Bayesian Analysis of Server Uptime Across Different Companies

Fabio Rizzi
s308770@studenti.polito.it

Abdirashid Chorshanbiyev
s314800@studenti.polito.it

Nicholas von Butovitsch
s321806@studenti.polito.it

Mirkomil Egamberdiev
s315875@studenti.polito.it

February 28, 2024

1 Problem statement

These servers are in use at 10 different companies with a different businesses. The variables u_{ij} represent the i-th uptime of a server belonging to the j-th company. We assume the uptimes of the servers belonging to company $j \in 1, \ldots, 10$ are independent exponential random variables with parameters λ_j. The parameters λ_j are different from company to company, but we assume they are independent realizations from a common Gamma population with parameters (a, b) to be estimated. You can fix as priors the way you prefer, for example to the following distributions.

$$\pi(a) \sim \text{Gamma}(\alpha_a, \beta_a)$$

$$\pi(b) \sim \text{Gamma}(\alpha_b, \beta_b)$$

Please choose the parameters of the prior so to make them little informative, except for one of the companies that has studied the topic already in the past, finding that for their servers the mean uptime was about 100 days, with a standard deviation of 30 days (approximately).

- Simulate a synthetic dataset which follows the model above. The dataset should contain 50 measurements for each of the 10 companies. The parameters can be chosen by yourself.
- Write down the explicit expression of the posterior distribution (adapting the results explained in class to the Gaussian case).
- Calculate the full conditionals.
- Write down a code to perform Bayesian inference either with a simple Gibbs approach, if possible, or with some Metropolis within Gibbs step if necessary.
- Report the results of your analysis.

2 Method

2.1 Simulate the synthetic dataset

In order to simulate a synthetic dataset following the requirements of 50 measurements for each company, it requires us to also choose certain parameters ourselves. Specifically the initial value of a, b, α_a, β_a, α_b and β_b. This is often done as the finding strong prior information is helpful in improving precision and perhaps decreasing the time to convergence. The mean up-time as defined by the problem is 100 days. Therefore, the initial values follow a relatively simple relationship in obtaining the constraint on the choice of parameters. Along with the knowledge that both parameters follow a gamma distribution, thereby the mean of the two parameters will be $\frac{a}{b}$. In this case the parameters $a = 10$ and $b = 0.1$ were chosen as they satisfied the constraint. The choice of α_a, β_a, α_b and β_b also depends on the mean as they also follow a gamma distribution, but as none is given to us these values were chosen somewhat arbitrarily.

2.2 Posterior Distribution

Before beginning to solve this problem, it is vital to first observe the hierarchical nature of the situation. In its simplified form; the problem can be visualized as such:

$$\text{data}(u_{ij})\leftarrow\lambda_j\leftarrow\text{the priors for } \lambda_j \text{ (e.g. } a,b) \leftarrow \text{hyper-priors for } a,b$$

Therefore, in order to build the desired model, it is necessary to not only estimate the λ parameters but also estimate a and b. Although these variables are not given as such, we believe that they follow some sort of distribution. This means that we need to sample λ, a, b from their respective distributions.
The posterior distribution describes the new updated degree of belief with respect to our parameters θ, based on the data and is ultimately the distribution which is desired. This posterior distribution can be explicitly constructed. The priors of parameters a and b are directly given to us as samples from the gamma distribution as noted in the problem, rewritten here again for reference:
$\pi(a) \sim \text{Gamma}(\alpha_a, \beta_a)$
$\pi(b) \sim \text{Gamma}(\alpha_b, \beta_b)$
The likelihood of the up-time of the i-th server belonging to the j-th company, given the company j is as follows: $\mathcal{L}(\{u_{ij}\}|\{\lambda_j\})$ As λ_j is dependent on a and b, the prior of lambda is then given as: $\pi(\{\lambda\}|a, b)$ Finally, with the trivial knowledge of the remaining priors, this gives us the entire posterior as:

