STAT 5370 – Decision Theory

Homework 3

Matthew Aaron Looney

7/2/2018

Contents

Problem 1: Multivariate Normal with unknow μ and Σ

Let $X \sim N(\mu, \Sigma)$ be a $p \times 1$ random vector distributed as a multivariate normal with mean μ and variance-covariance Σ, where both μ and Σ are unknown.

(a) Write out a candidate conjugate prior for (μ, Σ) motivating your choice.

Fortunately, this problem is very related to our group project paper…

We start by writing the likelihood function for the multivariate normal:

$$p(X|\mu, \Sigma) \propto \Sigma^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right\}$$

$$p(X|\mu, \Sigma) \propto \Sigma^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}tr(\Sigma^{-1}S)\right\}$$

where S is the matrix of sum of squares (sometimes called the scatter matrix).

$$S = \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'$$

The natural conjugate prior is normal-inverse-wishart:

$$\Sigma \sim \mathrm{IW}_{v0}\left(\Lambda_0^{-1}\right)$$

$$\mu|\Sigma \sim N(\mu_0, \Sigma/k_o)$$

$$p(\mu, \Sigma) = NIW(\mu_0, k_0, \Lambda_0, v_0)$$

$$p(\mu, \Sigma) = \frac{1}{Z}\Sigma^{-[(v_0+df)/2+1]} \exp\left\{-\frac{1}{2}tr(\Lambda_0\Sigma^{-1}) - \frac{k_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)\right\}$$

where $z = \frac{2^{v_0 df/2}\Gamma_{df}(v_0/2)(2\pi/k_0)^{df/2}}{\Lambda_0{}^{v_0/2}}$ is a constant and can be omitted from the model if we consider the prior up to a proportionality constant, as is typical.

(b) Derive the posterior joint distribution of μ and Σ associated to the prior you propose: is it in the same family of the prior?

$$p(\mu, \Sigma|X, \mu_0, k_0\Lambda_0, v_0) = NIW(\mu, \Sigma|\mu_n, k_n, \Lambda_n, v_n)$$

where, $u_n = \frac{k_0\mu+n\bar{x}}{k_n}$, $k_n = k_0+n$, $v_n = v_0+n$, $\Lambda_n = \Lambda_0 + S + \frac{k_0 n}{K_0+n}(\bar{x}-\mu_0)(\bar{x}-\mu_0)'$.

Therefore,

$$p(\mu,\Sigma|X) \propto L(X|\mu,\Sigma)P(\mu,\Sigma)$$

$$p(\mu,\Sigma|X) \propto \Sigma^{-\frac{n}{2}\exp}\left\{-\frac{n}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{X}-\mu)\right\}\exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right)\right\}\cdot$$

$$\cdot\Sigma^{-[\{(v_0+df)/2\}+1]}\exp\left\{-\frac{1}{2}tr(\Lambda_0\Sigma^{-1})-\frac{k_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)\right\}$$

$$\propto \Sigma^{-\{(v_n+df)/2\}+1}\exp\left\{-\frac{1}{2}tr(\Phi_n\Sigma^{-1})-\frac{k_n}{2}(\mu-\mu_n)'\Sigma^{-1}(\mu-\mu_n)\right\}$$

The posterior distribution is $p(\mu,\Sigma|X) \sim Normal-InverseWishart(\mu_n, k_n^{-1}, \Phi_n^{-1}, v_n)$

Or more succinctly,

$$\mu|\Sigma,x,y \sim \; N_2\left[(\bar{x},\bar{y})', n^{-1}\Sigma\right]$$

$$\Sigma|x,y \sim IW_2\left(S^{-1},n\right)$$

This is the same family as the prior.

