ISYE 6420: Bayesian Statistics **Spring 2020**

Homework #4

Instructor: Roshan Vengazhiyil, Brani Vidakovic *Name:* Nick Korbit*, gtID:* 903263968

Problem 1

a). If we observe n i.i.d. data points distributed as $(X, Y) \sim \mathcal{MVN}_2(\mathbf{0}, \Sigma)$, we get the following likelihood:

$$h(\rho|x, y) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right) \right\} \right) =$$

$$= \left(\frac{1}{2\pi\sqrt{1-\rho^2}} \right)^n \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right) \right\}$$

Applying log transform to the likelihood we obtain:

$$logh(\rho|x, y) = -nlog\left(2\pi\sqrt{1-\rho^2}\right) - \frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right)$$

We can write un-normalized posterior as:

$$\pi(\rho|x, y) \propto \prod_{i=1}^{n} \left(\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right) \right\} \frac{1}{(1-\rho^2)^{3/2}} \right)$$

$$\pi(\rho|x, y) \propto \prod_{i=1}^{n} \left(\frac{1}{2\pi(1-\rho^2)^2} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right) \right\} \right)$$

$$\pi(\rho|x, y) \propto \left(\frac{1}{2\pi(1-\rho^2)^2} \right)^n \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum \left(x_i^2 - 2\rho x_i y_i + y_i^2\right) \right\}$$

Applying log transform to the posterior we obtain:

$$log(\pi|x, y) \propto -nlog\left(2\pi\left(1-\rho^2\right)^2\right) - \frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right)$$

$$log(\pi|x, y) \propto -nlog2\pi - 2nlog\left(1-\rho^2\right) - \frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right)$$

b). Before devising steps for the Metropolis-Hastings procedure we make two observations.
First, our proposal distribution is uniform $\mathcal{U}(\rho_{i-1} - 0.1, \rho_{i-1} + 0.1)$. We can calculate the density $f(x)$ of the uniform distribution as $f(x) = \frac{1}{b-a}$ or in our case:

$$f(x) = \frac{1}{\rho_{i-1} + 0.1 - (\rho_{i-1} - 0.1)} = \frac{1}{0.2} = 5$$

So that our proposal distribution does not dependent on ρ.
Second, knowing the statistics after 100 observations, we can simplify $log(\pi|x, y)$ as

$$\sum_{i=1}^{n} \left(x_i^2 - 2\rho x_i y_i + y_i^2\right) = \sum_{i=1}^{n} x_i^2 - 2\rho \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 = 220.8903 - 165.0494\rho$$

$$\pi(\rho|x, y) \propto \left(\frac{1}{2\pi(1-\rho^2)^2} \right)^{100} \exp\left\{ -\frac{220.8903 - 165.0494\rho}{2(1-\rho^2)} \right\}$$

We now specify the Metropolis-Hastings procedure:

1. Start with arbitrary ρ_0
2. Generate proposal ρ' from $\mathcal{U}\left(\rho_{i-1} - 0.1, \rho_{i-1} + 0.1\right)$
3. Calculate acceptance probability α:

$$\alpha\left(\rho_n, \rho'\right) = \min\left\{1, \frac{\pi\left(\rho'\right)}{\pi\left(\rho_n\right)} \frac{q\left(\rho_n|\rho'\right)}{q\left(\rho'|\rho_n\right)}\right\}$$

   Taking into account our observations we can simply α:

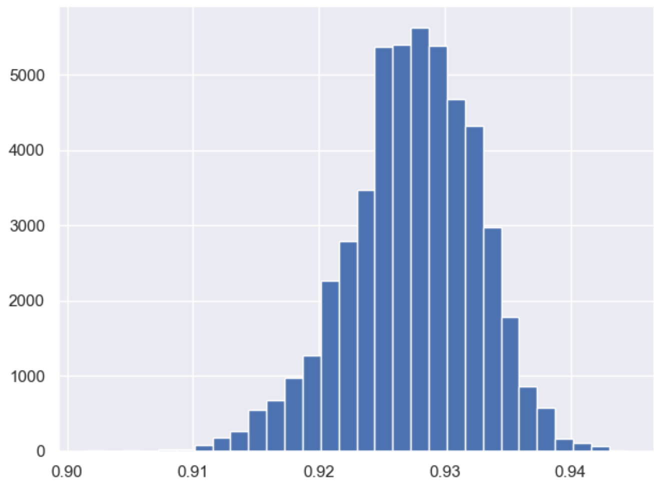
$$\alpha\left(\rho_n, \rho'\right) = \min\left\{1, \frac{\pi\left(\rho'\right)}{\pi\left(\rho_n\right)}\right\}$$

