LUND UNIVERSITY | LTH FACULTY OF ENGINEERING

Home Assignment 3

Felix E. Slothower
Theodora M. Gaiceanu

March 7, 2022

Task 1

a

To calculate the marginal posteriors $f(\theta|\boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})$ $f(\boldsymbol{\lambda}|\theta, \boldsymbol{t}, \boldsymbol{\tau})$ $f(\boldsymbol{\tau}|\theta, \boldsymbol{\lambda}, \boldsymbol{t})$ we introduce the following expression for conditional probabilities;

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

From this we can express the marginal posteriors in the following way;

$$f(\theta|\boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau}) = \frac{f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})}{f(\boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})} \tag{2}$$

$$f(\boldsymbol{\lambda}|\theta, \boldsymbol{t}, \boldsymbol{\tau}) = \frac{f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})}{f(\theta, \boldsymbol{t}, \boldsymbol{\tau})} \tag{3}$$

$$f(\boldsymbol{t}|\theta, \boldsymbol{\lambda}, \boldsymbol{\tau}) = \frac{f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})}{f(\theta, \boldsymbol{\lambda}, \boldsymbol{\tau})} \tag{4}$$

We notice that each of these expressions share a common numerator. We will try to find an expression for this numerator by expanding it into many parts so that we can apply the distributions given to us in the problem statement. We do this by using the same relation in equation 1 but re-written such that;

$$P(A \cap B) = P(A|B)P(B)$$

and now expand this property to a probability conditioned on four variables;

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

We will now expand $f(\theta, \boldsymbol{\lambda}, \boldsymbol{t}, \boldsymbol{\tau})$ in the numerator of equations 2,3,4 using the same technique as above, trying to extract probabilities which are given to us in the problem statement.

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

From this we can see that $f(\theta)$ and $f(\boldsymbol{\lambda}|\theta)$ are both given. Also, because $\boldsymbol{t}$ is not dependent on any other variable, $f(\boldsymbol{t}|\boldsymbol{\lambda},\theta)$ can be reduced to $f(\boldsymbol{t})$. Finally $f(\boldsymbol{\tau}|\boldsymbol{t},\boldsymbol{\lambda},\theta)$ can be modified to $f(\boldsymbol{\tau}|\boldsymbol{\lambda},\boldsymbol{t})$ since $\boldsymbol{\tau}$ does not depend on θ. Re-writing this new expression, and introducing the know distributions yields;

$$f(\boldsymbol{\tau},\boldsymbol{t},\boldsymbol{\lambda},\theta) = f(\boldsymbol{\tau}|\boldsymbol{t},\boldsymbol{\lambda})f(\boldsymbol{t})f(\boldsymbol{\lambda}|\theta)f(\theta) \tag{5}$$

$$= \exp\left(-\sum_{i=1}^{d}\lambda_i\left(t_{i+1}-t_i\right)\right)\prod_{i=1}^{d}\lambda_i^{n_i(\tau)}\prod_{i=1}^{d}\left(t_{i+1}-t_i\right)\prod_{i=1}^{d}\left(\frac{\theta^2}{\Gamma(2)}\lambda_i e^{-\theta\lambda_i}\right)\frac{\Psi^2}{\Gamma(2)}\theta e^{-\Psi\theta} \tag{6}$$

Since we are not asked to find expressions for the normalizing constants and are therefore only concerned with finding proportionality, we find that the marginal posteriors are proportional to the above expression where any distributions not dependent on the variable in question are discarded. For example with $f(\theta|\boldsymbol{\lambda},\boldsymbol{t},\boldsymbol{\tau})$ we will remove terms in equation 5 which do not depend on θ.

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

This expression can be algebraically rearranged to resemble the probability density function of a single gamma distribution as follows;

$$\prod_{i=1}^{d}\left(\frac{\theta^2}{\Gamma(2)}\lambda_i e^{-\theta\lambda_i}\right)\frac{\Psi^2}{\Gamma(2)}\theta e^{-\Psi\theta} = \left[\prod_{i=1}^{d}(\lambda_i)\frac{\Psi^2}{\Gamma(2)}\right]\theta^{2d+1}e^{-(\Psi+\sum_{i=1}^{d}\lambda_i)\theta}$$

From this we can see a resemblance to the gamma function with parameters;

$$f(\theta|\boldsymbol{\lambda},\boldsymbol{t},\boldsymbol{\tau}) \sim \Gamma(2d+2, \Psi+\sum_{i=1}^{d}\lambda_i)$$

The same procedure can be conducted for the two other marginal posteriors. Starting with $f(\boldsymbol{\lambda}|\theta,\boldsymbol{t},\boldsymbol{\tau})$.

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

Which can be rewritten in the form;

$$\exp\left(-\sum_{i=1}^{d}\lambda_i\left(t_{i+1}-t_i\right)\right)\prod_{i=1}^{d}\lambda_i^{n_i(\tau)}\prod_{i=1}^{d}\left(\frac{\theta^2}{\Gamma(2)}\lambda_i e^{-\theta\lambda_i}\right) = \tag{7}$$

$$\prod_{i=1}^{d}\left[\frac{\theta^2}{\Gamma(2)}\right]e^{-\lambda_i(t_{i+1}-t_i+\theta)}\lambda_i^{n_i(\tau)+1} \tag{8}$$

We can again find a resemblance to the gamma distribution in equation 8, this time with respect to individual intensities λ_i;

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

Finally, a proportional relationship can be found for $f(\boldsymbol{\tau}|\theta,\boldsymbol{\lambda},\boldsymbol{t})$ where;

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

For this marginal posterior we could not identify a distribution. It has the elements of a Gamma distribution with respect to the intensity $\boldsymbol{\lambda}$ but not for the parameter $\boldsymbol{t}$.

b

A hybrid MCMC algorithm is used in order to sample from the posterior $f(\theta,\lambda,t|\tau)$. This means that one can use Gibbs sampling in order to update $f(\theta|\lambda,t,\tau)$ and $f(\lambda|\theta,t,\tau)$ (the marginal posteriors have a Gamma distribution). But for the breakpoints $\mathbf{t}$, one should use Metropolis-Hastings sampler, as $f(t|\lambda,\tau,\theta)$ does not have a known distribution.

