Home Assignment 3

Monte Carlo and Empirical Methods for Stochastic Inference

Group 1

Adrian Murphy
ad5880mu-s@student.lu.se

Daniel Larsson
da6625la-s@student.lu.se

March 27, 2023

1 Coal Mine Disasters

The first part of this assignment deals with coal mine disasters in England from 1658 to 1980. We wish to examine the intensity of disasters and how it changes over time. Specifically, we want to find breakpoints in time where the intensity of disasters changes, perhaps due to technological advances, increased mining activities or improvements in labor rights.

We collect our d breakpoints as $\mathbf{t} = (t_1, \ldots, t_{d+1})$ where $t_1 = 1658$ and $t_{d+1} = 1980$ are called the endpoints. The disaster intensity in each interval $[t_i, t_{i+1})$ is λ_i and we denote $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$. We use a dataset with $n = 751$ recorded disasters in time points $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$. We model the accidents as an inhomogenous Poisson process, where the times between the accidents have an exponential distribution with intensity λ_i for accidents in the interval $[t_i, t_{i+1})$.

Per the instructions, we use a prior $\lambda \sim \Gamma(2, \theta)$ and a hyperprior $\theta \sim \Gamma(2, \Psi)$. In addition, we have a prior

$$f(\mathbf{t}) \propto \begin{cases} \prod_{i=1}^{d}(t_{i+1} - t_i), & \text{for } t_1 < t_2 < \ldots < t_d < t_{d+1} \\ 0, & \text{otherwise} \end{cases}$$

on the breakpoints, implying

$$f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \mathbf{t}) = \exp\left(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i)\right) \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})}$$

where $n_i(\boldsymbol{\tau})$ is the number of disasters in the interval $[t_i, t_{i+1})$

$$n_i(\boldsymbol{\tau}) = \sum_{j=1}^{n} \mathbb{1}_{[t_i, t_{i+1})}(\tau_j)$$

1.1 Marginal Posteriors

We firstly want to calculate the marginal posteriors $f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})$, $f(\boldsymbol{\lambda}|\theta, \mathbf{t}, \boldsymbol{\tau})$ and $f(\mathbf{t}|\theta, \boldsymbol{\lambda}, \boldsymbol{\tau})$. We use the formula for conditional probability to rewrite the posteriors as

$$\begin{aligned} f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau}) &= \frac{f(\theta, \boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})}{f(\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})} \\ f(\boldsymbol{\lambda}|\theta, \mathbf{t}, \boldsymbol{\tau}) &= \frac{f(\boldsymbol{\lambda}, \theta, \mathbf{t}, \boldsymbol{\tau})}{f(\theta, \mathbf{t}, \boldsymbol{\tau})} \\ f(\mathbf{t}|\theta, \boldsymbol{\lambda}, \boldsymbol{\tau}) &= \frac{f(\mathbf{t}, \theta, \boldsymbol{\lambda}, \boldsymbol{\tau})}{f(\theta, \boldsymbol{\lambda}, \boldsymbol{\tau})} \end{aligned} \tag{1}$$

In general, we have that

$$\begin{aligned} f(A, B, C, D) &= f(A|B, C, D)f(B, C, D) \\ &= f(A|B, C, D)f(B|C, D)f(C, D) \\ &= f(A|B, C, D)f(B|C, D)f(C|D)f(D) \end{aligned} \tag{2}$$

Using equation 2, we see that we can rewrite $f(\boldsymbol{\tau}, \mathbf{t}, \boldsymbol{\lambda}, \theta)$ as

$$f(\boldsymbol{\tau}, \mathbf{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{\tau}|\mathbf{t}, \boldsymbol{\lambda}, \theta)f(\mathbf{t}|\boldsymbol{\lambda}, \theta)f(\boldsymbol{\lambda}|\theta)f(\theta)$$

We know $f(\boldsymbol{\lambda}|\theta)$ and $f(\theta)$ from the problem statement. Moreover, $f(\mathbf{t}|\boldsymbol{\lambda}, \theta) = f(\mathbf{t})$ since $\mathbf{t}$ does not depend on $\boldsymbol{\lambda}$ and θ. Since $\boldsymbol{\tau}$ also is independent of θ, we can rewrite $f(\boldsymbol{\tau}|\mathbf{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{\tau}|\mathbf{t}, \boldsymbol{\lambda})$.