$$\pi(\{\mu_i j\}, \{\lambda_j\}, a, b) = \pi(\{\lambda_j\}, a, b|\{\mu_i j\}) = \mathcal{L}(\{u_{ij}\}|\{\lambda_j\})\pi(\{\lambda\}|a, b)\pi(a)\pi(b)$$

where:
$\mathcal{L}(\{u_{ij}\}|\{\lambda_j\}) = \Pi_i^{50}\Pi_j^{10}\lambda_j e^{-\lambda_j\mu_{ij}}$
$\pi(\{\lambda\}|a, b) = \Pi_{k=1}^{n}\frac{b^a}{\Gamma(a)}\lambda_j^{a-1}e^{-b\lambda}$
$\pi(a) = \frac{\beta_a^{\alpha_a}}{\Gamma(\alpha_a)}a^{\alpha_a-1}e^{-\beta_a a}$

$\pi(b) = \frac{\beta_b^{\alpha_b}}{\Gamma(\alpha_b)} b^{\alpha_b - 1} e^{-\beta_b b}$
The three priors are directly derived assuming independence of the different companies along with the PDF of the gamma distribution which is given as:
$f(x; \alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$
While the likelihood of the data given the parameters can be derived knowing that the λ parameters share an exponential distribution. The probability density function (PDF), which represents the probability of observing an up-time u_{ij} given the parameter λ_j, is directly applicable to this problem. Along with an assumption of independence of different up-time observations and companies the likelihood can thus be derived.

2.3 Full Conditional

In order to draw samples from the joint posterior distribution it requires the computation of the full conditionals for all the parameters where samples should be retrieved from. Specifically, we attempt to take the full conditionals of λ_j, a, and b.
However, because of:
$\pi(a|b, \{u_{ij}\}, \{\lambda_j\}) \propto \left(\prod_{j=1}^{10} \frac{b^a}{\Gamma(a)} \lambda_j^{a-1} e^{-b\lambda_j}\right) \times \frac{\beta_a^{\alpha_a}}{\Gamma(\alpha_a)} a^{\alpha_a - 1} e^{-\beta_a a}$
can be further simplified to:
$\pi(a|b, \{u_{ij}\}, \{\lambda_j\}) \propto \left(\prod_{j=1}^{10} \frac{b^a}{\Gamma(a)} \lambda_j^{a-1}\right) \times a^{\alpha_a - 1} e^{-\beta_a a}$
As normalizing constants from all terms can be removed.
$\pi(a|b, \{u_{ij}\}, \{\lambda_j\}) \propto \left(\frac{b^{10a}}{\Gamma(a)^{10}} (\prod_{j=1}^{10} \lambda_j)^{a-1}\right) \times a^{\alpha_a - 1} e^{-\beta_a a}$
As seen from these calculations, the full conditional takes no basic distribution form and hence it is difficult to sample from and thus we must sample with the Metropolis-Hastings method.
It is, in contrast to the previous parameter, possible to directly sample for b and the λ parameters since their full conditionals can be derived and thus Gibbs sampling can be utilized. The full conditional of λ and b are demonstrated below.
$\pi(\{\lambda_j\}|a, b, \{u_{ij}\}, \{\lambda_j\}) \propto \left(\Pi_i^{50} \Pi_j^{10} \lambda_j e^{-\lambda_j \mu_{ij}}\right) \times \left(\prod_{j=1}^{10} \frac{b^a}{\Gamma(a)} \lambda_j^{a-1} e^{-b\lambda}\right)$
We can simplify this to:
$\pi(\{\lambda_j\}|a, b, \{u_{ij}\}, \{\lambda_j\}) \propto \left(\Pi_i^{50} \lambda e^{-\lambda \mu_i}\right) \times \left(\frac{b^a}{\Gamma(a)} \lambda_j^{a-1} e^{-b\lambda}\right)$
j can be ignored as each λ parameter lies in the same distribution with all the other λ parameters
$\pi(\{\lambda_j\}|a, b, \{u_{ij}\}, \{\lambda_j\}) \propto \lambda^{50} \times \left(e^{-\lambda(50 \sum u_j + b)}\right)$
Set $U = \sum u$ and further set:
$B = 50U + b$
$A = 50 + a$
Finally we get:
$\pi(\{\lambda_j\}|a, b, \{u_{ij}\}, \{\lambda_j\}) \propto \frac{B^A}{\Gamma(A)} \lambda^{A-1} e^{-\lambda B}$
This follows the form of a Gamma distribution meaning that this full conditional can be sampled from directly.
For b the full conditional is found as:
$\pi(b|a, \{u_{ij}\}, \{\lambda_j\}) \propto \left(\prod_{j=1}^{10} \frac{b^a}{\Gamma(a)} \lambda_j^{a-1} e^{-b\lambda_j}\right) \times \left(\Pi_i^{50} \Pi_j^{10} \lambda_j e^{-\lambda_j \mu_{ij}}\right) \times \frac{\beta_b^{\alpha_b}}{\Gamma(\alpha_b)} b^{\alpha_b - 1} e^{-\beta_b b}$

