STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

MASTER OF SCIENCE IN BIOMATHEMATICS

END OF SEM EXAMINATION

STA 8201-BAYESIAN MODELLING AND DATA ANALYSIS

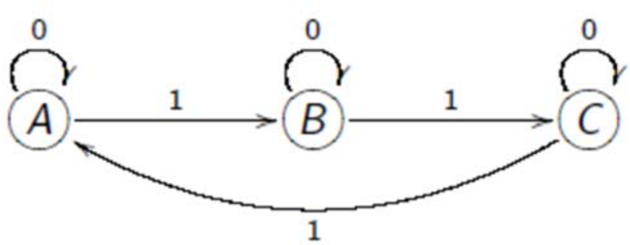
DATE: 18th April 2024 Time: 3 Hours

Instructions

1. This examination consists of **FIVE** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question One (20 Marks)

a. Explain what it means by aperiodic chain, and hence state whether or not the chain below is aperiodic (4 marks)



b. Describe Gibbs sampling algorithm (5marks)

c. A lifetime X of a particular machine is modelled by an exponential distribution with unknown parameter θ. If the parameterization is $f(x|\theta) = \theta e^{-\theta x}$, $x \geq 0, \theta > 0$, the MLE estimator for θ is $\hat{\theta}_{MLE}$=$\frac{1}{\hat{X}}$ on basis of a sample X_1,…,X_n. The lifetimes (in years) of X_1=5, X_2=6 and X_3=4 are observed.

i. Write down the MLE of θ for those observations. (3 marks)

Assume now that an expert believes that θ should have exponential distribution as well and that, on average θ shouls be $\frac{1}{3}$.

ii. Elicit a prior according to expert's belief. (2 marks)

iii. For the prior in (ii), find the posterior. Is the problem conjugate? (4 marks)

iv. Find the Bayes estimator $\hat{\theta}_{Bayes}$, and compare it with the MLE estimator from (i). Discuss (2 marks)

Question TWO (20 Marks)

Suppose that $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Bernoulli(\theta)$ and that θ has a $Beta(a, b)$ prior.

(a) Find the posterior Bayes estimator of θ. Show that it is a weighted average of the sample mean and the prior mean.

(b) Find the predictive distribution for X_{n+1}. Give a point estimate of X_{n+1} given the previous data.

Question THREE (20 Marks)

Sociologists have long been interested in *social mobility* – the transition of individuals between social classes defined on the basis of income or occupation. Consider a society with three social classes. Each individual may belong to the lower class (state 1), the middle class (state 2), or the upper class (state 3). Thus, the social class occupied by an individual in generation t may be denoted by $s_t \in \{1,2,3\}$. suppose that *intergenerational mobility* is described by the transition matrix P

$$\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$$

a. Determine the transition diagram from this transition matrix. (4 marks)

b. Find the transition probabilities after 3 years? (6 marks)

c. Write and run an **R code** to find the long term trend of the transition matrix. (10 marks)

Question FOUR (20 Marks)

Let y be the number of heads in n spins of a coin, whose probability of heads is θ.

a. If your prior for θ is a Beta(α, β) distribution show that your posterior mean for θ always lies between the prior mean α/(α+β) and the observed proportion of heads y/n. (7 marks)

b. Show that if the prior distribution for θ is uniform (Beta(1,1)) then the posterior variance is always less than the prior variance. (7 marks)

c. Give an example of a Beta(α, β) prior and data y; n for which them posterior variance of θ is higher than the prior variance. (6 marks)

Question FIVE (20 Marks)

A biologist is interested in the proportion, θ, of badgers in a particular area which carry the infection responsible for bovine tuberculosis. The biologist's prior distribution for θ is a beta(1, 19) distribution.

a)

i. Find the biologist's prior mean and prior standard deviation for θ. (3 marks)
ii. Find the cumulative distribution function of the biologist's prior distribution and hence find values θ_1, θ_2 such that, in the biologist's prior distribution, $\Pr(\theta < \theta_1) = \Pr(\theta > \theta_2) = 0.05$. (3 marks)

(b) The biologist captures twenty badgers and tests them for the infection. Assume that, given θ, the number, X, of these carrying the infection has a binomial(20, θ) distribution. The observed number carrying the infection is x = 2.

i. Find the likelihood function. (3 marks)
ii. Find the biologist's posterior distribution for θ. (4 marks)
iii. Find the biologist's posterior mean and posterior standard deviation for θ. (3 marks)
iv. Write R code to plot a graph showing the biologist's prior and posterior probability density functions for θ (4 marks)