We then have

$$f(\boldsymbol{\tau}, \mathbf{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{\tau}|\mathbf{t}, \boldsymbol{\lambda})f(\mathbf{t})f(\boldsymbol{\lambda}|\theta)f(\theta)$$

which is equivalent to our nominator in equation 1. We can use the distributions given in the assignment and define

$$f(\boldsymbol{\tau}, \mathbf{t}, \boldsymbol{\lambda}, \theta) \propto \exp\left(-\sum_{i=1}^{d} \lambda_i(t_{i+1} - t_i)\right) \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \prod_{i=1}^{d} (t_{i+1} - t_i) \prod_{i=1}^{d} \left(\frac{\theta^2}{\Gamma(2)} \lambda_i e^{-\theta\lambda_i}\right) \frac{\Psi^2}{\Gamma(2)} \theta e^{-\Psi\theta}$$

1.1.1 $f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})$

We are only interested in finding a proportional expression and not the normalizing constant. We can therefore remove the distributions from $f(\boldsymbol{\tau}, \mathbf{t}, \boldsymbol{\lambda}, \theta)$ that do not depend on θ (since they will simply scale the the sought distribution). We thus have

$$
\begin{aligned}
f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau}) &\propto \prod_{i=1}^{d} \left(\frac{\theta^2}{\Gamma(2)} \lambda_i e^{-\theta\lambda_i} \right) \frac{\Psi^2}{\Gamma(2)} \theta e^{-\Psi\theta} \\
&\propto \theta^{2d+1} \exp\left(-(\Psi + \sum_{i=1}^{d} \lambda_i)\theta \right) \\
&\propto \Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)
\end{aligned}
$$

In other words, the marginal posterior is proportional to a gamma distribution.

1.1.2 $f(\boldsymbol{\lambda}|\theta, \mathbf{t}, \boldsymbol{\tau})$

By the same argument,

$$
\begin{aligned}
f(\boldsymbol{\lambda}|\theta, \mathbf{t}, \boldsymbol{\tau}) &\propto \exp\left(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \right) \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \prod_{i=1}^{d} \left(\frac{\theta^2}{\Gamma(2)} \lambda_i e^{-\theta\lambda_i} \right) \\
&\propto \prod_{i=1}^{d} \exp(-\lambda_i (t_{i+1} - t_i + \theta)) \lambda_i^{n_i(\boldsymbol{\tau})+1}
\end{aligned}
$$

Then, we have that the marginal posterior is once again proportional to a gamma distribution

$$
f(\lambda_i|\theta, \mathbf{t}, \boldsymbol{\tau}) \propto \Gamma(n_i(\boldsymbol{\tau}) + 2, t_{i+1} - t_i + \theta)
$$

1.1.3 $f(\mathbf{t}|\theta, \boldsymbol{\lambda}, \boldsymbol{\tau})$

Finally,

$$
f(\mathbf{t}|\theta, \boldsymbol{\lambda}, \boldsymbol{\tau}) \propto \exp\left(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \right) \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \prod_{i=1}^{d} (t_{i+1} - t_i)
$$

In this case, we are unable to find a known distribution which the marginal posterior is proportional to.

1.2 Hybrid MCMC Algorithm

We now wish to construct a hybrid MCMC algorithm that samples from $f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})$. As shown in the previous exercise, $f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})$ and $f(\boldsymbol{\lambda}|\theta, \mathbf{t}, \boldsymbol{\tau})$ are proportional to some gamma distribution,

which means that we can use a Gibbs sampler to sample from them. However, $f(\mathbf{t}|\theta, \boldsymbol{\lambda}, \boldsymbol{\tau})$ is not proportional to a known distribution and we instead employ a Metropolis-Hastings algorithm to sample from this marginal posterior.

