

Family Name
First Name
Student Number

UNIVERSITY OF TECHNOLOGY SYDNEY

SCHOOL OF MATHEMATICAL AND PHYSICAL SCIENCES

37457 Advanced Bayesian Methods

Practice Final Examination 2023

Time Allowed: 10 minutes reading time and 2 hours writing time
Number of Questions: 4

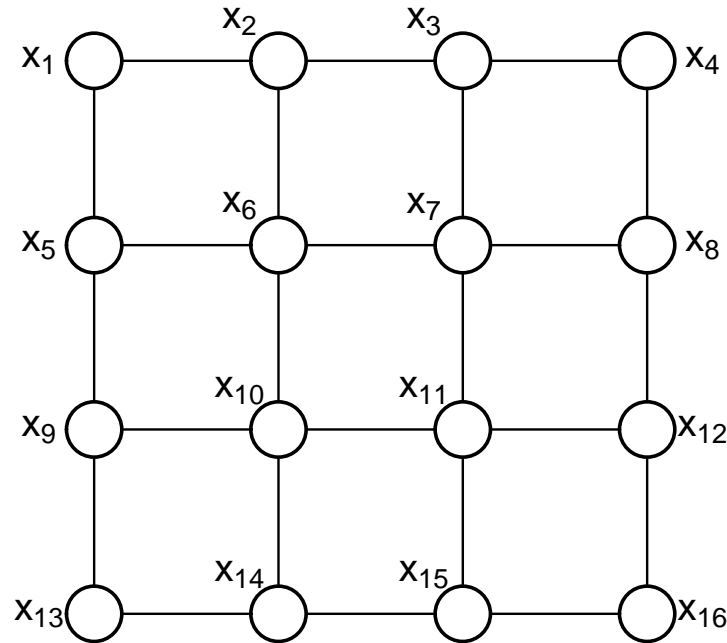
DIRECTIONS TO CANDIDATES

1. Each question is to be attempted.
2. All questions are of equal value.
3. All relevant working should be shown.
4. Calculators are NOT allowed.

EXAMINATION MATERIALS/AIDS ALLOWED

Pens
Ruler

1. (a) In the undirected graph shown immediately below each maximal clique contains exactly the same number of nodes. How many nodes are there in each maximal clique?



- (b) Suppose that the undirected graph shown above is a probabilistic graph with nodes corresponding to the continuous random variables x_1, \dots, x_{16} and joint density function $p(x_1, \dots, x_{16})$. The full conditional density function of x_1 is

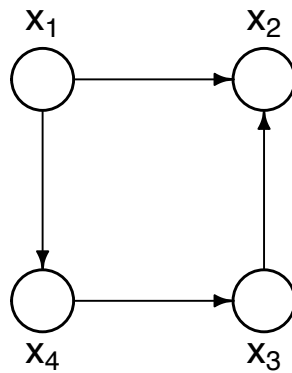
$$p(x_1 | \text{rest}) = p(x_1 | x_2, x_3, \dots, x_{16}). \quad (1)$$

Write down a simplified equivalent form of (1) that conditions on as few of the $\{x_2, x_3, \dots, x_{16}\}$ as possible.

- (c) Find the simplest forms of the full conditional density functions of:
- i. x_8 ,
 - ii. x_{11} ,
 - iii. x_{13} .

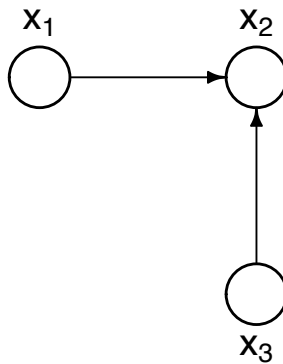
NOTE: This question continues on the next page.

(d) i. Consider the following probabilistic directed acyclic graph:



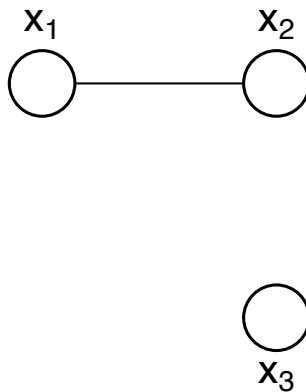
Is it true that $x_2 \perp\!\!\!\perp x_4 | \{x_1, x_3\}$? Justify your answer.

ii. Consider the following probabilistic directed acyclic graph:



Is it true that $x_1 \perp\!\!\!\perp x_3$? Justify your answer.

