Lund University
Faculty of Engineering (LTH)

FMSN50

Monte Carlo and Empirical Methods for Stochastic Inference

Home Assignment 3

Authors:
Justin Ma (941123-7812), tfy14jma@student.lu.se
Erik Norlander (940601-7831), nat13eno@student.lu.se

The report was submitted on: March 6, 2018

Lunds universitet
Lunds Tekniska Högskola

Contents

Introduction

This report has two parts. The first concerns coal mining disasters, where we will construct a hybrid MCMC algorithm to analyze the data and try and determine breakpoints in accident frequency. In the second part we study the 100-year Atlantic wave and try to estimate how high it can become using bootstraping.

Some essential probability theory

Throughout this report the law of conditional probability (1) will be used extensively. Therefore we present it here as essential background.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{1}$$

Of course this rule not only holds for probabilities but for probability density functions (PDF's) as well. Specifically it will be used with multiple events (or variables). Moreover, Baye's theorem (2) which can be derived from (1) will also be used.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{2}$$

1 Coal mine disasters

In this section we're going to analyze data from British coal mining disasters between 1658 and 1980. Naturally the conditions have changed during this long period, and therefore there's probably changes in disaster intensity. We want to find the breakpoints of these changes.

1.1 (a) The marginal posteriors

First, we want to compute the probability density functions $f(\theta \mid \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})$, $f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\tau})$ and $f(\boldsymbol{t} \mid \theta, \boldsymbol{\lambda}, \boldsymbol{\tau})$. Using the law of conditional probability (1) we rewrite them as

$$\begin{cases} f(\theta \mid \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau}) = \dfrac{f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})}{f(\boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})} \\ f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\tau}) = \dfrac{f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})}{f(\theta, \boldsymbol{t}, \boldsymbol{\tau})} \\ f(\boldsymbol{t} \mid \theta, \boldsymbol{\lambda}, \boldsymbol{\tau}) = \dfrac{f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})}{f(\theta, \boldsymbol{\lambda}, \boldsymbol{\tau})} \end{cases} \tag{3}$$

and recognize that the numerator $f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})$ is instrumental in computing all three conditional PDF's. Written in the assignment [1] is that $\lambda \sim \Gamma(2,\theta)$ and $\theta \sim \Gamma(2,\Psi)$. We also know that

$$f(\boldsymbol{t}) \propto \begin{cases} \prod_{i=1}^{d}(t_{i+1} - t_i), & \text{for} \quad t_1 < t_2 < ... < t_d < t_{d+1}, \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

and

$$f(\boldsymbol{\tau} \mid \boldsymbol{\lambda}, \boldsymbol{t}) = \exp\Big(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \Big) \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \tag{5}$$

Meaning, that $f(\boldsymbol{t})$, $f(\theta)$, $f(\boldsymbol{\lambda},\theta)$ and $f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda})$ are all given in the background of the assignment [1]. Therefore, we want to express the numerator $f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})$ in these probability density functions.

Using the rule of conditional probability (1) for 4 individual events we can rewrite it as

$$\begin{aligned} \mathbb{P}(A,B,C,D) &= \mathbb{P}(A \mid B,C,D)\mathbb{P}(B,C,D) = \mathbb{P}(A \mid B,C,D)\mathbb{P}(B \mid C,D)\mathbb{P}(C,D) \\ &= \mathbb{P}(A \mid B,C,D)\mathbb{P}(B \mid C,D)\mathbb{P}(C \mid D)\mathbb{P}(D). \end{aligned} \tag{6}$$

Of course, the same rules apply for a PDF. If we replace $\{A,B,C,D\}$ in (6) with $\{\boldsymbol{\tau}, \boldsymbol{t}, \boldsymbol{\lambda}, \theta\}$ and consider the same case for a PDF we arrive at the expression

$$f(\boldsymbol{\tau}, \boldsymbol{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}, \theta) f(\boldsymbol{t} \mid \boldsymbol{\lambda}, \theta) f(\boldsymbol{\lambda} \mid \theta) f(\theta). \tag{7}$$

Now, since we want to rewrite (7) with known PDF's we recognize that (7) includes $f(\theta)$ and $f(\boldsymbol{\lambda} \mid \theta)$ which are known. Moreover, we note that $\boldsymbol{t}$ is independent of $\boldsymbol{\lambda}$ and θ, meaning $f(\boldsymbol{t} \mid \boldsymbol{\lambda}, \theta) = f(\boldsymbol{t})$. We also note that $\boldsymbol{\tau}$ is independent of θ and therefore $f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda})$. This amounts the the final rearrangement of (7) into equation (8).

$$\begin{aligned} &f(\boldsymbol{\tau}, \boldsymbol{t}, \boldsymbol{\lambda}, \theta) = f(\boldsymbol{\tau} \mid \boldsymbol{t}, \boldsymbol{\lambda}) f(\boldsymbol{t}) f(\boldsymbol{\lambda} \mid \theta) f(\theta) \\ &= \exp\Big(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \Big) \prod_{i=1}^{d} \lambda_i^{n_i(\boldsymbol{\tau})} \prod_{i=1}^{d} (t_{i+1} - t_i) \prod_{i=1}^{d} \Big(\frac{\theta^2}{\Gamma(2)} \lambda_i e^{-\theta\lambda_i} \Big) \frac{\Psi^2}{\Gamma(2)} \lambda_i \theta e^{-\Psi\theta} \end{aligned} \tag{8}$$