According to [3], the Gibbs sampler forms a Markov chain on X by simulating a sequence of values (X_k). It uses several conditional distributions in order to approximate a joint distribution. The steps are:

- draw $X_{k+1}^1 \sim f_1(x^1|X_k^2,...,X_k^m)$
- draw $X_{k+1}^2 \sim f_1(x^2|X_{k+1}^1,X_k^3...,X_k^m)$
- ...
- draw $X_{k+1}^m \sim f_m(x^m|X_{k+1}^1,X_{k+1}^2,...,X_{k+1}^{m-1})$

This mechanism is used to update the marginal posteriors $f(\theta|\lambda,t,\tau)$ and $f(\lambda|\theta,t,\tau)$, using their expressions from Task 1a). The distribution for $f(\lambda|\theta,t,\tau)$ is implemented in the *lambda_posterior* function.

Since the distribution of $f(t|\lambda,\tau,\theta)$ is not known, one uses the Metropolis-Hastings algorithm to simulate a sequence of values X_k (in our case (t_k)), forming a Markov chain on X (on t), through the following mechanism: given X_k (t_k),

- generate $X^* \sim r(z|X_k)$ or generate $t^* \sim r(z|t_k)$

  According to [4], $r(z|X_k)$ is the transition density from which we sample, and it is referred to as the proposal kernel. In this assignment, a random walk proposal one at a time was used. This means we update one breakpoint at a time and, for each breakpoint t_i, a candidate t_i^* is generated according to $t_i^* = t_i + \epsilon$, with $\epsilon \sim U(-R,R)$ and $R = \rho(t_{i+1} - t_{i-1})$. **According to [4], this proposal is a symmetric one, which means that it holds that $r(t^*|t) = r(t|t^*)$, with t and t^* from our parameter space. This means that the acceptance probability reduces to $alpha = 1 \wedge \frac{f(t^*)}{f(t)}$. This means that areas with higher density $f(t^*) > f(t)$, are always accepted.**

- set

$$
t_{k+1} = \begin{cases} t^* & \text{w pr } \alpha = 1 \wedge \frac{f(t^*)}{f(t)} \\ t_k & \text{otherwise} \end{cases}
$$

  Here $f(t)$ is the marginal distribution $f(t|\lambda,\tau,\theta)$. The random walk proposal makes the chain aperiodic and irreducible and ensures the convergence of the MCMC algorithm [7]. Considering this and the comments about the proposal from above, the random walk proposal one at a time yields a correct MH algorithm. The steps described are implemented in the *metropolisHastings* function, which returns the updated breakpoints and the number of accepted probabilities. Also, before computing the acceptance probability, we checked that the candidate t^* is inside the interval (t_{i-1},t_{i+1}). This is very important because we should not have accepted candidates that are outside the interval.

c

Here we use a sample size of 10000 and a burn in size of 1000. We consider the years starting from 1670, because, as it can be seen in Figure 1, the evolution of disasters during the first decades is almost constant. We chose the values 0.01 for ρ and 20 for ψ. We compute the following steps for 1, 2, 3, 4, 5 breakpoints:

- we compute an initial breakpoints array, starting at 1670 and ending at 1980
- we do the burn in step, meaning that we do the hybrid MCMC algorithm for 100 samples (Gibbs sampling for updating $f(\theta|\lambda, t, \tau)$ and $f(\lambda|\theta, t, \tau)$ and Metropolis-Hastings for updating the breakpoints). The burn in step is important because it ensures the stationary behaviour of the chain.
- we do the hybrid MCMC algorithm for 1000 samples.

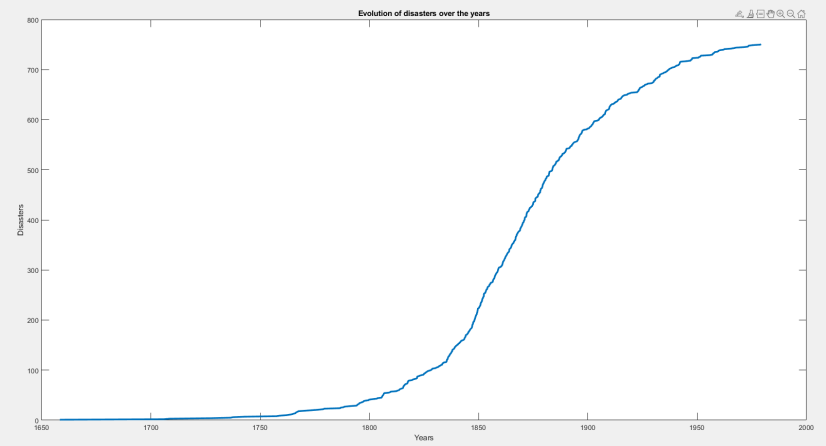
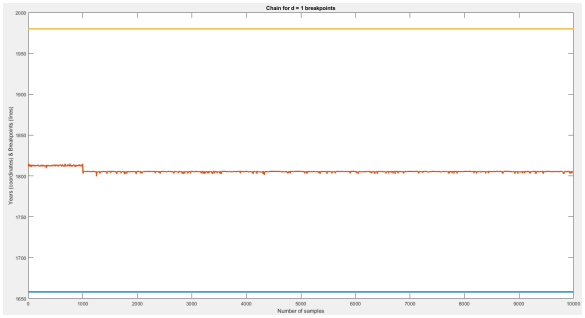


