HA3 for Monte Carlo and Empirical Methods

Jingmo Bai Zuoyi Yu

March 6, 2023

1 Coal mine disasters

1.1 Question a

Firstly, we conclude what we know from the given details. We know the intensities $\lambda \sim \Gamma(2, \theta)$ and the hyperprior $\theta \sim \Gamma(2, \Psi)$, where Ψ is a fixed hyperparameter. We also know the prior of breakpoints

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

which implies that

$$f(\tau|\lambda, t) = exp(-\sum_{i=1}^{d} \lambda_i(t_{i+1} - t_i)) \prod_{i=1}^{d} \lambda_i^{n_i(\tau)} \tag{2}$$

where

$$n_i(\tau) = number of disasters in the sub-interval[t_i, t_{i+1}) = \sum_{j=1}^{n} \mathbb{1}_{[t_i, t_{i+1})}(\tau_j). \tag{3}$$

Then, we can change the target marginal posteriors using conditional probability:

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

We can find all these marginal posteriors have the same numerator, so we begin with calculating $f(\theta, \lambda, t, \tau)$. From the rules of conditional probability, we can rewrite $f(\theta, \lambda, t, \tau)$ as:

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

From the information above, we know the distributions of $f(\theta)$ and $f(\lambda|\theta)$. And t is independent on λ and θ, so $f(t|\theta, \lambda) = f(t)$. τ is also independent on θ so $f(\tau|\theta, \lambda, t) = f(\tau|\lambda, t)$. Therefore, Equation 5 can be rewitten as

$$f(\theta, \lambda, t, \tau) = f(\tau|\lambda, t)f(t)f(\lambda|\theta)f(\theta) \tag{6}$$

Then, we can find this equation can be rewritten with what we have known

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

As we just need to compute the marginal posteriors up to normalizing constants and these posteriors are some distributions proportional to Equation 7, we can remove the terms that are independent on the target variable. For example, the first marginal posteriors can be written as:

$$f(\theta|\lambda,t,\tau) \propto f(\lambda|\theta)f(\theta) \tag{8}$$

So this marginal posterior has the form

$$f(\theta|\lambda,t,\tau) \propto \prod_{i=1}^{d}(\frac{\theta^2}{\Gamma(2)}\lambda_i e^{-\theta\lambda_i})\frac{\Psi^2}{\Gamma(2)}\theta e^{-\Psi\theta} = \prod_{i=1}^{d}(\lambda_i)\frac{\Psi^2}{\Gamma(2)}\theta^{2d+1}e^{-(\Psi+\sum_{i=1}^{d}\lambda_i)\theta} \tag{9}$$

We can see this is a Gamma function with parameters

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

The other two marginal posteriors could also be calculated as above,

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

where we do not know the distribution of t for $f(t|\theta,\lambda,\tau)$ to simplify.

1.2 Question b

To sample from the posterior $f(\theta,\lambda,t|\tau)$ we need construct a hybrid MCMC algorithm. For $f(\theta|\lambda,t,\tau)$ and $f(\lambda|\theta,t,\tau)$, as the distributions of them are known, we can update them using Gibbs sampling. For the breakpoints $f(t|\theta,\lambda,\tau)$ which has unknown distribution, we need to use a Metropolis-Hastings(MH) sampler combined with Gibbs sampler to update the breakpoints. From Page 15 of the slides for Lecture 10, we know the Gibbs sampler simulate a sequence of values with the following mechanism: Given X_k,

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

Therefore, we can use this mechanism to update the posteriors $f(\theta|\lambda,t,\tau)$ and $f(\lambda|\theta,t,\tau)$. Their distributions are found in the previous question. Then, we use MH algorithm to update the unknown distribution $f(t|\theta,\lambda,\tau)$. From Page 16 of the slides for Lecture 9, we know the MH algorithm simulates a sequence of values through the following mechanism: given X_k,