$\pi(b|a,\{u_{ij}\},\{\lambda_j\}) \propto \left(\prod_{j=1}^{10} \frac{b^a}{\Gamma(a)}\lambda_j^{a-1}e^{-b\lambda_j}\right) \times \frac{\beta_b^{\alpha_b}}{\Gamma(\alpha_b)} b^{\alpha_b-1}e^{-\beta_b b}$
$\pi(b|a,\{u_{ij}\},\{\lambda_j\}) \propto \left(\prod_{j=1}^{10} b^a e^{-b\lambda_j}\right) \times b^{\alpha_b-1}e^{-\beta_b b}$
$\pi(b|a,\{u_{ij}\},\{\lambda_j\}) \propto \left(b^{10a}e^{-10b\lambda_j}\right) \times b^{\alpha_b-1}e^{-\beta_b b}$
$\pi(b|a,\{u_{ij}\},\{\lambda_j\}) \propto \left(b^{10a+\alpha_b-1}e^{-b(10\bar{\lambda}+\beta_b)}\right) \propto \frac{10a+\alpha_b}{(\frac{10a+\alpha_b}{\beta_b+10\bar{\lambda}})^{10a+\alpha_b}} \left(b^{10a+\alpha_b-1}e^{-b(10\bar{\lambda}+\beta_b)}\right)$
This subsequently brings us to a Weibull distribution with the PDF as reference:

