Homework Assignment 3

Alfred Bornefalk
Simen Söderberg

6 March 2022

Contents

1 Coal Mine Disasters **3**
1.1 Problem A: The Marginal Posteriors 3
1.2 Problem B: Constructing a Hybrid MCMC Algorithm 4
1.3 Problem C: The Behavior of the Chain for Different Choices of Breakpoints 5
1.4 Problem D: Posterior Sensitivity for Different Ψ's 9
1.5 Problem E: Sensitivity of the Mixing and the Posteriors for ρ . . 13

2 The 100-year Atlantic Wave **17**
2.1 Problem A: Inverse of the Gumbel Distribution 17
2.2 Problem B: Parametric Bootstrap 18
2.3 100-year Return Value 18

3 References **20**

1 Coal Mine Disasters

1.1 Problem A: The Marginal Posteriors

To calculate the marginal posteriors, we begin by rewriting them using the formula for conditional probability:

$$\begin{cases} \mathrm{f}(\theta \mid \lambda, t, \tau) = \frac{f(\theta,\lambda,t,\tau)}{f(\lambda,t,\tau)} \\ \mathrm{f}(\lambda \mid \theta, t, \tau) = \frac{f(\theta,\lambda,t,\tau)}{f(\theta,t,\tau)} \\ \mathrm{f}(\mathrm{t} \mid \theta, \lambda, \tau) = \frac{f(\theta,\lambda,t,\tau)}{f(\theta,\lambda,\tau)} \end{cases} \tag{1}$$

From equation (1), we can see that all conditional probabilities have the same numerator, i.e. $f(\theta, \lambda, t, \tau)$. From the assignment, we have the prior $\lambda \sim \Gamma(2, \theta)$ and the hyperprior $\theta \sim \Gamma(2, \Psi)$, where Ψ is a fixed hyperparameter. The following two probability density functions are also given

$$f(t) \propto \begin{cases} \prod_{i=1}^{d}(t_{i+1} - t_i), & t_1 < t_2 < ... < t_d < t_{d+1} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$f(\tau \mid \lambda, t) = exp(-\sum_{i=1}^{d} \lambda_i(t_{i+1} - t_i)) \prod_{i+1}^{d} \lambda_i^{n_i(\tau)} \tag{3}$$

where,

$$n_i(\tau) = \text{number of disasters in the sub-interval } [t_i, t_{i+1}) = \sum_{j=1}^{n} \mathbb{1}_{[t_i, t_{i+1})}(\tau_j)$$

We can now express our nominator from (1) in our known probability density functions. By rewriting the nominator with some basic conditional probability rules we get

$$f(\theta, \lambda, t, \tau) = f(\tau \mid \lambda, t, \theta) f(\lambda, t, \theta) = f(\tau \mid \lambda, t, \theta) f(t \mid \lambda, \theta) f(\lambda, \theta) =$$

$$f(\tau \mid \lambda, t, \theta) f(t \mid \lambda, \theta) f(\lambda \mid \theta) f(\theta)$$

From the equation above we recognize that $f(\theta)$ and $f(\lambda \mid \theta)$ are known. From the exercise, we know that $f(t \mid \lambda, \theta) = f(t)$ since t is independent of λ and θ. τ is also independent of θ, thus we can rewrite $f(\tau \mid \lambda, t, \theta)$ as $f(\tau \mid \lambda, t)$. Consequently, we can rewrite the nominator from (1) as

$$f(\theta, \lambda, t, \tau) = f(\tau \mid \lambda, t) f(t) f(\lambda \mid \theta) f(\theta) \tag{4}$$

which we can rewrite as

$$f(\theta, \lambda, t, \tau) =$$

$$exp(-\sum_{i=1}^{d} \lambda_i(t_{i+1} - t_i)) \prod_{i=1}^{d} \lambda_i^{n_i(\tau)} \prod_{i=1}^{d} (t_{i+1} - t_i) \prod_{i=1}^{d} \frac{\theta^2}{\Gamma(2)} \lambda_i exp(-\theta\lambda_i) \frac{\Psi^2}{\Gamma(2)} \lambda_i exp(-\Psi\lambda_i)$$

To compute the marginal posteriors in (1), we note that we can remove the terms that is independent of the variable. We begin by observing the marginal posterior of θ:

$$f(\theta \mid \lambda, t, \tau) \propto \prod_{i=1}^{d} \frac{\theta^2}{\Gamma(2)} \lambda_i exp(-\theta\lambda_i) \frac{\Psi^2}{\Gamma(2)} \lambda_i exp(-\Psi\lambda_i)$$

$$\propto exp(-(\Psi + \sum_{i+1}^{d} \lambda_i)\theta)\theta^{2d+1} \propto \Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)$$

Thus, we can conclude that the marginal posterior $f(\theta \mid \lambda, t, \tau)$ is proportional to $\Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)$. Moving on with λ, we get

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

$$\propto \prod_{i=1}^{d} exp(-\lambda_i(t_{i+1} - t_i + \theta))\lambda_i^{n_i(\tau)+1} \propto \Gamma(n_i(\tau) + 2, t_{i+1} - t_i + \theta)$$

We can once again conclude that one of our marginal posteriors is proportional to a gamma distribution. Lastly, we observe the last marginal posterior:

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

Unfortunately, we can not find a known distribution for t.