As described in lecture 10 [1], the Gibbs sampler constructs a Markov chain by simulating a sequence

$$
\begin{aligned}
&\texttt{given } X_k \\
&\texttt{draw } X_{k+1}^1 \sim f_1(x^1|X_k^2, \ldots, X_k^m) \\
&\ldots \\
&\texttt{draw } X_{k+1}^m \sim f_m(x^m|X_{k+1}^1, \ldots, X_{k+1}^{m-1})
\end{aligned}
$$

We also use Metropolis-Hastings, described in lecture 9 [2], for $f(\mathbf{t}|\theta, \boldsymbol{\lambda}, \boldsymbol{\tau})$ according to

$$
\begin{aligned}
&\texttt{given } X_k \\
&\texttt{draw } X^* \sim r(z|X_k) \\
&\texttt{set } X_{k+1} = \begin{cases} X^* \texttt{ with prob. } & \alpha(X_k, X^*) = 1 \wedge \frac{f(X^*)r(X_k|X^*)}{f(X_k)r(X^*|X_k)} \\ X_k \texttt{ otherwise} & \end{cases}
\end{aligned}
$$

For the proposal kernel $r(z|X_k)$ we use the *random walk proposal one at a time* where we update one breakpoint at a time and, for each breakpoint t_i, generate a candidate t_i^* according to

$$
\begin{aligned}
t_i^* &= t_i + \varepsilon, \ \varepsilon \sim \mathcal{U}(-R, R) \\
R &= \rho(t_{i+1} - t_{i-1})
\end{aligned}
$$

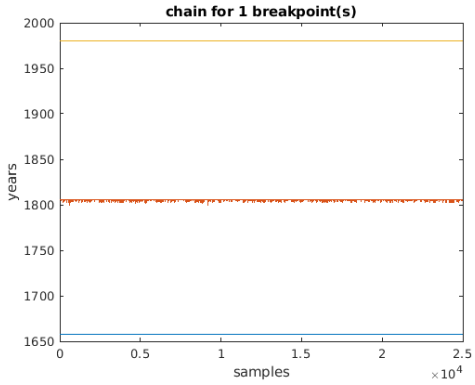
where ρ is a tuning parameter. Since $\mathcal{U}(-R, R)$ is symmetric, the proposal kernel is also symmetric, implying that $r(z|X_k) = r(X_k|z)$. Then, the acceptance probability in the Metropolis-Hastings algorithm simply is

$$
\alpha(X_k, X^*) = 1 \wedge \frac{f(X^*)}{f(X_k)}
$$

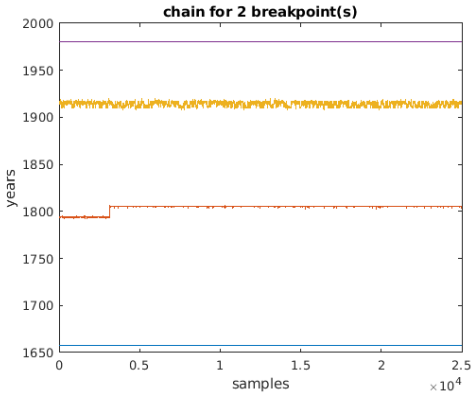
As described in lecture 9 [3], a symmetric proposal is one of the three main classes for proposal kernels yielding an aperiodic and irreducible chain. We thus have a correct Metropolis-Hastings algorithm.

1.3 Chain Behavior

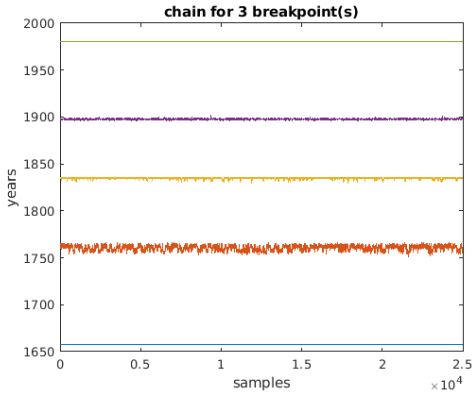
We use our hybrid MCMC algorithm to simulate a chain for one to five breakpoints. We use a burn-in with 10^4 iterations (to reach a somewhat stationary distribution) and then a sample size of $2.5 \cdot 10^4$. We use $\Psi = 20$ and $\rho = 0.01$. In the plots, the two completely straight lines correspond to the start and end years, 1658 and 1980.



