} \sum (Y_i - \mu)^2} \mathbb{I}(0 < B \leq 1)$$

, as $V$ is known and given.

Thus, the log-likelihood function for $\mu$ and $B$ is given by,

$$\boxed{l(\mu, B) = \frac{k}{2} \ln(B) - \frac{B}{2V} \sum (Y_i - \mu)^2 + (\text{const. in } \mu, B),\ 0 < B \leq 1}$$

Then,

$$\frac{\partial}{\partial \mu} l(\mu, B) = \frac{B}{V} \sum (Y_i - \mu) \implies \frac{\partial^2}{\partial \mu^2} l(\mu, B) = -\frac{kB}{V}$$

So,

$$p_\mu(\mu) \propto \left( -\mathbb{E}\left[ \frac{\partial^2}{\partial \mu^2} l(\mu, B) \right] \right)^{1/2} \propto \sqrt{\frac{kB}{V}} \implies \boxed{p_\mu(\mu) \propto 1}$$

Similarly,

$$\frac{\partial}{\partial B} l(\mu, B) = \frac{k}{2B} - \frac{\sum (Y_i - \mu)^2}{2V} \implies \frac{\partial^2}{\partial B^2} l(\mu, B) = -\frac{k}{2B^2}$$

So,

$$p_B(B) \propto \left( -\mathbb{E}\left[ \frac{\partial^2}{\partial B^2} l(\mu, B) \right] \right)^{1/2} \propto \sqrt{\frac{k}{2B^2}} \implies \boxed{p_B(B) \propto \frac{1}{B},\ 0 < B \leq 1}$$

Joint-Jeffreys'prior: $p_{\mu,B}(\mu, B) \propto \dfrac{1}{B},\ 0 < B \leq 1$

b) Assuming $p(\mu, B) \propto 1/B, 0 < B < 1$, give the marginal posterior distribution of $B \mid y$, the conditional posterior distribution of $\mu \mid y, B$, and of $\theta_i$, for $i = 1, \dots, k$. Show what would be involved in working out the marginal posterior densities for $\mu \mid y$ or $\theta_1, \dots, \theta_k \mid y$. The alternative is to simulate a large sample from the joint posterior density. Demonstrate for the WNBA or wordle data. Along with estimating posterior means, try commenting on the rankings. For example, what is the probability the Minnesota Lynx have the highest $\theta_i$, or that CRANE has the lowest $\theta_i$?

**Solution:**

The joint-posterior density of $\mu$ and $B$ is,

$$
\begin{aligned}
f_{\mu,B|y}(\mu, B \mid y, V) &\propto L(\mu, B) \cdot p(\mu, B) \\
&\propto V^{-k/2} B^{k/2-1} \exp\left[-\frac{B}{2V} \sum (Y_i - \mu)^2\right], \quad 0 < B \le 1 \\
&\propto V^{-k/2} B^{k/2-1} \exp\left[-\frac{B}{2V} \sum (Y_i - \mu + \bar{Y} - \bar{Y})^2\right] \\
&\propto \left(V^{-(k-1)/2} B^{(k-1)/2-1} \exp\left[-\frac{B}{2V} \sum (Y_i - \bar{Y})^2\right]\right)\left((V/B)^{-1/2} \exp\left[-\frac{kB}{2V}(\mu - \bar{Y})^2\right]\right)
\end{aligned}
$$

So, the marginal posterior distribution of $B \mid y$ is constrained Gamma$\left(\frac{k-1}{2}, \frac{\sum (Y_i - \bar{Y})^2}{2V}; 1\right)$ as we only allow $B \in (0, 1]$ and force the density to be zero for all $B > 1$. This gives,

$$
\boxed{f_{B|Y}(B \mid Y, V) = \frac{B^{(k-1)/2-1} e^{-BS/(2V)}}{\int_0^1 t^{(k-1)/2-1} e^{-tS/(2V)}\, dt}, \quad 0 < B \le 1, \ S = \sum_{i=1}^{k} (Y_i - \bar{Y})^2}
$$