If we study the relations (3) closer we realize that to get the proportionality of the marginal posterior we can simply remove the terms that does not depend on it since these would only act as scaling factors. For θ this becomes

$$\begin{aligned} f(\theta \mid \boldsymbol{\tau}, \boldsymbol{t}, \lambda) &\propto \prod_{i=1}^{d} \Big(\frac{\theta^2}{\Gamma(2)} \lambda_i e^{-\theta\lambda_i} \Big) \frac{\Psi^2}{\Gamma(2)} \lambda_i \theta e^{-\Psi\theta} \\ &\propto \theta^{2d+1} e^{-(\Psi + \sum_{i=1}^{d} \lambda_i)\theta} \\ &\propto \Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i). \end{aligned} \tag{9}$$

Meaning that $f(\theta \mid \boldsymbol{\tau}, \boldsymbol{t}, \boldsymbol{\lambda})$ is proportional to a Gamma distribution $\Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)$. Using the same method on the second relation in (3) we get

$$f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\tau}) \propto \prod_{i=1}^{d} e^{-\boldsymbol{\lambda_i}(t_{i+1} - t_i)} \cdot \lambda_i^{n_i(\tau)} \cdot \lambda_i e^{-\theta\lambda_i} = \prod_{i=1}^{d} e^{-\lambda_i(t_{i+1} - t_i + \theta)} \cdot \lambda_i^{n_i(\tau)+1}. \tag{10}$$

Which is recognized to also be a Gamma-distribution with $\Gamma(n_i(\tau)+1, t_{i+1} - t_i + \theta)$. Now, for the final relation in (3) we get

$$f(\boldsymbol{t} \mid \boldsymbol{\lambda}, \boldsymbol{\tau}, \theta) \propto \exp\Big(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \Big) \prod_{i=1}^{d} \lambda_i^{n_i(\tau)} \prod_{i=1}^{d} (t_{i+1} - t_i). \tag{11}$$

This is not a known distribution.

1.2 (b) A hybrid MCMC algorithm

In order to sample from the posterior $f(\theta, \boldsymbol{\lambda}, \boldsymbol{t} \mid \boldsymbol{\tau})$ a MCMC-algorithm needs to be implemented. The marginal posteriors (9) and (10) has known distributions and therefore can be sampled by a Gibbs sampler. However, as just stated at the end of last section, (11) is not a known distribution.

Therefore, a hybrid Markov Chain Monte Carlo-algorithm (MCMC) has to be used. This algorithm will consist of a Gibbs sampler and a Metropolis-Hasting's algorithm. Giving us the ability to sample the marginal posterior $f(\boldsymbol{t} \mid, \theta, \boldsymbol{\lambda}, \boldsymbol{\tau})$ with Metropolis-Hastings so that finally $f(\theta, \boldsymbol{\lambda}, \boldsymbol{t} \mid \boldsymbol{\tau})$ can be sampled.

Metropolis Hastings

The MH-algorithm simulates a sequence of X_k, forming a Markov chain, through the following scheme:

- given X_k,
- generate $X^* \sim r(z \mid X_k)$ and
- set

$$X_{k+1} = \begin{cases} X^* & \text{w. prob.} \quad \alpha(X_k, X^*) \equiv 1 \wedge \dfrac{f(X^*) r(X_k \mid X^*)}{f(X_K) r(X^* \mid X_k)} \\ X_k & \text{otherwise} \end{cases}$$

Here $r(z \mid X_k)$ is the transition density which we sample from, and is referred to as the proposal kernel. In our case a random walk was used, which has the following scheme: Update one breakpoint at a time and for each breakpoint t_i we generate a candidate t_i^* according to

$$t_i^* = t_i + \epsilon \quad \text{with} \quad \epsilon \sim \mathcal{U}(-R, R),$$

where $R = \rho(t_{i+1} - t_{i-1})$ and ρ is a tuning parameter of the proposal distribution. The proposal kernel of the random walk is symmetric since the $\mathcal{U}(-R,R)$-distribution is. Meaning, $r(t^* \mid t) = r(t \mid t^*), \forall (t^*, t) \in \mathbb{T}^2$, where $\mathbb{T}^2$ is our parameter space, and therefore cancel each other out in the MH-scheme. Leaving us with a simplified final step:

- set

$$X_{k+1} = \begin{cases} X^* & \text{w. prob.} \quad \alpha(X_k, X^*) \equiv 1 \wedge \dfrac{f(X^*)}{f(X_K)} \\ X_k & \text{otherwise} \end{cases}$$

f in our case refers to the marginal distribution $f(\boldsymbol{t} \mid \boldsymbol{\lambda}, \boldsymbol{\tau}, \theta)$ equation (11). The random walk proposal is a symmetrical proposal, which is one of the three main classes of proposal kernels r. This means that our chain becomes aperiodic and irreducible. For further details we refer to lecture [2] and the course literature [6, p.202]. Also, for further details on Metropolis Hastings we refer to the lectures [3].

Gibbs sampler

The Gibbs sampler uses multiple conditional distributions to approximate a joint distribution by following the scheme

- draw $X_1^j \sim f(X_1 \mid X_2^{j-1}, ..., X_k^{j-1})$
- draw $X_2^j \sim f(X_2 \mid X_1^j, ..., X_k^{j-1})$
- and so on until
- draw $X_k^j \sim f(X_k \mid X_1^j, ..., X_{k-1}^j)$
- at which point $(X_1, X_2, ..., X_k)$ approaches $f(X_1, X_2, ..., X_k)$

The Gibbs sampler will be used to generate $f(\theta, \boldsymbol{\lambda}, \boldsymbol{t} \mid \boldsymbol{\tau})$ in this assignment. To do so we simply follow the scheme for $\theta, \boldsymbol{\lambda}$ and $\boldsymbol{t}$. This will yield a density function conditioned on $\boldsymbol{\tau}$. For further details we refer to the lectures [3].