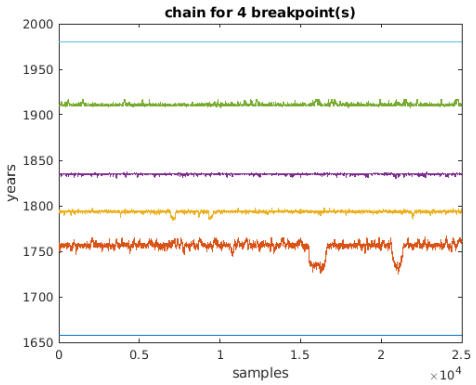
(a) Chain behavior for one breakpoint



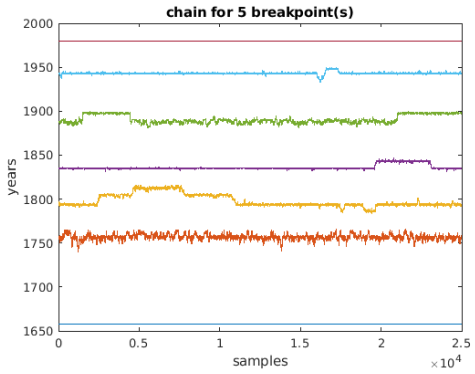
(b) Chain behavior for two breakpoints



(c) Chain behavior for three breakpoints



(d) Chain behavior for four breakpoints



(e) Chain behavior for five breakpoints

Figure 1: Chain behavior for different numbers of breakpoints

When adding new breakpoints, we notice that the old breakpoints are often preserved (for example, we still have a breakpoint around year 1800 when going from 1 to 2 breakpoints), indicating that the algorithm finds appropriate breakpoints. We may also plot the breakpoints together with the number of accidents.

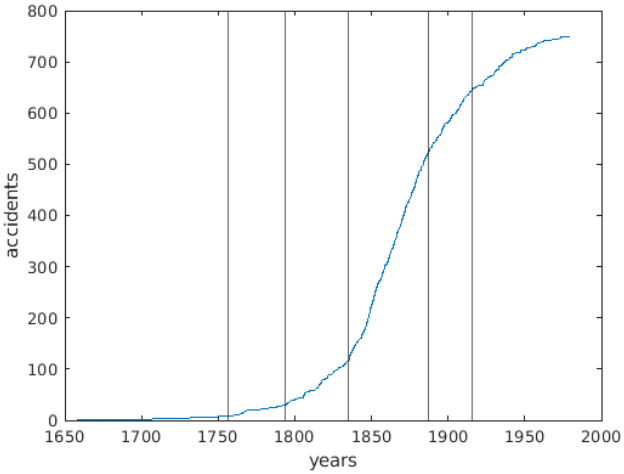
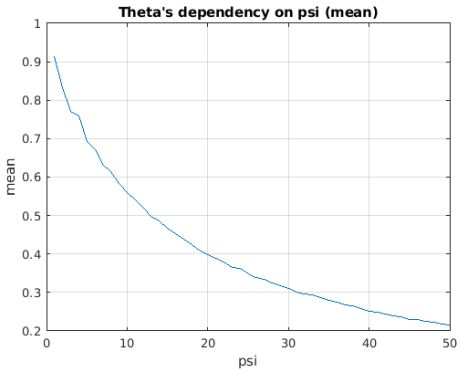