1.2 Problem B: Constructing a Hybrid MCMC Algorithm

In part B of this exercise, the objective is to implement a hybrid MCMC algorithm which samples from the posterior $f(\theta, \lambda, t \mid \tau)$. We have that the distributions for both $f(\theta \mid \lambda, t, \tau)$ and $f(\lambda \mid \theta, t, \tau)$ are known, which means that we can sample them using a Gibbs sampler. The distribution for $f(t \mid \theta, \lambda, \tau)$ is not known and we need to sample from it using an MCMC algorithm. More specifically, we will use a Gibbs sampler combined with a Metropolis-Hastings algorithm. To construct the Gibbs sampler, we simulate a sequence of values (X_k) which forms a Markov chain on X according to the following algorithm:

Given X_k
draw $X_{k+1}^1 \sim f_1(x^1 \mid X_k^2, ..., X_k^m)$
draw $X_{k+1}^2 \sim f_2(x^2 \mid X_{k+1}^1, X_k^3, ..., X_k^m)$
draw $X_{k+1}^3 \sim f_3(x^3 \mid X_{k+1}^1, X_{k+1}^2, X_k^4 ..., X_k^m)$
...
draw $X_{k+1}^m \sim f_m(x^m \mid X_{k+1}^1, X_{k+1}^2, ..., X_{k+1}^{m-1})$

To construct the MH-algorithm that is used when sampling from the unknown distribution $f(t \mid \theta, \lambda, \tau)$ as in the Gibbs sampler, we simulate a sequence of values (X_k) which forms a Markov chain on X according to this algorithm:

Given X_k
generate $X^* \sim r(z \mid X_k)$
set $X_{k+1} = \begin{cases} \text{X}^* \text{ with probability } \alpha(X_k, X^*) \stackrel{\text{def}}{=} 1 \wedge \frac{f(X^*)r(X_k \mid X^*)}{f(X_k)r(X^* \mid X_k)} \\ \text{X}_k \text{ otherwise} \end{cases}$

$r(z \mid X_k)$ is the transition density or the proposal kernel on X which we sample from. f is the marginal distribution for t, i.e. $f(t \mid \theta, \lambda, \tau)$. In this assignment, the proposal kernel is given and it is a random walk proposal one at a time. We define it as: Update one breakpoint at a time. For each breakpoint t_i we generate a candidate t_i^* through

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

$$R = \rho(t_{i+1} - t_{i-1})$$

where ρ is a tuning parameter of the proposal distributions. We note that the proposal kernel is symmetric, since the distribution $\mathbb{U}(-R, R)$ is symmetric. Then, it holds that $r(z \mid X_k) = r(X_k \mid z), \forall (X_k, z) \in X^2$. This property leads to the following for the acceptance probability that we can utilize in the last step of the MH-algorithm

$$\alpha(X_k, X^*) \stackrel{\text{def}}{=} 1 \wedge \frac{f(X^*)r(X_k \mid X^*)}{f(X_k)r(X^* \mid X_k)} = \frac{f(X^*)}{f(X_k)}$$

1.3 Problem C: The Behavior of the Chain for Different Choices of Breakpoints

In this problem, we investigate the behavior of the chain for different numbers of breakpoints. We decide to implement our algorithm for $d = 1, 2, 3, 4, 5$. However, we begin by plotting the number of accumulated number of accidents over the period 1658-1980. The plot can be seen in figure 1. For all simulations in this assignment, we are using a burn-in of $1,000$ and a sample size of $10,000$.

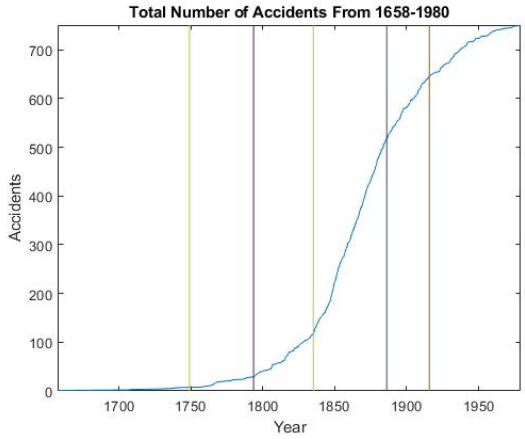

Figure 1: Plot of the total number of accidents from 1658-1980. The vertical lines are the five different breakpoints for $d = 5$. The blue graph showcases the cumulative number of accidents in the investigated period.