   where

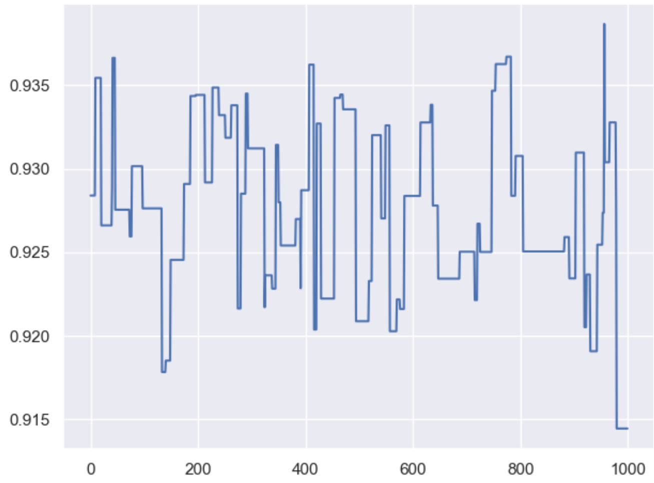
$$\pi\left(\rho\right) = \left(\frac{1}{\left(1-\rho^2\right)^2}\right)^{100} \exp\left\{-\frac{220.8903 - 165.0494\rho}{2\left(1-\rho^2\right)}\right\}$$

4. Determine ρ_{n+1} as
   - $\rho_{n+1} = \rho'$ with probability α
   - $\rho_{n+1} = \rho_n$ with probability $1 - \alpha$
5. Set $n = n + 1$ and go to Step 2.

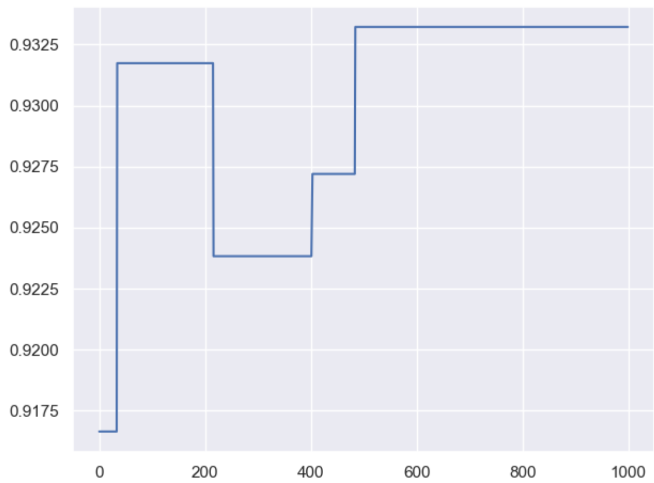
c). We now simulate 51000 samples from the posterior. Burning the first 1000 samples we plot the histogram of the last 50000 samples:



We see that the empirical distribution is unimodal, with a peak around 0.93 value.
We also plot the last 1000 observations.



We notice that the majority of samples are -1 with a small spike with values $(-1, -0.6)$.
Let's calculate Bayes' estimators: the mean of the simulated data is 0.9274 and the mode is 0.9307.
d). Let's switch proposal distribution to the uniform $\mathcal{U}(-1, 1)$. Plotting last 1000 samples will yield:



We see that using a pure sample from $\mathcal{U}(-1, 1)$ will result in discarding the majority of the samples in later stages of the simulation. Thus, we might stick to the original distribution in order to generate samples that are accepted more often and calculate more reliable statistics.

Problem 2

a). Let's start with a joint distribution $f(T_i, \lambda, \mu)$:

$$f(T_i, \lambda, \mu) = \left\{ \prod_{i=1}^{n} f\left(T_i | \lambda, \mu\right) \right\} \pi(\lambda)\pi(\mu) = \left\{ \prod_{i=1}^{n} \lambda\mu \exp\left\{-\lambda\mu t_i\right\} \right\} \lambda^{c-1}\mu^{d-1} \exp\{-\alpha\lambda - \beta\mu\} =$$

$$= \lambda^n \mu^n \lambda^{c-1} \mu^{d-1} \exp\{-\alpha\lambda - \beta\mu\} \exp\left\{-\lambda\mu \sum_{i=1}^{n} t_i\right\}$$

We now find the full conditionals for λ and μ by selecting the terms from $f(T_i, \lambda, \mu)$ that contain λ and μ respectively and then normalize:

$$f(\lambda | T_i, \mu) \propto \lambda^{n+c-1} \exp\left\{-\lambda \left(\alpha + \mu \sum_{i=1}^{n} t_i\right)\right\}$$

We can recognize a $\mathcal{G}a(a,b)$ distribution of a general form $x^{a-1}e^{-bx}$ where in our case $a = n + c$ and $b = \alpha + \mu \sum_{i=1}^{n} t_i$. So that

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

Using the same derivation we can also state that

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

b). We now develop a Gibbs sampler procedure for our case. But first we substitute n, c, α, β, $\sum_{i=1}^{n} t_i$ with experimental data and hyperparameters. So that we can sample from:

$$[\lambda|\mu, t_1, \ldots, t_n] \sim \mathcal{G}a\left(23, 512\mu + 100\right)$$

$$[\mu|\lambda, t_1, \ldots, t_n] \sim \mathcal{G}a\left(21, 512\lambda + 5\right)$$