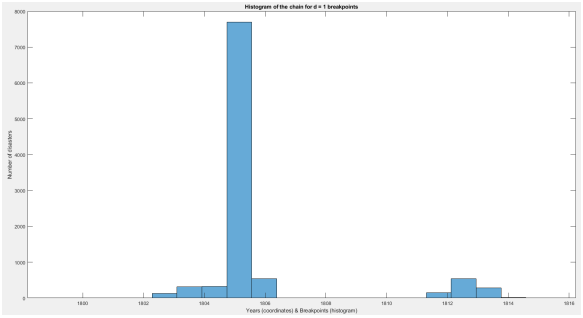
Figure 1: The evolution of disasters during the years

The behaviour of the chain for 1, 2, 3, 4, 5 breakpoints can be analyzed both visually and numerically. One wants to separate the distribution of the disasters using breakpoints. The intervals between the breakpoints should have different intensities, so one should notice a difference in what happens before and what happens after a breakpoint. In other words, one should see that the values for the intensities λ_i are different. If the values for the intensities λ_i are close to each other, than it means that some breakpoints are not necessary. In Table 1, one can observe that the intensities have different values even when using 4 ($d = 5$) and 5 ($d = 6$) breakpoints, although the tendency is to get closer values when increasing the number of breakpoints. The values for the breakpoints can be analyzed in Table 2. First of all, the values are ordered ($t_1 < t_2 < t_3 < t_4 < t_5$), which means that our MH algorithm for updating the breakpoints works fine. But, similarly to the values of the intensities, the breakpoints tend to have closer values when increasing their number. Also, when using more breakpoints, some values remain almost the same, a new value being added at the end. This is visible when using 5 breakpoints. The first three elements are almost the same as in the situation with four breakpoints. Table 3 shows the results for the θ updates for different number of breakpoints.

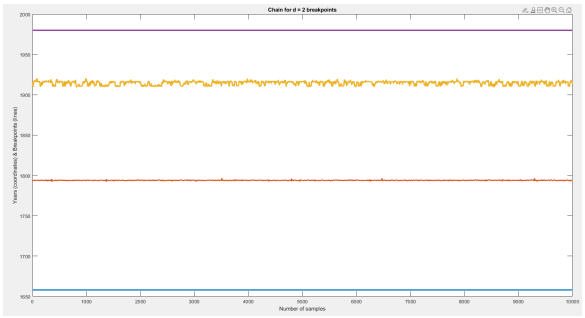
Figure 2 presents the breakpoints estimates over the number of samples and the histograms of the separated disasters distributions. As it was expected, when using 1 or 2 breakpoints, one gets distributions with high variance. The distributions start to have a lower variance when increasing the number of breakpoints. In some situations, there is a possibility for the chain to slightly jump to another value (Figures 2a, 2b and Figures 2i and 2j). The jumps are reflected also in the histograms, as they tend to be bimodal distributions, unlike the distributions without jumps, which are unimodal. Anyway, there are no overlapping distributions, which means that we do not use too many breakpoints.



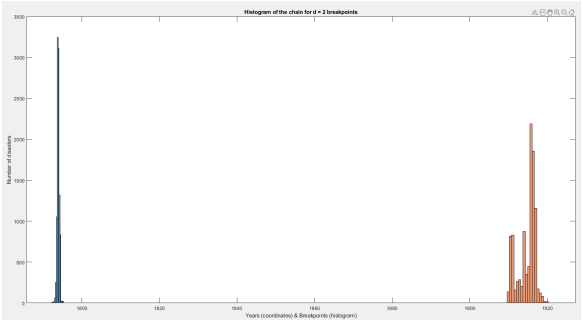
(a) Chain for one breakpoint



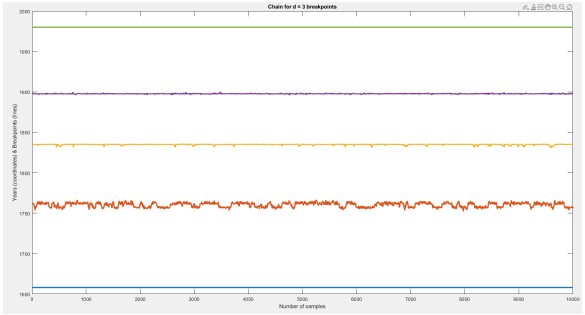
(b) Histogram of the chain for one breakpoint



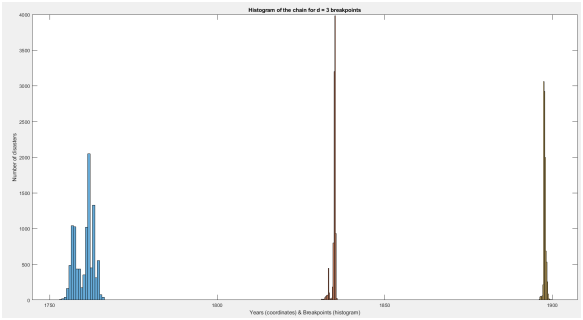
(c) Chain for two breakpoints



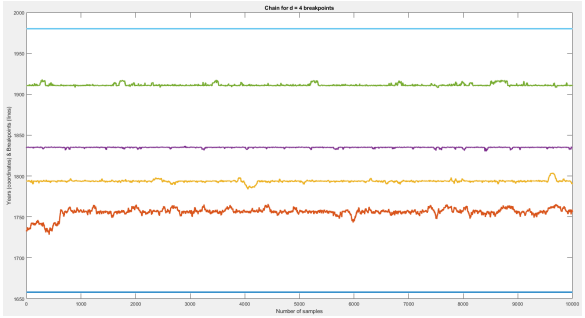
(d) Histogram of the chain for two breakpoints