1.3 (c) Behaviour of the chain for different d

To begin with the file `'coal_mine_disasters.mat'` was loaded which contains all information needed about the intensities of disasters between 1658 and 1980. To investigate the behaviour of the chain for $d = 1, 2, 3, 4, 5, 6$ number of breakpoints, we set the initial values by dividing the interval in d-parts. From figure 3 we can see that setting a initial points lower than approximately 1690 is not preferable, as we note that the intensity of accidents is constant for the first few decades.

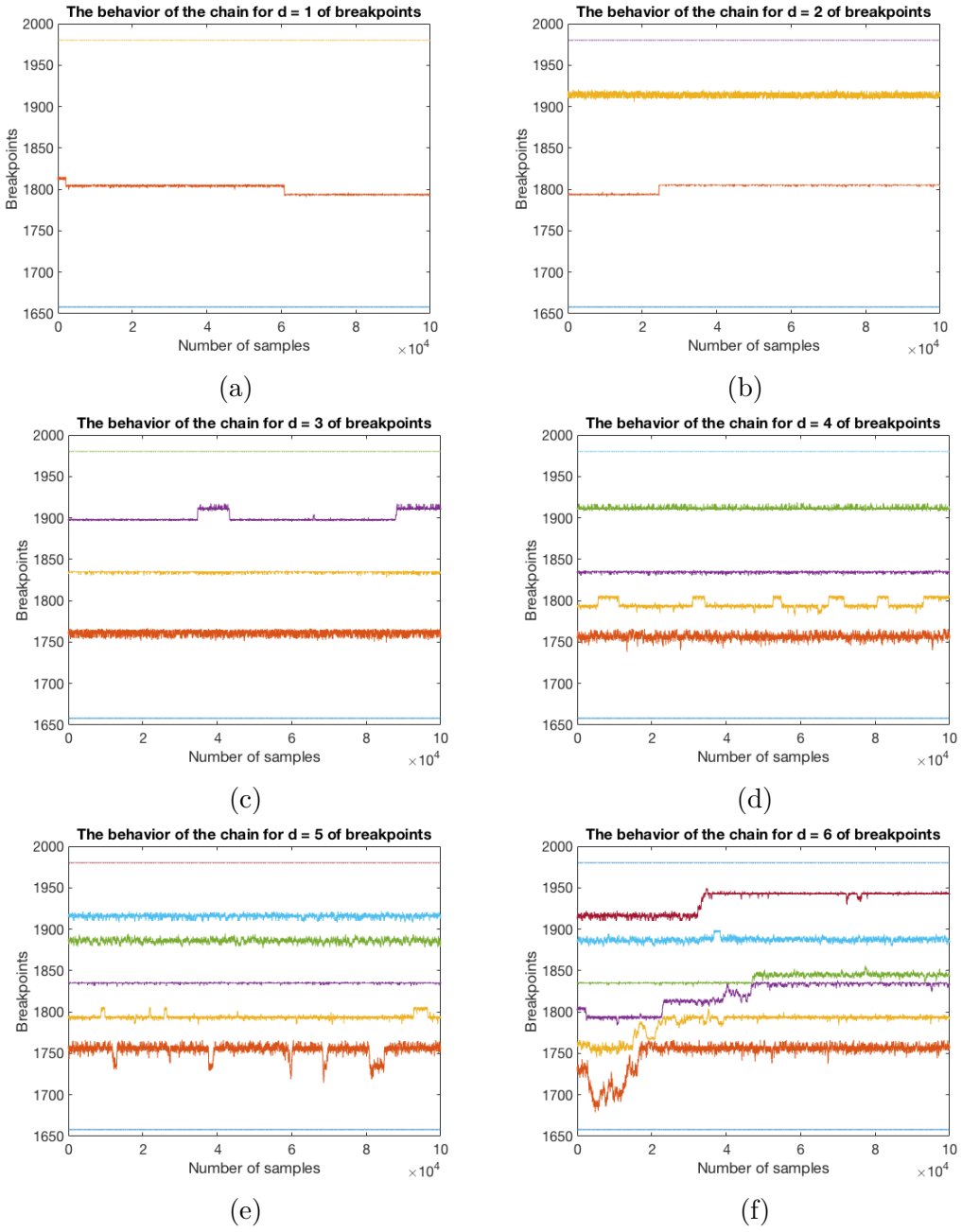


Figure 1: Behaviour of the breakpoint estimations for different values of d. The parameters were set to $\psi = 20$ and $\rho = 0.01$ for all breakpoints. The burn-in was set to 5000 and sample size to 100000

