

LTTC: APPLIED BAYESIAN METHODS

Take-home exam

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 (20), A2 (20), B1 (8), B2 (32), B3(10) and B4 (10). 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**, $\text{Beta}(\alpha, \beta)$, 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 $\Gamma(\cdot)$ is the Gamma function.

The **Gamma distribution**, $\text{Gamma}(\alpha, \beta)$, 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**, $\text{Normal}(\theta, \tau^{-1})$, 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 **Negative Binomial distribution**, $\text{NegBin}(r, \theta)$, has probability mass function

$$p(x \mid \theta) = \binom{x+r-1}{x} \theta^r (1 - \theta)^x, \quad x \in \{0, 1, 2, \dots\}.$$

The **Exponential distribution** with rate θ , $\text{Exponential}(\theta)$, has probability density function

$$p(x \mid \theta) = \theta \exp(-\theta x), \quad x > 0, \theta > 0.$$

The **Binomial distribution**, $\text{Binomial}(n, \theta)$, has probability mass function

$$p(y \mid \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad y = 0, 1, \dots, n.$$

TURN OVER

SECTION A

A1 For a new test for pregnancy, θ is the proportion of pregnant women who will test positive and λ is the probability that a non-pregnant woman tests positive. In the population of women who take the test, we assume that 100q% are pregnant.

- (a) A woman decides to use the test to see whether she is pregnant. She tests negative. What is the probability that she is in fact pregnant? [5]
- (b) The same woman, having seen that the first test was negative, decides to repeat the test, in order to provide more reassurance that she is not pregnant. Assuming that repeat tests on the same woman are conditionally independent given her true pregnancy status, show that the probability that the second test will also be negative is

$$\frac{A(1 - \lambda)^2 + (1 - \theta)^2}{(1 - \theta) + A(1 - \lambda)} .$$

What is the value of A ? [5]

- (c) A scientist thought that a $\text{Beta}(\alpha, \beta)$ prior was reasonable for θ . Then the new pregnancy test is used for 20 pregnant women, out of whom 18 test positive (denoted by $y = 18$). Derive the posterior distribution $p(\theta|y)$. Show that it can be written as a Beta distribution and state its parameters. [5]
- (d) Now the scientist wishes to predict the number (denoted by $\tilde{y}$) of positive results out of a new group of 10 pregnant women who are going to take the test. Derive the predictive distribution $p(\tilde{y}|y)$. [5]

A2 Suppose that observations Y_i , $i = 1, \dots, n$, are independent and exponentially distributed with rate θ :

$$Y_i \mid \theta \sim \text{Exponential}(\theta), \quad i = 1, \dots, n.$$

- (a) Derive Jeffreys' prior for θ and show that it can be written as a $\text{Gamma}(0,0)$ distribution. [5]
- (b) Given this Jeffreys' prior, derive a transformation $\phi = g(\theta)$ such that a uniform prior for ϕ is assumed. [5]
- (c) Using the prior from part (a) and the data $\mathbf{y} = (y_1, \dots, y_n)$ for $(Y_1, \dots, Y_n)$, show that the posterior distribution of θ is a $\text{Gamma}(n, n\bar{y})$ distribution, where $\bar{y}$ denotes the sample mean. [5]
- (d) Derive the predictive distribution $p(\tilde{Y} = \tilde{y}|\mathbf{y})$ for a new observation $\tilde{y}$, where $\tilde{Y} \mid \theta \sim \text{Exponential}(\theta)$. [5]

CONTINUED

SECTION B

B1 Considering the following hierarchical model:

$$\begin{aligned} Y_i | \theta_i &\sim \text{NegBin}(r, \theta_i) , \quad i = 1, 2 \\ \theta_i | \alpha, \beta &\sim \text{Beta}(\alpha, \beta) , \quad i = 1, 2 \\ \alpha &\sim \text{Exponential}(0.01) \\ \beta &\sim \text{Exponential}(0.01) \end{aligned}$$

where r is a constant; α, β, θ_1 and θ_2 are unobserved random variables; the observed values of Y_1 and Y_2 are y_1 and y_2 , respectively.

- (a) Draw the DAG for this model. [5]
- (b) Show that the full-conditional distribution of θ_1 is a Beta distribution. [6]
- (c) Suppose that you have worked out the full-conditional distributions of all four unobserved random variables, α, β, θ_1 and θ_2 . Explain in detail how to use these distributions in the Gibbs sampler to obtain a sample from the posterior distribution of θ_1 . [5]

B2 Suppose we use a Bayesian point estimator $t = T(y)$ (with y denoting observed data) to estimate an unknown parameter θ .

- (a) Prove the following equation:

$$E_{\theta|y}(t - \theta)^2 = \text{var}_{\theta|y}(\theta) + \{E_{\theta|y}(\theta) - t\}^2 ,$$

where $t = T(y)$, and $E_{\theta|y}(\cdot)$ and $\text{var}_{\theta|y}(\cdot)$ denote operators for the posterior mean and the posterior variance, respectively. [6]

- (b) Suggest an optimal estimator $T(y)$ for θ that minimises the posterior expected loss, $E_{\theta|y}[L(\theta, T(y))] = \int L(\theta, T(y))p(\theta|y)d\theta$, under squared-error loss, $L(\theta, T(y)) = (\theta - T(y))^2$. Justify your suggestion. [4]

B3 (a) A nationwide survey is carried out, in which 50 voters in each of 20 different English cities are questioned. Let Y_i be the number of sampled voters in city i ($i = 1, \dots, 20$) who intend to vote Conservative, and let θ_i be the true proportion of voters in city i who intend to vote Conservative. Let y_i denote the observed value of Y_i and $\mathbf{y} = (y_1, \dots, y_{20})$.

The following two models are proposed for these data.

- Model M1:

$$\begin{aligned} Y_i | \theta_i &\sim \text{Bin}(50, \theta_i) \quad i = 1, \dots, 20, \\ \log \left(\frac{\theta_i}{1 - \theta_i} \right) &= \gamma_i , \\ \gamma_i &\sim \text{Normal}(0, 10^6) . \end{aligned}$$

TURN OVER

– Model M2:

$$\begin{aligned} Y_i \mid \theta_i &\sim \text{Bin}(50, \theta_i) \quad i = 1, \dots, 20, \\ \log\left(\frac{\theta_i}{1 - \theta_i}\right) &= \gamma_i, \\ \gamma_i &\sim \text{Normal}(\mu_\gamma, \tau^{-1}) \end{aligned}$$

with μ_γ and τ having non-informative priors.

- (i) Explain the difference between models M1 and M2. Which model do you think is more realistic, and why? [6]
- (ii) The two models are fitted using Markov chain Monte Carlo (MCMC) algorithms. The data, y_i , for cities $i = 1, 6$ and 8 are given in the table below. Also given in the table are the posterior means of θ_i , $E[\theta_i \mid \mathbf{y}]$, from the two models, M1 and M2. Which of the two sets of posterior means below denoted as Set A and Set B comes from Model M1 and which comes from Model M2? Justify your answers. [4]

i	y_i	Set A	Set B
		$E[\theta_i \mid \mathbf{y}]$	$E[\theta_i \mid \mathbf{y}]$
1	15	0.32	0.30
6	35	0.54	0.70
8	25	0.44	0.50