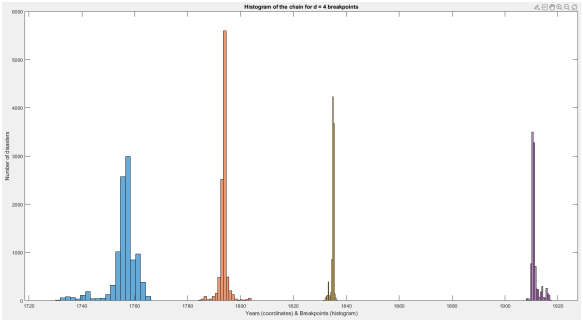
(e) Chain for three breakpoints



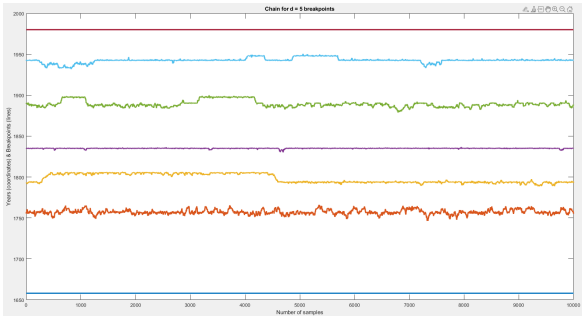
(f) Histogram of the chain for three breakpoints



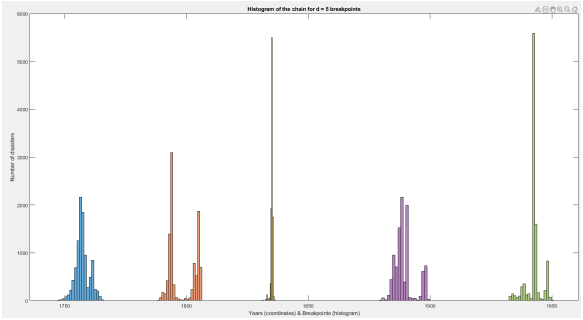
(g) Chain for four breakpoints



(h) Histogram of the chain for four breakpoints



(i) Chain for five breakpoints



(j) Histogram of the chain for five breakpoints

Figure 2: The behaviour of the chain when using a different number of breakpoints. The figures on the left show the breakpoint estimate over the number of samples. The figures on the right show the histograms of the separated disasters distributions centered in the breakpoints.

d	6	5	4	3	2
λ_1	0.0829	0.0304	0.0558	0.2076	0.3913
λ_2	0.5793	0.5965	1.54786	5.2818	3.8040
λ_3	1.8611	2.1517	7.8946	1.4835	-
λ_4	7.3584	7.7573	2.3008	-	-
λ_5	3.2191	1.8335	-	-	-
λ_6	1.1560	-	-	-	-

Table 1: The intensities λ for different number of breakpoints (the number of breakpoints is $d-1$)

breakpoints	5	4	3	2	1
t_1	1755	1756	1757	1793	1804
t_2	1792	1792	1835	1916	-
t_3	1834	1834	1898	-	-
t_4	1886	1910	-	-	-
t_5	1942	-	-	-	-

Table 2: The values for the breakpoints

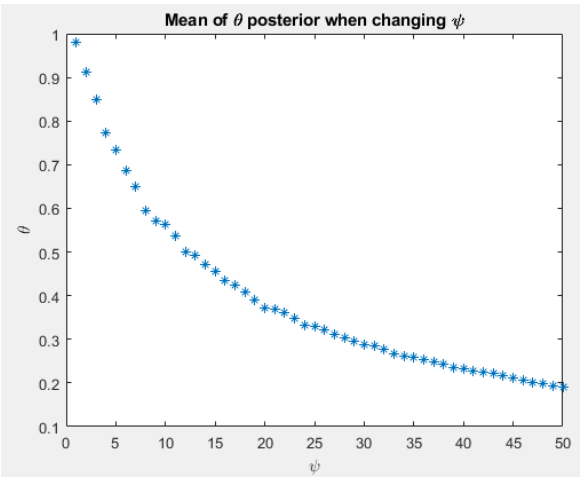


Figure 3: The dependence of the mean of the θ posterior on ψ. x axis: mean of θ posterior, y axis: different values of ψ

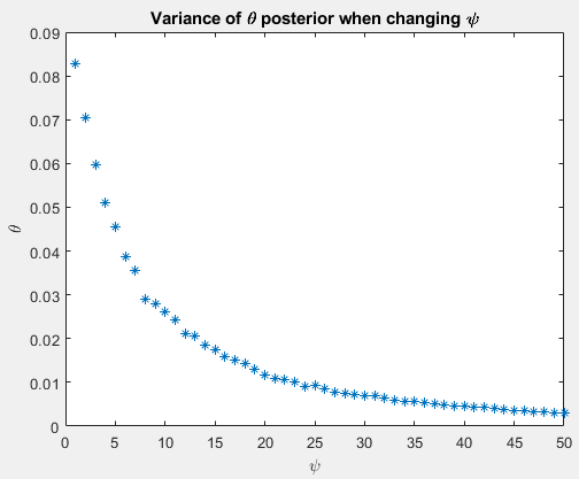


Figure 4: The dependence of the variance of the θ posterior on ψ. x axis: variance of θ posterior, y axis: different values of ψ

d

For this task we considered 4 breakpoints, a value of 0.01 for ρ and a range of values from 1 to 50 for ψ. We consider again 10000 samples and 1000 samples for the burn-in. In order to analyze the sensitivity of the posteriors to the choice of the hyperparameter ψ, one computes the hybrid MCMC algorithm for different values of ψ.

From Task 1a), one knows that $f(\theta|\lambda,t,\tau) \propto \Gamma(2+2d, \sum_{i=1}^{d}\lambda_i + \psi)$. Consequently, there is a dependence for θ of ψ.

The dependence of the mean of $f(\theta|\lambda,t,\tau)$ posterior on ψ can be observed in Figure 3. It seems that one could say that $\mathbb{E}(\theta) \propto \psi^{-1}$. The dependence of the variance of $f(\theta|\lambda,t,\tau)$ posterior on ψ can be observed in Figure 4. It holds that $\mathbb{V}(\theta) \propto \psi^{-2}$.

From Task 1a) it holds that $f(\lambda|\theta,t,\tau) \propto \Gamma(n_i(\tau)+1, \tau + t_{i+1} - t_i)$. So the posterior $f(\lambda|\theta,t,\tau)$ is dependent on θ. We know that θ is dependent on ψ. This means that the posterior $f(\lambda|\theta,t,\tau)$ has also a dependency on ψ.