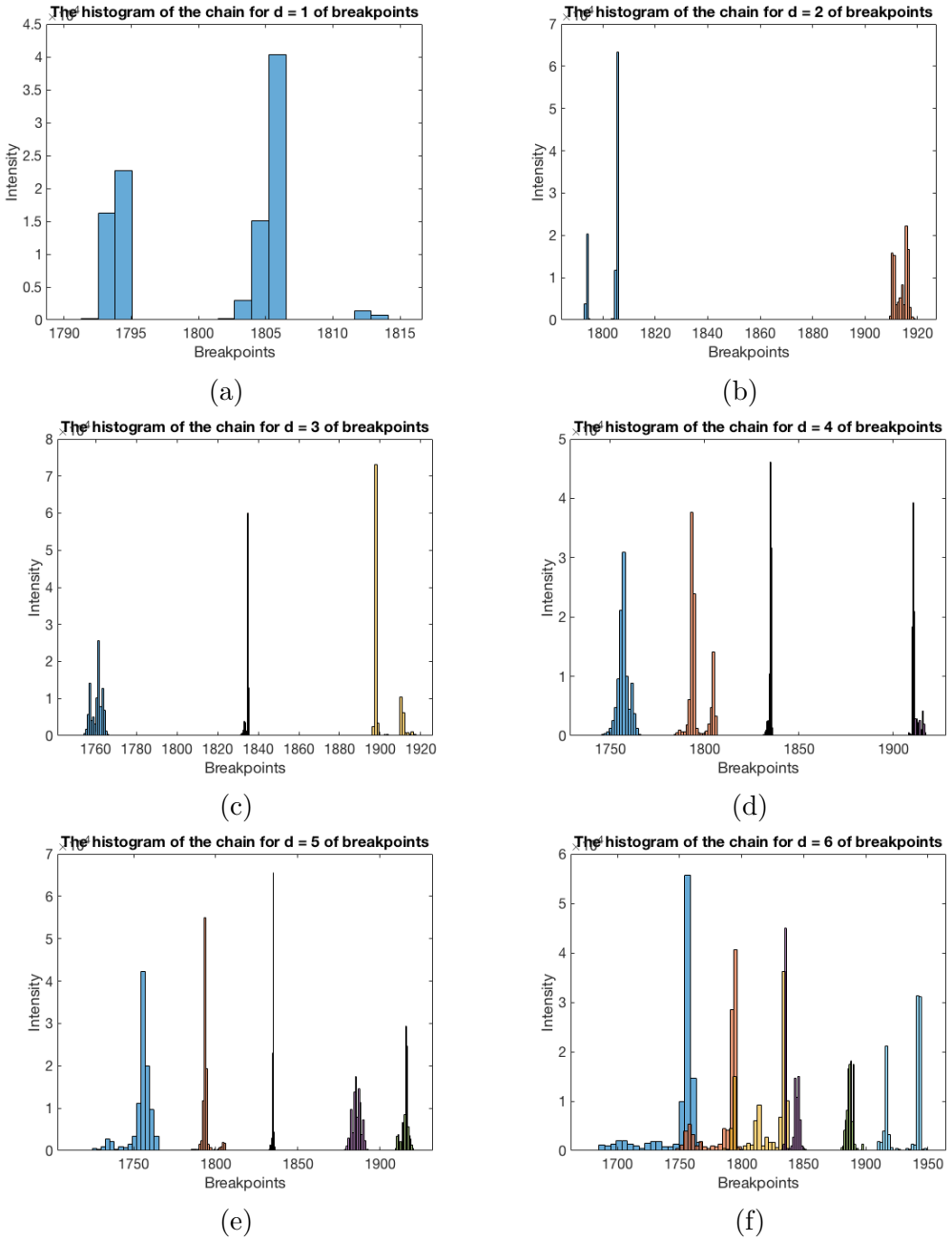


Figure 2: Behaviour of the breakpoint estimations for different values of d. The parameters were set to $\psi = 20$ and $\rho = 0.01$ for all breakpoints, plotted as histograms. The burn-in was set to 5000 and sample size to 100000

As we can see in figure 1, when increasing the number of breakpoints d, the chain always seems to converge to the same values. This suggests that the algorithm is capable of finding appropriate breakpoints given that the initial breakpoints is in the interval. Comparing e.g. figure 1c and figure 1d, we see that by adding a new breakpoint a new appropriate breakpoint is found while the other seems to stay the same. However when using 6 breakpoints, as seen in figure 1f, the first chain seem to experience difficulty finding the correct breakpoint. This seems to affect the lower chains following it as well. This could be evidence that 6 breakpoints are

one too many, as the algorithm seems to work fine for 5 breakpoints (figure 1e).

In figure 2 we see the histogram of the chains for different number of breakpoints. These figures corresponds to the figures in 1, and we can see that the width of the histograms are relatively low, i.e the variance is relatively low. In some figures, e.g. figure 2a and 2d we can see that for some chains the histogram is bimodal, meaning that there is a probability for the chain to jump between to breakpoints as we can confirm in 1a and 1d. For 6 breakpoints in figure 2f there is some overlap of the histograms around 1750 and 1800. This could be one reason why some of the lower chains in 1f jump to different breakpoints.

Table 1: The intensities for different number of breakpoints. The parameters $\psi = 20$ and $\rho = 0.01$ were used for all breakpoints, burn-in were set to 1000 and sample size to 10^5.

d	7	6	5	4	3	2	1
λ_1	0.0926	0.1127	0.0799	0.0945	0.0793	0.3230	0.1878
λ_2	0.1419	0.7144	0.7837	0.5640	1.4165	5.3456	3.8151
λ_3	0.6181	2.0573	2.5258	2.2383	7.8320	1.9149	
λ_4	2.1674	5.3144	7.3616	7.4047	2.1826		
λ_5	8.2216	8.6544	3.6105	2.3036			
λ_6	3.4056	3.9385	1.0248				
λ_7	3.4880	0.9443					
λ_8	1.0036						

The whole point of having breakpoints is to separate the distribution of accidents, see figure 3, in several intervals with different intensities, meaning that what happens before and after a breakpoint should be different. By looking at the table 1 we can see that for larger number of breakpoints some intensities are really close, which means that those particular breakpoints are unnecessary. This cutoff seem to happen at 5 breakpoints. Because for 6 and above some intensities are relatively close to each other (λ_1 and λ_2 with d = 6 for instance). We concluded from table 1 and figures 1 and 2 that 5 breakpoints seem to be a good amount of breakpoints.

From this conclusion we decided to plot the breakpoints on the curve received from the `'coal_mine_disasters.mat'` with the means of the breakpoints $t_1, ..., t_5$. As can be seen in the figure 3 we could roughly draw straight lines between the intersections of the breakpoints and the curve which would create a decent approximation of the curve. Moreover, the breakpoints seem to intersect points of change for the curve which is what we wanted to find.