From figure 1, we note that the number of accidents are close to zero for the first years. Thus, we decide to delimit the data set by setting the starting point at $t = 1690$. Furthermore, the rate of accidents dramatically increase in the end. This leads to us setting the ending point at $t = 1960$. We plot the behavior of the chain for the d different number of breakpoints below.

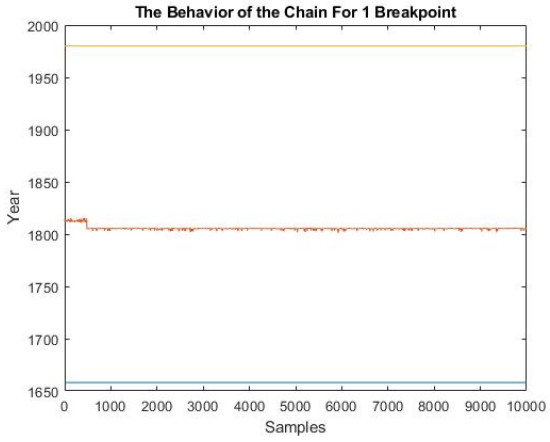

Figure 2: The behavior of the chain for $d = 1$, $\psi = 20$, and $\rho = .01$. The blue horizontal line represents the starting point 1690, whereas the red line corresponds to t_1, and the yellow line to the ending point 1960.

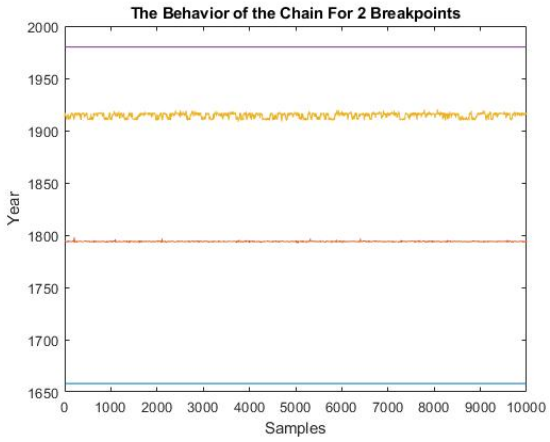

Figure 3: The behavior of the chain for $d = 2$, $\psi = 20$, and $\rho = .01$. The blue horizontal line represents the starting point 1690, whereas the red line corresponds to t_1, the yellow line to t_2, and the purple line to the ending point 1960.

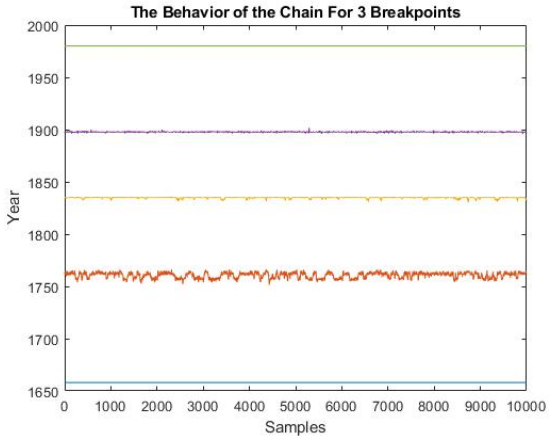

Figure 4: The behavior of the chain for $d = 3$, $\psi = 20$, and $\rho = .01$. The blue horizontal line represents the starting point 1690, whereas the red line corresponds to t_1, the yellow line to t_2, the purple line to t_3, and the green line to the ending point 1960.

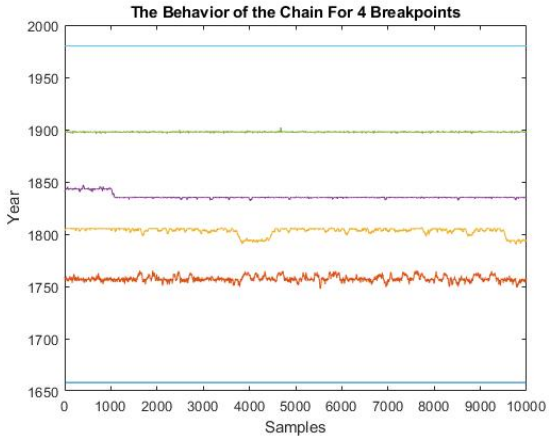

Figure 5: The behavior of the chain for $d = 4$, $\psi = 20$, and $\rho = .01$. The blue horizontal line represents the starting point 1690, whereas the red line corresponds to t_1, the yellow line to t_2, the purple line to t_3, the green line to t_4, and the cyan line to the ending point 1960.

