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.

The **Binomial distribution**, Binomial(n, θ), has probability mass function

$$p(y \mid \theta) = \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y} \;, \quad y = 0, 1, \ldots, n \;.$$

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

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

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 lie detector test has probability of η of being correct regardless of the truth; that is, η of tests provide positive results when people lie, and η of tests provide negative results when people are honest.

Suppose we believe 90% of people taking the test are actually honest.

(a) A person tests negative on the test. What is the probability, in terms of η, that he actually lies? [5]

(b) A statistician argues that one should use a uniform prior for the log odds $\log \frac{\eta}{1-\eta}$. Derive the resulting prior for η. [5]

A2 (a) Suppose that, for the parameter θ of Binomial(n, θ), the posterior distribution is $p(\theta|\mathbf{y})$. Suggest how to use this posterior distribution to test the hypothesis that $H_0 : \theta < \theta_0$ vs. $H_1 : \theta \geq \theta_0$, when the losses for type I and II errors are different, and the loss for a correct decision is zero. [5]

(b) Prove that a Beta$(0, \frac{1}{4})$ prior for the parameter θ of the Binomial(n, θ) distribution is an improper prior. [5]

A3 Suppose that a random variable $Y|\theta \sim$ Gamma$(\alpha, 2\theta)$, where α is a known constant while θ is an unknown parameter.

(a) Show that the distribution of $Y|\theta$ is a one-parameter exponential family distribution. [4]

(b) Using the results from part (a), derive a conjugate prior for θ. [4]

(c) Suppose with data $\mathbf{y}$ observed for Y, we have obtained the posterior distribution $p(\theta|\mathbf{y})$ as a Gamma(α_1, β_1) distribution. Now we wish to predict the value (denoted by $\tilde{y}$) for a new observation of Y. Derive the predictive distribution $p(\tilde{y}|\mathbf{y})$. [6]

(d) Derive the Jeffreys prior for θ. [6]

SECTION B

B1 (a) If $\theta_1, \theta_2, \ldots$ are exchangeable, write down the general representation theorem which implies a hierarchical model for them. [5]

(b) Prove that two random variables Y_1 and Y_2 are exchangeable if they are independent and identically distributed. Justify each step of your proof. [5]

(c) Suppose we have collected n independent observations Y_i from a Normal(μ, τ^{-1}) distribution with unknown mean μ and unknown precision τ; that is,

$$Y_i \mid \mu, \tau \sim \text{Normal}(\mu, \tau^{-1}) \ , \quad \text{for} \quad i = 1, \ldots, n \ .$$

Two independent priors are assumed for μ and τ:

$$\begin{aligned} \mu &\sim \text{Normal}(0, 10) \ , \\ \tau &\sim \text{Gamma}(0.1, 0.1) \ . \end{aligned}$$

Let $\mathbf{y}$ denote the values $(y_1, \ldots, y_n)$ of the observations. Derive two full-conditional distributions $p(\mu|\tau, \mathbf{y})$ and $p(\tau|\mu, \mathbf{y})$, and write down one of them as a Normal distribution and the other as a Gamma distribution with corresponding distribution parameters. [10]

B2 Suppose we have the following generalised linear mixed model for data Y_1 and Y_2:

$$\begin{aligned} Y_i \mid \theta_i &\sim \text{Poisson}(2\theta_i) \ , \ i = 1, 2 \ , \\ \log \theta_i &= \beta_0 + \beta_1 X_i + \lambda_i \ , \\ \lambda_i \mid \tau &\sim \text{Normal}(0, \tau^{-1}) \ , \\ \beta_0 &\sim \text{Normal}(1, 10) \ , \\ \beta_1 &\sim \text{Normal}(1, 10) \ , \\ \tau &\sim \text{Gamma}(0.1, 0.1) \ , \end{aligned}$$

where X_i are observed covariates.

(a) Draw the directed acyclic graph (DAG) for this model, using dashed arrows for deterministic dependency and solid arrows for stochastic dependency. [3]

(b) By moralising the graph, determine whether the conditional-independence statement $(Y_1 \perp\!\!\!\perp Y_2 \mid \beta_0, \beta_1)$ is true. [6]

(c) Derive the full-conditional distribution for τ. [5]

(d) Describe how you could use a 'batching' method to estimate the Monte Carlo standard error (MCSE) of the posterior mean of θ_2, using a sample $(\theta_2^{(M+1)}, \ldots, \theta_2^{(N)})$ from the posterior distribution of θ_2 obtained by a Gibbs sampler, for some appropriately chosen $M, N > 0$. [6]

B3 Suppose there is a probabilistic model represented by a DAG with random variables $X_1, \ldots, X_K$.

(a) Write down the factorisation theorem for these random variables. [5]

(b) Factorise the full-conditional distribution of X_k (up to a constant of proportionality), by using the elements of the Markov blanket of X_k. [5]

(c) Show how your expression in part (b) can be derived, justifying carefully each step in your derivation. [10]

END OF PAPER