Table 2: The values for the breakpoints in figure 3.

t_1	t_2	t_3	t_4	t_5
1752	1792	1835	1886	1915

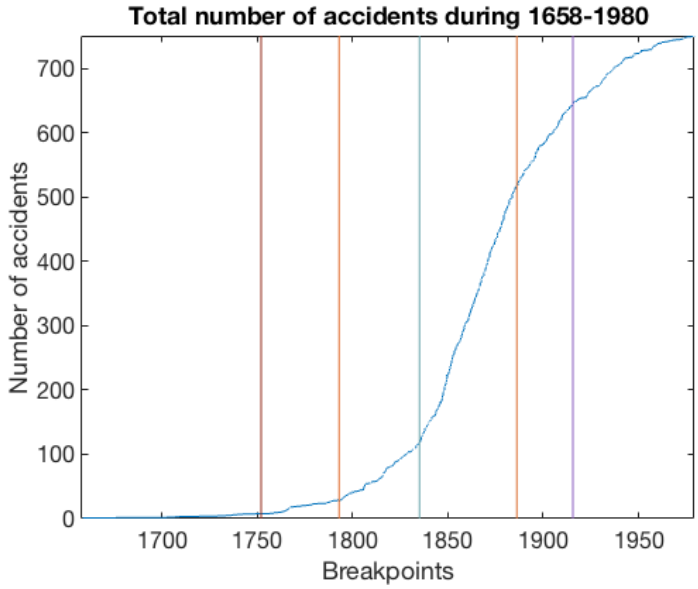
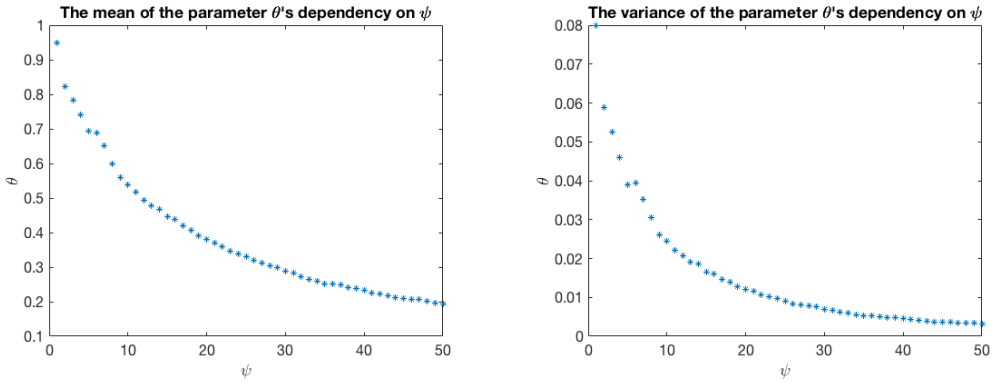


Figure 3: The figure illustrates the number of accidents for the time period 1658-1980. The data is taken from the file `'coal_mine_disaster.mat'`. The straight lines are the markings for the breakpoints. In this case we have d = 5 breakpoints which yields 6 lambda (intensity) values. The breakpoint values can be found in table 2.

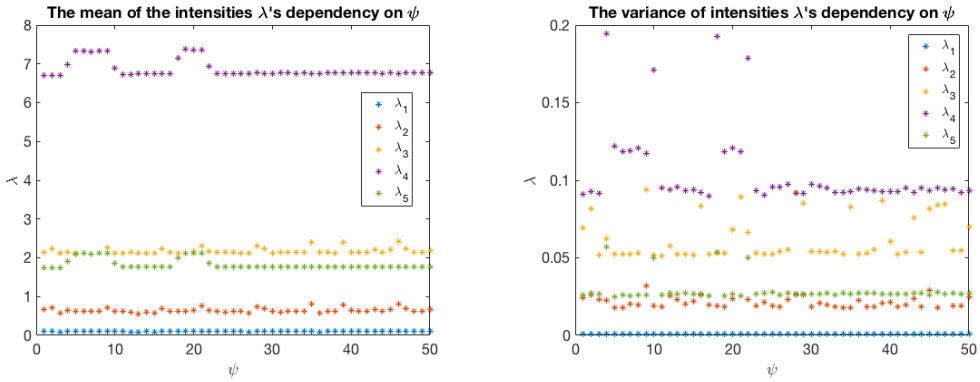
1.4 (d) Sensitivity of posteriors for choice of Ψ

Analyzing which posteriors depend on the hyperparameter Ψ we recall from equation (9) that $f(\theta \mid \boldsymbol{\tau}, \boldsymbol{t}, \boldsymbol{\lambda}) \propto \Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)$ from [1]. This means that there is a dependence for the mean and variance of θ on Ψ as $\mathbb{E}(\theta) \propto \Psi^{-1}$ and $\mathbb{V}(\theta) \propto \Psi^{-2}$ assuming everything else remains constant. In our simulations this seems to be the case as figure 4 shows.



(a) (b)

Figure 4: Displaying how the mean (a) and the variance (b) of θ depend Ψ. The parameter $\rho = 0.025$, burn-in was set to 1000 and sample size to 10000 for $d = 4$ breakpoints.