(c) What is the marginal posterior distribution for the vector of means μ?

$$p(\mu,\Sigma|X) \propto \Sigma^{-(v_n/2)+1}\exp\left\{-\frac{1}{2}tr(\Phi_n\Sigma^{-1})-\frac{k_n}{2}(\mu-\mu_n)'\Sigma^{-1}(\mu-\mu_n)\right\}$$

$$\propto \Sigma^{-(v_n/2)+1}\exp\left\{-\frac{1}{2}tr\left(\Phi_n+k_n(\mu-\mu_n)(\mu-\mu_n)'\right)\Sigma^{-1}\right\}$$

where if we assume $\Phi_n+k_n(\mu-\mu_n)(\mu-\mu_n)'=A$ then we arrive at the following result:

$$p(\mu,\Sigma|X) \propto \Sigma^{-(v_n/2)+1}\exp\left\{-\frac{1}{2}tr\left(A\Sigma^{-1}\right)\right\}$$

We next need to integrate out Σ to arrive at the marginal distribution for μ.

$$p(\mu|X) \propto \int_A \Sigma^{-(v_n/2)+1}\exp\left\{-\frac{1}{2}tr\left(A\Sigma^{-1}\right)\right\}d\Sigma$$

$$p(\mu|X) \propto t_{v_n-df+1}\left[\mu_n, \frac{\Lambda_n}{v_n-df+1}\right]$$

This marginal is a multivariate t-distribution with mean μ and degrees of freedom equal to (v_n-df+1).

Problem 2: Gibbs Sampling Problem

We investigate the effects of a bimodal posterior on the performance of Gibbs sampling. Suppose we have a statistical model with two-dimensional parameter $\theta = (\theta_1, \theta_2)'$, and say θ has the following posterior distribution:

$$\theta|Data \ \frac{1}{2}N(\mu_1, \Sigma) + \frac{1}{2}N(\mu_2, \Sigma)$$

Consider using Gibbs sampling to generate a sample from this posterior. Given the current state $\theta^{(t)}$, $\theta^{(t+1)}$ is generated through the following scheme:

(a) Sample $\theta_1^{(t+1)}$ from $p\theta_1|\theta_2^t, Data)$;

(b) Sample $\theta_2^{(t+1)}$ from $p\theta_2|\theta_1^{t+1}, Data)$

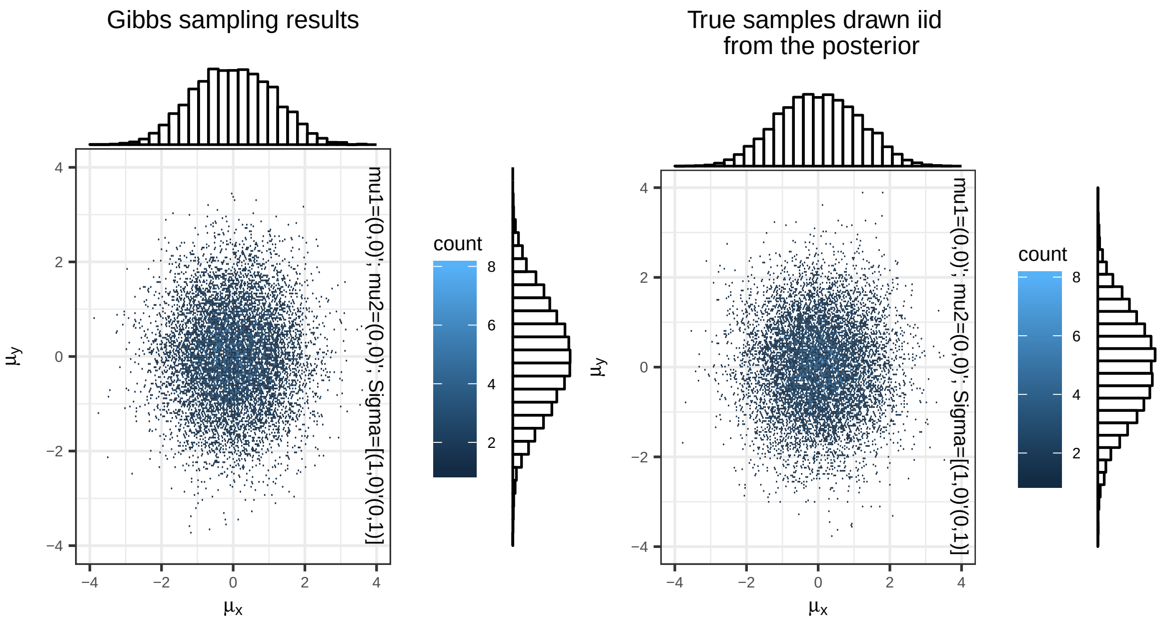
(c) Save $\theta^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)})'$

Derive the Gibbs updates and implement the sampling procedure using your R software. Run your procedure 10, 000 using $\mu_1 = (0,0)'$ and $\mu_2 = (a,a)'$, and $\Sigma = [(1,0)', (0,1)]$. Try using small and large values for a, for example $a \in (0, 1.5, 10)$. Compare the Gibbs sampling results to true samples drawn iid from the posterior and comment on your findings. Repeat the experiment using $\Sigma = [(1,0)', (0,\sigma^2)]$ for a large a large and σ; comment on the results.

