STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

MASTER OF SCIENCE IN STATISTICAL SCIENCE

END OF SEMESTER EXAMINATION

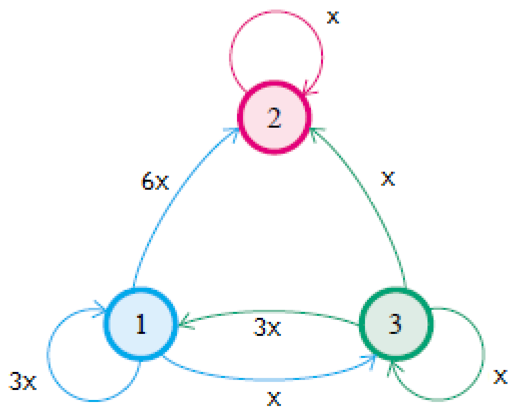
STA 8201 BAYESIAN MODELLING AND DATA ANALYSIS

Date: 20th December, 2021 **Duration:** 3 Hours

Answer Question ONE and any other Two questions:

Question 1 (30 marks)

a. Find the transition matrix from the transition diagram below (4 marks)



b. Briefly describe Gibbs sampler for parameters θ_1, θ_2, and θ_3, with joint posterior distribution $p(\theta_1, \theta_2, \theta_3|y)$ (5 marks)

c. The number of the lions y=1, 2, 3 … breaking out of a Nairobi national park within the last one month follow the distribution $f(y|\theta) = \theta(1-\theta)^{y-1};\ y = 1,2,3\ldots;\ 0 < \theta < 1$. Find the Jeffrey's prior distribution of θ and hence or otherwise its posterior distribution. (10 marks)

d. Suppose that $x_1, x_2, ..., x_n$ is a random sample from a geometric distribution with pdf

$$p(x|\theta) = \begin{cases} \theta(1-\theta)^x & 0 < \theta \leq 1 \quad x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}.$$

Show that $p(x|\theta)$ is an exponential distribution and hence find a conjugate prior for θ. Find the posterior distribution for. θ. (6 marks)

e. Find the long term trend of the transition matrix (5 Marks)

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} = P.$$

<u>Question 2 (15 marks)</u>

a. A doctor is called to see a sick child. The doctor has prior information that 90% of sick children in that neighbourhood have the flu, while the other 10% are sick with 1 measles. Let F stand for an event of a child being sick with flu and M stand for an event of a child being sick with measles. Assume for simplicity that F ∪ M = Ω, i.e., that there no other maladies in that neighbourhood. A well-known symptom of measles is a rash (the event of having which we denote R). Assume that the probability of having a rash if one has measles is P(R | M) = 0.95. However, occasionally children with flu also develop rash, and the probability of having a rash if one has flu is P(R | F) = 0.08. Upon examining the child, the doctor finds a rash. What is the probability that the child has measles?

(5 marks)

b. Consider a bivariate normal posterior distribution of the parameters θ_1 and θ_2:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

i. Determine the full conditionals of $\theta_1|\theta_2$ and $\theta_2|\theta_1$. (5 Marks)