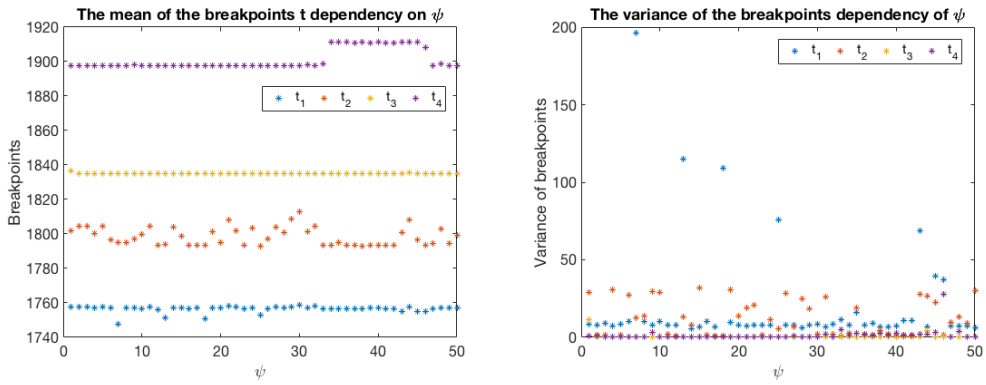


(a) (b)

Figure 5: Displaying how the mean (a) and the variance (b) of the intensity λ depend on Ψ. The parameter $\rho = 0.01$, burn-in was set to 1000 and sample size to 10000 for $d = 4$ breakpoints.

If θ is dependent on Ψ, then by consequence, it's reasonable to assume that λ is too. Recalling equation (10) $f(\boldsymbol{\lambda} \mid \theta, \boldsymbol{t}, \boldsymbol{\lambda}) \propto \Gamma(n_i(\tau) + 1, t_{i+1} - t_i + \theta)$ which obviously has a dependency on θ. However, investigating the Ψ-dependence on the mean of λ we arrive at figure 5a which does not display any large dependency on Ψ. Considering how small θ is in comparison to $t_{i+1} - t_i$ which is the difference between two breakpoints this is a reasonable result. The variance of λ will have a even smaller dependency on Ψ due to the same reasoning as for θ, seen in figure 5b.

Considering how Ψ affects the breakpoints t we recall equation (11). The posterior distribution of $\boldsymbol{t}$ has a dependency on λ, but since λ's dependency on θ was so small we do not expect any sizeable shift for the breakpoints.



(a) (b)

Figure 6: The t dependency on $\Psi = 1 : 50$. The parameters $\rho = 0.025$, burn-in was set to 1000 and sample size to 10000, for $d = 4$ breakpoints

Our simulation can be seen in figure 6 where it's obvious that there's no notable dependence on Ψ for the mean of the breakpoints (figure 6a). The variance of t dependent on Ψ in figure 6b doesn't seem to have a dependence on Ψ either.

1.5 (e) Mixing and posteriors in relation to ρ

The only algorithm directly dependent on ρ is the sampling of $\boldsymbol{t}$ using Metropolis Hastings. Studying the acceptance ratio we can get an understanding for how it will affect the draws of t. According to the lectures we want to have an acceptance ratio of 30% to get good mixing. However this proved rather difficult for some breakpoints. Considering figure 7 one would think that picking $0.02 < \rho < 0.04$ would guarantee an acceptance rate of about 30%.

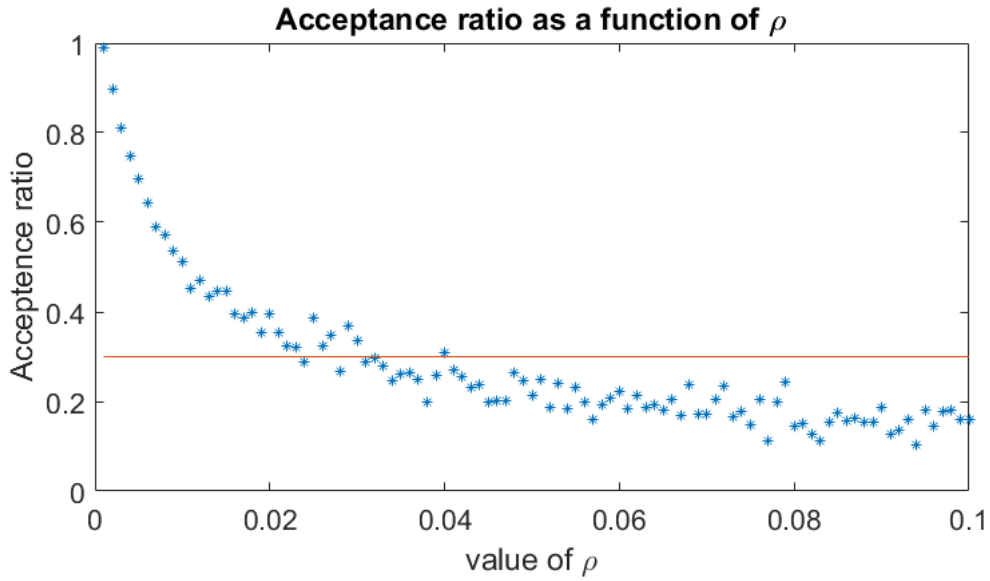


Figure 7: Average acceptance ratio as a function of ρ. 100 samples for each ρ was taken. To get good mixing we should have an acceptance ratio of about 30%.

In figure 8 we see the autocorrelation function for a certain breakpoint, in this case we chose t_1, for four different values of the tuning parameter ρ. We can clearly see that for a very small choice of ρ the dependency for the breakpoint sequence is very large. From figure 7 and 8 we can conclude that a good value for the tuning parameter should be somewhere between $0.02 < \rho < 0.04$ for this particular breakpoint t_1.

From figure 9 we can conclude that the tuning parameter ρ affects the breakpoints differently, thus one should choose a specific ρ-value for a specific breakpoint to achieve the optimal result.