And, the conditional posterior distribution of $\mu \mid y, B$ is,

$$
\boxed{\mu \mid y, B \sim N\left(\bar{Y}, \frac{V}{kB}\right)}
$$

To calculate the conditional posterior distribution of $\theta_i$'s, using Bayes' Rule,

$$
\begin{aligned}
f_{\theta_i|Y_i}(\theta_i \mid y_i, \mu, A) &\propto f_{Y_i|\theta_i}(y_i \mid \theta_i, \mu, A) \times f_{\theta_i|\mu,A}(\theta_i \mid \mu, A) \\
&\propto \exp\left[-\frac{1}{2V}(y_i - \theta_i)^2 - \frac{1}{2A}(\theta_i - \mu)^2\right] \\
&\propto \exp\left(-\frac{1}{2}\left[\left(\frac{1}{V} + \frac{1}{A}\right)\theta_i^2 - 2\theta_i\left(\frac{y_i}{V} + \frac{\mu}{A}\right) + \left(\frac{y_i^2}{V} + \frac{\mu^2}{A}\right)\right]\right)
\end{aligned}
$$

Comparing the exponent with standard Normal,

$$
\frac{1}{\text{var}(\theta_i \mid y_i, \mu, A)} = \frac{1}{V} + \frac{1}{A} \implies \text{var}(\theta_i \mid y_i, \mu, A) = \frac{VA}{V + A}
$$

And,

$$
\frac{\mathbb{E}[\theta_i \mid y_i, \mu, A]}{\text{var}(\theta_i \mid y_i, \mu, A)} = \frac{y_i}{V} + \frac{\mu}{A} \implies \mathbb{E}[\theta_i \mid y_i, \mu, A] = \frac{Ay_i + V\mu}{A + V}
$$

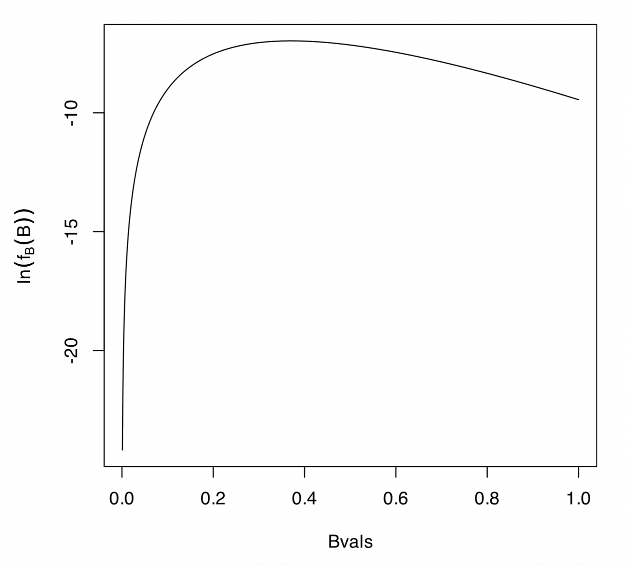
Hence,

$$
\boxed{\theta_i | y_i, \mu, A \sim N\left(\frac{Ay_i + V\mu}{A + V}, \frac{VA}{V + A}\right)}
$$

Similarly, rewriting mean and variance in terms of $B$,

$$\boxed{\theta_i | y_i, \mu, B, V \sim N\left(B\mu + (1-B)y_i, (1-B)V\right), \quad i = 1, \ldots, k}$$

