LTCC: Applied Bayesian Methods

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

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

$$p(y \mid \theta) = \begin{pmatrix} n \\ y \end{pmatrix} \theta^y (1-\theta)^{n-y} \ , \quad y = 0, 1, \ldots, n \ .$$

The **Gamma distribution**, $\text{Gamma}(\alpha, \beta)$, has probability density function

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

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

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

$$p(y \mid \theta) = \binom{y+r-1}{y} \theta^r (1-\theta)^y \ , \quad y = 0, 1, \ldots, n \ ,$$

and mean $\frac{r(1-\theta)}{\theta}$.

The **Normal distribution**, $\text{Normal}(\theta, \tau^{-1})$, has probability density function

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

1 The Negative Binomial distribution, NegBin(r, θ), describes the distribution of the number of failures (denoted by Y) before the r-th success in an experiment that consists of a sequence of independent and identically distributed Bernoulli trials, where each trial has a probability θ of success. We collect a sample of n independent observations Y_i, $i = 1, \ldots, n$, with $Y_i \mid \theta \sim \text{NegBin}(r, \theta)$, where θ is unknown and the value of r is known.

(a) Show that the Negative Binomial distribution of $Y_i \mid \theta$ belongs to the one-parameter exponential family.

(b) Using the results from part (a), derive the conjugate prior for θ and show that it can be written as a Beta distribution. Propose three non-informative priors expressed as a Beta distribution.

(c) Suppose that a $Beta(a, b)$ prior is considered reasonable for θ. Derive the posterior distribution $p(\theta \mid \mathbf{y})$ for θ, where $\mathbf{y} = (y_1, \ldots, y_n)$ denotes the values of the n observations. Show that the posterior distribution can be written as a Beta distribution.

(d) Suppose that we want to predict the value of a new observation (denoted by $\tilde{y}$). Derive the predictive distribution $p(\tilde{y}|\mathbf{y})$.

2 A statistician wants to model the numbers (denoted by Y_1 and Y_2) of cases of childhood leukaemia in two cities, using the following generalised linear mixed model:

$$\begin{aligned}
Y_i|\theta_i &\sim \text{Poisson}(\theta_i) \ , \ \ i = 1, 2 \ , \\
\log \theta_i &= \beta_0 + \beta_1 X_i + \lambda_i \ , \\
\lambda_i|\tau &\sim \text{Normal}(0, \tau^{-1}) \ , \\
\beta_0 &\sim \text{Normal}(0, 100) \ , \\
\beta_1 &\sim \text{Normal}(0, 100) \ , \\
\tau &\sim \text{Gamma}(0.001, 0.001) \ ,
\end{aligned}$$

where X_i is the value of the covariate for the i^{th} city.

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

(b) Determine whether the conditional-independence statement $(Y_1 \perp\!\!\!\perp Y_2 \mid \beta_0, \beta_1, \lambda_1)$ is true.

(c) Assume that you wish to perform a Gibbs sampler to obtain samples from the joint posterior density of the parameters of the model. Are all the full conditional densities of a known form? Write down the full-conditional distribution for τ.

3 A scientist is interested in the proportion of female horseshoe crabs that have at least one male crab (called satellites) residing nearby. The scientist investigates whether the proportion is affected by two factors: the female crab's colour (denoted by L with $L = 1$ if the colour is light and $L = 0$ otherwise) and the female crab's width (denoted by W with $W = 1$ if the width is larger than 25cm and $W = 0$ otherwise). The data collected by the scientist are arranged in the 12 cells of the table below, according to a 2-by-2 factorial layout, where y_i and n_i are the number of female crabs which have satellites and the total number of female crabs, respectively, in the ith cell, $i = 1, \ldots, 12$, from top to bottom and then left to right.

$W = 0$						$W = 1$					
$L = 0$			$L = 1$			$L = 0$			$L = 1$		
i	y_i	n_i	i	y_i	n_i	i	y_i	n_i	i	y_i	n_i
1	3	10	4	20	42	7	4	11	10	86	109
2	11	39	5	35	69	8	9	28	11	152	196
3	8	31	6	31	65	9	11	32	12	133	158

The scientist is interested in modelling and learning about the probability of having satellites (denoted by θ_i for the ith cell) in each cell. Your task is to design an appropriate random effects logistic regression model for the scientist. The model should use the colour L, width W and their interaction as explanatory variables, and also allow for over-dispersion.

(a) Write down an appropriate Bayesian hierarchical model for these data.

(b) Suppose that, using a Gibbs sampler, you have obtained a sample from the posterior distribution of the parameters of your model. In order to compare the probabilities of having satellites in the first (i=1) and the second (i=2) cell, explain how to use this sample to estimate the posterior distribution of the ratio of the first cell probability against the second cell probability of having satellites.

End of Paper