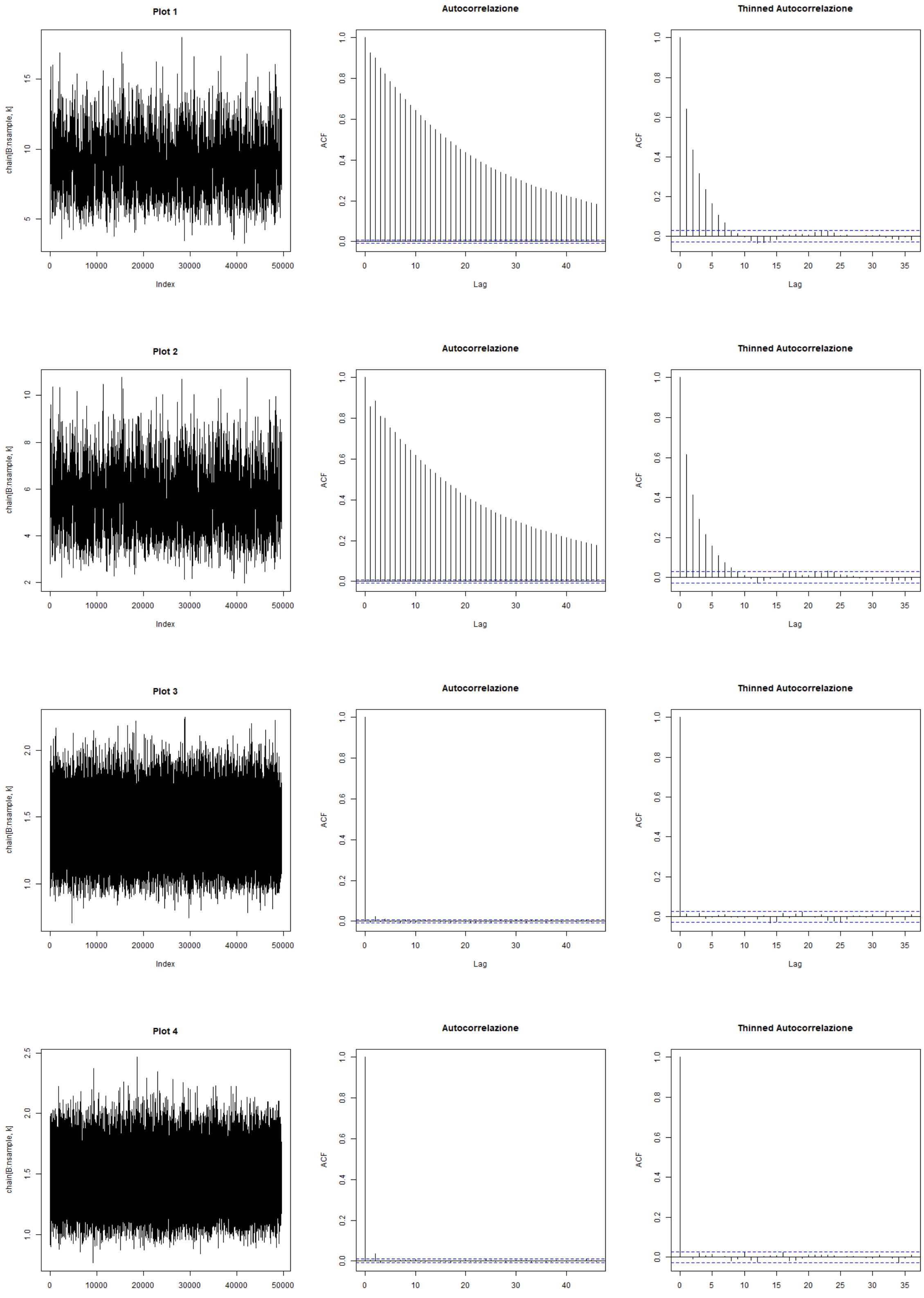
$$D_{it} = \begin{cases} f(x; K_{weibull}, \lambda_{weibull}) = \frac{K_{weibull}}{\lambda_{weibull}} \left(\frac{x}{\lambda_{weibull}}\right)^{K_{weibull}-1} e^{-(x/\lambda_{weibull})^{K_{weibull}}} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases} \quad (1)$$

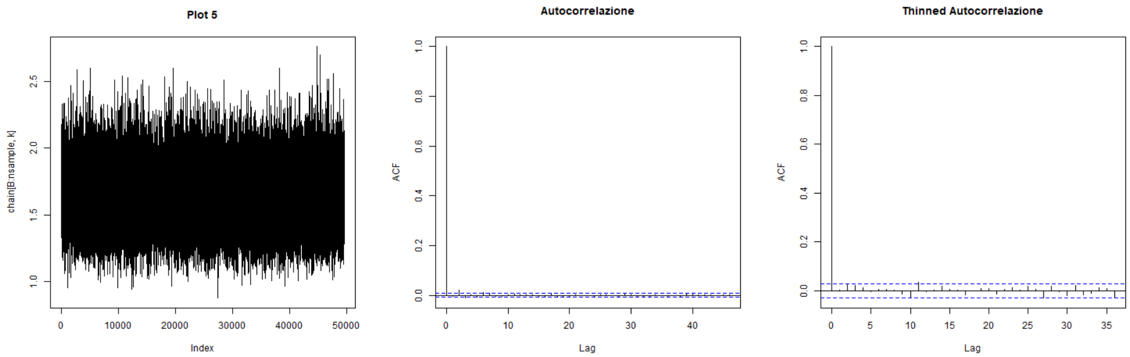
with parameters K and λ. Solving for these parameters we get:
$k_{weibull} = 10a + \alpha_b$ and
$\lambda_{weibull} = \frac{10a+\alpha_b}{\beta_b+n*\bar{\lambda}}$
and thus a viable PDF which we can sample from for the λ and b parameters are found. The a parameter's full conditional does not follow any known distribution meaning that it is difficult to sample from it directly.