Now, plotting the log-posterior density of $B \mid y$,




caption*Figure: Before Scaling

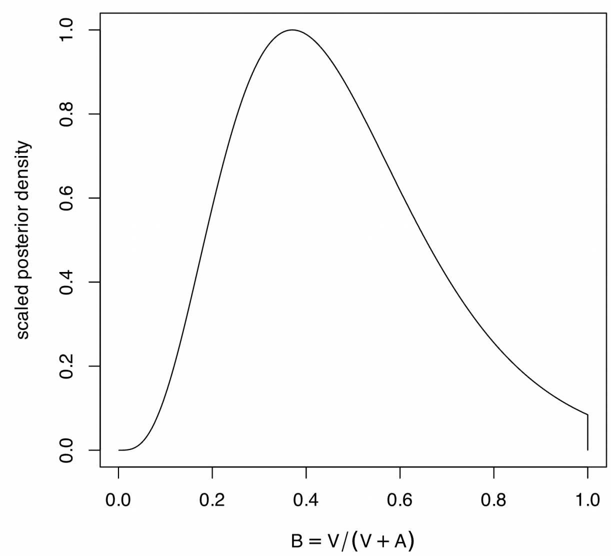
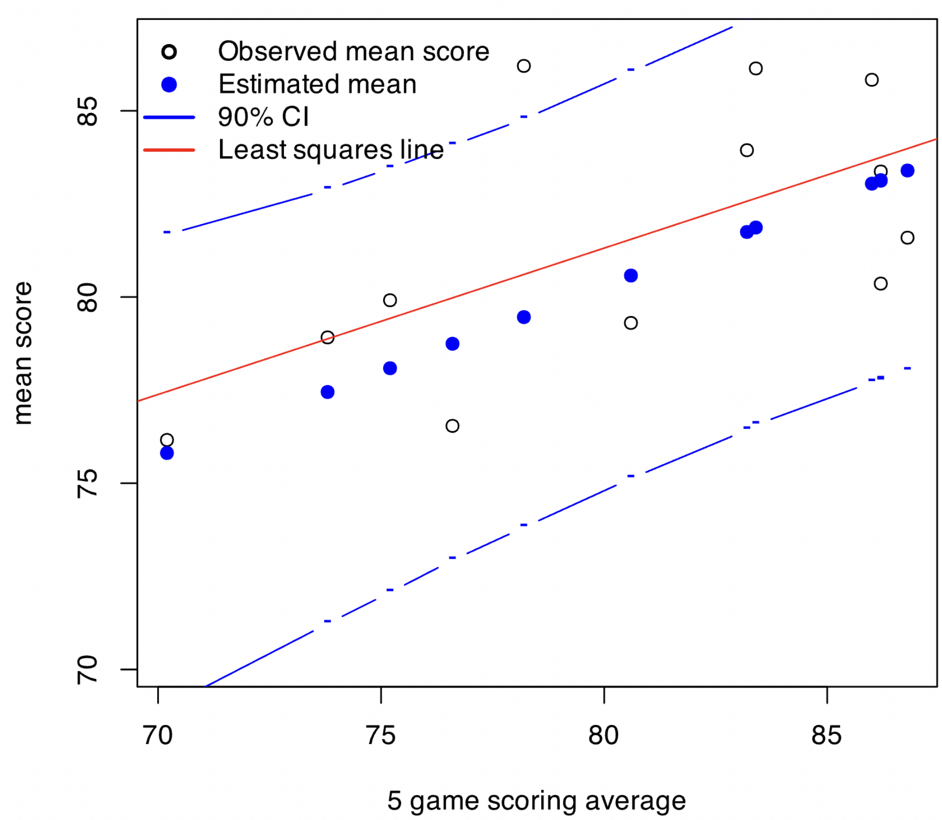



Figure: After Scaling

From the simulation, the confidence interval plot for WNBA data is,



Finally, the probabilities that a team has highest $\theta_i$ is,

| Team | $\mathbb{P}(\theta = \max(\theta))$ |
|---|---|
| Aces | 0.170 |
| Dream | 0.003 |
| Fever | 0.022 |
| Liberty | 0.089 |

| Team | $\mathbb{P}(\theta = \max(\theta))$ |
|---|---|
| Lynx | 0.206 |
| Mercury | 0.178 |
| Mystics | 0.010 |
| Sky | 0.013 |

| Team | $\mathbb{P}(\theta = \max(\theta))$ |
|---|---|
| Sparks | 0.007 |
| Storm | 0.176 |
| Sun | 0.042 |
| Wings | 0.085 |

c) Often our $Y_i$'s are group averages, and we have $V = \dfrac{\sigma^2}{n}$. Suppose we have a within-groups variance estimate $s^2 \mid \sigma^2 \sim \text{Gamma}(m, m/\sigma^2)$ that is independent of the $Y_i$'s. Assume the prior distribution $p(\sigma^2) \sim 1/\sigma^2$ and show how the simulation algorithm can be expanded to allow for an unknown $\sigma^2$. Demonstrate and show how the posterior intervals change.