The general algorithm for the Gibbs updating is as follows:

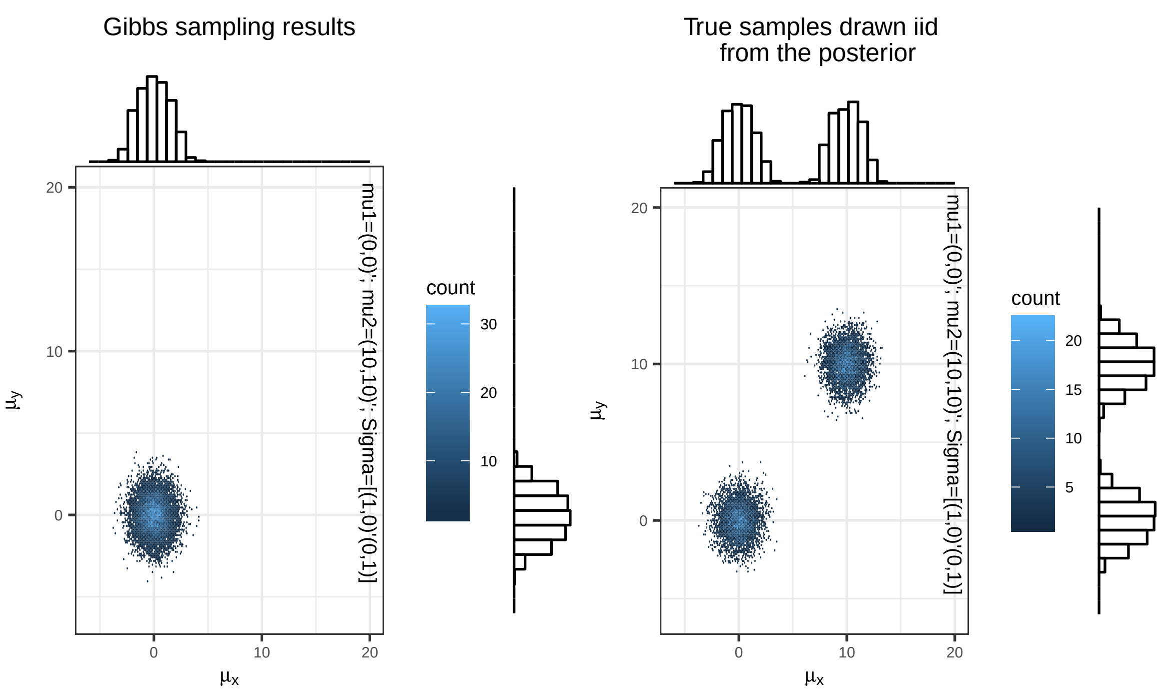
- Initialize $\theta^0 \sim q(\theta)$
- for iteration $i = 1, 2, ..., N$...
- $\theta_1^{(1)} \sim p(\Theta_1 = \theta_1 | \Theta_2 = \theta_2^{(i-1)}, \Theta_3 = \theta_3^{(i-1)}, ..., \Theta_N = \theta_N^{(i-1)})$
- $\theta_2^{(1)} \sim p(\Theta_2 = \theta_2 | \Theta_1 = \theta_1^{(i-1)}, \Theta_3 = \theta_3^{(i-1)}, ..., \Theta_N = \theta_N^{(i-1)})$
- $\vdots$
- $\theta_N^{(1)} \sim p(\Theta_N = \theta_N | \Theta_1 = \theta_1^{(i-1)}, \Theta_2 = \theta_2^{(i-1)}, ..., \Theta_N = \theta_N^{(i-1)})$
- End

The above algorithm is used to iterate back and forth between multivariate (bivariate in this case) distributions until convergence has been achieved. In general the Gibbs sampler is a good starting point when utilizing computation methods to uncover distributions. There are occasions where the Gibbs sample will fail. We will see in this problem that this is such a case. . .



