



Strathmore
UNIVERSITY

**STRATHMORE UNIVERSITY
INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN STATISTICAL SCIENCE
END OF SEMESTER EXAM 2019/2020
STA 8201-BAYESIAN MODELLING AND DATA ANALYSIS**

Date: 31st August, 2020

Duration: 3½ Hours

Instructions to candidates

- i. *This is an open book virtual examination*
- ii. *Attempt Question any **THREE** questions*

Question 1 (20 marks)

- a. Assume that you want to investigate the proportion (θ) of defective items manufactured at a production line. A researcher takes at random sample of 36 items. Six were defective in the sample. Assuming a Jeffrey prior for θ , determine the posterior distribution of θ . (13 marks)
- b. The researcher now tells you that he did not decide on the sample size before the sampling was performed. Her sampling plan was to keep sampling items until she had found six defective ones. Redo the posterior calculations this time under the new sampling scheme. Comment on the results. (7 marks)

Question 2 (20 marks)

- a. How do Metropolis-Hastings (M-H) algorithm and Gibbs sampling relate? (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 $\theta_2|\theta_1$. (7 Marks)
- ii. Sketch a **R code** for Gibbs sampling from the full conditionals in (i.) above. (8 Marks)

Question 3 (20 marks)

A Supreme Court considered the use of Bayes Theorem to combine items of evidence in *OJ Simpsons* case. Before any prosecution or defence evidence is presented, it might be reasonable to assume that the culprit is a male aged between 24 and 59 years old. There were approximately 120 000 males of this age who lived locally to where the crime took place. Of course, the culprit may not be local. This increases the hypothesized number of possible men who could be the culprit from around 120 000 to 160 000, say.

The prosecution case rested on forensic evidence, which will be called E:

E = DNA match between Simpsons and a sample, accepted as being from the culprit, taken from the victim.

The prosecutor's forensic expert testified that for a randomly chosen person, the probability that their DNA matched that of the sample was 1/160 000 000. The defence argued that this probability was in fact 1/1 600 000.

- What is the naïve probability that Simpsons is guilty? (3 marks)
- The prosecutor's forensic expert testified argued that

$$P(\text{Simpsons not guilty} \mid E) = 1/160\,000\,000.$$

- Do you support the definition and interpretation of this conditional probability? Support your argument. (5 marks)
- The judge and jury require an estimate of the probability that Simpsons is guilty, given the evidence E. Calculate estimate of this probability. (9 marks)
 - Would you recommend use of Bayes Theorem in a similar case? Give reason for your answer. (3 marks)

Question 4 (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

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

- Determine the transition diagram from this transition matrix (3 marks)
- What will be the proportions of lower, middle and upper class in long run? (10 marks)
- Write a **R code** to find the long term trend of the transition matrix (7 Marks)