1. **Gibbs sampling for the Normal Hierarchical Model**

Consider the following hierarchical model. For now, treat the level-1 variances \( V_i \) as known (but unequal). For example, we might have \( V_i = \sigma^2/n_i \), where \( \sigma^2 \) is estimated with many degrees of freedom.

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

\[
\theta_i \mid \mu, A \stackrel{i.i.d.}{\sim} N(\mu, A), \quad p_{\mu,A}(\mu, A) \propto 1.
\]

a) Derive the marginal posterior density for \( A \), and note that it is positive when \( A = 0 \) (meaning we don’t want a prior density for \( A \) that goes to \( \infty \) as \( A \to 0 \)). Also find the conditional posterior \( f_{\theta_i \mid y,B_i}(\theta_i \mid y, B_i) \), where \( B_i = V_i/(V_i + A) \).

b) Write out the conditional densities needed for Gibbs sampling: \( f_A \mid y, \theta, \mu \), \( f_\mu \mid y, \theta, A \) and \( f_{\theta_i} \mid y, \mu, A \). Implement a Gibbs sampling algorithm to estimate the parameters for the 2007 NBA data, predicting the season average points scored based on the first 7 to 12 games (the numbers played before a particular date that season). Compare the estimated posterior means for each team’s \( \theta_i \) to the sample estimates \( y_i \).

c) Show how to modify the algorithm to allow for an unknown \( \sigma^2 \) (where \( V_i = \sigma^2/n_i \)). Assume \( p(\sigma^2) \propto 1/\sigma^2 \), and that you have an estimate \( \hat{\sigma}^2 \) (e.g., the MSE from individual game scores).

d) You can allow for team covariates by changing the level-2 model to be \( \theta_i \mid \beta, A \sim N(X_i'\beta, A) \). For example, we have each team’s scoring average from the previous season, and perhaps that information could be useful for predicting the 2007 averages.

2. **Direct Simulation**

Assuming known (but not necessarily equal) \( V_i \)'s, it is possible to generate independent posterior draws using rejection sampling. The advantage of this over Gibbs sampling is that there is no worry about convergence, and the resulting draws are independent. With equal \( V_i \)'s we found it useful to work with the transformation \( B = V/(V + A) \). With unequal \( V_i \)'s we can define \( B_o = V_o/(V_o + A) \), where \( V_o \) is chosen as a “typical” \( V_i \). With \( V_i = \sigma^2/n_i \), this is equivalent to choosing a typical sample size \( n_o \).

a) Write out the posterior density for \( B_o \mid y \) and show that it is approximately a constrained Gamma distribution (we showed in week 12 that this is exact if the \( V_i \)'s are all equal).

b) Choose a \( V_o \) for the NBA data and implement an EM algorithm to find the posterior mode of \( B_o \).

c) Define a rejection sampling algorithm by fitting a constrained Gamma density \( f_o(B_o) \) with the same mode as the true posterior density \( f_{B_o}(B_o \mid y) \). Try it out with the NBA data and compare the estimated posterior means for each team’s \( \theta_i \) to the sample estimates \( y_i \).

d) Now adjust the algorithm to allow for level-2 covariates (see 1d).
3. Unequal and Unknown Variances

We see from presentation 2 that we can make exact inference in the Normal hierarchical model (NHM) when the level-1 variances $V_i$ are known. To handle the unknown $V_i$ problem, it is sufficient to find a method for simulating $V_1, \ldots, V_k$ from their posterior distribution, given $k$ sample variances. These $V_i$’s may then be passed to the $V$-known algorithm for simulating posterior $\theta_i$’s, with a new set of $V_i$’s taken as known for each simulation.

If the $Y_i$’s are averages for each of $k$ groups, then $V_i = \sigma_i^2 / n_i$. To simplify notation, let $X_i = (n_i - 1)s_i^2$, $\eta_i = 1/(2\sigma_i^2)$ and $m_i = (n_i - 1)/2$.

\[
\begin{align*}
X_i \mid \eta_i & \sim \text{Gamma}(m_i, \eta_i), \quad i = 1, \ldots, k \\
\eta_i \mid \alpha, \lambda & \sim \text{Gamma}(\alpha, \lambda), \quad p(\alpha, \lambda) \propto \lambda^{-1}, \quad \lambda > 0, \quad \alpha > 0.
\end{align*}
\]

\[
\eta_i \mid x_i, \alpha, \lambda \sim \text{Gamma}(m_i + \alpha, x_i + \lambda), \quad i = 1, \ldots, k.
\]

a) Work out the joint posterior density for $\alpha$ and $\lambda$. Show that this is equivalent to the posterior density for the Poisson-Gamma hierarchical model, where at level-1 we have $X_i \mid \eta_i \sim \text{Poisson}(m_i \eta_i)$, with the same level-2 model.

b) Show that, if we could simulate draws from $f_{\alpha,\lambda|x}$, it would be easy to generate draws from $V_i \mid x, \alpha, \lambda$, $i = 1, \ldots, k$.

c) Derive a Newton Raphson algorithm to find the posterior mode of $\alpha$ and $\lambda$. Apply the algorithm to the $k = 30$ sample variances for the NBA data. Does a bivariate Normal approximation seem appropriate?

d) Assuming the $m_i$’s are all integers, we can express the ratio of Gamma functions in the posterior distribution as a polynomial in $\alpha$ (although finding the coefficients takes some effort). Show how this allows us to find a marginal posterior density for $\lambda$. How would you simulate from $f_{\alpha|x,\lambda}$?

e) The $n_i$’s must all be integers, but $n_i$’s that are even will yield non-integer $m_i$’s. To handle this complication, use the approximation $\Gamma(\alpha + 1/2)/\Gamma(\alpha) \approx \sqrt{\alpha}$ to get an enveloping candidate density $f_{\alpha}$.}

There are no turn-in problems this week. The take-home final will be available the Monday after classes end (possibly earlier). It will cover material from the entire semester and may require you to make graphs and get numerical summaries for some real data. You may have the test for three days (e.g., pick up some time Friday and return it some time on Monday). I don’t expect you to work on it non-stop during that time, but I want you to have plenty of time to work the problems (and to take breaks from working on them). You should not talk to anyone other than me about the test until after the finals period (or whenever everyone has finished taking it). You may use your text and notes and any other non-living resources.