(e) Consider the following undirected graph:



and suppose that the potentials over the two maximal cliques of the graph are set to be

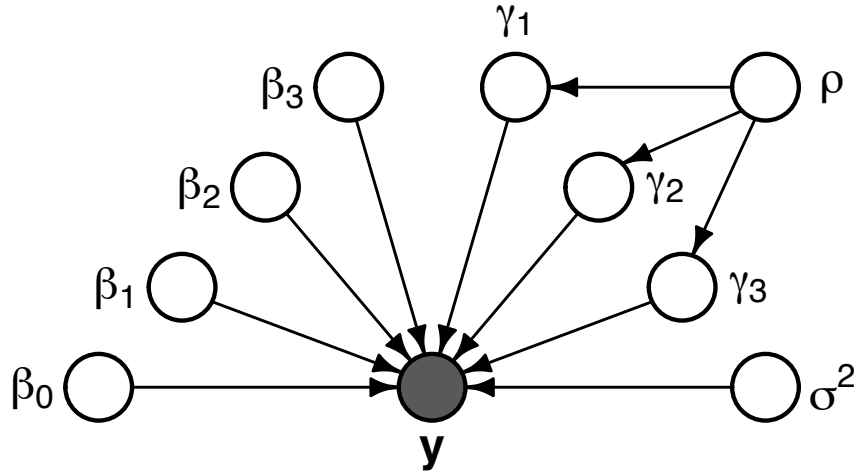
$$\psi_{12}(x_1, x_2) = x_1 + x_2 + 1, x_1, x_2 \in \{0, 1\}$$

and

$$\psi_3(x_3) = 6^{x_3} \quad x_3 \in \{0, 1\}.$$

Determine the joint probability mass function $p(x_1, x_2, x_3)$ corresponding to these potential functions.

2. Consider the following directed acyclic graph:



Hint: Parts of this question may benefit from the **POSSIBLY USEFUL FACTS** sheet on the last page of this examination.

- (a) Write down the Markov blanket of ρ .
- (b) Write down the Markov blanket of σ^2 .

Now consider the Bayesian statistical model based on a probabilistic version of this directed acyclic graph where \mathbf{y} contains the data y_1, \dots, y_n , which is modelled according to:

$$p(\mathbf{y} | \beta_0, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \sigma^2) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-(y_i - \beta_0 - \beta_1\gamma_1 x_{1i} - \beta_2\gamma_2 x_{2i} - \beta_3\gamma_3 x_{3i})^2}{2\sigma^2} \right\} \right]$$

where x_{1i} , x_{2i} and x_{3i} , $1 \leq i \leq n$, are predictor values that are assumed to be fixed (non-random). The prior distributions of the β_j 's are:

$$\beta_0, \beta_1, \beta_2, \beta_3 \text{ independently distributed as } N(0, 10^{10})$$

and the prior distribution of σ^2 is

$$\sigma^2 \sim \text{Inverse-Gamma}(0.01, 0.01).$$

The random variables γ_1, γ_2 and γ_3 are each discrete taking only the values 0 and 1 and with joint probability mass function

$$p(\gamma_1, \gamma_2, \gamma_3 | \rho) = \prod_{j=1}^3 \rho^{\gamma_j} (1 - \rho)^{1-\gamma_j}, \quad \gamma_j = 0, 1.$$

The prior distribution of ρ is

$$\rho \sim \text{Beta}(25, 40).$$

NOTE: This question continues on the next page.

(c) Show that the full conditional distribution of σ^2 is

$$\sigma^2 | \text{rest} \\ \sim \text{Inverse-Gamma} \left(0.01 + \frac{n}{2}, 0.01 + \frac{1}{2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \gamma_1 x_{1i} - \beta_2 \gamma_2 x_{2i} - \beta_3 \gamma_3 x_{3i})^2 \right).$$

All steps should be shown and justified.

