Biostatistics 778: Advanced Statistical Computing

Homework 3

Due date: 2013-12-20

Problems

1. *Rejection sampling* (5 pts). Let $Y_i \sim$ Exponential(β) for $i = 1, \ldots, n$ (with mean $1/\beta$) and let β have a half-Normal (σ) prior distribution, i.e. the prior density of β is

$$\pi(\beta \mid \sigma) = \sqrt{\frac{2}{\pi\sigma^2}} \exp(-\beta^2/2\sigma^2)$$

   for $\beta > 0$.

   Write a function named `postsample` which takes an input vector `y`, a sample size `N`, and a value for the parameter σ and uses rejection sampling to simulate a sample of size `N` from the posterior distribution of $\beta \mid y_1, \ldots, y_n$. Your function should have the following prototype:

```
postsample <- function(y, N, sigma) {
        ## Your code here
        ## Return a numeric vector containing the simulated samples
}
```

   Specifically, produce a sample of size 1,000 for $\sigma = 0.5$ and the following ys:

```
20.100306  2.272066  3.796734  2.265275  3.480183
```

2. *Importance sampling* [OPTIONAL] (5 pts). Write a function named `postmean` which takes as input

   - a posterior sample from the distribution of $\beta \mid y_1, \ldots y_n$,
   - a lower bound for σ, and
   - an upper bound for σ,

   and uses importance sampling/reweighting to plot the posterior mean of β as a function of σ. You should use the following prototype for your function:

```
postmean <- function(beta, lower, upper) {
        ## Your code here
        ## Generate a plot of the beta vs. sigma
}
```

3. *Markov chain Monte Carlo* (10 pts). Hierarchical models are sometimes used to "combine evidence" across multiple studies or study units. For example, when studying air pollution and health, it is common to estimate the association between daily changes in air pollution

and some health outcome for many cities and then combine the estimates across cities via an hierarchical model. The model can be written as

$$\begin{aligned} Y_{it} &\sim \text{Poisson}(\mu_{it})] \\ \log \mu_{it} &= \alpha_i + \beta_i x_{it} \\ \beta_i &\sim \mathcal{N}(\mu, \tau^2) \\ \alpha_i &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

where i indexes the city and t indexes time. The prior distributions are

$$\begin{aligned} \mu &\sim \mathcal{N}(0, A) \\ \tau &\sim IG(a, b) \\ \sigma &\sim IG(c, d) \end{aligned}$$

Use the data from the Courseplus web site to generate samples from the posterior distribution of μ and τ using a hybrid Gibbs sampler. Write a function called `postpollution` that has the form

```
postpollution <- function(y, x, g, N = 1000, burn = 1000) {
        ## Return a list with elements 'mu' and 'tau' which are
        ## vectors containing N posterior samples after
        ## 'burn' burn-in
}
```

Put all of your functions in an R package and make sure the package passes `R CMD check`.