Figure 2: Accidents over time and breakpoints as vertical lines

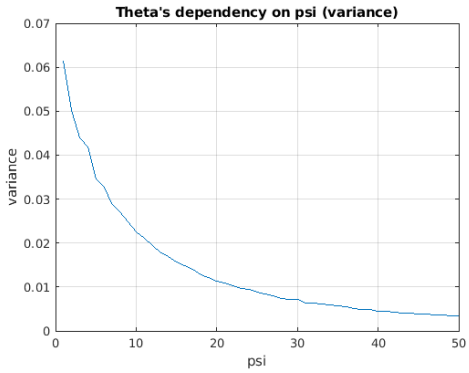
Ideally, we would like the intensity of accidents to be the same between to breakpoints. We seem to catch the intensities quite well with these five breakpoints, especially the accidents between 1840 to 1890 (approximately) which almost seem to grow at a fixed linear rate.

1.4 Sensitivity to Ψ

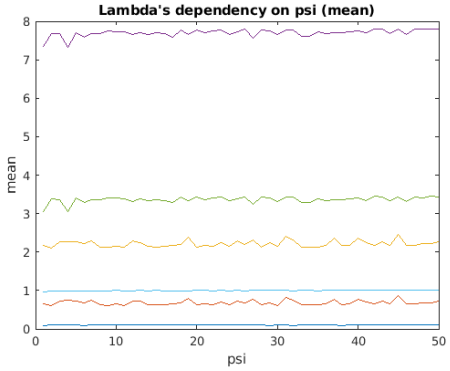
We wish to test how sensitive the posteriors are to choices of Ψ. We accomplish this by examining how the mean and variance of the posteriors change with $1 \leq \Psi \leq 50$ and, still, $\rho = 0.01$. As in the previous exercise, we use a burn-in of 10^4 and a sample size of $2.5 \cdot 10^4$ when simulating with five breakpoints.



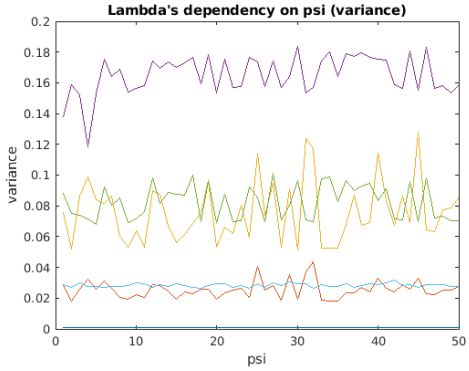
(a) Mean values of θ depending on Ψ



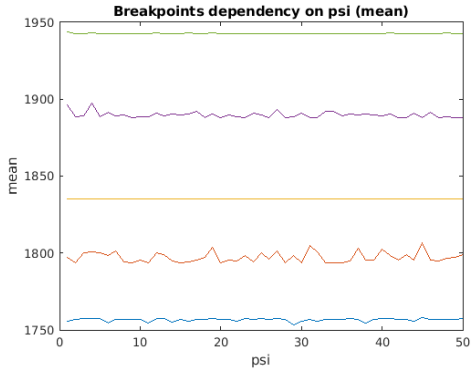
(b) Variance of θ depending on Ψ



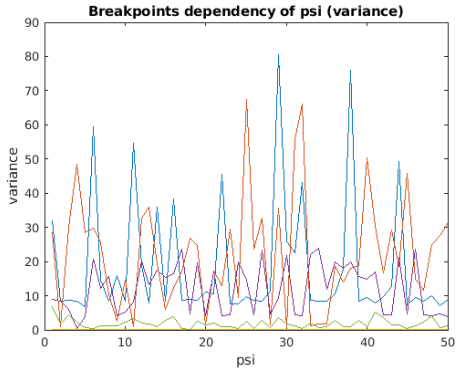
(c) Mean values of the five λ depending on Ψ



(d) Variance of the five λ depending on Ψ



(e) Mean values of the five breakpoints depending on Ψ



(f) Variance of the five breakpoints depending on Ψ

Figure 3: Sensitivity of posteriors depending on the choice of Ψ

Since $f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})$ is proportional to $\Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)$, we should have $\mathbb{E}(\theta) \propto \Psi^{-1}$ and $\mathbb{V}(\theta) \propto \Psi^{-2}$ (from the Gamma distribution). This seems reasonable when considering figure 3a and 3b.