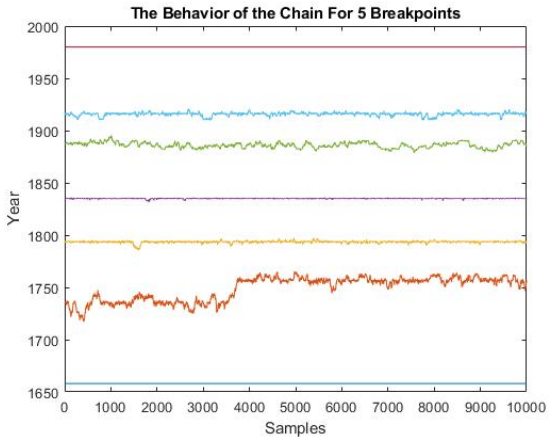

Figure 6: The behavior of the chain for $d = 5$, $\psi = 20$, and $\rho = .01$. The blue horizontal line represents the starting point 1690, whereas the red line corresponds to t_1, the yellow line to t_2, the purple line to t_3, the green line to t_4, the cyan line to t_5, and the pink line to the ending point 1960.

Observing the figures 2 to 6 above, the chain seems to converge to similar values when we increment d. When comparing this result to figure 1, it is deemed

reasonable to henceforth set $d = 5$. The observed steepness of the curve fits those breakpoints fairly well. The chosen breakpoints for d are displayed in table 1.

Table 1: The estimated breakpoints t_i, $i = \{1, 2, 3, 4, 5\}$, for $d = 5$.

Breakpoint	Year
t_1	1752
t_2	1792
t_3	1835
t_4	1886
t_5	1915

1.4 Problem D: Posterior Sensitivity for Different Ψ's

In this exercise, the aim is to analyze how sensitive the posteriors are for different values of the hyperparameter Ψ. We do this by studying how the mean and variance changes for values of Ψ where we set $\Psi = \{1, 2, ..., 50\}$. We start by observing how sensitive θ is for changes in Ψ. The following two plots shows how the mean and variance changes.

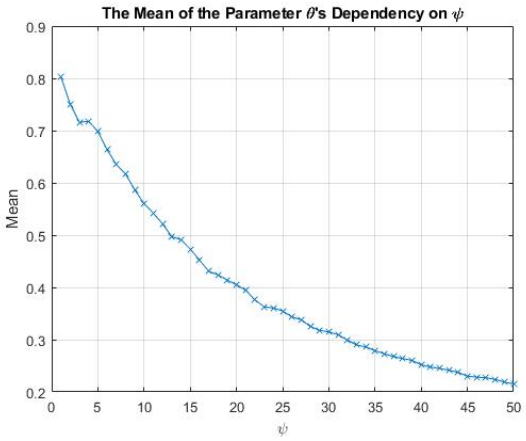

Figure 7: The mean of the θ parameter for $\Psi = \{1, 2, ..., 50\}$, and $\rho = .01$.

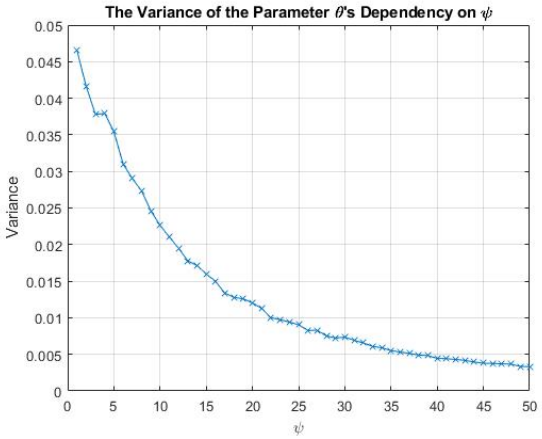

Figure 8: The variance of the θ parameter for $\Psi = \{1, 2, ..., 50\}$, and $\rho = .01$.

Since the marginal posterior $f(\theta|\lambda, t, \tau)$ is proportional to $\Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)$, we note that the mean and variance of θ related to Ψ, ceteris paribus, is $\mathbb{E}[\theta] = \frac{1}{\Psi}$ and $\mathbb{V}[\theta] = \frac{1}{\Psi^2}$. This seems reasonable when studying figure 7 and 8. Moving on, we study how sensitive λ_i, $i = \{1, 2, ..., 6\}$, is for Ψ. The following plots display how the mean and variance changes.

