MAST90104: Introduction to Statistical Learning

Week 12 Lab and Workshop

1 Lab

1. In practice, in the hatching eggs problem rather than using one observation x of the number of eggs in a region, there are likely to be m, say $\mathbf{x} = (x_1, \cdots, x_k)$ in k different parts of the one region. There are now different $N_i, i = 1, \cdots, k$ for the k different parts of the region and these random variables are assumed independent. Also rather than simulating from the full joint density, it is desired to simulate from the posterior distribution for $p, \mathbf{N} = (N_1, \cdots N_k)$ given the data $\mathbf{x}$.

   (a) What are the conditional densities than now need to be simulated at each step?

   (b) Alter the code for the sampler from lectures

   ```
   gibbs.f2 = function(x0, p0, N0, m, J){
   # Genearates m samples of (x,p,N) values by the Gibbs sampler.
   # In total J+m samples are generated but the first J are discarded.
   # (x0,p0,N0) is the initial value.
   x.seq <- p.seq <- N.seq <- rep(-1, J+m+1)
   x.seq[1] <- x0; p.seq[1] <- p0; N.seq[1] <- N0
   for(j in 2:(J+m+1)) {x.seq[j] <- rbinom(1, N.seq[j-1], p.seq[j-1])
   p.seq[j] <- rbeta(1, (x.seq[j] + 2), (N.seq[j-1] - x.seq[j] + 4))
   N.seq[j] <- rpois(1, 16 * (1 - p.seq[j])) + x.seq[j]}
   result <- list(X=x.seq[(J+2):(J+m+1)],p=p.seq[(J+2):(J+m+1)],
   N=N.seq[(J+2):(J+m+1)])
   result
   }
   ```

   to simulate the posterior distribution for p, N given the data $x_1, \cdots, x_m$.

   (c) Implement your code for data $x_1, = 5, x_2 = 4, x_3 = 6$ with a variety of burn in values and sample sizes after the burn in. Comment on the results using similar plots to lectures.

   (d) Implement your code with 3 different x-values, say $x_1 = 1, x_2 = 0, x_3 = 1$ and $x_1 = 20, x_2 = 25, x_3 = 15$. How does this effect the answer to the previous question and the resulting plots. (Note that with $x_1 = 20, x_2 = 25, x_3 = 15$ you may need to alter the prior distribution for $\mathbf{N}$ because the joint density is only positive if $N. > x..$)

2. Suppose that $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$.

   (a) What is the conditional distribution of $X_1 | X_2 = x_2$?

   (b) Write an R function that uses the Gibbs sampler to generate a sample of size $n = 1000$ from the $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}\right)$ distribution.
   Plot traces of X_1 and X_2.

   (c) Use your simulator to estimate $\mathbb{P}(X_1 \geq 0, X_2 \geq 0)$. To get a feel for the convergence rate, calculate the estimate using samples $\{1, \ldots, k\}$, for $k = 1, \ldots, n$, and then plot the estimates against n.

   (d) Now change $\boldsymbol{\Sigma}$ to $\begin{pmatrix} 4 & 2.8 \\ 2.8 & 4 \end{pmatrix}$ and generate another sample of size 1000.
   What do the traces/estimates look like now?

2 Workshop

3. Suppose $g(u)$ is a decreasing function of u. Let a random number X be generated by the following algorithm:

   1° Generate U from Unif$(0, 1)$ and V from Unif$(0, 1)$ independently.

   2° If $U + V < 1$, then deliver $X = g(U)$; otherwise, go to 1°.

   Show that the distribution function of X is given by

   $$F(x) = P(X \leq x) = (1 - g^{-1}(x))^2, \quad g(1) \leq x \leq g(0).$$

4. Consider a random sample $X_1, \cdots, X_n$ satisfying $X_i \sim$ pois(θ). We assume a gamma prior pdf $\theta \sim$ gamma(β, κ) with known β and κ, i.e. $p(\theta) = \dfrac{1}{\beta^\kappa \Gamma(\kappa)} \theta^{\kappa-1} e^{-\theta/\beta}$; $\theta > 0$. Find the posterior distribution of θ and its mean.

5. The *precision* of a normal distribution is the reciprocal of its variance. Show that the gamma distribution is conjugate for the normal distribution with a known mean and unknown precision. Identify the posterior parameters in terms of the data and the prior parameters.

6. The *normal-gamma* family of distributions for a random vector (M, R) has a marginal density for R which is gamma and the conditional density of M given $R = r$ as normal fixed mean and precision proportional to r. Show that the *normal-gamma* family is conjugate for normal data with unknown mean and precision. Identify the posterior parameters in terms of the data and the prior parameters. Interestingly, the marginal distribution the normal-gamma family is a shifted and scaled t-distribution whose degrees of freedom depend on the shape parameter of the gamma prior and the sample size, so the t-distribution is relevant to Bayesian as well as frequentist inference - but this is not part of this exercise.

7. Let $X_1, \cdots, X_n$ be a random sample from an exponential distribution with pdf $f(x|\theta) = \theta e^{-\theta x}$, $x > 0$. (So the mean of X_i is $\frac{1}{\theta}$ instead of θ.) Assume the prior pdf of θ is $p(\theta) = \beta e^{-\beta\theta}$ where β is known. Find the posterior distribution for θ and its mean.