Furthermore, we know that $\boldsymbol{\lambda}$ also depends on Ψ since it depends on θ which in turn depends on Ψ, $f(\lambda_i|\theta, \mathbf{t}, \boldsymbol{\tau}) \propto \Gamma(n_i(\boldsymbol{\tau}) + 2, t_{i+1} - t_i + \theta)$. However, when considering figure 3c and 3d, showing how the mean and variance of λ_i, $1 \leq i \leq 5$ change given $1 \leq \Psi \leq 50$, we can not really see a trend. This could be because $t_{i+1} - t_i$ is, comparatively, much larger than θ, effectively rendering

the dependency insignificant.

$\mathbf{t}$ depends on $\boldsymbol{\lambda}$, but given how insignificant the dependence of $\boldsymbol{\lambda}$ on Ψ was, we do not expect to see a large dependence of $\mathbf{t}$ on Ψ either. This is confirmed when considering figure 3e and 3f.

1.5 Sensitivity to ρ

We wish to test how sensitive the posteriors are to the choice of ρ. We note that it is only $\mathbf{t}$ that depends on ρ. For our Metropolis-Hastings algorithm, we want an acceptance ratio around 30% to achieve a good balance between mixing and convergence [4]. We start by plotting the acceptance ratio as a function of $0.01 \leq \rho \leq 0.1$ with $\Psi = 20$.

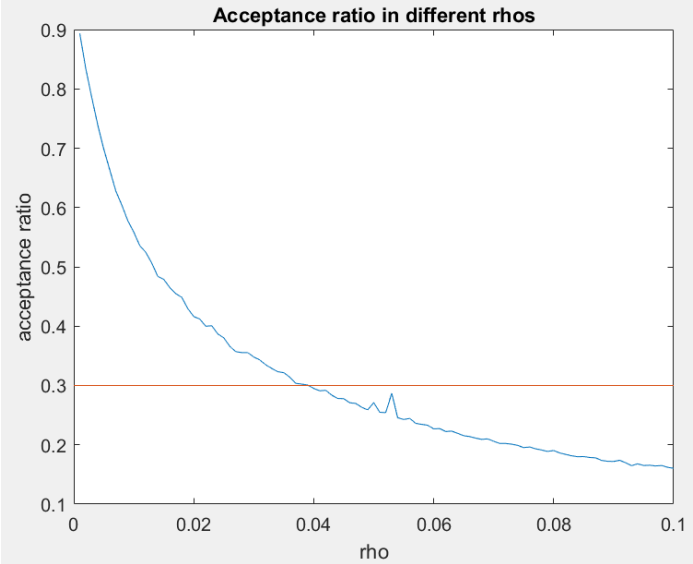


Figure 4: Acceptance ratio as a function of ρ

Considering figure 4, $\rho \approx 0.04$ seems like a more appropriate choice than the previously used $\rho = 0.01$. We can plot the auto-correlation of the breakpoints for our previous $\rho = 0.01$ and a new $\rho = 0.04$

