TMA4300 Computer Intensive Statistical Methods

Exercise 2, Spring 2012

Note: The solution of problems A and B should be handed in no later than March 23rd 2012.

In this exercise you can use the built-in random number functions in R to generate realisations from standard univariate distributions (like `runif`, `rnorm`, `rbeta` and `rgamma`) or you can use the corresponding functions you coded in Exercise 1. However, remember that there are two common parameterisations for the gamma distribution, so if you use the function `rgamma` you must first check what parameterisation that is used by that function.

Problem A: MCMC for a change point problem

In this problem we consider the following model. We have observations $x_1, \dots, x_n$ of (ordered) time points for coal mine disaster events between 1851 and 1963. Download the file `coalmine.txt` and load and plot the data:

```
coal = read.table("coalmine.txt")$V1
plot(coal, 1:length(coal))
```

This displays the accumulated number of events as a function of time. It appears that the relative frequency of events changed at least once during the observed 112 years.

We assume that the number of events in an interval $(a, b]$ is Poisson distributed with expectation $\int_a^b \lambda(t)dt$, where $\lambda(t)$ is a piecewise constant intensity function, with independence between disjoint intervals. This is a simple inhomogeneous Poisson process. Denote the breakpoints between the constant intensities T_k, with $1851 = T_0 < T_1 < \cdots < T_N = 1963$, and let

$$\lambda(t) = \begin{cases} \lambda_1, & T_0 \le t < T_1 \\ \lambda_k, & T_{k-1} \le t < T_k,\ k = 1, \dots, N-1 \\ \lambda_N, & T_{N-1} \le t \le T_N \end{cases}$$

Let $y_k(T.)$ denote the number of events in interval (T_{k-1}, T_k). By subdividing the observation period into short intervals, one can derive the data likelihood as

$$f(x|T., \lambda.) = \exp\left(-\int_{T_0}^{T_N} \lambda(t)dt\right) \prod_{k=1}^{N} \lambda_k^{y_k} = \exp\left(-\sum_{k=1}^{N} \lambda_k (T_k - T_{k-1})\right) \prod_{k=1}^{N} \lambda_k^{y_k(T.)}$$

Assume a Gamma prior distribution for each λ_k, and a uniform prior for the change points, all independent.

In this exercise, you only have to consider the case $N = 2$, but extending the algorithms to general (but fixed) N is straighforward. With $N = 2$, the unknown quantities are $\theta = (T_1, \lambda_1, \lambda_2)$.

1. Write an expression for the joint posterior density (unnormalised) for θ given x, $\pi(\theta) \propto p(\theta) f(x|\theta)$.

2. Calculate the full conditional densities for λ_1 and λ_2, showing that they are independent and Gamma distributed conditionally on T_1.

3. Implement an MCMC algorithm for $\pi(\theta)$ using Gibbs sampling for λ_1 and λ_2, and a simple random walk Metropolis step for T_1 (i.e. propose a value for $T_1^{(n+1)}$ from $T_1^{new} \sim \text{Unif}(T_1^{(n)} - w, T_1^{(n)} + w)$ and accept with probability $\alpha(\theta, \theta^{new}) = \min(1, \pi(\theta^{new})/\pi(\theta))$.

   Practical implementation note: It is rarely necessary or desired to calculate the actual density values; whenever possible, use the logarithms. This avoids practical problems with unnormalised densities. For example, accept moves when $\log U \le \log \pi(\theta^{new}) - \log \pi(\theta)$. A particularly useful function is `lgamma`, that can be used to calculate $\log(k!) = \log \Gamma(k+1)$. Convince yourself that the $\min(1, \cdot)$-part is handled correctly when using this method.

4. Run the algorithm for different values of the hyper-parameters in the prior distribution. Estimate the marginal posterior distributions $\pi(T_1|x)$, $\pi(\lambda_1|x)$, and $\pi(\lambda_2|x)$ by making histograms of the simulated values. Remember to discard an appropriate burn-in period! Also use simulated values to estimate $\mathrm{E}(T_1|x)$, $\mathrm{E}(\lambda_1|x)$, $\mathrm{E}(\lambda_2|x)$, and $\mathrm{Corr}(\lambda_1, \lambda_2|x)$. Repeat the simulation experiment for different values of the hyper-parameters and study how these values influence the results. Can you intuitively understand what you observe?

5. Run the algorithm for different values of the tuning parameter w and observe how this influences the length of the burn-in period and mixing properties of the Markov chain, but do not influence the limiting distribution.