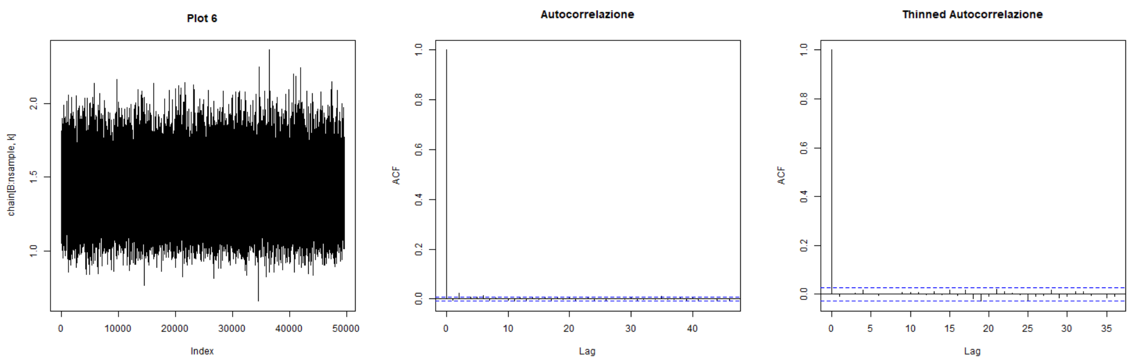
2.4 Conclusion for this section

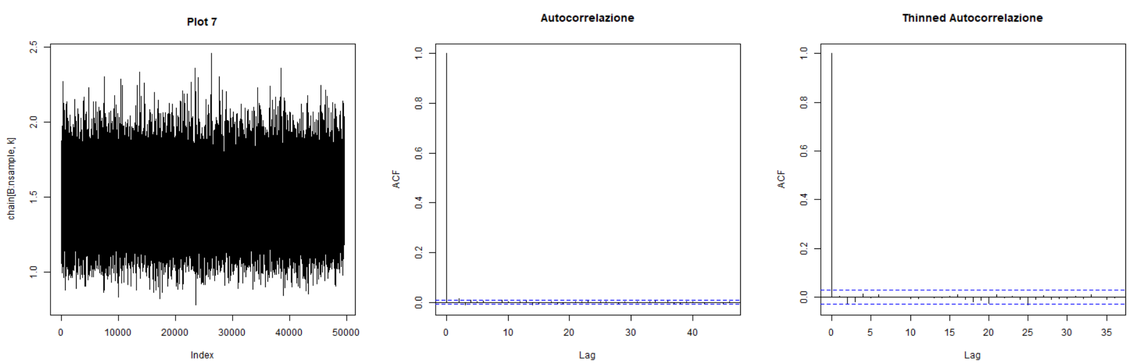
Using all this, we can thus use a Metropolis-within-Gibbs-sampling with parameter a making use of the subroutine of the Metropolis-Hastings algorithm and the rest of the parameters, where a full conditional exists, will use Gibbs sampling directly.
In order to get reliable samples from the Metropolis-within-Gibbs-sampling method it helps to have a burn-in period which helps to allows the algorithm to 'mature' and reach its distribution equilibrium. To do so, the initial distribution samples are discarded and in this case the first 500 samples are discarded from a total iteration of 50000.
When sampling from parameter a using the Metropolis Hastings algorithm it is important to choose both a proposal value and an acceptance probability. In this project the proposal distribution is based on the current state along with some randomness. Specifically, a random value is generated from a uniform distribution between -2 and 2 and added to the current state of the parameter a. Then an acceptance probability is calculated through $\min(1, \frac{proposed_prob}{current_prob})$. This follows the random walk proposal with a uniform distribution due to a perceived symmetry of a. We assume a constant variance of 0.2 and get an acceptance of around 60%. According to the slides this isn't ideal however despite significant changes to the proposal variance it does not seem to have a great impact on the acceptance rate.