- generate $X^* \sim r(z|X_k)$ and
- set $X_{k+1} = \begin{cases} \text{X}^* \text{ w.pr. } \alpha(X_k,X^*) = 1 \wedge \frac{f(X^*)r(X_k|X^*)}{f(X_k)r(X^*|X_k)}, \\ \text{X}_k \text{ otherwise.} \end{cases}$

where the notation $a \wedge b = min\{a, b\}$, $r(z|X_k)$ is the transition density of what we sample.
The proposals are given as *Random walk proposal* and we choose *Random walk proposal one at a time*, which means to update one breakpoint at a time. To be specific, for each breakpoint t_i, its candidate t_i^* is generated through

$$t_i^* = t_i + \epsilon, \epsilon \sim U(-R, R) \tag{10}$$

where $R = \rho(t_{i+1} - t_{i-1})$. As $U(-R, R)$ is symmetric, this proposal kernel is symmetric, which means $r(z|X_k) = r(X_k|z)$. Therefore, $\alpha(X_k, X^*)$ could be reduced to:

$$\alpha(X_k, X^*) = 1 \wedge \frac{f(X^*)}{f(X_k)} \tag{11}$$

1.3 Question c

What we want to do in this part is to find a suitable number of breakpoints. A breakpoint is a point that separates two different distributions, and this could be found through both figures and data.
We begin with observing the total number of disasters during the years. From Figure 1, we can see the number of accidents happened before 1700 is almost 0, so we choose the start of years from 1690 and the end year is still 1980.

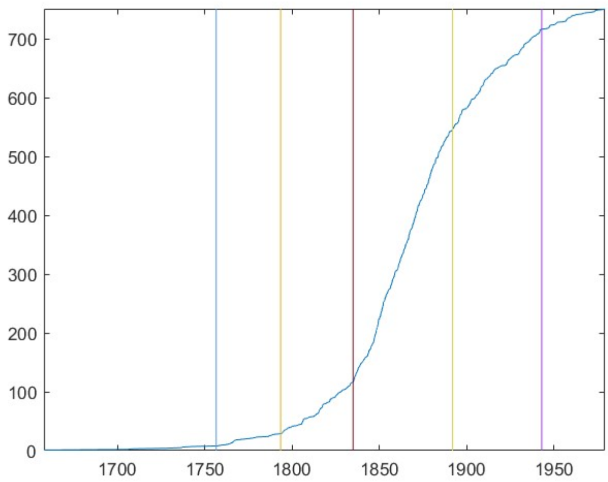
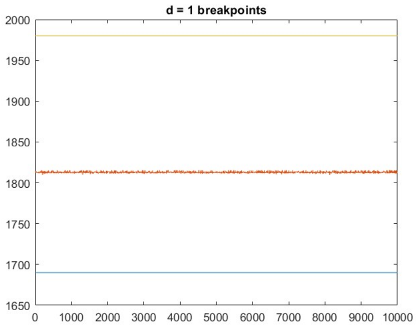
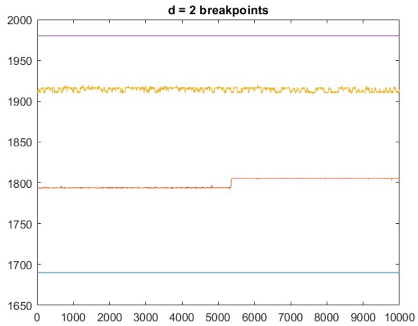


Figure 1: Plot of the number of disasters happened from 1650 to 1980, the vertical lines are breakpoints.

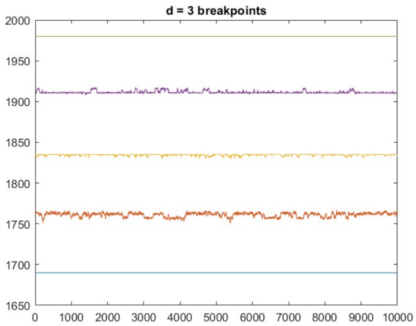
Then, we choose burn-in with size of 1000 and samples with size of 10000, $\rho = 0.01, \Psi = 20$, and investigate the behaviours of d = 1, 2, 3, 4, 5. Figure 2 shows the behaviours of the chain with a different number of breakpoints. We can see as the number of breakpoints increasing, the chain tends to stay at the same values. Combined the vertical lines in Figure 1, we can see using 5 breakpoints is quite reasonable to separate the curve of the total number of disasters.