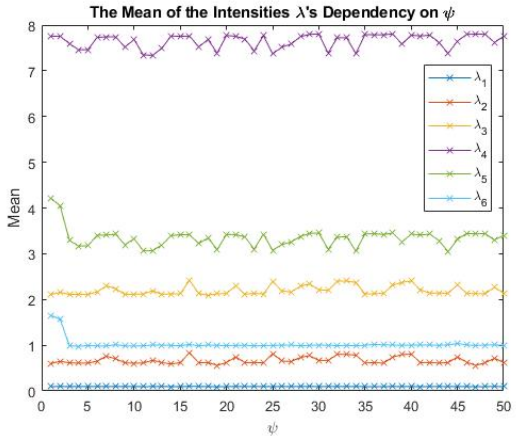

Figure 9: The mean of the intensities λ_i, $i = \{1, 2, ..., 6\}$, for $\Psi = \{1, 2, ..., 50\}$, and $\rho = .01$. The legends bar connects the graphs to the corresponding λ_i.

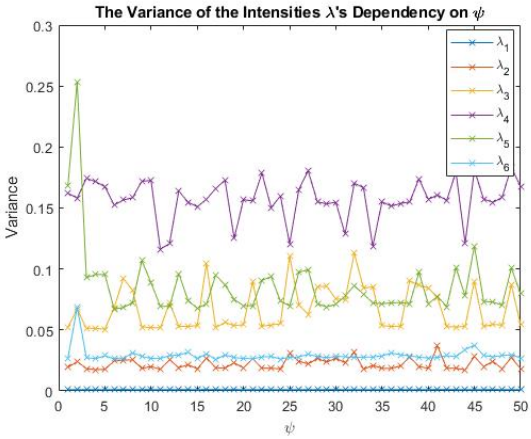

Figure 10: The variance of the intensities λ_i, $i = \{1, 2, ..., 6\}$, for $\Psi = \{1, 2, ..., 50\}$, and $\rho = .01$. The legends bar connects the graphs to the corresponding λ_i.

From figure 9 and 10, it is quite clear that the mean and variance of λ is not substantially sensitive to changes in Ψ. The marginal posterior $f(\lambda \mid \theta, t, \tau)$ is proportional to $\Gamma(n_i(\tau) + 2, t_{i+1} - t_i + \theta)$. We note that λ is dependent on θ, which has been shown earlier to be quite heavily affected to the changes in Ψ. However, $t_{i+1} - t_{t-1}$ is much larger than θ, thus affecting the mean and variance much more, leading to changes in Ψ having a small effect. Next, we study how changes in Ψ affect the marginal distribution $f(t \mid \theta, \lambda, \tau)$. The following plots shows how the mean and variance of t vary for different values of Ψ.

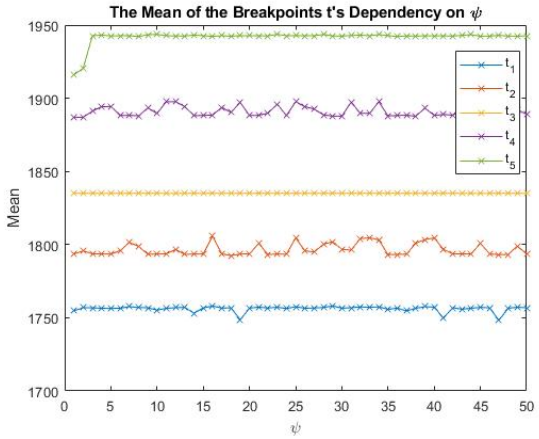

Figure 11: The mean of the breakpoints t_i, $i = \{1, 2, 3, 4, 5\}$, for $\Psi = \{1, 2, ..., 50\}$, and $\rho = .01$. The legends bar connects the graphs to the corresponding t_i.

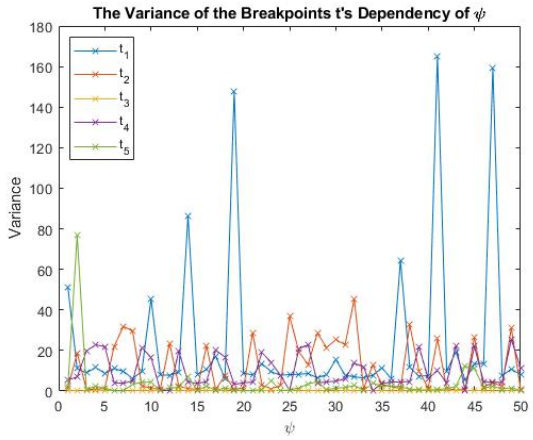

Figure 12: The variance of the breakpoints t_i, $i = \{1, 2, 3, 4, 5\}$, for $\Psi = \{1, 2, ..., 50\}$, and $\rho = .01$. The legends bar connects the graphs to the corresponding t_i.