(d) Determine the full conditional distribution of ρ .

3. Carla Qing is an entomologist studying the *graphium ambrax* butterfly that first was discovered in a remote Tasmanian gorge in March 2014. Her current research concerns estimation and statistical inference for the mean lifetime, in days, of this particular genus. Her budget allows for 10 butterflies to be monitored and have their lifetimes measured.

Let

$$\mathbf{x} = (x_1, \dots, x_{10})$$

be the sample of lifetimes in days. Carla postulates the following model:

the x_i s are a random sample with density function

$$p(x) = \theta e^{-\theta x}, \quad x > 0,$$

where $1/\theta$ is the mean lifetime. Because of the small sample size and prior knowledge available to Carla regarding lifetimes of similar butterfly populations she decides to adopt a Bayesian approach which involves placing the following prior density function on θ :

$$p(\theta) = \frac{5^{40}}{\Gamma(40)} \theta^{39} e^{-5\theta}, \quad \theta > 0.$$

According to this Bayesian approach, the joint density function of the sample is

$$p(\mathbf{x}|\theta) = \prod_{i=1}^{10} (\theta e^{-\theta x_i}), \quad \theta > 0.$$

Hint: Parts of this question may benefit from the **POSSIBLY USEFUL FACTS** sheet on the last page of this examination.

- Determine the posterior density function of θ in terms of the x_1, \dots, x_{10} and simplify as much as possible.
- Determine the Bayes estimate of θ in terms of the x_1, \dots, x_{10} .
- After collecting her data and typing it into a spreadsheet Carla then uses the statistics package SASQWATCH to obtain the following summary:

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
7.07	8.52	9.89	9.96	11.23	12.37

Suppose that you are in Carla's position. Obtain, in unsimplified form, the numerical value of the Bayes estimate based on the observed data for reporting to colleagues at an upcoming laboratory meeting.

NOTE: This question continues on the next page.

- (d) A few days later, Carla presents her estimate at the laboratory meeting. One of Carla's colleagues, Ashley Finkoff, who took some advanced Statistics subjects whilst at university, says that her analysis would benefit from adding a 90% credible interval for θ . Ashley explains Carla to that the required calculations benefit from a function in SASQWATCH named `quantile_gamma`. Carla then looks up the help page for this function obtains the following:

Description:

```
Quantile function for the Gamma distribution with
parameters 'shape' and 'rate'.
```

Usage:

```
quantile_gamma(probability, shape, rate)
```

Arguments:

```
probability: a number between 0 and 1.
shape: shape parameter. Must be strictly positive.
rate: rate parameter. Must be strictly positive.
```

Details:

```
The Gamma distribution with parameters 'shape' = s
and 'rate' = r has density
      p(x) = r^s / (Gamma(s)) x^(s-1) e^-(x*r)
for x > 0, s > 0 and r > 0.
```

Then Carla looks up "Quantile function" on the Internet encyclopedia Webipedia and discovers that the quantile function Q of a random variable is the inverse of its cumulative distribution F ; that is, $Q = F^{-1}$. The Webipedia page gives the following example for the random variable Z having a $N(0, 1)$ distribution:

Suppose we require c so that $0.95 = P(Z \leq c)$. Then $F(c) = 0.95$ where F is the $N(0, 1)$ cumulative distribution function. It then follows that $c = F^{-1}(0.95) = Q(0.95)$. To three decimal places the solution is $c = 1.645$.

Describe the computations that should Carlo do in SASQWATCH to obtain a 90% credible interval for θ .

- (e) Carla correctly obtains a 90% credible interval for θ of the form (L, U) where L and U are positive numbers such that $L < U$. She shows it to Ashley who tells her that she should also obtain a 90% credible interval for $1/\theta$ since $1/\theta$ is the mean lifetime according to Carla's model. Again suppose that you are in Carla's position. Derive a 90% credible interval for $1/\theta$ in terms of L and U .

