# Homework 4

**ISyE 6420**

Fall 2023

Use of unsolicited electronic and printed resources is allowed except the communication that violates Georgia Tech Academic Integrity Rules (e.g., direct communication about solutions with a third party, use of HW-solving sites, and similar). [1]

**Fall23 HW4.1.** Pairs $(X_i, Y_i), i = 1, \ldots, n$ consist of correlated standard normal random variables (mean 0, variance 1) forming a sample from a bivariate normal $\mathcal{MVN}_2(\mathbf{0}, \Sigma)$ distribution, with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

The density of $(X, Y) \sim \mathcal{MVN}_2(\mathbf{0}, \Sigma)$ is[2]

$$f(x, y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right\},$$

with $\rho$ as the only parameter. Take prior on $\rho$ by assuming Jeffreys' prior on $\Sigma$ as $\pi(\Sigma) = \frac{1}{|\Sigma|^{3/2}} = \frac{1}{(1-\rho^2)^{3/2}}$, since the determinant of $\Sigma$ is $1 - \rho^2$. Thus

$$\pi(\rho) = \frac{1}{(1-\rho^2)^{3/2}} \mathbf{1}(-1 \le \rho \le 1).$$

(a) If $(X_i, Y_i), i = 1, \ldots, n$ are observed, write down the likelihood for $\rho$. Write down the expression for the posterior, up to the proportionality constant (that is, un-normalized posterior as the product of likelihood and prior).

(b) Since the posterior for $\rho$ is complicated, develop a Metropolis-Hastings algorithm to sample from the posterior. Assume that $n = 100$ observed pairs $(X_i, Y_i)$ gave the following summaries:

$$\sum_{i=1}^{100} x_i^2 = 115.9707, \quad \sum_{i=1}^{100} y_i^2 = 105.9196, \quad \text{and} \quad \sum_{i=1}^{100} x_i y_i = 84.5247.$$

In forming a Metropolis-Hastings chain take the following proposal distribution for $\rho$: At step $i$ generate a candidate $\rho'$ from the uniform $\mathcal{U}(\rho_{i-1} - 0.1, \rho_{i-1} + 0.1)$ distribution. Why the proposal distribution cancels in the acceptance ratio expression?



[2] See (6.1) on page 243 in `http://statbook.gatech.edu`.

(c) Simulate 51000 samples from the posterior of $\rho$ and discard the first 1000 samples (burn in). Plot two figures: the histogram of $\rho$s and the realizations of the chain for the last 1000 simulations. What is the Bayes estimator of $\rho$ based on the simulated chain?

(d) Replace the proposal distribution from (b) by the uniform $\mathcal{U}(-1, 1)$ (independence proposal). Comment on the results of MCMC.

**Fall23 HW4.2.** Exponentially distributed lifetimes have constant hazard rate equal to the rate parameter $\lambda$. When $\lambda$ is a constant hazard rate, a simple way to model heterogeneity of hazards is to introduce a multiplicative frailty parameter $\mu$, so that lifetimes $T_i$ have distribution[3]

$$T_i \sim f(t_i|\lambda, \mu) = \lambda\mu \exp\{-\lambda\mu t_i\}, \ t_i > 0, \ \lambda, \mu > 0.$$

The prior on $(\lambda, \mu)$ is

$$\pi(\lambda, \mu) \propto \lambda^{c-1}\mu^{d-1} \exp\{-\alpha\lambda - \beta\mu\},$$

that is, $\lambda$ and $\mu$ are apriori independent with distributions $\mathcal{G}a(c, \alpha)$ and $\mathcal{G}a(d, \beta)$, respectively. The hyperparameters $c, d, \alpha$ and $\beta$ are known (elicited) and positive.

Assume that lifetimes $t_1, t_2, \dots, t_n$ are observed.

(a) Show that full conditionals for $\lambda$ and $\mu$ are gamma,

$$[\lambda|\mu, t_1, \dots, t_n] \sim \mathcal{G}a\left(n + c, \mu\sum_{i=1}^{n} t_i + \alpha\right),$$

and by symmetry,

$$[\mu|\lambda, t_1, \dots, t_n] \sim \mathcal{G}a\left(n + d, \lambda\sum_{i=1}^{n} t_i + \beta\right).$$

(b) Using the result from (a) develop Gibbs Sampler algorithm that will sample 21000 pairs $(\lambda, \mu)$ from the posterior and burn-in the first 1000 simulations. Assume that $n = 20$ and that the sum of observed lifetimes is $\sum_{i=1}^{20} t_i = 522$.

Assume further that the priors are specified by hyperparameters $c = 3, d = 1, \alpha = 100$, and $\beta = 5$. Start the chain with $\mu = 0.1$.

(c) From the produced chain, plot the scatterplot of $(\lambda, \mu)$ as well as histograms of individual components, $\lambda$, and $\mu$. Estimate posterior means and variances for $\lambda$ and $\mu$. Find 95% equitailed credible sets for $\lambda$ and $\mu$.

(d) A frequentist statistician estimates the product $\lambda\mu$ as $\frac{n}{\sum_{i=1}^{n} t_i} = 20/522 = 0.0383$. What is the Bayes estimator of this product? (Hint: It is not the product of averges, it is the average of products, so you will need to save products in the MCMC loop).

[3]This is a toy example. In realistic applications the multiplicative frailty depends on the $i$, or on a subclass of population as $\mu_{j(i)}$.