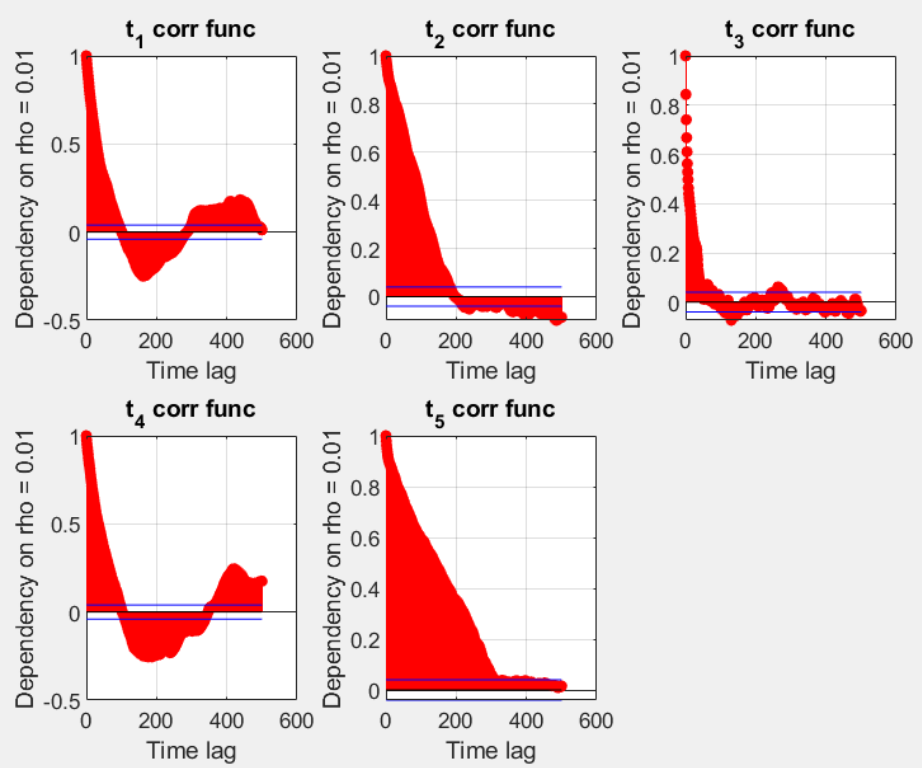


Figure 5: Auto-correlation functions for the breakpoints using $\rho = 0.01$

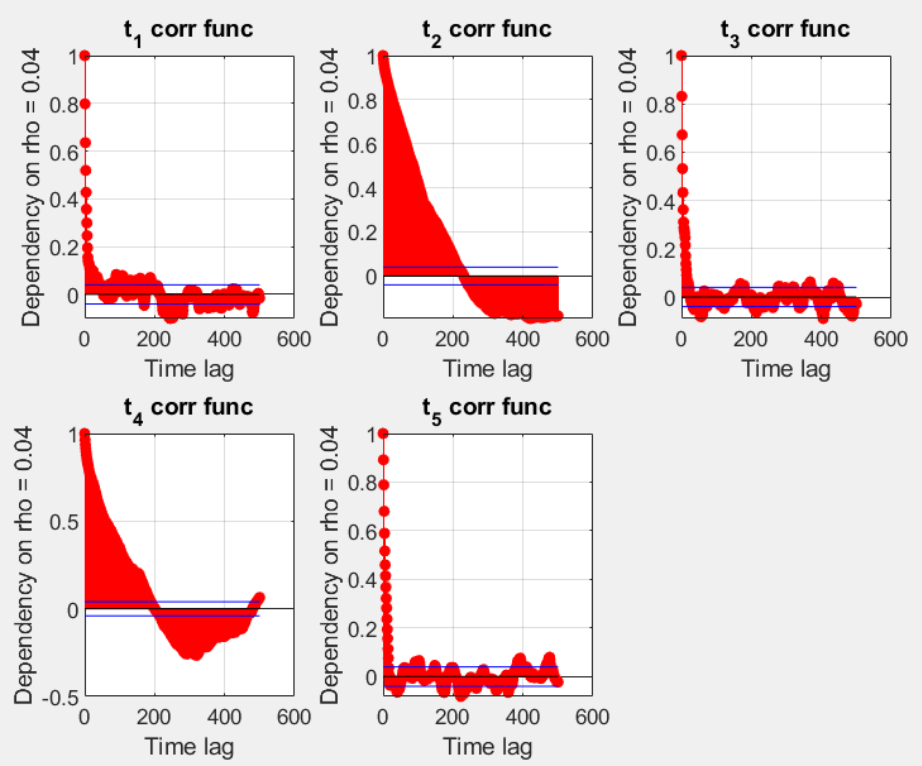


Figure 6: Auto-correlation functions for the breakpoints using $\rho = 0.04$

We can see that the time-dependency for larger time lags, at least for t_1 and t_5, is lower since the auto-correlation function goes to zero much faster for $\rho = 0.04$ than 0.01. This is a desired behavior for our chain, further indicating that $\rho = 0.04$ is an improved parameter value.