The dependency of the mean of $f(\lambda|\theta,t,\tau)$ posterior on ψ can be analyzed in Figure 5. But one can conclude that $f(\lambda|\theta,t,\tau)$ does not have such a big dependence on ψ. This is actually expected, as the term θ is very small in comparison to the term $t_{i+1} - t_i$. The dependency of the variance of $f(\lambda|\theta,t,\tau)$ posterior on ψ can be analyzed in Figure 6. Here, it can be noticed that the variance $\mathbb{V}(\lambda)$ is even less dependent on ψ.

breakpoints	θ
5	0.4617
4	0.2571
3	0.3211
2	0.0921
1	0.1191

Table 3: The values for the θ posterior when using different number of breakpoints

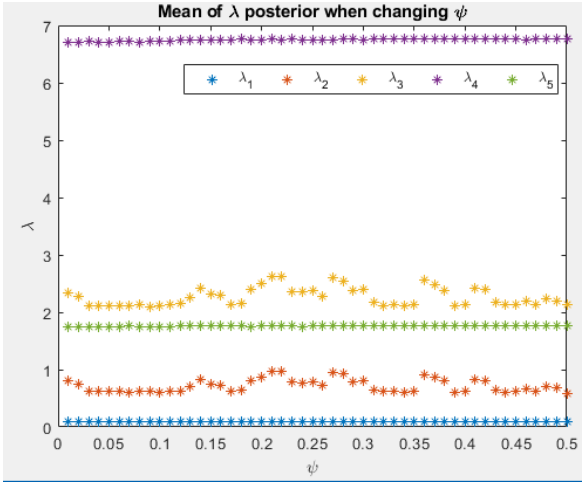


Figure 5: The dependence of the mean of the λ posterior on ψ. x axis: mean of λ posterior, y axis: different values of ψ

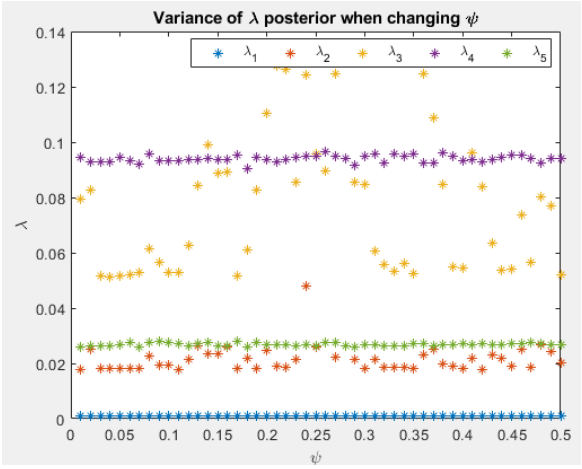


Figure 6: The dependence of the variance of the λ posterior on ψ. x axis: variance of λ posterior, y axis: different values of ψ

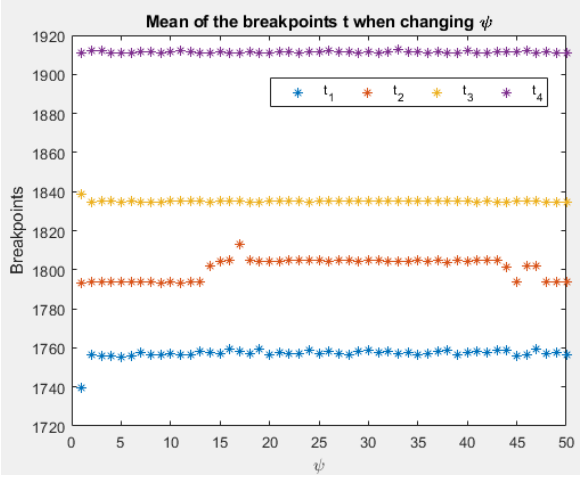


Figure 7: The dependence of the mean of the breakpoints on ψ x axis: mean of t posterior, y axis: different values of ψ

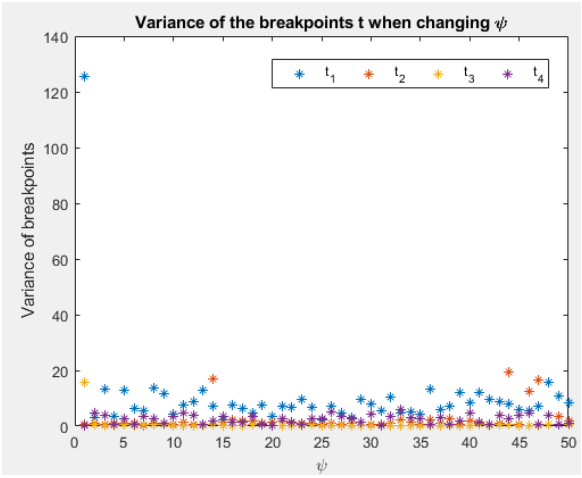


Figure 8: The dependence of the variance of the breakpoints on ψ x axis: variance of t posterior, y axis: different values of ψ

From Task 1a), it is known that $f(t|\lambda,\tau,\theta) \propto \exp(-\sum_{i=1}^{d}(t_{i+1}-t_i)\lambda_i)\prod_{i=1}^{d}\lambda_i^{n_i(\tau)}\prod_{i=1}^{d}(t_{i+1}-t_i)$. So the posterior $f(t|\lambda,\tau,\theta)$ is dependent on λ, which does not have such a big dependence on ψ. Therefore, we do not expect a substantial dependence of $f(t|\lambda,\tau,\theta)$ on ψ either.

As it can be seen in Figure 7, the mean of the breakpoints t does not have a strong dependence on ψ. By analyzing Figure 8, the variance of the breakpoints t does not have a visible dependence on ψ either.

Concluding, the only posterior sensitive to the change of ψ is $f(\theta|\lambda,t,\tau)$.