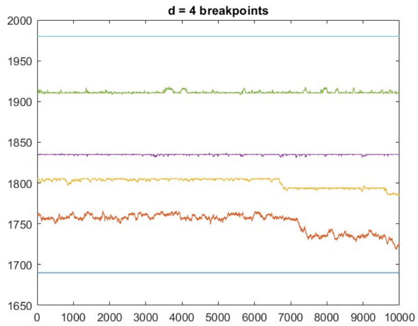
(a) d = 1



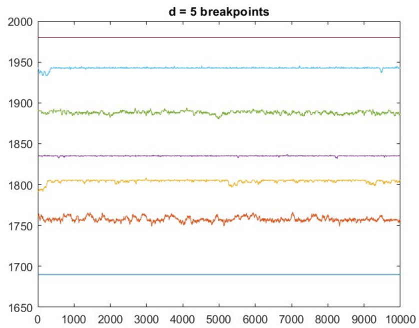
(b) d = 2



(c) d = 3



(d) d = 4



(e) d = 5

Figure 2: Plot of the behaviour of the chain with different number of breakpoints.

To make a robust conclusion, we also observe the data of different number of breakpoints. As the intensity λ implies the different exponential distribution between accidents, we could observe λ to find out if our breakpoints are reasonable or not. As we can see in Table 1, all of the values of λ are different, which means all of the intervals between breakpoints have different distribution. And compared with smaller d, d = 6(5 breakpoints) seems to be more accurate for separating. So we think d = 6 is a really good choice, the final value of the breakpoints are shown in Table 2.

Table 1: The intensities λ of different number of breakpoints

d	2	3	4	5	6
λ_1	0.4215	0.3977	0.1191	0.1172	0.1905
λ_2	3.9389	5.4315	1.2267	0.3846	0.7061
λ_3	x	1.5966	6.6330	2.3329	2.5587
λ_4	x	x	1.6717	6.9667	7.3918
λ_5	x	x	x	1.8533	3.3026
λ_6	x	x	x	x	1.0366

Table 2: The values of breakpoints

Breakpoint	Year
t_1	1757
t_2	1796
t_3	1835
t_4	1888
t_5	1942

1.4 Question d

In this problem, we will investigate how sensitive the parameters of the posteriors λ, θ, t to the choice of Ψ. We observe the changes of mean and variance of these parameters while changing Ψ. We set a range of 1 to 50 for Ψ. Firstly, we concern the sensitivity of the parameter θ. In 1.1, we know that$f(\theta|\lambda, t, \tau) \sim \Gamma(2d+2, \Psi + \sum_{i=1}^{d} \lambda_i)$, so we can get $\mathbb{E}(\theta) = 1/\psi, \mathbb{V}(\theta) = 1/\Psi^2$. Therefore, as the value of Ψ increasing, the values of mean and variance should decrease, which could be proved by Figure 3.

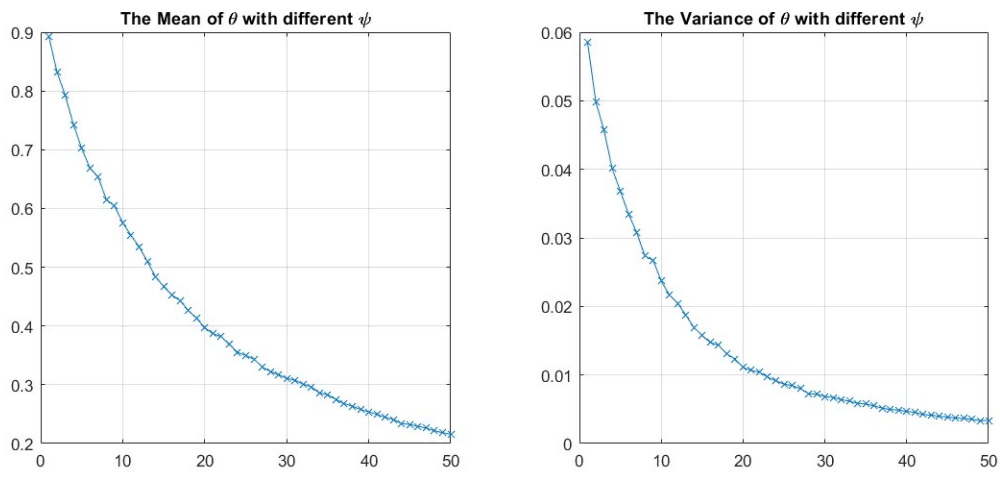


Figure 3: Plot of the mean and variance of θ.