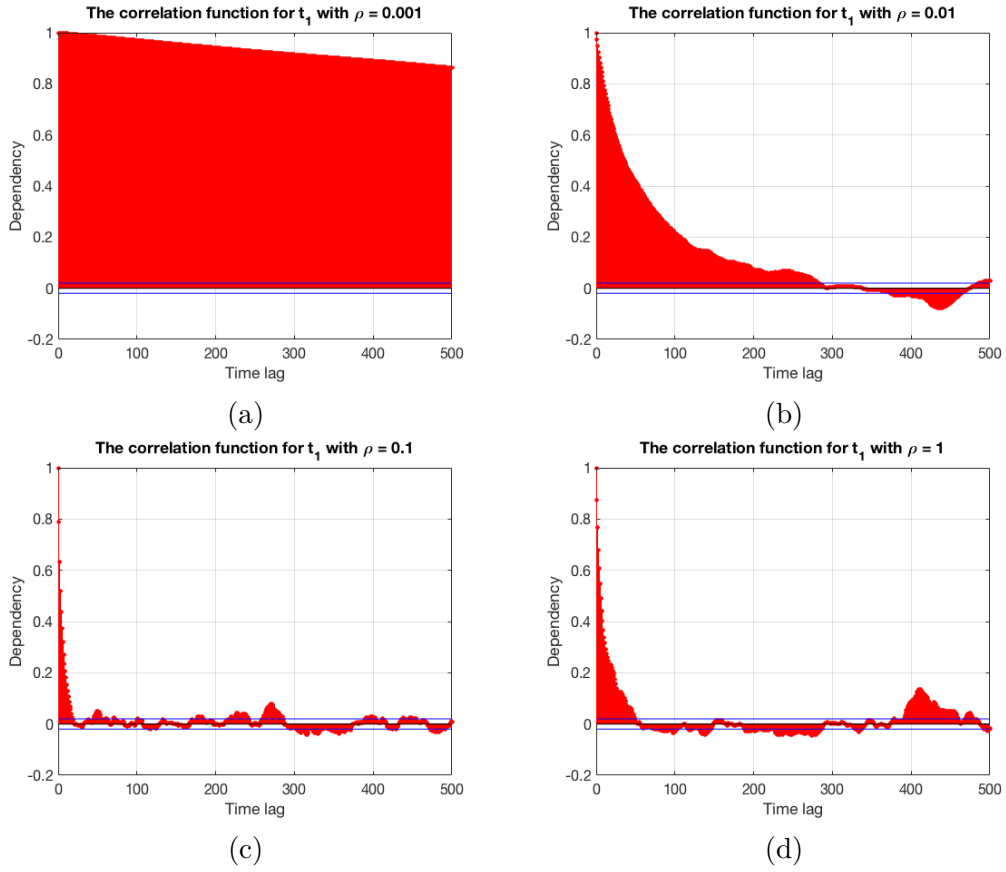


Figure 8: The autocorrelation functions for the breakpoint t_1 with varying ρ. The time lag was set 500.

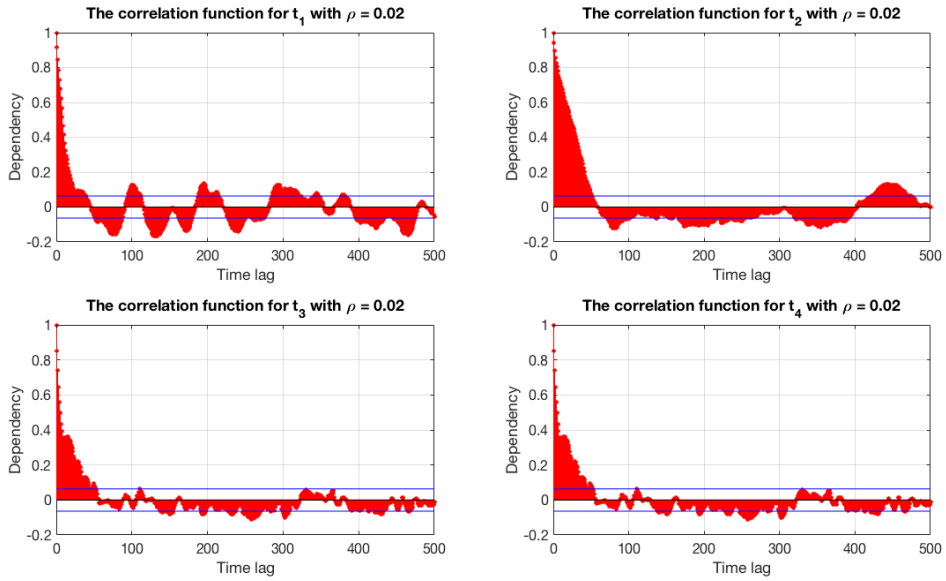
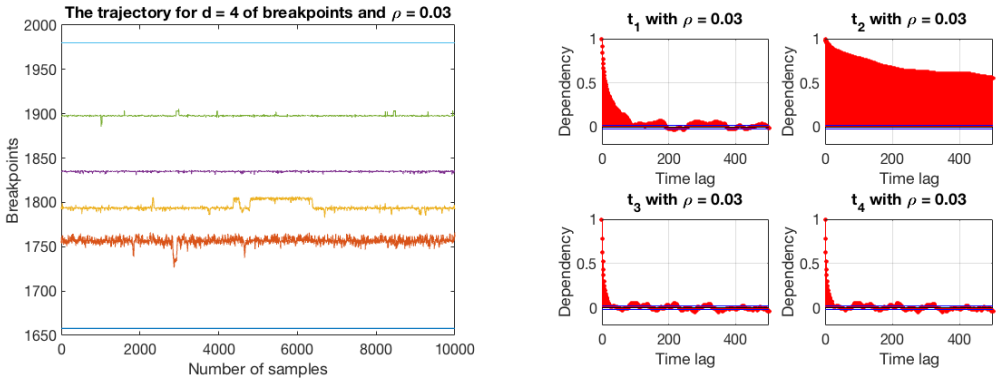