However, when testing for $\rho = 0.02$ we found that the auto-correlation function for t_2 approached zero much faster than for either $\rho = 0.01$ or $\rho = 0.04$. Therefore, it might be desirable to have different ρ_i corresponding to different breakpoints t_i.

2 The 100-year Atlantic Wave

The second part of the assignment is about the height of the waves in the north Atlantic. The height has been measured 14 times a month during the winter months, having the following Gumbel distribution found in the assignment instructions fit well to this data

$$F(x; \mu, \beta) = \exp(-\exp(-\frac{x-\mu}{\beta})),\ x\ \mu \in \mathbb{R},\ \beta > 0 \tag{3}$$

2.1 Gumbel distribution inverse

We can calculate the inverse of above equation 3 as $F^{-1}(u; \mu, \beta)$, setting $u = F(x; \mu, \beta)$ and then solving for x.

$$u = \exp(-\exp(-\frac{x-\mu}{\beta})) \iff \ln(-\ln(u)) = -\frac{x-\mu}{\beta} \iff \mu - \beta \cdot \ln(-\ln(u)) = x$$

As such, we see that $F^{-1}(u; \mu, \beta) = \mu - \beta \cdot \ln(-\ln(u))$.

2.2 Parametric bootstrapped 95% confidence intervals

In this part we estimate a 95% confidence interval for the parameters β and μ using parametric bootstrapping as described in lecture 14 [5]. First, we estimate $\hat{\beta}$ and $\hat{\mu}$ using the Atlantic wave data and the Gumbel distribution given to us, which gives the resulting parameters $\hat{\beta} = 1.4858$ and $\hat{\mu} = 4.1477$.

Using these parameters, we can get new Atlantic wave data samples $Y_b^*, b \in \{1, 2, ...B\}$ through the inversion method, which can then be used in the Gumbel distribution to estimate new parameters β and μ for these bootstrapped samples.

These new parameters β and μ are compared to the actual parameters $\hat{\beta}$ and $\hat{\mu}$, calculating the error between the two as $\Delta_b^* = \beta(Y_b^*) - \hat{\beta}$ and $\Delta_b^* = \mu(Y_b^*) - \hat{\mu}$. Using these to calculate the 95% confidence interval, we get the intervals $I_{\hat{\beta}} = (1.3947, 1.5782)$ and $I_{\hat{\mu}} = (4.023, 4.2579)$

2.3 One-sided 95% confidence interval for the 100-year return value

For this task we want to estimate a one sided upper bounded parametrically bootstrapped 95% confidence interval for the 100-year return value. First of all, we note that the return value of the T:th observation will be $F^{-1}(1-1/T, \beta, \mu)$, and that there will be $T = 3*14*100 = 4200$ observation in total (since we make 14 observations every month, for 3 months every year, for 100 years).

The actual return value of the T:th observation will be $F^{-1}(1 - 1/T, \hat{\beta}, \hat{\mu}) = 16.5436$, which will be compared to the return value of all bootstrapped samples $F^{-1}(1 - 1/T, \beta(Y_b^*), \mu(Y_b^*)), b \in \{1, 2, ..., B\}$. After making these comparisons in a similar fashion to the above task (calculating the difference between real and bootstrapped return values), we can estimate that the upper bound of this one sided 95% confidence interval will be 17.2519.

References

[1] M. Wiktorsson, *The Gibbs sampler (cont.)*, February 2023, lecture 10, slide 15

[2] M. Wiktorsson, *The Metropolis-Hasting (MH) algorithm*, February 2023, lecture 9, slide 16

[3] M. Wiktorsson, *different types of proposal kernels*, February 2023, lecture 9, slide 21

[4] Wikipedia, *Metropolis–Hastings algorithm*. Gathered Mars 2023. Available at https://en.wikipedia.org/wiki/Metropolis–Hastings_algorithm

[5] M. Wiktorsson, *Parametric bootstrap*, March 2023, lecture 14, slide 19