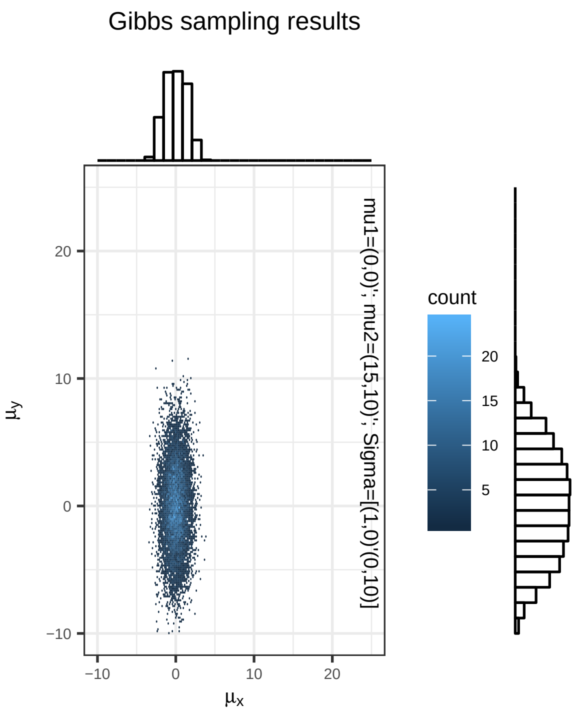
The above plots show the results from running a Gibbs sampler and a plot from taking true samples drawn iid from the posterior distribution. The parameter values for each plot are $\mu_1 = (0,0)'; \mu_2 = (0,0)'$; $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

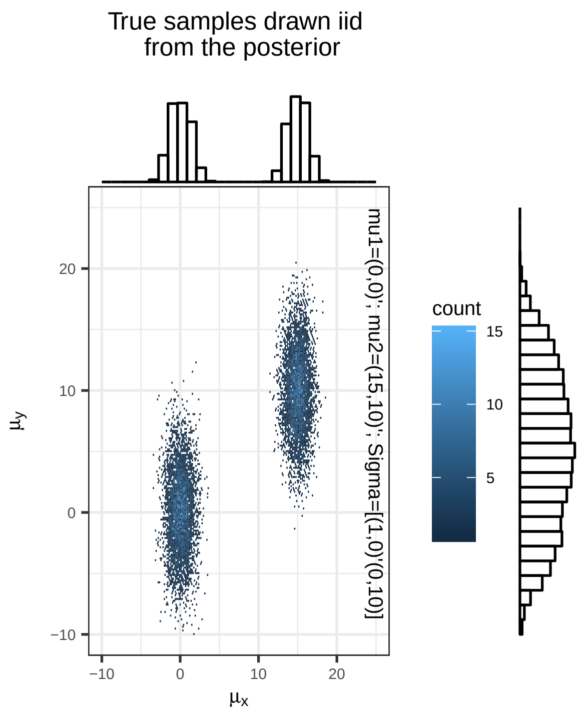
The below plots show the results from running a Gibbs sampler and a plot from taking true samples drawn iid from the posterior distribution. The parameter values for each plot are $\mu_1 = (0,0)'; \mu_2 = (10,10)'$; $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. We can clearly see how the Gibbs sampler fails to uncover the true bivariate density while the true samples drawn iid from the posterior are well resolved into two distributions.



The below plots show the results from running a Gibbs sampler and a plot from taking true samples drawn iid from the posterior distribution. The parameter values for each plot are $\mu_1 = (0,0)'$; $\mu_2 = (15,10)'$; $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix}$. We can again clearly see how the Gibbs sampler fails to uncover the true bivariate density while the true samples drawn iid from the posterior are well resolved into two distributions. We can see the means of the two distributions from the iid drawn samples. The density has spread out due to the larger variance but the Gibbs sample does not resolve into two separate densities, it simply spreads out.

The Gibbs sampler fails in this case.





Appendix: R Code