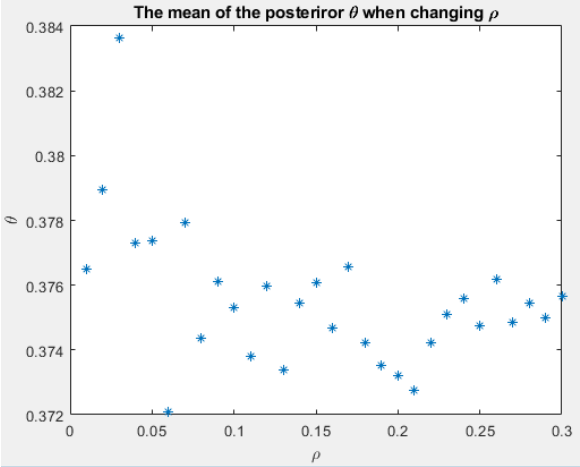


Figure 9: The dependence of the mean of the θ posterior on ρ. x axis: mean of θ posterior, y axis: different values of ρ

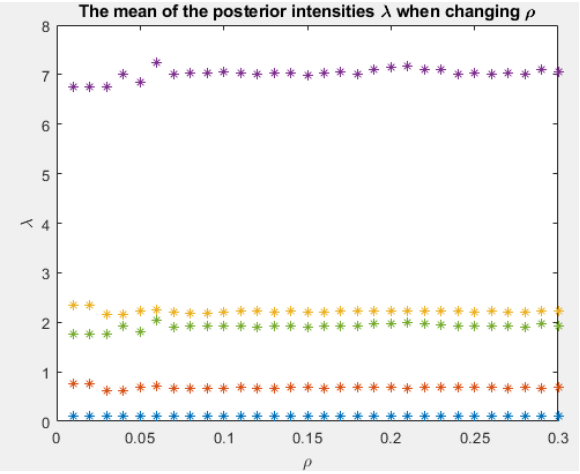


Figure 10: The dependence of the mean of the λ posterior on ρ. x axis: mean of λ posterior, y axis: different values of ρ

e

First, we will analyze the sensitivity of $f(\theta|\lambda, t, \tau)$ and $f(\lambda|\theta, t, \tau)$ to the choice of ρ in the proposal distribution.

We used 10000 samples and 1000 samples for the burn-in. The number of breakpoints used is 4 and the value of ψ is 20. We used a range of values from 0.01 to 0.3 for ρ. In order to analyze the sensitivity, we did the hybrid MCMC algorithm for the range of ρ values and we stored the mean of $f(\theta|\lambda, t, \tau)$ and $f(\lambda|\theta, t, \tau)$ for each value.

By analyzing Figures 9 and 10, one can not observe any dependence of the posteriors $f(\theta|\lambda, t, \tau)$ and $f(\lambda|\theta, t, \tau)$ on ρ. Therefore, one can conclude that the posteriors $f(\theta|\lambda, t, \tau)$ and $f(\lambda|\theta, t, \tau)$ are not so sensitive to the change of ρ.

In [5], it is said that the asymptotic variance of τ_N is given by:

$$\sigma^2 = r(0) + 2\sum_{l=1}^{\infty} r(l)$$

with

$$r(l) = \lim_{\infty} \mathbb{C}(\phi(X_{n+l}), \phi(X_n)),$$

where r is the transition density (dependent on the proposal), (X_k) is the Markov chain, $\mathbb{C}$ is the covariance function. Mixing is actually the speed at which $r(l)$ tends to zero. In [1] and [5], it is said that a good mixing is one that has a fast decreasing $r(l)$ and that the mixing is dependent on the choice of the proposal.

In [5], it is said that a good acceptance rate for a good mixing is around 30%. Therefore, we analyzed the acceptance rate for a range of values for ρ between 0.001 and 0.1. We used 4 breakpoints, a value of 20 for ψ.

From Figure 11, it can be observed that the ρ values that ensure a good acceptance rate are between 0.01 and 0.045 (the acceptance rate is between 40% and 25% for these range of values). Therefore, setting ρ between 0.01 and 0.045 would result in a good acceptance rate.

In [1], another way to analyze the mixing is to observe how the correlation in the sequence of $X^(t)$ (where (X_k) is the chain) behaves at different time iteration lags. Consequently, we will analyze the autocorrelation of the breakpoints t for different values of ρ. Considering the comments with respect to the acceptance rate, we chose four values for ρ: 0.01, 0.02, 0.03 and 0.04. We used 4 breakpoints, a value of 20 for ψ, 10000 samples and 1000 samples for the burn-in set. We computed the hybrid MCMC algorithm for the foru values of ρ.

From Figure 12, we can see that the autocorrelation is rapidly decreasing especially for ρ 0.01 and 0.04. But we can see that whenever is a jump in a breakpoint, the autoccorelation function is affected

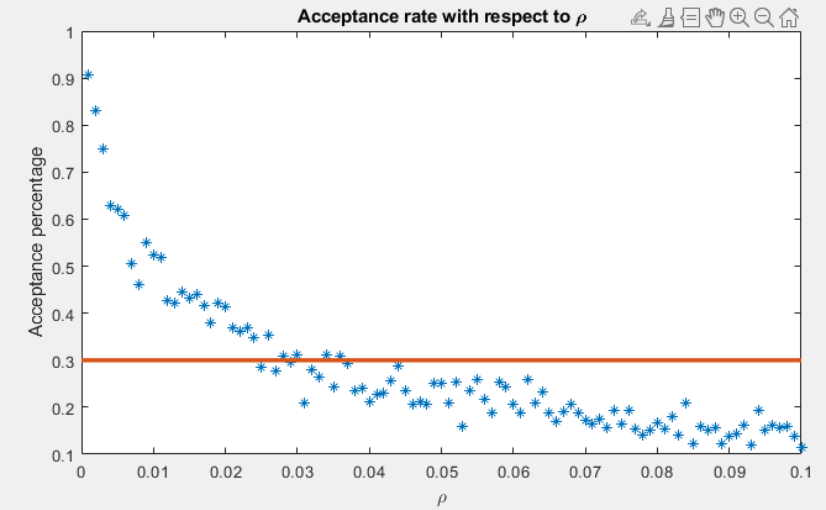


Figure 11: The evolution of the acceptance rate with respect to ρ. x axis: the acceptance rate, y axis: the values of ρ

also (see 12c and the autocorrelation function for t_1 and t_2, or 12e and the autocorrelation function for t_1 and t_2).

