HW3 STT 465
(MSU, Fall, 2015)

Due: Wednesday, Oct 14th, 2015 in class (hard copy) or by e-mail (pdf).

Note: Please provide a report as neat as you can, with clear answers. Include the derivation as part of your answers and the code at the end of each answer.

Data. For this HW we will use the data set `LDL.txt` (the data set is in the GitHub repository). This data set contains measurements of Low-density lipoprotein (LDL cholesterol) in female and male patients. The overall goal of the analysis is to determine whether there are systematic differences in the mean and variance of LDL cholesterol between male and female patients.

1. Descriptive statistics

1.1. Provide descriptive statistics by group and a boxplot of LDL cholesterol versus sex.
1.2. Summarize your findings.

2. Likelihood analyses

Assume that LDL cholesterol follows a normal distribution with mean and variances that are sex-specific, that is,

$$[y_i|Gender_i = F] \sim N(\mu_F, \sigma_F^2) \quad \text{and} \quad [y_i|Gender_i = M] \sim N(\mu_M, \sigma_M^2)$$

2.1. Write down the likelihood function.
2.2. Derive the Maximum Likelihood estimators of the means and variances.
2.3. Using the data available in GitHub compute and report the MLE estimate of the mean and variances of each of the groups, and provide an approximate 95% CI for the <u>means</u> of the groups.

3. Bayesian Analyses

Complete the Bayesian model by assuming

$$p(\mu_F, \mu_M, \sigma_F^2, \sigma_M^2) = p(\mu_F)p(\mu_M)p(\sigma_F^2)p(\sigma_M^2)$$

where:

$$p(\mu_F) = \frac{Exp\left(-\frac{(\mu_F-\mu_0)^2}{2\sigma_0^2}\right)}{\sqrt{2\pi\sigma_0^2}} \ , \ \ p(\mu_M) = \frac{Exp\left(-\frac{(\mu_M-\mu_0)^2}{2\sigma_0^2}\right)}{\sqrt{2\pi\sigma_0^2}},$$

$$p(\sigma_F^2) = \frac{\left(\frac{S_0}{2}\right)^{\frac{df_0}{2}} Exp\left(-\frac{S_0}{2\sigma_F^2}\right)}{\Gamma\left(\frac{df_0}{2}\right) [\sigma_F^2]^{\left[1+\frac{df_0}{2}\right]}}, \ \ p(\sigma_M^2) = \frac{\left(\frac{S_0}{2}\right)^{\frac{df_0}{2}} Exp\left(-\frac{S_0}{2\sigma_M^2}\right)}{\Gamma\left(\frac{df_0}{2}\right) [\sigma_M^2]^{\left[1+\frac{df_0}{2}\right]}}$$

3.1. Derive the fully conditional distributions for the mean and variances.
3.2. Describe, conceptually, a Gibbs sampler for drawing samples from the posterior distribution of the mean and variances.

Run the Gibbs sampler with 50,000 iterations collected after discarding 1,000 samples and burn in. From the Gibbs ampler report:

  3.3. Estimates of the posterior means, posterior standard deviations and 95% (Highest-Posterior Desnity) Bayesian Credibility Regions.

  3.4. Convergence diagnostics for each of the unknowns, including: trace plot, estimated number of effective samples, auto-correlations, and MC standard errors.

  3.5. Summarize your findings in light of the objective above described and results obtained from your analyses.