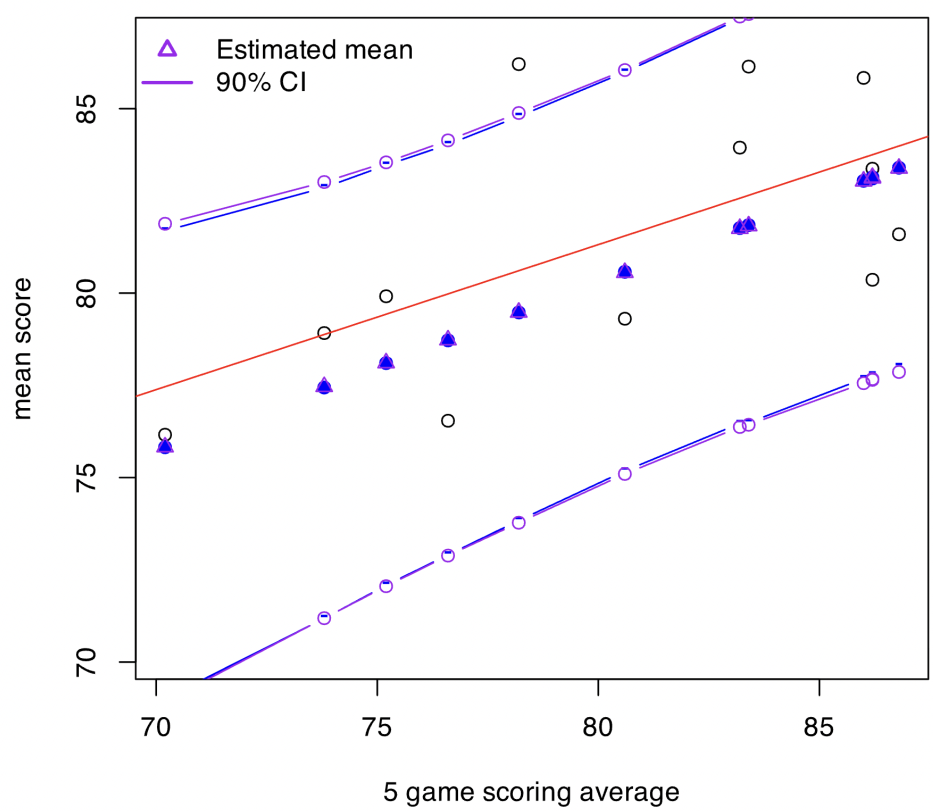
**Solution:**

Using Jeffreys's prior, the pivot,

$$W = \frac{(N-k)\text{MSE}}{\sigma^2} \sim \chi^2_{n-k}$$

whether we treat $\sigma^2$ as known and $s^2$ as random, or if we condition on $s^2$ and treat $\sigma^2$ as random. Then, we generate new $\sigma^2$ using: W=rchisq(nsim, N-k) and sigsq = (N-k)*mse/W

And, set our new value of $V$ to be new$V = \dfrac{\text{sigsq}}{n}$ and use the same algorithm as in part (b) to calculate new values for the confidence intervals and estimated means. See plot below.



d) Explain why things become much more complicated if the $V$'s are unequal, and how equal $V$'s requires equal sample sizes as well as equal $\sigma$'s.

**Solution:**

If each $Y_i$ has it's own sampling variance $V_i$, there would be no single shrinkage factor and each group will have a different ratio,

$$B_i = \frac{V_i}{V_i + A}$$

The partial pooling of $\theta_i$ toward $\mu$ will no longer be uniform across the groups $i$, so groups with smaller $V_i$(i.e. more precise data) would shrink less, while groups with larger $V_i$ (i.e. less precise data) would shrink more. This results in heavier computation, as we must keep track of all the $V_i$, and the posterior for $\mu, A$ can still be sampled but will not have a clean closed-form formula like the one in the case with equal $V$'s.

Now,

For each individual group, $\text{var}(Y_i \mid \theta_i, \sigma^2) = \dfrac{\sigma^2}{n_i}$, so if we want,

$$\text{var}(Y_i \mid \theta_i) = \frac{\sigma^2}{n_i} = V, \quad \text{for all } i$$

We will need equal sample sizes as well as equal $\sigma$'s