TMA4300 Computer Intensive Statistical Methods
Interactive lecture 4, Spring 2018

Problem A: MCMC for a hierarchical Bayesian model

In this problem we will consider a hierarchical Bayesian model for analysing observed life times for a specific type of electronic components. Assume we have available such components from N different producers. We number the producers from 1 to N. For producer number $i \in \{1, \ldots, N\}$ we test n_i components, and number these components from 1 to n_i. For component number j from producer number i we observe how long time x_{ij} this component works before it fails. We assume x_{ij} to be exponentially distributed with intensity λ_i, i.e.

$$x_{ij}|\lambda_i \sim \text{Exponential}(\lambda_i).$$

The intensity λ_i is thus characterising the quality of the components produced by producer number i. We assume the various x_{ij}'s to be conditionally independent given the intensities $\lambda_1, \ldots, \lambda_N$.

Apriori we assume the λ_i's to be conditionally independent and identically distributed given two hyper-parameters α and β, and we assume

$$\lambda_i \sim \text{Gamma}(\alpha, \beta).$$

As the last step in the hierarchical model we assume α and β to be apriori independent, β to be inverse gamma distributed,

$$\beta \sim \text{InvGamma}(a, b),$$

where a and b are treated as fixed constants, and α is assumed to have an (improper) uniform distribution on $[0, \infty)$.

The gamma distribution Gamma(α, β) has density

$$p(x|\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{for } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

and the inverse gamma distribution InvGamma(a, b) has density

$$p(x|a,b) = \begin{cases} \frac{1}{b^a \Gamma(a)} \frac{e^{-\frac{1}{xb}}}{x^{a+1}} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

1. Write down an expression for the posterior distribution $p(\lambda_1, \ldots, \lambda_N, \alpha, \beta | x)$, where $x = (x_{ij}; j = 1, \ldots, n_i, i = 1, \ldots, N)$ is the observed failure times. It is sufficient to find an expression that is proportional to the posterior distribution.

   Derive the full conditional distributions for each component in the parameter vector $(\lambda_1, \ldots, \lambda_N, \alpha, \beta)$. If possible specify what parametric family each full conditional belongs to, and their parameter values.

2. Use pseudo code to outline how you would generate samples from the posterior distribution using Markov chain Monte Carlo (MCMC). Specify in particular what your proposal distributions are and simplify as much as possible the expressions for the corresponding acceptance probabilities.

Assume you have run the MCMC algorithm for M iterations and denote the generated states by $\{(\lambda_1^m, \ldots, \lambda_N^m, \alpha^m, \beta^m)\}_{m=0}^M$. In particular $(\lambda_1^0, \ldots, \lambda_N^0, \alpha^0, \beta^0)$ is the initial state. It is of interest to use the MCMC output to estimate the following three quantities.

- The posterior mean of λ_i, i.e.

$$\mathrm{E}[\lambda_i | x].$$

- The posterior probability that the quality of the components from producer number i is better than the quality of the components from producer number j, i.e.

$$P(\lambda_i < \lambda_j | x).$$

- The posterior probability that a new component from producer number i will work at least until a given time t, i.e.

$$P(x_{i,\text{new}} > t | x),$$

where x is the observed failure times and $x_{i,\text{new}}$ is the failure time for a new component from producer number i.

3. Specify how you would estimate each of the three quantities above based on the output from the MCMC algorithm.

4. Specify how you would estimate 95% credible intervals for λ_i and for $x_{i,\text{new}}$ from the output of the MCMC algorithm.

5. If you have time: Implement and run the Markov chain when we have observations from $N = 4$ producers with $n_1 = 78$, $n_2 = 22$, $n_3 = 45$ and $n_4 = 4$ and the observed life times for products from the different producers can be found in the following files:

   - 1: `https://www.math.ntnu.no/emner/TMA4300/2018v/interactive/data1.txt`
   - 2: `https://www.math.ntnu.no/emner/TMA4300/2018v/interactive/data2.txt`
   - 3: `https://www.math.ntnu.no/emner/TMA4300/2018v/interactive/data3.txt`
   - 4: `https://www.math.ntnu.no/emner/TMA4300/2018v/interactive/data4.txt`

   Choose for example $a = 9/4$ and $b = 4/5$ so that apriori $\mathrm{E}[\beta] = 1$ and $\mathrm{SD}[\beta] = 2$. Use the simulation output to estimate the quantities discussed above.