Then, consider about λ. We know $f(\lambda|\theta, t, \tau) \sim \Gamma(n_i(\tau) + 2, t_{i+1} - t_i + \theta)$. So λ is dependent on θ, which means the mean and variance should be affected by Ψ significantly. However, from Figure 4, we can see indeed the value of Ψ does not affect the mean and variance a lot. This is caused by the term $t_{i+1} - t_i$, which is much larger than θ, so the sensitivity of λ is dominated by $t_{i+1} - t_i$. Therefore, the mean and variance is not sensitive to the change of Ψ.

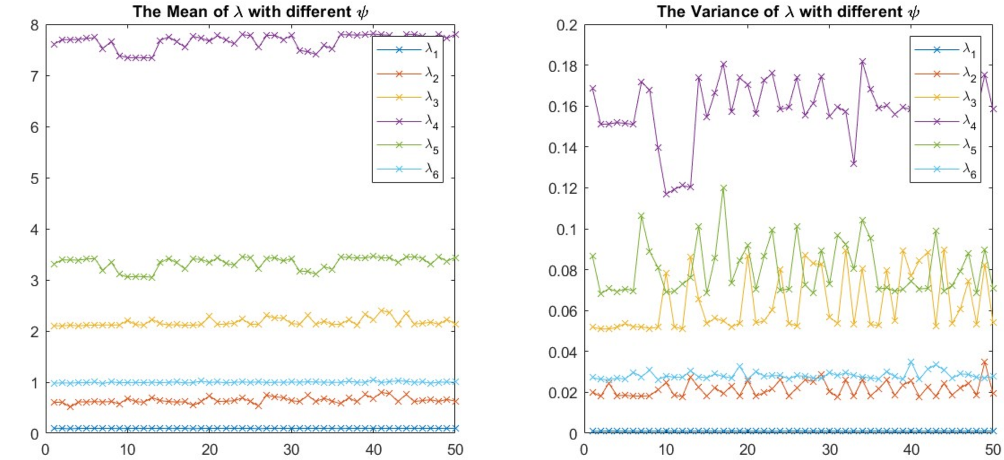


Figure 4: Plot of the mean and variance of λ.

Finally, we consider about the breakpoints t. As $f(t|\theta, \lambda, \tau) \propto \prod_{i=1}^{d}(t_{i+1}-t_i)e^{-\lambda_i(t_{i+1}-t_i)}\lambda_i^{n_i(\tau)}$, t is dependent on λ. And from the previous analysis, λ is not affected by the value of Ψ. So the mean and variance of t should not change a lot as well. This could be verified in Figure 5.

