Bayesian Inference

Tutorial 6

1. Radio-tagging data involves placing a radio-tag on a number of individuals and (assuming no radio failures) recording the number of deaths that occur at a series of successive "capture" times. We assume that only a single radio-tagging event occurs where a total of n lambs are "tagged". We let x_t denote the number of sheep that are subsequently recorded as having died within the interval $(t-1, t]$ (assuming tagging occurs at time 0), for $t = 1, \ldots, T$. We let x_{T+1} denote the number of individuals that survive until time T (i.e. the end of the study). The corresponding likelihood function is a function of the survival probabilities of the sheep. We assume two distinct survival probabilities: ϕ_1 corresponding to first-year survival probability and ϕ_a the "adult" survival probability (i.e. older than first-years). The likelihood is given by,

$$f(\boldsymbol{x}|\phi_1, \phi_a) \propto \prod_{i=1}^{T} p_i^{x_i}$$

where,

$$p_i = \begin{cases} 1-\phi_1 & i = 1 \\ \phi_1(1-\phi_a) & i = 2 \\ \phi_1\phi_a^{i-2}(1-\phi_a) & i = 3, \ldots, T \\ \phi_1\phi_a^{T-1} & i = T+1. \end{cases}$$

Without any prior information on ϕ_1 and ϕ_a we specify $\phi_1 \sim U[0,1]$ and $\phi_a \sim U[0,1]$, independently. Describe how the Gibbs sampler can be used to obtain a sample from the posterior distribution, $\pi(\phi_1, \phi_a|\boldsymbol{x})$. Do you notice anything of interest here? (The answer is obviously yes here - so what is it that is interesting?!).

2. (From May 2007 exam) Let $X_1, \ldots, X_n$ be i.i.d. random variables such that $X_i \sim N(\mu, \omega^{-1})$, and let both μ and ω be unknown. μ is given Jeffreys' prior, and ω is given the prior $\Gamma(\alpha, \beta)$, where α and β are known values (*Note that the prior for μ is of the form $p(\mu) \propto 1$*). We observe data $x_1, \ldots, x_n$, which we denote $\boldsymbol{x}$.

   (a) Write down the joint posterior distribution $\pi(\mu, \omega|\boldsymbol{x})$ up to a proportionality constant. [1]

   (b) Now write down the conditional posterior distributions $\pi(\mu|\omega, \boldsymbol{x})$ and $\pi(\omega|\mu, \boldsymbol{x})$, up to proportionality constants. Both are of standard forms, enabling you to obtain the full conditional posterior distributions. Give the names of these distributions and their parameter values. [4]

   (c) There is an algorithm that can be used to simulate dependent draws from the joint posterior distribution in cases like this, where the conditional posterior distributions are of standard form. What is it called? [1]

   (d) Outline the steps required to obtain draws from $\pi(\mu, \omega|\boldsymbol{x})$ using this algorithm. Note that initial draws may not be from the target distribution, so your outline should contain guidance for determining which draws should be used in making inferences about $\pi(\mu, \omega|\boldsymbol{x})$ and which (if any) should be discarded. [5]

   (e) How should the algorithm be used to obtain estimates of the posterior expectations $\mathbb{E}_\pi(\mu|\boldsymbol{x})$ and $\mathbb{E}_\pi(\omega|\boldsymbol{x})$? [1]

3. Suppose that we wish to use the Metropolis-Hastings algorithm to generate a sample from $f(x) = N(0, \sigma^2)$, and that we use the proposal $q(y|x) = N(ax, \tau^2)$ for $-1 < a < 1$.

   (a) What is the corresponding acceptance probability $\alpha(x, y)$?

   (b) For what value of τ^2 would this particular sampler never reject the candidate value?

   (c) What happens when $a = 0$?

4. Show that the Metropolis-Hastings algorithm generates a reversible Markov chain, i.e.,

$$\pi(x)\mathcal{K}(x, y) = \pi(y)\mathcal{K}(y, x),$$

   and show that $\pi(x)$ is the stationary distribution of the algorithm i.e.,

$$\int \pi(x)\mathcal{K}(x, y)dx = \pi(y).$$

   Note: Remember that for the M-H algorithm, $\mathcal{K}(x, y) = q(y|x)\alpha(x, y)$, where $\mathcal{K}(x, y)$ corresponds to the event that we sample y from a current state x.

5. Suppose that we observe data $\boldsymbol{x} = (x_1, ..., x_N)$, such that each $x_i \stackrel{iid}{\sim} Pareto(\alpha, \beta)$, where,

$$\alpha \sim Exp(m); \quad \beta \sim Exp(n),$$

   for known m and n. The $Pareto(\alpha, \beta)$ distribution has density

$$p(x) = \beta\alpha^{\beta}x^{-(\beta+1)},$$

   where $0 < \alpha \leq x$ and $\beta > 0$. Describe how we can obtain a sample from the posterior distribution of the parameters, given the data.