4. Leroy Meng is a nutritional health researcher studying the health effects of diet and smoking on lung disease. He recruits 814 subjects and records the following data:

$x_{1i} = 1$ if the i th subject is a smoker and 0 otherwise,

x_{2i} = concentration of dietary saturated fat in the i th subject,

$x_{3i} = 1$ if the i th subject is female and 0 otherwise,

and y_i = number of carcinogen-DNA complexes in the lungs of the i th subject

for $1 \leq i \leq 814$. However, 263 of the subjects did not make available their smoking status and the x_{1i} data has 263 missing values and $814 - 263 = 551$ observed values.

Normally Leroy would fit a non-Bayesian Poisson regression model but he is concerned about the high proportion of missing values and consults with a biostatistician in his unit named Annika Chadwick. Annika recommends use of Bayesian inference engine software to analyse Leroy's data, starting with a simple missing completely at random model for smoking status. The full model is, with $1 \leq i \leq 814$,

$y_i | \beta_0, \beta_1, \beta_2, \beta_3, x_{1i}$ independently distributed as $\text{Poisson} \left(\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}) \right)$,

$\beta_0, \beta_1, \beta_2, \beta_3$ independently distributed as $N(0, 10^{10})$,

$x_{1i} | \rho$ independently distributed as $\text{Bernoulli}(\rho)$, $1 \leq i \leq 814$,

ρ distributed as $\text{Beta}(1, 1)$.

Note that, in this model, the x_{1i} predictor values are treated as random variables whilst the x_{2i} and x_{3i} predictor values are treated as fixed.

Let

\mathbf{y}

be the vector of length 814 containing the y_i values. Also let

$\mathbf{x}_{1,\text{obs}}$

be the vector of length 551 containing the observed smoking status data. The vector of length 263 of the x_{1i} random variables that are not observed is denoted by

$\mathbf{x}_{1,\text{mis}}$.

- (a) Draw a directed acyclic graph representation of Annika's Bayesian model with nodes consisting of the model parameters

