STK 4021 - Applied Bayesian Analysis and Numerical Methods
Thordis L. Thorarinsdottir
Fall 2014

# Problem Set 7

**Problem 1 (Jeffreys' prior for the multivariate normal model).** For the multivariate normal model, Jeffreys' rule for generating a prior distribution on $(\boldsymbol{\theta}, \Sigma)$ gives $p_J(\boldsymbol{\theta}, \Sigma) \propto |\Sigma|^{-(p+2)/2}$.

(a) Explain why the function $p_J$ cannot actually be a probability density for $(\boldsymbol{\theta}, \Sigma)$.

(b) Let $p_J(\boldsymbol{\theta}, \Sigma|\mathcal{D})$ be the probability density that is proportional to $p_J(\boldsymbol{\theta}, \Sigma) \times p(\mathcal{D}|\boldsymbol{\theta}, \Sigma)$. Obtain the form of $p_J(\boldsymbol{\theta}, \Sigma|\mathcal{D})$, $p_J(\boldsymbol{\theta}|\Sigma, \mathcal{D})$ and $p_J(\Sigma|\boldsymbol{\theta}, \mathcal{D})$.

**Problem 2 (Australian crab data).** The files `bluecrab.dat` and `orangecrab.dat` contain measurements of body depth ($Y_1$) and rear width ($Y_2$), in millimeters, made on 50 male crabs from each of two species, blue and orange. We will model these data using a bivariate normal distribution.

(a) For each of the two species, obtain posterior distributions of the population mean $\boldsymbol{\theta}$ and covariance matrix $\Sigma$ as follows: Using the semiconjugate prior distributions for $\boldsymbol{\theta}$ and $\Sigma$,

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Lambda_0), \quad \Sigma^{-1} \sim \text{Wishart}(\nu_0, \mathbf{S_o^{-1}}),$$

set $\boldsymbol{\mu}_0$ equal to the sample mean of the data, $\Lambda_0$ and $\mathbf{S_0}$ equal to the sample covariance matrix and $\nu_0 = 4$. Obtain 10000 posterior samples of $\boldsymbol{\theta}$ and $\Sigma$.

(b) Plot values of $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ for each group and compare. Describe any size differences between the two groups.

(c) From each covariance matrix obtained from the Gibbs sampler, obtain the corresponding correlation coefficient. From these values, plot posterior densities of the correlations $\rho_{blue}$ and $\rho_{orange}$ for the two groups. Evaluate differences between the two species by comparing these posterior distributions. In particular, obtain approximation to $\mathbb{P}(\rho_{blue} < \rho_{orange}|\mathcal{D})$. What do the results suggest about differences between the two populations?

Solutions will be discussed in class on October 17.