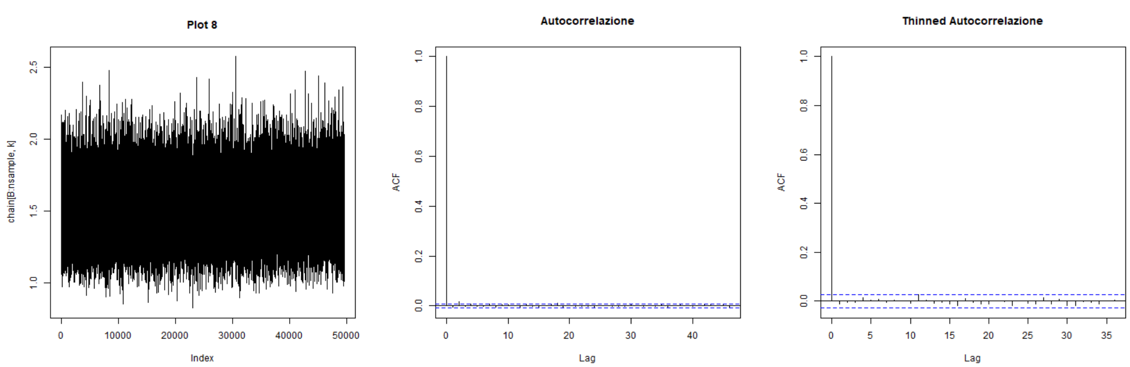
3 Results

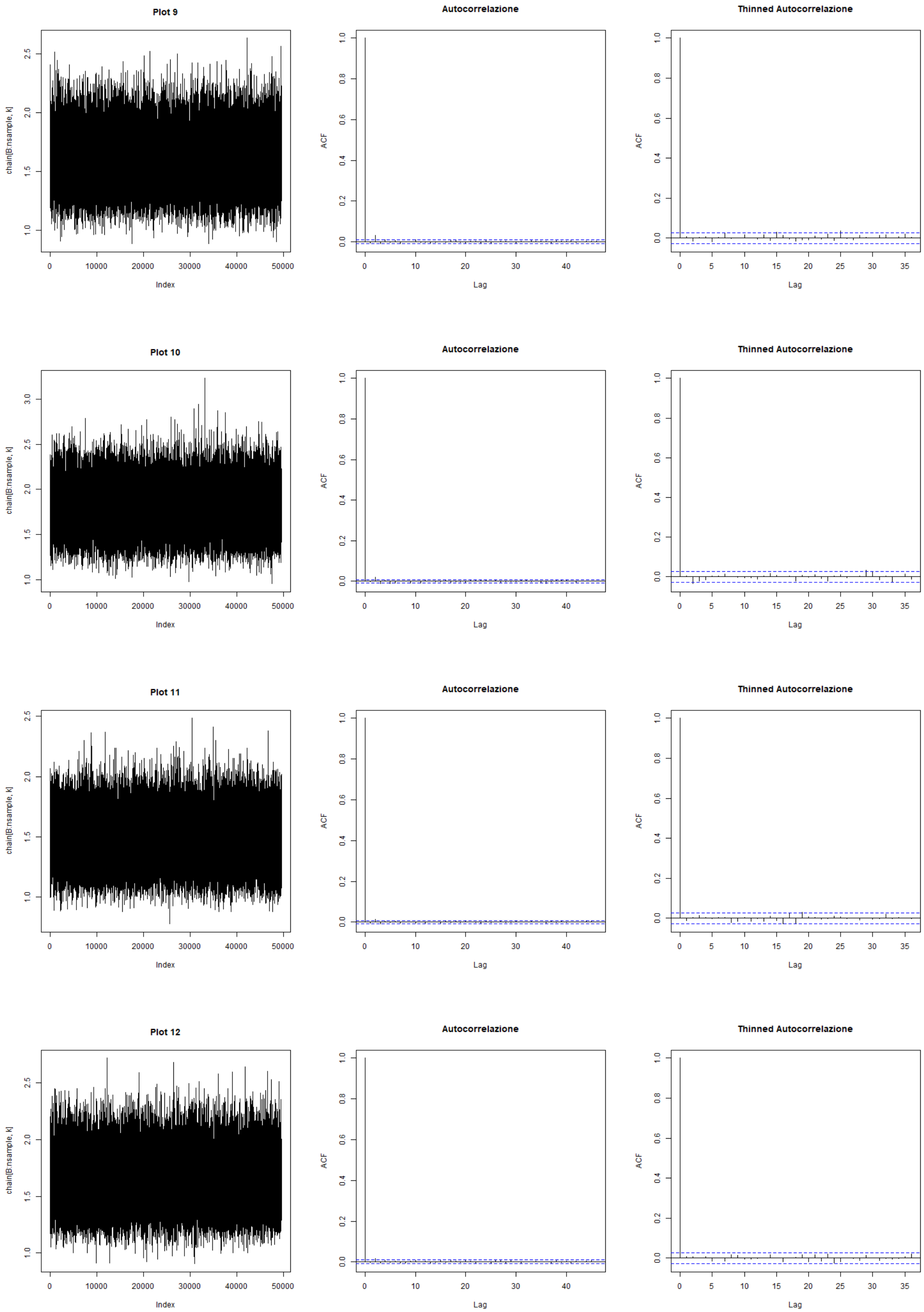

Plot 1
Autocorrelazione
Thinned Autocorrelazione
Plot 2
Autocorrelazione
Thinned Autocorrelazione
Plot 3
Autocorrelazione
Thinned Autocorrelazione
Plot 4
Autocorrelazione
Thinned Autocorrelazione
chain[B:nsample, k]
ACF
Index
Lag


Plot 5
chain[B:nsample, k]
Index
Autocorrelazione
ACF
Lag
Thinned Autocorrelazione
ACF
Lag


Plot 6
chain[B:nsample, k]
Index
Autocorrelazione
ACF
Lag
Thinned Autocorrelazione
ACF
Lag


Plot 7
chain[B:nsample, k]
Index
Autocorrelazione
ACF
Lag
Thinned Autocorrelazione
ACF
Lag


Plot 8
chain[B:nsample, k]
Index
Autocorrelazione
ACF
Lag
Thinned Autocorrelazione
ACF
Lag



From the first column the plots, for each parameter, reveal a random sampling from a

distribution that hopefully represents the target distribution. On the second column an Auto correlation plot demonstrates its correlation with past samples. As can be noted, the ACF decreases very quickly after the initial lags. Hence, it has very little dependence on its past and therefore signifies that the average amount of samples which are used to approximate the posterior mean is higher than it should be if instead the samples were taken from independent samples from the posterior distribution.
Finally, on the third column there are plots indicating the thinned Auto correlation where every 10^{th} sample was used. This is utilized as simply sampling it normally will not be independent as can be noticed from the ACF column. Taking a thinned ACF approximates an independent sequence far better than a pure ACF.
A table with each parameter's Mean and Standard Deviation accompanies the results as such:

Table 1: Mean and Standard Deviation (SD)

	Mean	SD
a	8.812	1.8970
b	5.413	1.1705
λ_1	1.406	0.6550
λ_2	1.455	0.3936
λ_3	1.636	0.4750
λ_4	1.402	0.3347
λ_5	1.470	0.4477
λ_6	1.522	0.3874
λ_7	1.597	0.7701
λ_8	1.786	0.4715
λ_9	1.474	0.3419
λ_{10}	1.649	0.3283

4 Conclusion

As can be noted from the results, the parameter estimates seem to be accurate as sampling from the parameters seem to be independent from each other and thus can derive á close approximation of the posterior distribution.