Also, from Figure 12, it is obvious that the parameter ρ influences the behaviour of the t posterior. Analyzing all the plots in Figure 12, one could say that the optimal choice for four breakpoints would be $\rho = 0.01$ or $\rho = 0.04$ (as in in Figures 12b and 12h all the breakpoints t_i have a very fast decreasing autocorrelation function and a small time dependency).

In conclusion, taking into account the acceptance rate and the autocorrelation for the breakpoints, a good choice for ρ would be a value between 0.01 and 0.04, the optimal ones being 0.01 or 0.04.

Task 2

a

The Gumbel distribution has the distribution function;

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

To solve for the inverse we will use the following relations;

$$F\left(F^{-1}(u)\right) = u \Rightarrow F^{-1}(F(u)) = u$$

Using this we can manipulate the function for the Gumbel distribution to find it's inverse;

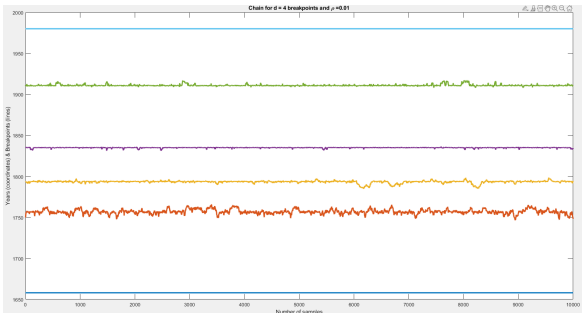
$$u = \exp\left(-\exp\left(-\frac{F^{-1}(u:\mu,\beta)-\mu}{\beta}\right)\right)$$

$$\ln\left(-\ln\left(u\right)\right) = -\frac{F^{-1}(u:\mu,\beta)-\mu}{\beta}$$

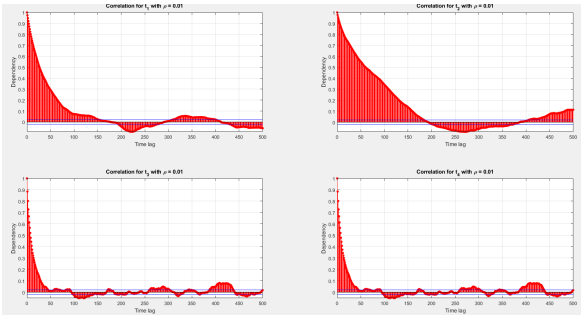
$$-\beta\ln\left(-\ln\left(u\right)\right) + \mu = F^{-1}(u:\mu,\beta)$$

Re-writing to a standard form, the inverse of the Gumbel distribution becomes;

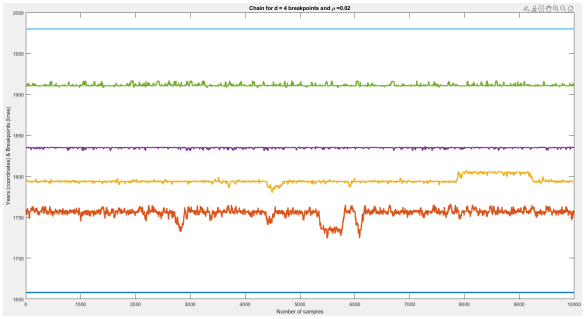
$$F^{-1}(u:\mu,\beta) = \mu - \beta\ln\left(-\ln\left(u\right)\right) \tag{9}$$



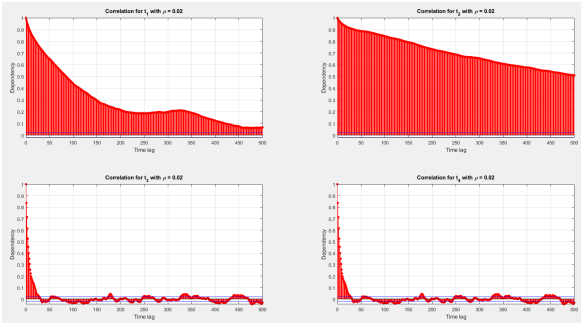
(a) Chain for four breakpoints and $\rho = 0.01$



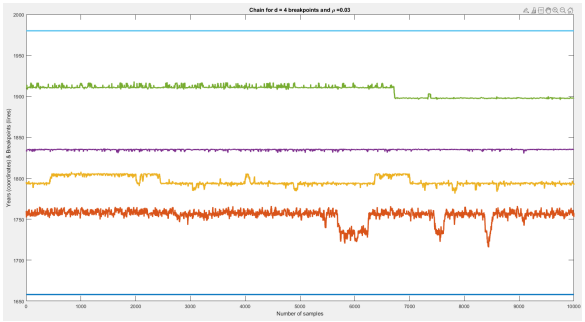
(b) Autocorrelation for the breakpoints for $\rho = 0.01$



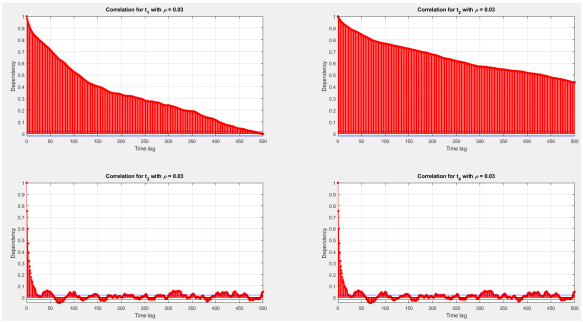
(c) Chain for four breakpoints and $\rho = 0.02$