The algorithm is then:

1. Start with $\mu_0 = 0.1$
2. Sample initial λ_0 from $\mathcal{G}a(23, 151.2)$
3. Sample
   - μ_{n+1} from $[\mu|\lambda, t_1, \ldots, t_n] \sim \mathcal{G}a\left(21, 512\lambda_n + 5\right)$
   - λ_{n+1} from $[\lambda|\mu, t_1, \ldots, t_n] \sim \mathcal{G}a\left(23, 512\mu_{n+1} + 100\right)$
4. Set $n = n + 1$ and go to Step 3.

c). We now simulate 51000 paths using the Gibbs sampler. We plot the histograms for λ and μ and the scatter plot linking those variables:

Figure 1: Distribution of Lambdas.

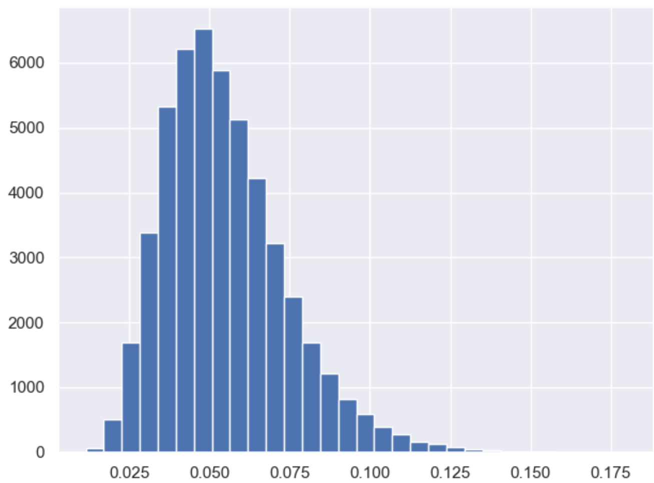


Figure 2: Distribution of Mus.

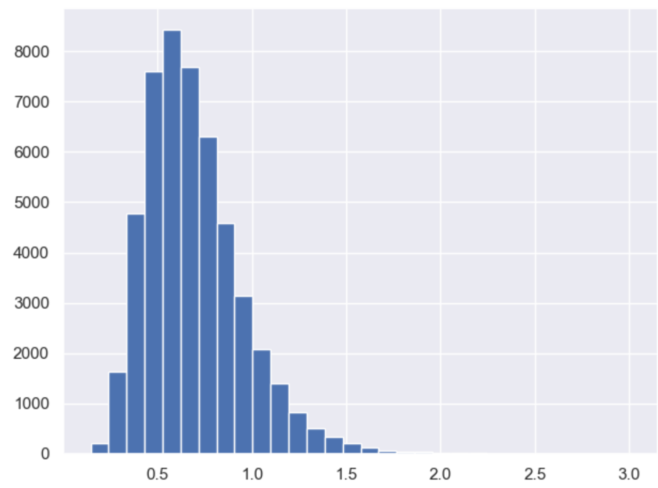
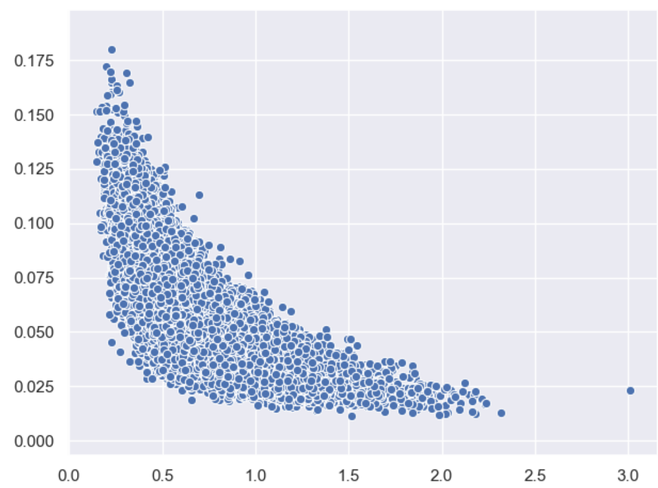


Figure 3: Scatter of Mus (x-axis) and Lambdas (y-axis).



Let's calculate the statistics based on the simulated λ and μ:

- $E[\lambda] \approx 0.0548$
- $Var[\lambda] \approx 0.0003667$
- 95-% Equitailed credible set for λ is [0.02569, 0.10013]
- $E[\mu] \approx 0.69033$
- $Var[\mu] \approx 0.06504$
- 95-% Equitailed credible set for λ is [0.3141, 1.30128]

d). We can also calculate Bayes estimator by taking the average of the product of λ and $\mu \approx 0.03426$.
Note. The code for solving Q1 and Q2 is implemented in *hw4.py*. Just run *python hw4.py*.

References

[1] Engineering Biostatistics: An Introduction using MATLAB and WinBUGS. Brani Vidakovic - Wiley Series in Probability and Statistics.