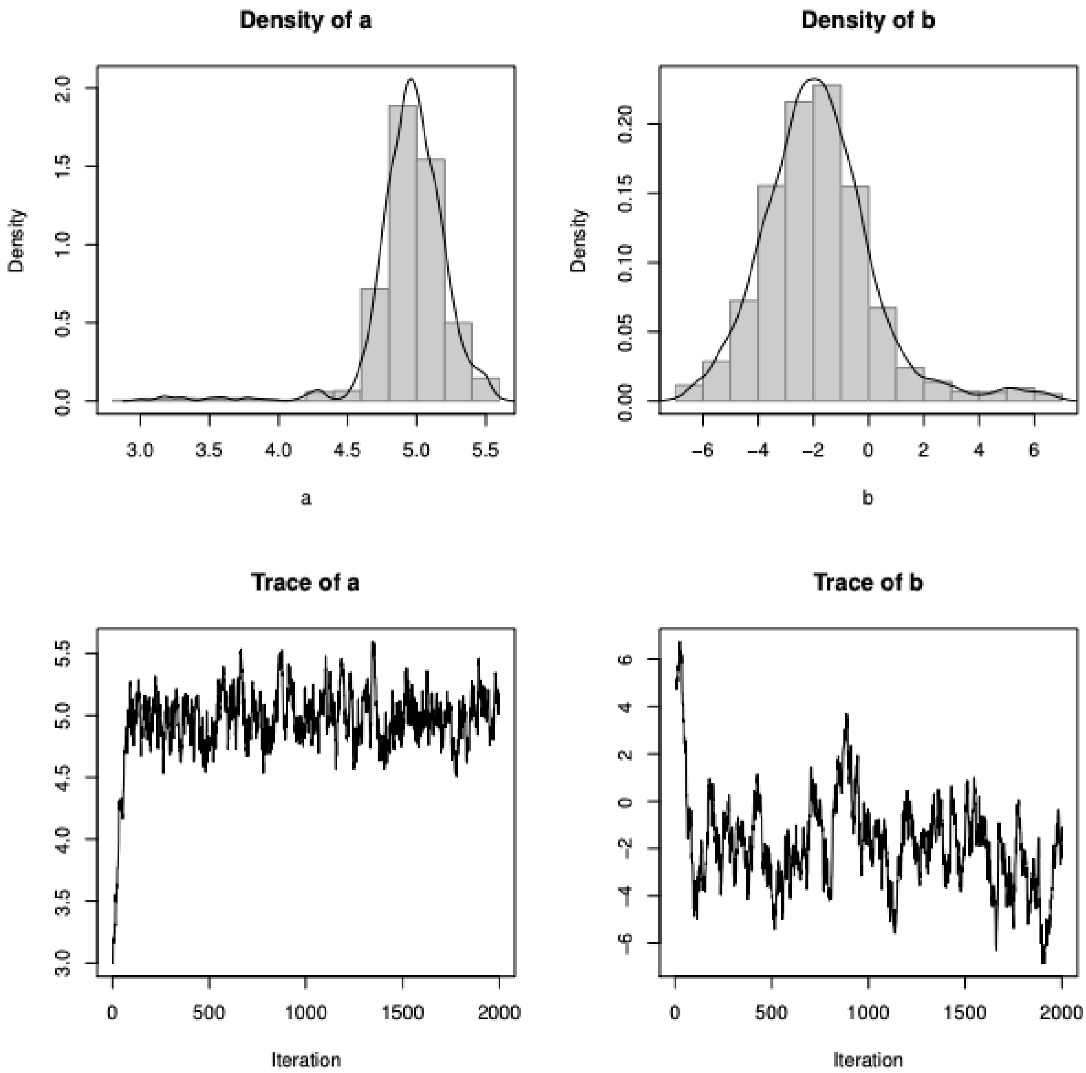
ii. Write a **R code** for Gibbs sampling from the full conditionals in (a.) above. (5 Marks)

<u>Question 3 (15 marks)</u>

a. In a study, physicians were asked what the odds of breast cancer would be in a woman who was initially thought to have a 1% risk of cancer but who ended up with a positive mammogram result (a mammogram accurately classifies about 80% of cancerous tumors and 90% of benign tumors.) 95 out of a hundred physicians estimated the probability of cancer to be about 75%. Do you agree? (5 marks)

b. How does Markov Chain Monte Carlo differ fundamentally from ordinary Monte Carlo? Explain the terms "thinning" and "burn-in" and the purpose of each in the context of the following plots of output from a Gibbs sampler. (4 marks)

c. Explain the terms "thinning" and "burn-in" and the purpose of each in the context of the following plots of output from a Gibbs sampler. What thinning and burn-in would you recommend for this problem? (6 marks)



Question 4 (15 marks)

a. It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%.
(5 marks)

b. The WinBUGS model below concerns the proportion of seeds that germinated on each of 21 plates arranged according to a 2 by 2 factorial layout by seed and type of root extract. Let r_i and n_i respectively be the number of germinated and the total number of seeds on the i th plate, i =1,...,N.

```
model { for (i in 1:K) {
        for (j in 1:n) {
                Y[i, j] ~ dnorm(eta[i, j], tauC)
                eta[i, j] <- phi[i, 1] / (1 + phi[i, 2] * exp(phi[i, 3] * x[j]))
        }
        phi[i, 1] <- exp(theta[i, 1])
        phi[i, 2] <- exp(theta[i, 2]) - 1
        phi[i, 3] <- -exp(theta[i, 3])
        for (k in 1:3) {
                theta[i, k] ~ dnorm(mu[k], tau[k])
                }
        }
        tauC ~ dgamma(1.0E-3, 1.0E-3)
        sigmaC <- 1 / sqrt(tauC)
        varC <- 1 / tauC
        for (k in 1:3) {
                mu[k] ~ dnorm(0, 1.0E-4)
                tau[k] ~ dgamma(1.0E-3, 1.0E-3)
                sigma[k] <- 1 / sqrt(tau[k])
                }
        }
}
```

i. Describe using simple mathematical equation the likelihood and the priors from the WinBUGS code above (7 marks)

ii. Sketch the densities of parameter "tauC" under Uniform and Gamma priors?
(3 marks)