STATG004: Applied Bayesian Methods

2 hours

Answer ALL questions. Section A carries 40% of the total marks and Section B carries 60%. The relative weights attached to each question are as follows: A1 (10), A2 (10), A3 (20), B1 (20), B2 (20) and B3 (20). The numbers in square brackets indicate the relative weight attached to each part question.

You may use the following notation and results:

The **Beta distribution**, Beta(α, β), has probability density function

$$p(y \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \ , \quad 0 < y < 1 \ ,$$

where $\Gamma(.)$ is the Gamma function, mean $\frac{\alpha}{\alpha+\beta}$ and, for α, $\beta > 1$, mode $\frac{\alpha-1}{\alpha+\beta-2}$.

The **Gamma distribution**, Gamma(α, β), has probability density function

$$p(y \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \ , \quad y > 0 \ ,$$

and mean $\frac{\alpha}{\beta}$.

The **Normal distribution**, Normal(θ, τ^{-1}), has probability density function

$$p(y \mid \theta, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\left[-\frac{\tau}{2}(y-\theta)^2\right] \ , \quad -\infty < y < \infty \ .$$

The **Poisson distribution**, Poisson(λ), has probability mass function

$$p(y \mid \lambda) = \frac{\lambda^y}{y!} e^{-\lambda} \ , \quad y = 0, 1, 2, \ldots$$

SECTION A

A1 A blood test T_1 to diagnose a disease D has probability of 80% of being correct regardless of the truth, i.e. 80% is the probability of the test being positive when a person truly has the disease, and 80% is also the probability of the test being negative when he/she is normal. If this test is positive, a different test T_2 is then carried out. The second test always correctly diagnoses if the person is in fact normal, but has a probability of 95% of being correct if he/she has the disease. Assume that, given the true disease status (i.e. having the disease or not) of a person, the second test is conditionally independent of the first test for the same person. Suppose that 10% of the people in the population under study is affected by the disease.

(a) A person tests positive on his/her first test. What is the probability that he/she really has the disease? [4]

(b) What is the probability that this person will test positive again on his/her second test, given that he/she tested positive on his/her first test? [6]

A2 Suppose that

$$Y \mid \theta \sim \text{Normal}(\theta, \tau^{-1}),$$

where θ is an unknown parameter and the value of τ is known.

(a) Show that the distribution of $Y \mid \theta$ belongs to the one-parameter exponential family. [5]

(b) Using the results from part (a), derive the conjugate prior for θ, clearly stating the prior mean and the prior precision of θ. [5]

A3 Suppose that the lifetime Y of a type of machine follows a Gamma$(4, \theta)$ distribution, i.e., $Y|\theta \sim \text{Gamma}(4, \theta)$, where θ is an unknown parameter.

(a) Suppose that we use a Gamma(α_0, β_0) prior for θ and that we record a random sample of the lifetimes $(y_1, \ldots, y_n)$ from n machines of this type. Let $\mathbf{y}$ denote the values $(y_1, \ldots, y_n)$ of the sample and $\bar{y}$ denote the sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Derive the posterior distribution for θ, clearly stating the parameters of the posterior. [5]

(b) Suppose that we wish to predict the lifetime (denoted by $\tilde{y}$) of a new machine of the same type. Derive the predictive distribution $p(\tilde{y}|\mathbf{y})$. [6]

(c) Derive the Jeffreys prior for θ. Show that it can be written as a Gamma distribution. [6]

(d) Is this Jeffreys prior proper? Justify your answer. [3]

SECTION B

B1 A study is carried out in which a count response variable is measured on n individuals at each of times $1, 2, \ldots, m$. Let Y_{ij} denote the response of person i at time j. Y_{ij} may take values $0, 1, 2, \ldots$. Consider the following model for these data:

$$\begin{aligned} Y_{ij} \mid \alpha_i, \beta &\sim \text{Poisson}(e^{\alpha_i + \beta(j-1)}) \ , \ i = 1, \ldots, n; \ j = 1, \ldots, m \ , \\ \alpha_i \mid \mu_\alpha, \tau_\alpha &\sim \text{Normal}(\mu_\alpha, \tau_\alpha^{-1}) \ , \\ \beta &\sim \text{Normal}(0, 10^6) \ , \\ \mu_\alpha &\sim \text{Normal}(0, 10^6) \ , \\ \tau_\alpha &\sim \text{Gamma}(0.001, 0.001) \ . \end{aligned}$$

(a) Draw the directed acyclic graph (DAG) for this model. [4]

(b) By moralising the graph, determine whether the following conditional independence statement is true: $Y_{i1} \perp\!\!\!\perp Y_{i2} \mid \alpha_i$. Justify briefly your answer. [5]

(c) Derive the full-conditional distribution of τ_α, clearly stating the parameters of the distribution. [5]

(d) Suppose, using a Gibbs sampler, after a burn-in of M iterations you have obtained samples $(\beta^{(M+1)}, \ldots, \beta^{(N)})$ from the posterior distribution of β. Explain how to use these samples to estimate the posterior mean, median, mode and the central 95% posterior credible interval of β. [6]

B2 (a) Prove that the posterior mode of $\theta|\mathbf{y}$ is an optimal point estimator $T(\mathbf{y})$ for θ that minimises the posterior expected loss under zero-one loss. [6]

(b) Assuming a Beta$(0, 0)$ prior for $\theta \in (0, 1)$, derive a transformation $\phi = \phi(\theta)$ such that a uniform prior for ϕ is assumed. [5]

(c) List the main advantage and three disadvantages of using the Jeffreys prior. [4]

(d) Prove that two random variables X_1 and X_2 are identically distributed if they are exchangeable. [5]

B3 The data given in the table below are the numbers of pump failures, $Y_i = y_i$, observed in t_i thousands of hours for 10 different systems (indexed by $i = 1, \ldots, 10$) in a nuclear power plant:

	System i									
	1	2	3	4	5	6	7	8	9	10
y_i	5	1	5	14	3	19	1	1	4	22
t_i	94.3	15.7	62.9	125.8	5.2	31.4	1.0	1.0	2.1	10.5

You are asked to design an appropriate hierarchical Bayesian model for the inference about system-specific pump failure rates (per thousand hours) θ_i, which are unknown. Assume that $Y_i \mid \theta_i \sim \text{Poisson}(\theta_i t_i)$.

(a) Write down all the probabilistic distributions for your model. [6]

(b) Draw the DAG for your model. [4]

(c) Suppose that the full-conditional distributions of all the unknown random variables in the model have been derived. Explain in detail how to use these distributions to develop a Gibbs sampler to obtain samples $\theta_6^{(M+1)}, \ldots, \theta_6^{(N)}$ from the posterior distribution for θ_6, where M is the number of iterations of burn-in and N is the total number of iterations. [6]

(d) The results from a hierarchical Bayesian model show that, although the posterior means of θ_5 and θ_6 are very similar, the 95% central posterior credible interval for θ_6 is much narrower than that for θ_5. Briefly comment on the reason for this. [4]

End of Paper