The marginal posterior of t has dependency on λ. However, since the dependency between λ and Ψ is very small, the dependency between t and Ψ should be small as well. Figure 11 and 12 supports this argument; t does not seem to be substantially dependent on Ψ.

1.5 Problem E: Sensitivity of the Mixing and the Posteriors for ρ

In this subsection, we analyze how sensitive our marginal posteriors are for different values of ρ. We note that t is the only parameter that is dependent on the tuning parameter ρ directly since it is sampled from the Metropolis-Hastings sampler. For the MH-algorithm, we aim to have an acceptance ratio $\alpha(X_k, X^*)$ between 20-30% (Oehm, 2018). Below, we plot the acceptance ratio as a function of ρ to find a suitable choice of ρ.

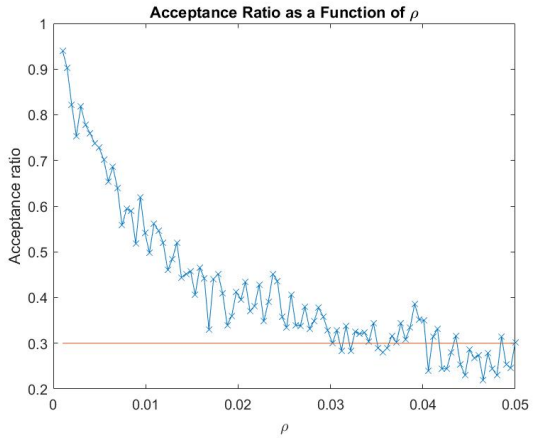

Figure 13: The acceptance ratio as a function of ρ, for $\psi = 20$, and $\rho = [.001, .05]$. The blue graph displays how the acceptance ratio varies with ρ, and the horizontal orange line holds a constant value of .30.

From figure 13, we see that a suitable value for ρ seems to be somewhere between .03 and .04. This indicates that the previously chosen ρ of .01 was not optimal. As previously mentioned, t was the only parameter directly dependent on ρ. Consequently, it is not a far-fetched hypothesis that θ and λ will not be significantly affected by different values for ρ. We plot how the mean of θ and λ changes with respect to ρ below.

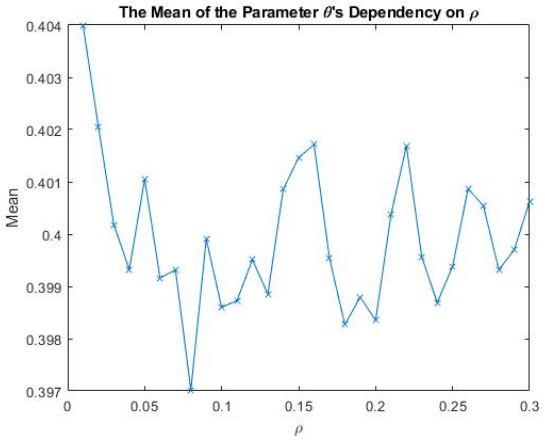

Figure 14: The mean of the parameter θ for $\psi = 20$, and $\rho = .01, .02, ..., .30$.

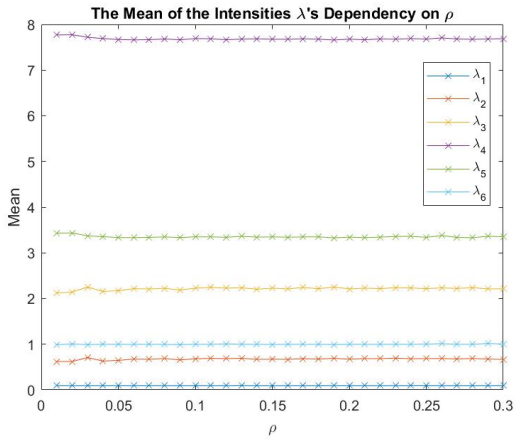

Figure 15: The mean of the intensities λ_i, $i = \{1, 2, ..., 6\}$, for $\psi = 20$, and $\rho = \{.01, .02, ..., .30\}$. The legends bar connects the (almost) horizontal lines to the corresponding λ_i.