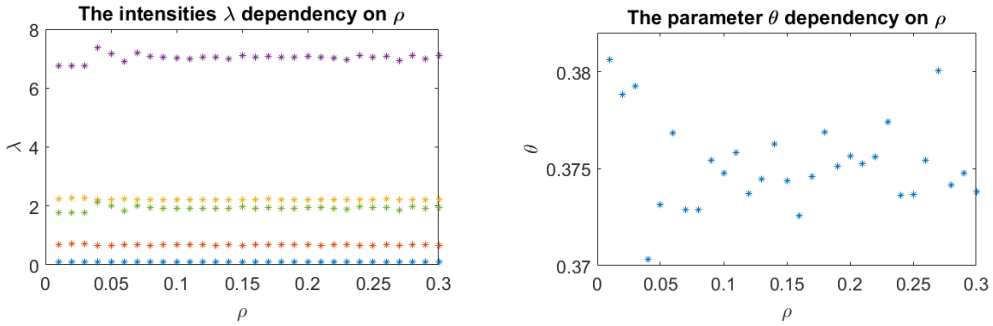
Figure 9: The correlation function for the breakpoints $t_{1:4}$ with $\rho = 0.02$. The time lags was set to 500.



(a) (b)

Figure 10: The trajectory plot for $\rho = 0.03$ can be seen in 10a and the correlation plots for $t_{1:4}$ can be seen in 10b. The time lag is set to 500.

In figure 10 we can observe the trajectory for the four breakpoints and the breakpoints correlation function. For t_2 we can see in the trajectory plot 10a that the chain jumps between two values, this makes the chain more time dependent, which affects its autocorrelation function.



(a) (b)

Figure 11: Plots describing the dependency on ρ for λ in figure 11a and θ 11b. $\Psi = 20$, burn in was set to 1000 and sample size to 10000.

We can conclude from figure 11 that the intensities and the parameter θ is more or less unaffected for different values on the tuning parameter ρ. Therefore, no further investigation was conducted.

2 The 100-year Atlantic wave

In this section we're going to analyze some properties of the 100-year Atlantic wave using the bootstrap method. The bootstrap method computes variation in statistics on the data itself. Meaning, the data is 'pulling itself up by its own bootstrap' [4]. We're going to use the parametric bootstrap. To realize this method you need a distribution that estimates the data you're trying to analyze. For the purposes of this assignment a Gumbel distribution was used, the CDF of which is described by

equation (13).

We wish to study the deviation from the mean of some parameters from the actual data and compare it to the mean of a random Gumbel-distributed sample. We will use (12) for our estimates.

$$\delta = \bar{x} - \bar{x}^* \tag{12}$$

Where $\bar{x}$ is a mean from the data in `'atlantic.txt'` and $\bar{x}^*$ is a mean of an estimate sampled from the Gumbel distribution. The idea is to compare the the actual mean of a measurement with the α-percentile of the difference between the means and randomly drawn samples from our test-distribution. Once these are drawn we simply sort our deltas and take the top and/or bottom α-percentile of the resulting vector to find our standard deviation.

2.1 (a) Finding the inverse

As stated earlier, the estimation distribution was considered known to be a Gumbel distribution (13).

$$F(x;\mu,\beta) = \exp\Big(-\exp\Big(-\frac{x-\mu}{\beta}\Big)\Big), \quad x \in \mathbb{R} \tag{13}$$

We aim to find it's inverse and recreate data samples from a uniformly distributed variable $u \in U(0,1)$. Denoting $F(x;\mu,\beta)$ as u and solving for x yields the inverse as equation (14)

$$\begin{aligned} &\ln\Big(\frac{1}{u}\Big) = \exp\Big(-\frac{x-\mu}{\beta}\Big) \Longleftrightarrow \ln\Big(\frac{1}{\ln(\frac{1}{u})}\Big) = \frac{x-\mu}{\beta} \\ &x = F^{-1}(u;\mu,\beta) = \beta \ln\Big(\frac{1}{\ln(\frac{1}{u})}\Big) + \mu \end{aligned} \quad . \tag{14}$$

2.2 (b) Estimating μ and β using bootstrap

We want to create a vector of δ's defined by (12) for μ and β where every index represents a sample compared with the mean and then sort it. To estimate $\bar{\mu}$ and $\bar{\beta}$ from the original data set a MATLAB-function called `'est_gumbel.m'` was used which accepts a dataset as input and return μ and β as if that dataset was a Gumbel-distribution.

Then a certain number of random samples was drawn from a Gumbel distribution using a uniformly distributed $u \in \mathcal{U}(0,1)$. This generated distribution was then plugged into `'est_gumbel.m'` to get draws for estimates of μ^* and β^* for this distribution. These estimates where subtracted from their respective means and all samples were put into a vector. Following the method described earlier for a double sided interval with $\alpha = 0.05$ the resulting confidence intervals can be found in table 3.

Table 3: Confidence intervals for the μ- and β- parameters using the bootstrap method.

Parameter	Lower Bound	Upper Bound
μ	1.3880	1.5786
β	4.0498	4.2405

2.3 (c) Estimating the upper bound of the 100-year wave

In this problem a very similar method to the earlier section was used but instead of the μ and β- parameters we're interested in the 100-year wave. We also know from the assignment [1] that the number of observations in 100 years is $T = 4200$. Meaning that if this value for T is plugged into our inverse Gumbel-distribution (14) will produce a random sample of the 100-year wave. Using the same scheme with a δ-vector as before, the upper bound was found to be **17.2575** meters with a one sided 95% confidence interval. We're only interested in the upper bound because of practical reasons. We want to find the worst case not the best case of a bad situation.

References

[1] *Home assignment 3*

[2] *Lecture 9*

[3] *Lecture 10*

[4] *Bootstrap confidence intervals*

[5] *Bootstrap estimate of bias*

[6] *Geof H. Givens and Jennifer A. Hoeting Computational Statistics Second Edition*