$\beta_0, \beta_1, \beta_2, \beta_3, \rho$

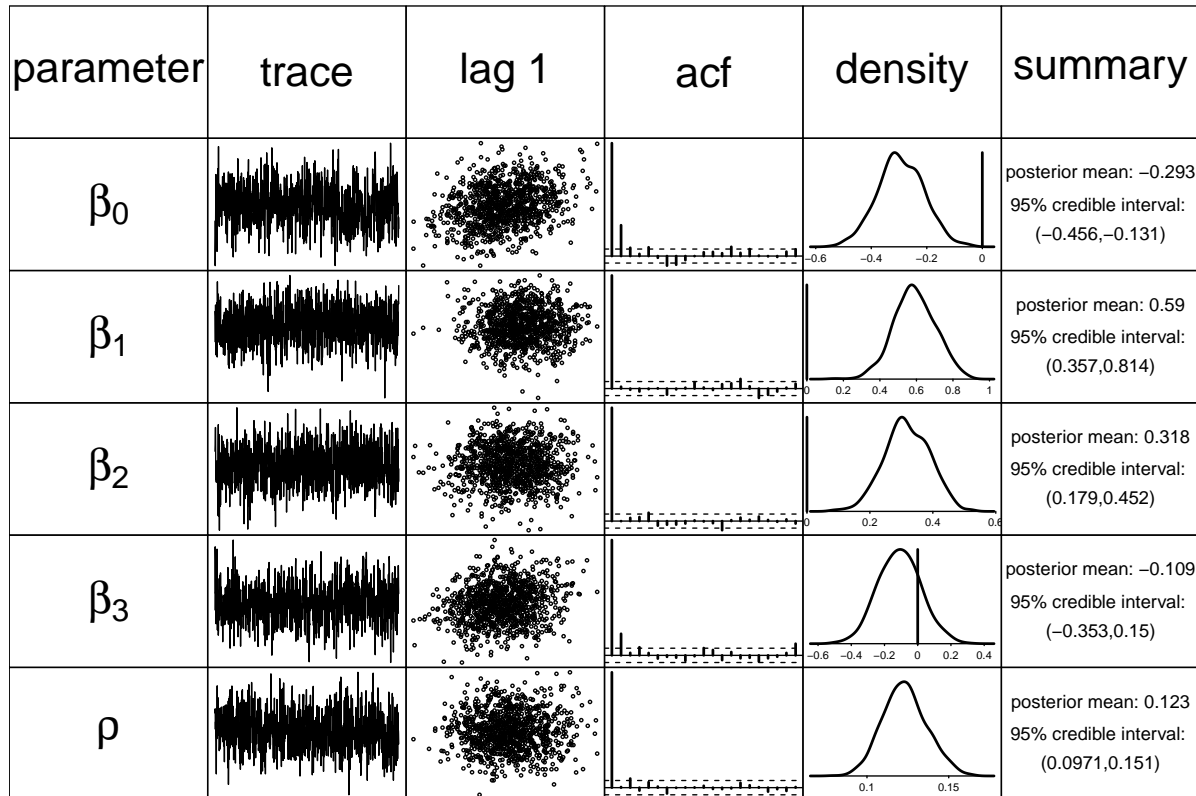
and the vectors

$\mathbf{y}, \mathbf{x}_{1,\text{obs}}, \mathbf{x}_{1,\text{mis}}$.

Use shading to indicate the observed data nodes.

NOTE: This question continues on the next page.

(b) Annika uses a Bayesian inference engine computer package to obtain Markov chain Monte Carlo samples from each of the parameters and unobserved smoking status variables in her model. The following graphic is produced based on the samples:



Suppose that you are in Annika's position. Write a short memorandum to Leroy, with length half a page or less, that summarises the effects of each of the three predictors on mean number of carcinogen-DNA complexes.

NOTE: This question continues on the next page.

- (c) Annika discusses her analysis with a senior epidemiologist in her group named Bryce Sharpe. Bryce expresses his concern about the plausibility of Annika's simple missing completely at random data model since he believes that non-smokers are more likely to respond to a smoking status question compared with non-smokers. He suggests introduction of the data r_1, \dots, r_{814} where

$$r_i = \begin{cases} 1, & \text{if } x_{1i} \text{ is observed,} \\ 0, & \text{if } x_{1i} \text{ is missing} \end{cases}$$

and modelling the r_i according to the logistic regression model:

$$r_i | x_{1i}, \phi_0, \phi_1 \text{ independently distributed as Bernoulli } \left(\frac{e^{\phi_0 + \phi_1 x_{1i}}}{1 + e^{\phi_0 + \phi_1 x_{1i}}} \right),$$

for $1 \leq i \leq 814$ with

$$\phi_0, \phi_1 \text{ independently distributed as } N(0, 10^{10}).$$

Let

$$\mathbf{r} \tag{2}$$

denote the vector of length 814 containing the r_i values.

Annika asks Bryce to draw a directed acyclic graph representation of his model so that she can better see how to implement it in the Bayesian inference engine software package. Suppose that you are in Bryce's position. Draw the appropriate directed acyclic graph with nodes for each of

$$\beta_0, \beta_1, \beta_2, \beta_3, \rho, \phi_0, \phi_1$$

and also nodes for each of the vectors

$$\mathbf{y}, \mathbf{x}_{1,\text{obs}}, \mathbf{x}_{1,\text{mis}}, \mathbf{r}.$$

Use shading to indicate the observed data nodes.

END OF EXAMINATION

This page is devoid of examination material so that the next page containing POSSIBLY USEFUL FACTS can be easily detached.

POSSIBLY USEFUL FACTS

Three distributions are defined by the following table for parameters $A > 0$ and $B > 0$. The mean of each distribution, as a function of A and B , is also given.

notation	density function	mean
Beta(A, B)	$\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} x^{A-1} (1-x)^{B-1}, \quad 0 < x < 1$	$A/(A+B)$
Gamma(A, B)	$\frac{B^A}{\Gamma(A)} x^{A-1} e^{-Bx}, \quad x > 0$	A/B
Inverse-Gamma(A, B)	$\frac{B^A}{\Gamma(A)} x^{-A-1} e^{-B/x}, \quad x > 0$	$B/(A-1)$ provided $A > 1$