Family Name/
Surname: ____________________

First Name/
Given Name: ____________________

ID number: |__|__|__|__|__|__|__|__|

Venue: ____________________

Seat Number: ______

Mathematics and Statistics

EXAMINATION

End-of-year Examinations, 2021

STAT314/461-21S2 (C) Bayesian Inference

Time allowed: **4 HOURS**

Number of questions: **4**

Number of pages: **7**

Instructions to Students:

- Answer all 4 questions.
- Write your name and student ID on top of EVERY page of your answer.
- Write your answers on blank paper then scan and upload on completion.
- There is a total of 60 marks.
- Show all workings.
- Open book exam

For Examiner Use Only

Question	Mark
Q1 /15	
Q2 /15	
Q3 /15	
Q4 /15	
Total /60	

Questions Start on Page 3

1. **Question 1 (15 marks)**

   A local basketball club has 20 players, 2 of whom are experts and 18 are newbies. An expert gets a ball into the basket (i.e. scores) on 90% of all attempts, whereas a newbie does it with probability 50%.

   (a) A random player has made 2 attempts and scored both times. What is the probability that they are an expert? (5 marks)

   (b) What is the probability that the next attempt will be a miss? (Hint: think of posterior predictive distribution.) (5 marks)

   (c) How many more goals (without any misses in between) would make you at least 95% certain that the player is an expert? (5 marks)

2. Question 2 (15 marks)

Continuing with the basketball example, let p denote the probability of scoring a goal on any one try. Then the probability that it took a player i exactly $x_i = x^*$ misses before scoring a goal is

$$Pr(x_i = x^* | p) = (1-p)^{x^*} p, \qquad \text{for } x^* = 0, 1, 2, \ldots$$

Assume you have observations $x_1, \ldots, x_n$ for n players.

(a) Derive the maximum likelihood estimator for p. (Don't forget the second order condition and assume the sample average $\bar{x}$ is not zero.) (3 marks)

(b) Within the Bayesian framework, assume a beta prior for p:

$$p \sim \text{Beta}(a, b),$$

and use Bayes' formula to derive the posterior distribution $f(p|x_1, \ldots, x_n)$. (3 marks)

(c) Find the posterior mean for p, $E(p|x_1, \ldots, x_n)$ based on the above distribution. What happens to it when the sample size increases? (Hint: think of what happens when the sample size goes to infinity.) (3 marks)

(d) For $n = 3$ players, the observed number of misses before a goal were 1, 8 and 0 respectively. Assuming a uniform prior for p, what is the posterior mean probability of scoring on any given attempt? (3 marks)

(e) Explain the difference between the classical p-value and the Bayesian posterior probability in the context of hypothesis testing. (3 marks)