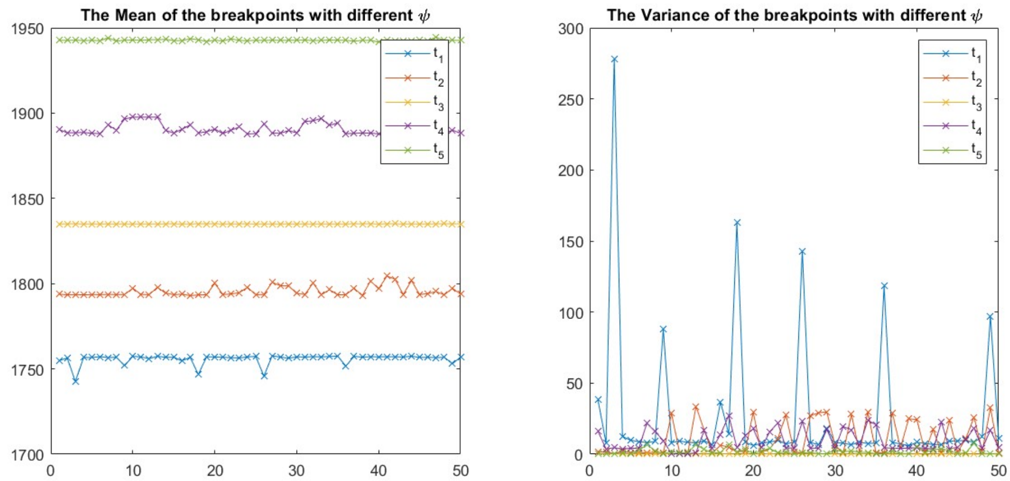


Figure 5: Plot of the mean and variance of breakpoint t.

1.5 Question e

In this problem, we are asked to investigate how sensitive the parameters of the posteriors λ, θ, t to the choice of ρ.
From Figure 6, it is hard to see any dependency between λ or θ and ρ. So we can say $\lambda and \theta$ are not sensitive to the change of ρ.

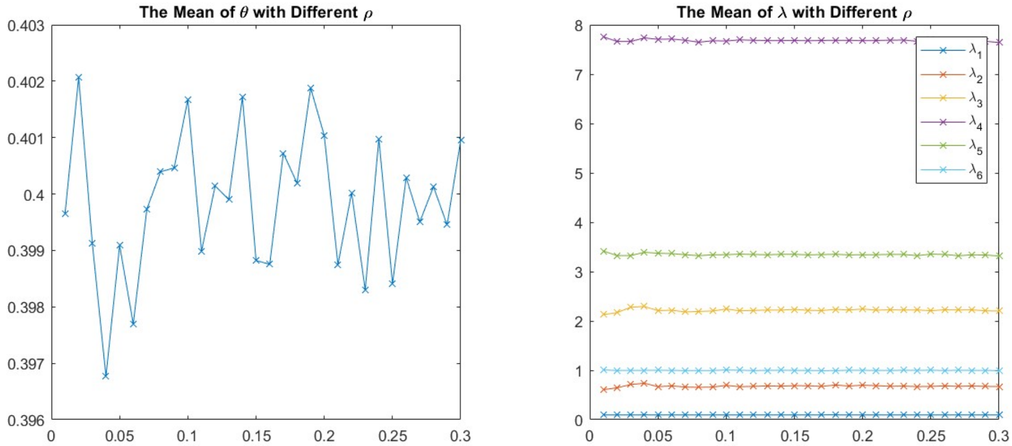


Figure 6: Plot of the mean of θ and λ

Then we consider about the breakpoints t. Since t is random walk proposal with MH sampler, it has direct dependence on ρ. We begin with observing the acceptance ratio first. From the slides for Lecture 12, we know a good acceptance rate is around 30%. From Figure 7, we could see a good acceptance rate is between 0.03 to 0.05.

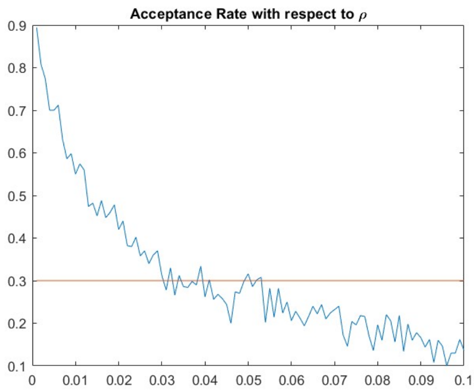
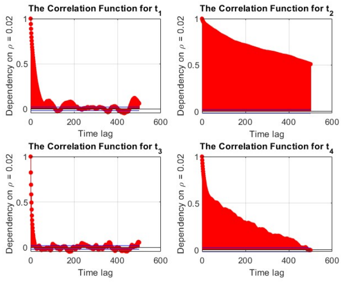
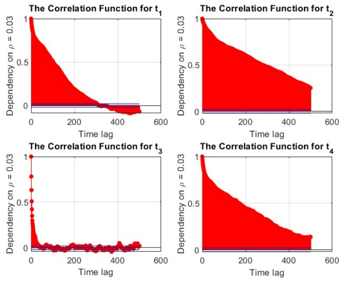


Figure 7: Plot of the average acceptance rate of ρ.