In line with our hypothesis, the results from figure 14 and 15 show that both θ and λ are more or less unaffected by different values of ρ. Therefore, we move on with a more rigorous analysis of t. Previously, we saw that a value of $\rho = .03$ was a suitable value with regards to a good acceptance ratio, yielding a ratio approximately equal to 30%. Thus, we plot how the behavior of the chain varies for $\rho = .03$ below.

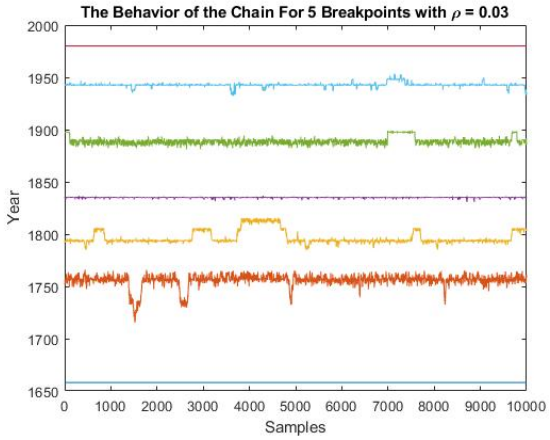

Figure 16: The behavior of the chain for $\psi = 20$, and $\rho = .03$. The blue horizontal line represents the starting point 1690, whereas the red line corresponds to t_1, the yellow line to t_2, the purple line to t_3, the green line to t_4, the cyan line to t_5, and the pink line to the ending point 1960.

From figure 16, we note that t_4 and t_5 shows a similar pattern. In the further analysis, we will therefore only consider the first four breakpoints. Considering the figure output, we expect t_3 to be the least time-dependent breakpoint. In order to confirm this hypothesis, we plot the auto-correlation functions for the first four breakpoints below.

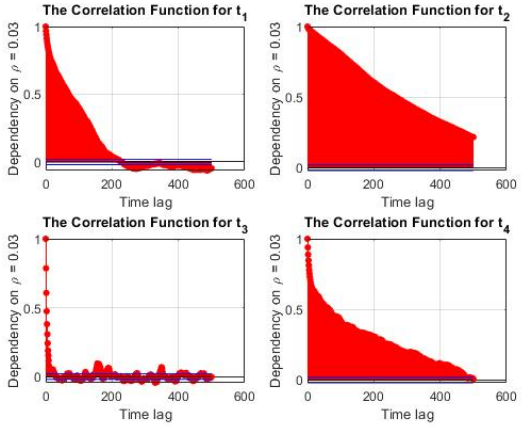

Figure 17: Subplot of the auto-correlation functions for the breakpoints t_i, $i = \{1, 2, 3, 4\}$, for $\psi = 20$, and $\rho = .03$.

Figure 17 supports the argument that t_3 is the least time-dependent since its auto-correlation functions quickly goes to zero. Additionally, with the same argument, t_2 is the most time-dependent breakpoint, which is further supported by analyzing figure 16. Moving on, we plot the auto-correlation functions for $\rho = \{.01, .02, .04\}$ to understand how the chain is affected by different ρ's.

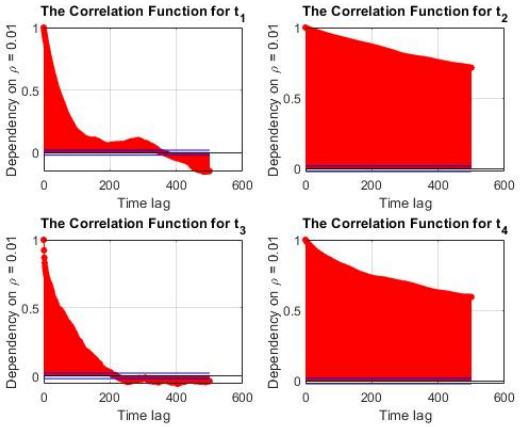

Figure 18: Subplot of the auto-correlation functions for the breakpoints t_i, $i = \{1, 2, 3, 4\}$, for $\psi = 20$, and $\rho = .01$.

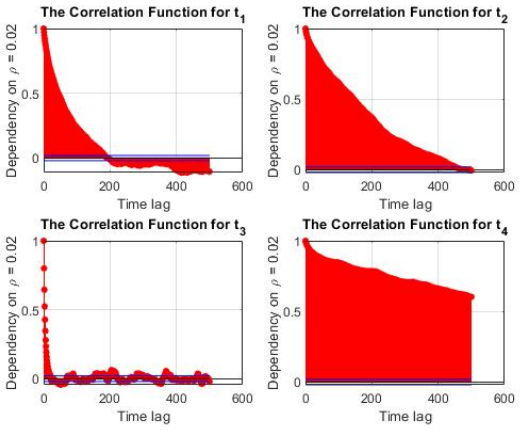

Figure 19: Subplot of the auto-correlation functions for the breakpoints t_i, $i = \{1, 2, 3, 4\}$, for $\psi = 20$, and $\rho = .02$.