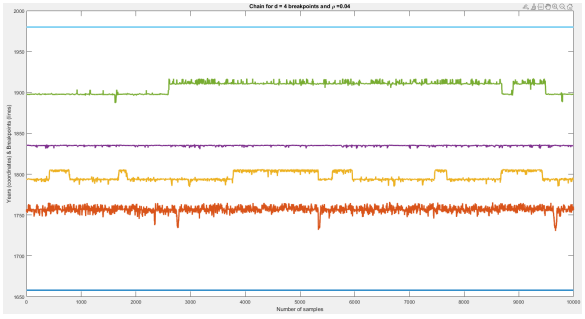
(d) Autocorrelation for the breakpoints for $\rho = 0.02$



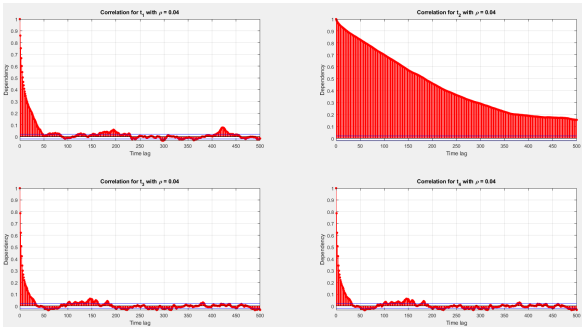
(e) Chain for four breakpoints and $\rho = 0.03$



(f) Autocorrelation for the breakpoints for $\rho = 0.03$



(g) Chain for four breakpoints and $\rho = 0.04$



(h) Autocorrelation for the breakpoints for $\rho = 0.04$

Figure 12: The behaviour of the chain when using four breakpoints and different values for ρ. The figures on the left show the breakpoint estimate over the number of samples. The figures on the right show the autocorrelation for the breakpoints over time lags for different values for ρ

b

In [2], the bootstrapping is a method used to simulate the distribution of a statistic. The observed data is repeatedly resampled, every time an empirical distribution function being produced from the resampled data. A new value of the statistic can be computed for each resampled data set. All these new values represents an estimate of the sampling distribution of the desired statistic. In other words, the boostrapping method allows the data to pull itself up by its own bootstrap [2].

Parametric boostrapping uses a distribution in order to estimate the data needed for analyzing. Here, the Gumbel distribution is used to fit the data and to estimate the parameters.

The goal is to compare the mean of the parameters from the true data tot the mean of a random sample from the Gumbel distribution. First, one constructs the empirical distribution from the data y. Then one simulates B new data sets Y_b^*, where each Y_b^* is obtained by drawing with replacement n times among the y_i's [6]. Then one computes the values $t(Y_b^*)$ of the estimator. Finally, one computes $\Delta_b^* = t(Y_b^*) - \tilde{\tau}$, where $\tilde{\tau}$ is the mean of the parameters from the true data.

In MATLAB, we estimated β and μ using the Gumbel distribution. Then, for 1000 samples, we computed the inverse $F^{-1}(u : \mu, \beta)$ according to Task 2a) using a uniformly distributed $u \in U(0, 1)$ and we estimated the new parameters β_{new} and μ_{new} using the Gumbel distribution for $F^{-1}(u : \mu, \beta)$. We saved these new values. We sorted the difference $\beta - \beta_{new}$ and $\mu - \mu_{new}$. Then we computed the 95% confidence interval. The values for the parameters and the 95% confidence intervals can be seen in Table 4.

Parameter	Estimate	Lower Bound	Upper Bound
μ	4.1477	4.0266	4.2811
β	1.4858	1.3926	1.5847

Table 4: Parametric bootstrapped 95% confidence intervals for the parameters μ and β

c

This task is actually very similar to the previous one, but this time we want to find a one sided parametric bootstrapped confidence interval just for the 100-year Atlantic wave.

We know that there are $T = 3 \cdot 14 \cdot 100 = 4200$ observations in total. The wave is given by $F^{-1}(1 - 1/T; \mu, \beta)$, so we compute the wave using the inverse Gumbel distribution according to Task 2a). Then, for 1000 samples, we compute the new estimate of the wave $wave_{new}$ using the new samples μ_{new} and β_{new} from the previous task and the inverse Gumbel distribution. Then we sort the array containing the difference $wave - wave_{new}$ and we compute the upper bound of the 95% confidence interval. We computed the upper bound as we are interested in the worst case scenario, therefore the highest wave.

The one sided parametric bootstrapped 95% confidence interval for the 100-year return value can be observed in Table 5.

Parameter	Estimate	Upper Bound
wave	16.5436	17.2336

Table 5: One sided parametric bootstrapped 95% confidence interval for the 100-year return value

References

[1] Hoeting J. A. Givens G. H. *Computational Statistics*. John Wiley & Sons, 2013. Chap. 7.

[2] Hoeting J. A. Givens G. H. *Computational Statistics*. John Wiley & Sons, 2013. Chap. 9.

[3] Wiktorsson M. *Monte Carlo and Empirical Methods for Stochastic Inference*. Lund University, 2022. Chap. Lecture 10.

[4] Wiktorsson M. *Monte Carlo and Empirical Methods for Stochastic Inference*. Lund University, 2022. Chap. Lecture 9.

[5] Wiktorsson M. *Monte Carlo and Empirical Methods for Stochastic Inference*. Lund University, 2022. Chap. Lecture 12.

[6] Wiktorsson M. *Monte Carlo and Empirical Methods for Stochastic Inference*. Lund University, 2022. Chap. Lecture 13.

[7] Fearnhead P. Sherlock C. and Roberts G. O. "The Random Walk Metropolis: Linking Theory and Practice Through a Case Study". In: *Statistical Science* 25.2 (2010), pp. 172–190. DOI: 10.1214/10-STS327. URL: https://arxiv.org/pdf/1011.6217.pdf.