STAT 532 Assignment 9 -2015
(Due Wednesday, November 4 by 1:00 p.m. if want it back by end of Friday)
(Due Friday, November 6 by 1:00 p.m. if want it back by end of Tuesday)

Show all work **neatly** and **in order** for full credit. Same instructions as previous homeworks...ask if you need clarification.

1. Let's apply the convergence criteria to some artificially constructed sets of chains with varying properties.

   (a) (5 pts) We will simulate 3 artificial chains under 5 different Scenarios to play around with assessing convergence. Simulate 1000 iterations per chain for each scenario (you do not have to worry about burn-in (or warm-up) for this one) and plot them all (plot the chains from the same Scenario on the same plot).

      i. Scenario 1: three from $N(0, 1)$ white noise

      ii. Scenario 2: one chain from $N(-1, 1)$, one chain from $N(0, 1)$, and one chain from $N(1, 1)$

      iii. Scenario 3: three from $MVN(0, \Sigma)$ with common $\sigma^2 = 1$ and $\rho = 0.8$ using `rmvnorm()` (Suppose this is three chains from dispersed starting values that have not yet converged)

      iv. Scenario 4: three from non-stationary correlated chains using `diffinv(rnorm(999))`.

      v. Scenario 5: three stationary chains with correlation using `filter(rnorm(1000), filter=rep(1,10), circular=TRUE)`

   (b) (3 pts) Are $\widehat{R}$ and *n.eff* are calculated in the newest version of the `coda` package in the same way they are described in the text book?

   (c) (3 pts) Calculate $\widehat{R}$ and *n.eff* for all Scenarios using the `coda()` package. Display the results using a graph or table.

   (d) (4 pts) Include results from two other convergence diagnostics as well (you choose). If they do not rely on multiple chains, combine the three chains we simulated above into a single chain.

   (e) (8 pts) Compare and discuss your results for the five scenarios and different convergence diagnostics. For these scenarios, do they seem to do what we want them to?

2. We want to use the following two-level hierarchical Normal-Normal model:

$$\begin{aligned} y_{ij} &\sim N(\theta_j, \sigma_y^2) \qquad i = 1, \ldots, n_j, \; j = 1, \ldots J \\ \theta_j &\sim N(\mu, \sigma_\theta^2) \end{aligned}$$

Typically enough information is available to estimate μ and σ_y^2 that one can use any reasonable priors for them (such as $p(\mu, \sigma_y) \propto 1$ or $p(\mu, log(\sigma_y)) \propto 1$ as we have seen). However, there has been much debate about what prior distributions are appropriate for the σ_θ^2 if the goal is to include little (if any) prior information. It turns out that choice of the prior distribution at this upper level matters (in terms of sensitivity of inference and getting proper posterior distributions), particularly when there is very small variability among the θ_j's (i.e. σ_θ^2 is close to zero) or the number of groups J is small. Here are some priors that have been used:

- Prior A: $\sigma_\theta^2 \sim Inv - Gamma(0.001, 0.001)$
- Prior B: $\log(\sigma_\theta) \sim Unif(-100, 100)$
- Prior C: $\sigma_\theta \sim Unif(0, 100)$
- Prior D: $\sigma_\theta \sim$ folded-t or half-t (absolute value of Student's t-distribution centered at 0)
- Prior E: $p(\sigma_\theta^2) \propto \frac{1}{\sigma^2}$
- Prior F: $p(\sigma_\theta) \propto 1I_{(0,-\infty)}$
- Prior G: $p(\sigma_\theta^2|\sigma_y^2) = \frac{a}{\sigma_y^2}(1+\frac{\sigma_\theta^2}{\sigma_y^2})^{(a+1)}$ where $a = 7$ (Prior from Gustafson, Hossain, MacNab (2006))

(a) Re-write Prior A as a scaled-Inverse-χ^2 distribution (*i.e.* find ν_0 and ^{2} as parameterized in our text)

(b) Plot the prior distributions on their original scales. Make sure to specify the df for the half-t. (Try to display them all on one page)

(c) Transform Prior B so that you can directly compare it graphically to Prior C. (You may do this computationally)

(d) Transform prior C so that you can directly compare it graphically to prior A. (You may do this computationally)

(e) How is Prior D related to a half-Cauchy distribution?

3. Read Gelman's (2006) paper *Prior distributions for variance parameters in hierarchical models.* Answer the following questions related to it. It may also be helpful to read section 11.7 in the text book, as well as Chapter 5.

   (a) (4 pts) Folded-t distribution: (You should do this BEFORE looking at the functions I wrote.)

      i. Write an R function to calculate the density function of the folded non-central t-distribution. Include your R-function (from a script, and not the console window) and a plot showing the density. You may check your work with my functions,

      ii. Write an R function to obtain random draws from the folded non-central t-distribution. Include your R-function and a plot showing a histogram of the draws with the density from (2) overlaid on top of it to convince yourself both functions are working. (See Page 582 in BDA3)

   (b) Write out the basic hierarchical model described in the paper. Assume you do *not* know σ_y^2.

   (c) Simulate artificial data from the basic hierarchical model described in the paper. Make sure to set the random seed using `set.seed()` so that you can repeat any analysis using the same data. Use the following values:

      - $J = 8$
      - $\mathbf{n} = (n_1, n_2, \ldots, n_J) = (5, 10, 30, 30, 20, 25, 50, 10)$
      - $\sigma_y^2 = 4$
      - $\mu = 20$
      - $\sigma_\alpha^2 = 2$

   (d) Write out the prior distributions you could use to fit the hierarchical model to your artificial data (assume μ, σ_y^2, and σ_α^2 are unknown).

   (e) Fit the model to your artificial data (you can program it yourself or use JAGS, BUGS, or Stan) under the following three priors for σ_α.

      - Improper uniform prior on σ_α
      - Inv-Gamma(ϵ, ϵ) prior on σ_α^2
      - Uniform(0,A) prior on σ_α

      i. Choose reasonable values for ϵ and A chosen with goal of choosing values such that the prior has negligible influence on the posterior. Provide them here with a very brief justification.

      ii. Compare the approximate posterior distributions of σ_α or σ_α^2 obtained from the three priors. Compare and discuss the results.

      iii. Also compare inferences for μ under the three different priors for the variance.

(f) Now, rewrite the model with the redundant multiplicative reparameterization. Use the following priors and discuss the resulting posterior inference for σ_α and μ (compare for these two and to those obtained in part (e))

- the folded normal distribution (half-normal centered at zero) for σ_α.
- the half-Cauchy for σ_α.

4. Repeat parts (c), (e), and (f) from Problem 3 using a new artificial data set constructed with $\sigma_\alpha^2 = 0.01$ (use the same random seed as in Problem 3).

   (a) Discuss any differences or general observations for the comparison of the two magnitudes of variance.

   (b) Using two of the prior choices, provide a caterpillar plot displaying posterior intervals for each of the group means. Show the overall mean and sample averages on the plot to examine the degree of shrinkage for the different groups. Discuss the results. (I provided some code in the *MHwithinGibbsExampleLogitNormal.R*

   (c) Now, make the corresponding plots for the Problem 3 artificial data. Discuss differences after comparing to those obtained in the previous plot. There should be 4 plots that you are comparing.

5. (8 pts) Work through the 8 Schools example in Chapter 5 and using the code I provided. We will be going through parts of it in class as well. Include a few comments and/or questions here.