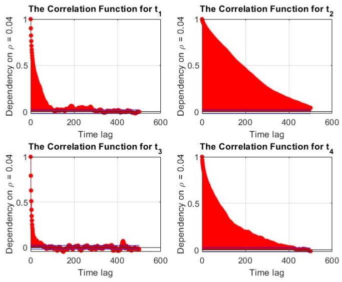
Then we check the effect on different breakpoint. From what we have found above, we test with $\rho = [0.02, 0.03, 0.04, 0.05]$. Figure 8 shows autocorrelation for different breakpoint, and we can see ρ affects the breakpoints differently. Therefore, to choose a different ρ for each breakpoint is a good idea, and a good choice of ρ is the one that decreases fast and has less time dependency.



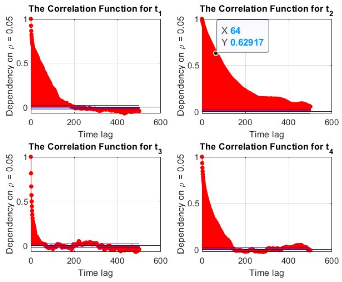
(a) $\rho = 0.02$



(b) $\rho = 0.03$



(c) $\rho = 0.04$



(d) $\rho = 0.05$

Figure 8: Plot of autocorrelation of breakpoints t.

2 Parametric bootstrap for the 100-year Atlantic wave

2.1 Question a

In this question, we aim to find the inverse of A Gumbel distribution. A Gumbel distribution with a distribution function defined as

$$F(x;\mu,\beta) = exp(-exp(-\frac{x-\mu}{\beta})), x \in \mathbb{R} \tag{12}$$

where $\mu \in \mathbb{R}$ and $\beta > 0$, is a good fit to the data. The parameters can be estimated using the given Matlab function. The Tth expected return value of the significant wave height is given by $F^{-1}(1-1/T;\mu,\beta)$.
We set $u = F(x;\mu,\beta)$ so $x = F^{-1}(u;\mu,\beta)$

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

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

$$x = \mu - \beta ln(-ln(u))$$

Thus, the inverse is obtained

$$F^{-1}(u;\mu,\beta) = \mu - \beta ln(-ln(u)) \tag{13}$$

2.2 Question b

In this question, we aim to calculate parametric bootstrapped 95% confidence intervals for the parameters μ and β.
First, we estimate the $\hat{\mu}$ and $\hat{\beta}$ using the given observations and the estimating function. The estimation of parameters are:

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

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

Then, we compute new data using the inverse $F^{-1}(u; \mu, \beta)$ with a uniform distribution $u \in U(0, 1)$, and we estimated the bootstrap parameters μ and β with the newly obtained data using the given estimating function. We do this 1000 times. Then we calculate the difference between the estimation of parameters from the bootstrap and the true data. And we sort these differences. Finally, we can obtain parametric bootstrapped 95% confidence intervals for the parameters μ and β.

Table 3: Parametric bootstrapped 95% confidence intervals for the parameters

Parameter	Lower Bound	Upper Bound
μ	4.0189	4.2700
β	1.3920	1.5801

2.3 Question c

In this question, we want to calculate the one-sided upper bounded parametric bootstrapped 95% confidence interval for the 100-year return value.
The Tth expected return value of the significant wave height is given by $F^{-1}(1 - 1/T; \mu, \beta)$, and we know that T = 3*14*100. First, we estimate the waves by the inverse of the Gumbel distribution with the estimate of parameters μ and β. Then, we estimate the new estimation of waves by using the bootstrap parameters from the previous problem. Then we can get the difference between the two estimations and sort it. Finally, we can obtain the one-sided upper bounded parametric bootstrapped 95% confidence interval for the 100-year value.

Table 4: One-sided upper bounded parametric bootstrapped 95% confidence interval

Estimate	Upper Bound
16.5436	17.2488