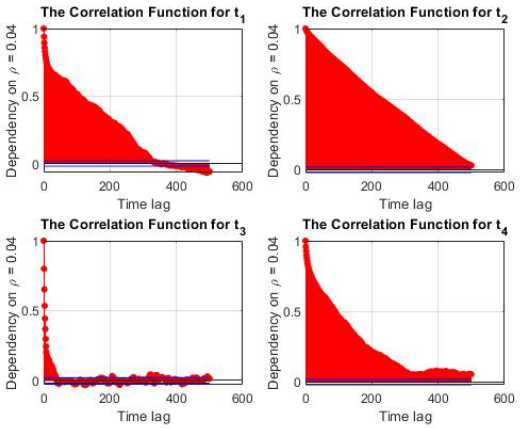

Figure 20: Subplot of the auto-correlation functions for the breakpoints t_i, $i = \{1, 2, 3, 4\}$, for $\psi = 20$, and $\rho = .04$.

From the figures above, we can conclude that the auto-correlations for $\rho = .01$ and $\rho = .02$ are more time-dependent overall compared to when we set $\rho = .03$ and $\rho = .04$. This further suggests that choosing ρ in the interval $[.03, .04]$ yields the most efficient proposal distribution.

2 The 100-year Atlantic Wave

2.1 Problem A: Inverse of the Gumbel Distribution

In this part of the exercise, we aim to find the inverse of the Gumbel distribution. It is defined as

$$F(x; \mu, \beta) = exp(-exp(-\frac{x-\mu}{\beta})), x \in \mathbb{R} \tag{5}$$

where $\mu \in \mathbb{R}$ and $\beta > 0$ is a good fit to the data. Our objective is then to find the following inverse $F^{-1}(u; \mu, \beta)$. We set $u = F(x; \mu, \beta)$ and solve (5) for x.

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

Thus, the following inverse has been found

$$F^{-1}(u; \mu, \beta) = \mu - \beta ln(-ln(u)) \tag{6}$$

2.2 Problem B: Parametric Bootstrap

Now, we are to estimate confidence intervals for β and μ using parametric bootstrapping. We begin by estimating $\hat{\beta}$ and $\hat{\mu}$ from the given observations in the file "atlantic.txt" and the function "est_gumbel.m", which returns the estimated parameters for our dataset. This yields the following estimates:

$$\hat{\beta} = 1.4858$$

$$\hat{\mu} = 4.1477$$

Thus, the expected values for our parameters have been estimated. We assume that the data comes from the Gumbel distribution in (5). We generate new bootstrapped samples Y_b^* from the inverse Gumbel distribution in (6) through the following algorithm:

for $i = 1 \to N$ do
    draw $u \sim U(0, 1)$
end for
set $Y_b^* \leftarrow F^{-1}(u; \hat{\mu}, \hat{\beta})$
return Y_b^*

From our recently bootstrapped samples Y_b^*, we perform our bootstrap estimates of the parameters β^* and μ^*. We do this B times, where B is set to 1000. From the bootstrapped sampled parameter estimations we calculate the difference between the estimated parameters from "atlantic.txt" and our bootstrap parameter estimations. The following errors are calculated:

$$\Delta_b^* = \hat{\beta}(Y_b^*) - \hat{\beta}$$

$$\Delta_b^* = \hat{\mu}(Y_b^*) - \hat{\mu}$$

From the error terms we can now form a 95% confidence interval for our parameters. The confidence interval for $\hat{\beta}$ is

$$I_{.05}^{\hat{\beta}} = (1.462, 1.507)$$

and the confidence interval for $\hat{\mu}$ is

$$I_{.05}^{\hat{\mu}} = (4.117, 4.175)$$

2.3 100-year Return Value

In the last subsection of the home assignment, we can utilize our method presented in the previous subsection with slight adjustments. We know that the total number of observations during a 100-year period is $T = 3 \cdot 14 \cdot 100 = 4200$. The T^{th} return value of the wave height is

$$F^{-1}(1 - \frac{1}{T}; \mu, \beta) = 16.544$$

We construct the confidence interval by sampling the wave heights with $u = 1 - \frac{1}{T}$ and our previously sampled $\hat{\beta}(Y_b^*)$ and $\hat{\mu}(Y_b^*)$ for each b. Using a one sided 95 % confidence interval, the upper bound is found to be 17.257.

3 References

Oehm, D 2018, 'Metropolis-Hastings Algorithm from Scratch', *Gradient Descending*.