3. Question 3 (15 marks)

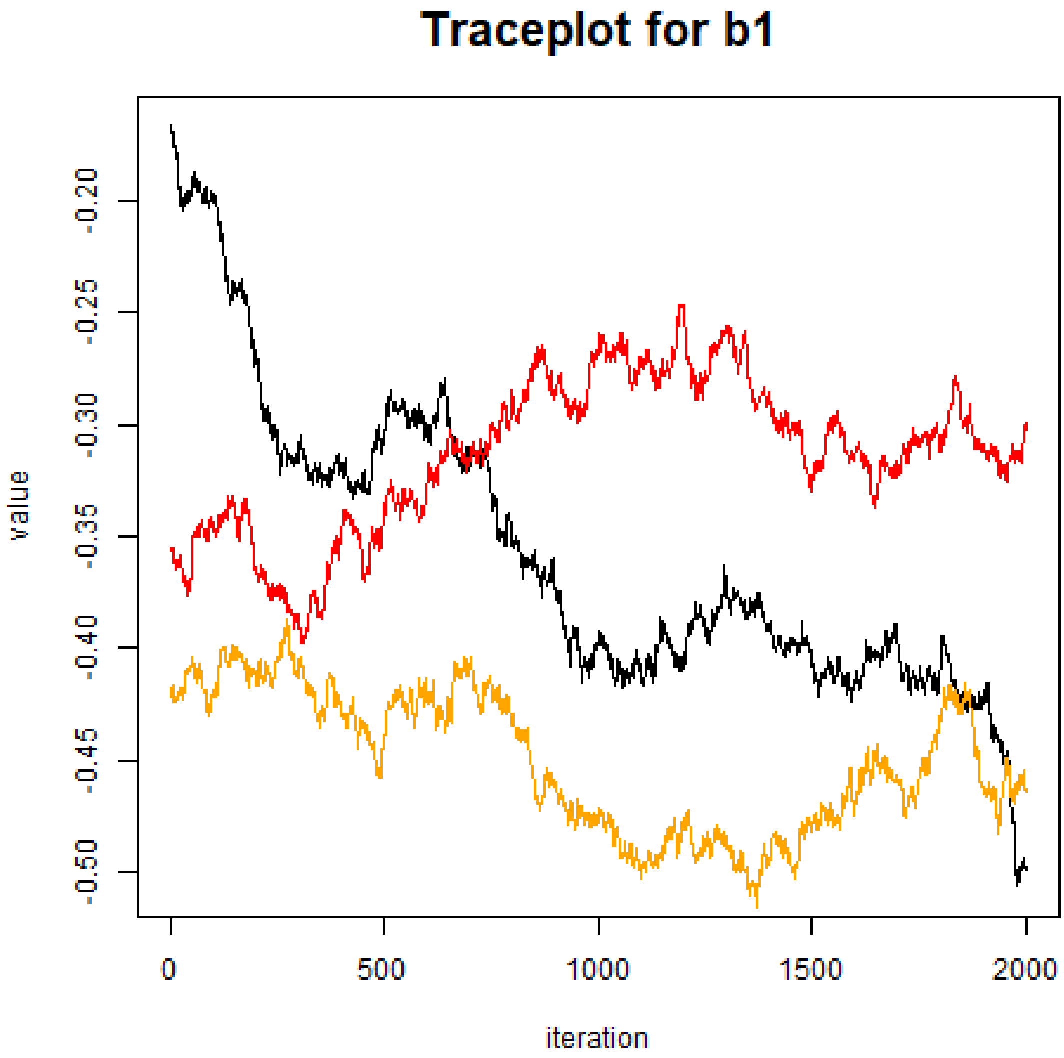

Figure 1: Convergence diagnostics: trace plots for three parallel Metropolis-Hastings chains for one of the parameters of a Poisson regression model

(a) Why is it important to check the convergence of MCMC procedures? (2 marks)

(b) Figure 1 displays traceplots for three parallel Metropolis-Hastings chains for one of the parameters of a Poisson regression model. The acceptance rates for the three chains were 96.5%, 95.1% and 94.6%. After discarding the first 1000 observations of each chain as burn-in the Gelman-Rubin convergence diagnostic $\hat{R}$ was 3.1. The effective Monte Carlo sample size, after discarding the first 1,000 iterations of each chain was $N_{\text{eff}} = 36.0$. Apart from increasing the overall (or nominal) Monte Carlo sample size, What could be done to improve convergence and increase the effective Monte Carlo sample size for this Metropolis-Hastings procedure? (2 marks)

(c) Again with respect to Figure 1, explain why the effective Monte Carlo sample size is so much lower than the the nominal Monte Carlo sample size of $3,000$. (2 marks)

(d) Explain why the effective Monte Carlo sample size for importance sampling is usually less than the nominal Monte Carlo sample size. (2 marks)

(e) Suppose a Metropolis-Hastings algorithm based on an asymmetric jumping density, $J_k(\theta^{(k)}|\theta^{(k-1)})$, has been running for some time and by iteration $k=(t-1)$ produces a draw from the posterior density for the model parameter θ. That is,

$$p(\theta^{(t-1)}=\theta_*)=p(\theta=\theta_*|\mathrm{data}),$$

where θ_* is any point in the parameter space.

Show that this implies that

$$p(\theta^{(t-1)}=\theta_a,\theta^{(t)}=\theta_b)=p(\theta^{(t-1)}=\theta_b,\theta^{(t)}=\theta_a)$$

for all θ_a and θ_b such that

$$p(\theta_b|\mathrm{data})J_t(\theta_a|\theta_b)\geq p(\theta_a|\mathrm{data})J_t(\theta_b|\theta_a).$$

(7 marks)

4. Question 4 (15 marks)

Suppose n people, sampled at random from the general population have been tested for Covid-19 with a new diagnostic test. Let $Y = 1$ denote a positive result and $Y = 0$ a negative result. The true disease status Z is not observed. $Z = 1$ denotes disease and $Z = 0$ denotes no disease. We let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and $\mathbf{Z} = (Z_1, \ldots, Z_n)$. One model for the test results is:

$$\begin{aligned}
Z_i|\phi &\sim \text{Bernoulli}(\phi), \text{ independently for } i = 1, \ldots, n \\
Y_i|Z_i = 1, \gamma &\sim \text{Bernoulli}(\gamma), \text{ independently for } i \text{ such that } Z_i = 1 \\
Y_i|Z_i = 0, \beta &\sim \text{Bernoulli}(1-\beta), \text{ independently for } i \text{ such that } Z_i = 0
\end{aligned}$$

Suppose the weakly informative priors:

$$\begin{aligned}
\phi &\sim \text{Beta}(1, 9) \\
\gamma &\sim \text{Beta}(8, 2) \\
\beta &\sim \text{Beta}(9, 1)
\end{aligned}$$

are adopted and *a priori* independence is assumed, so that $p(\phi, \gamma, \beta) = p(\phi)p(\gamma)p(\beta)$.

(a) Sketch a DAG for this model. (2 marks)

(b) Write down the steps of a Gibbs sampler for this problem. (4 marks)

(c) Derive the full conditional posterior distributions for this problem. You may make use of standard conjugacy results in these derivations.

(9 marks)

End of Examination