6. Implement a block random walk proposal Metropolis–Hastings algorithm for $\pi(\theta)$. Here, each iteration consists of first proposing a new value T_1^{new} for T_1 using the same method as before, and then also proposing new values for λ_1 and λ_2 from their full conditional distributions given T_1^{new}, i.e. from $\pi(\lambda_1|T_1^{new})$ and $\pi(\lambda_2|T_1^{new})$. Then, do a joint accept/reject step, accepting the new values with probability

$$\alpha(\theta, \theta^{new}) = \min\left(1, \frac{\pi(\theta^{new})q(\theta^{new}, \theta)}{\pi(\theta)q(\theta, \theta^{new})}\right)$$

where $q(\theta, \cdot)$ is the density of the joint proposal distribution.

7. Run the new algorithm for different values of the tuning parameter w and check that the generated posteriors match your previous results. Compare the burn-in and mixing properties of the two algorithms. For each w, also compare the overall acceptance probabilities for the two methods.

Problem B: Korsbetningen

"I Herrens år 1361, tredje dagen efter S:t Jacob, föll utanför Visbys portar gutarna i danskarnas händer. Här är de begravda. Bed för dem."

"In the year of our Lord 1361, on the third day after S:t Jacob, the Goth fell outside the gates of Visby at the hands of the Danish. They are buried here. Pray for them."

In 1361 the Danish king Valdemar Atterdag conquered Gotland[1] and captured the rich Hanseatic town of Visby. The conquest was followed by a plunder of Visby ("brandskattning"). Most of the defenders were killed in the attack and are buried in a field outside of the walls of Visby. In the 1920s the gravesite was subject to several archeological excavations. A total of 493 femurs[2] (256 left, and 237 right) were found. We want to figure out how many persons were likely buried at the gravesite. It must reasonably have been at least 256, but how many more?

To build a simple model for this problem, we assume that the number of left (x_1) and right (x_2) femurs are two independent observations from a $\mathrm{Bin}(N, \phi)$ distribution. Here N is the total number of people buried and ϕ is the probability of finding a femur, left or right. The unkown parameter vector is $\theta = (N, \phi)$. Assume a $\mathrm{Beta}(a, b)$ prior for ϕ, and a $\mathrm{Unif}(256, 2500)$ prior for N.

[1] Strategically located in the middle of the Baltic sea, Gotland had shifting periods of being partly self-governed, and in partial control by the Hanseatic trading alliance, Sweden, Denmark, and the Denmark-Norway-Sweden union, until settling as part of Sweden in 1645. Gotland has an abundance of archeological treasures, with coins dating back to Viking era trade routes via Russia to the Arab Caliphates.

[2] lårben (sv), femoral (no)

1. Write an expression for the joint posterior density (unnormalised) for θ given x_1 and x_2, $\pi(\theta) \propto p(\theta) f(x_1, x_2|\theta)$.

2. Calculate the full conditional density for ϕ, showing that the Beta prior is a conditionally conjugate distribution.

3. Implement an MCMC algorithm for $\pi(\theta)$ using Gibbs sampling for ϕ, and a simple random walk Metropolis step for N (i.e. propose a value for $N^{(n+1)}$ from $N^{new} \sim \text{Unif}(N^{(n)} - w, N^{(n)} + w)$ and accept with probability $\alpha(\theta, \theta^{new}) = \min(1, \pi(\theta^{new})/\pi(\theta))$.

4. Run the algorithm for different values of the hyper-parameters in the prior distribution. Plot the simulated values of $\phi^{(n)}$ against $N^{(n)}$, and estimate $\text{Corr}(N, \phi|x_1, x_2)$. Repeat the simulation experiment for different values of the Beta hyper-parameters and study how these values influence the results.

5. Implement a block random walk proposal Metropolis–Hastings algorithm for $\pi(\theta)$. Here, each iteration consists of first proposing a new value N^{new} for N using the same method as before, and then also proposing a new value for ϕ from its full conditional distribution given N^{new}, i.e. from $\pi(\phi|N^{new})$. Then, do a joint accept/reject step, accepting the new values with probability

$$\alpha(\theta, \theta^{new}) = \min\left(1, \frac{\pi(\theta^{new}) q(\theta^{new}, \theta)}{\pi(\theta) q(\theta, \theta^{new})}\right)$$

   where $q(\theta, \cdot)$ is the density of the joint proposal distribution.

6. Run the new algorithm for different values of the tuning parameter w. Compare the burn-in and mixing properties of the two algorithms. For each w, also compare the overall acceptance probabilities for the two methods.