Lab - Extreme Values

Johan Hellmark, jo5285he-s & Gustaf Sundell, gu0147su-s

May 19, 2021

1

Scatter plot of the data:

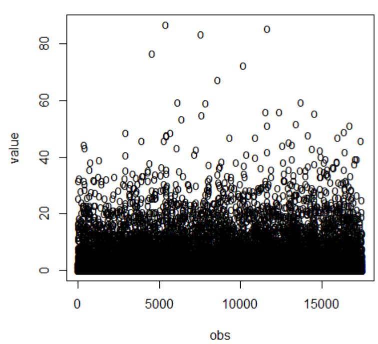

Figure 1: Scatter plot for question 1

2

A reasonable threshold value is $u_o = 30$ since the mean excess seem to be fairly linear after this value which it should be according to the theory. Before $u = 30$ the line seems to curve and is hence to linear. One could argue that the point $u = 60$ could be a reasonable choice but since very few measurements exceeds this limit and which would lead to a large variance.

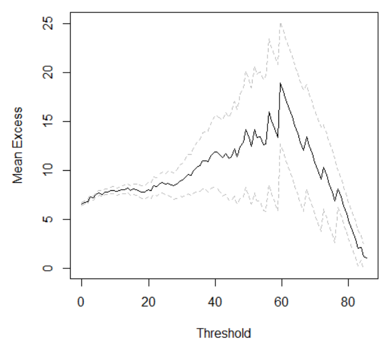

Figure 2: The mean residual life plot for the rain data.

3

Looking at the plot, we can see that both parameters become less stable after $u = 30$, indicating that $u_0 = 30$ still appears reasonable, as the stability of both parameters vastly decreases for $u > 30$.

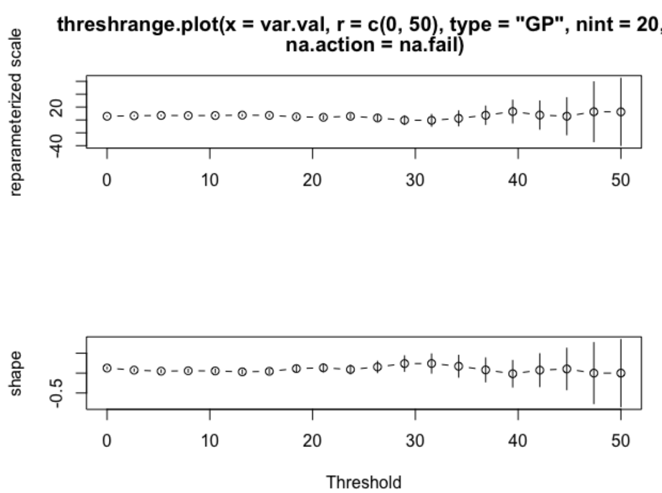

Figure 3: The plot of parameter choices depending on threshold.

parameter	value
scale	7.440252
shape	0.184498
negative log likelihood	485.0937

Table 1: The values of the parameters asked for in the instructions.

	scale	shape
scale	2.000796	12.79684
shape	12.796835	179.48456

Table 2: The hessian of the model parameters.

4

The task in the description was performed and the asked for results are presented below.

5

Running the suggested command gave the same estimated parameters, and comparing to the inverse hessian, the parameters ended up with the same covariance matrix.

6

The estimated parameter covariance matrix is presented below.

	scale	shape
scale	0.9187666	-0.06550540
shape	-0.0655054	0.01024179

7

Running the command solve(rain$models$fit1$results$hessian) gives the same covariance matrix as in question 6.

8

The fitted values as well as lower and upper endpoints for the 95 % confidence intervals of the 2 parameters of the GP model are presented below. Here it may be noted that the confidence interval of the shape parameter includes zero.

This indicates that the shape parameter cannot be significantly considered to be non-zero, suggesting it might not have earned its place in the model. This suggests a model without a shape parameter can be considered.

	lower	estimate	upper
scale	5.56158139	7.440252	9.3189230
shape	-0.01385379	0.184498	0.3828497

9

Same result as above.

10

The 95 % confidence intervals based on profile likelihood is presented in the table below. It is clear that these differ from the ones based on asymptotic normal distribution and for example zero is not included in the confidence interval of the shape.

	lower	upper
scale	5.9302	9.5062
shape	0.0139	0.4133

11

The result from running the suggested command (for the two parameters respectively) resulted in the 95 % Confidence Intervals; scale: (5.7563, 9.5092), shape: (0.015, 0.4145).

The result differ from the ones obtained in the previous question. The reason for this is the added range within which the profile likelihood is performed. When finding the profile likelihood confidence interval using the interface in2extRemes,the search ranges are chosen automatically, and hence differ from the results with the suggested method.

12

The expected return levels were calculated according to the formulas given in the course literature and the results obtained were: 10-year return level: 65.961 and 100-year return level: 106.342.

13

The following confidence intervals were obtained by use of the delta method in the interface:

Return period	lower	fit	upper
10 years	55.893	65.961	76.0299
100 years	65.6224	106.342	147.0622

Table 3: Confidence intervals for question 13

14

The 95% confidence interval for the 10-year return level based on profile likelihood was (58.6417, 81.1786). The 95% confidence interval for the 100-year return level based on profile likelihood was (80.973, 184.582).

15

As can be seen in the diagnostics plot in Figure 4, the model has a decent qq-plot, however exhibiting some larger deviances in the higher quantiles. Furthermore, the empirical quantiles fit the simulated data well. There is a big difference i the spike around 0 for the empirical and theoretical pdf, however apart from this they follow each other nicely.

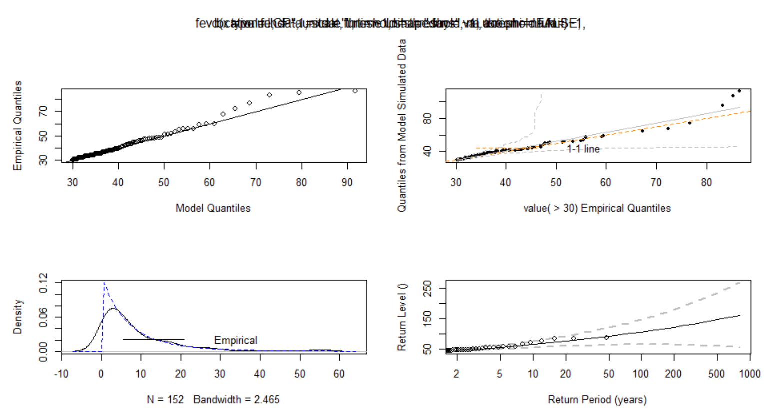

Figure 4: Diagnostics plot for question 15

16

The question is interpreted to mean that $\gamma = \xi = 0$, which yields a special case of the Generalized Pareto Model. Modelling the excess mean $y = X - u,\ y > 0$ with this assumption means that y will follow an exponential distribution, with pdf $f(y) = \frac{e^{-y/\tilde{\sigma}}}{\tilde{\sigma}}$, where $\tilde{\sigma} = \sigma + \xi(u - \mu) = \sigma$, as $\xi = 0$. The ML estimate of $\tilde{\sigma} = \sigma$ with exponential distributed y just so happens to coincide with the mean, i.e. $\hat{\sigma} = \bar{y} = 9.084211$.

17

Running the suggested command results in the same estimation of σ as in question 16.

18

The variance of the maximum likelihood estimate for the scale parameter was calculated according to $\hat{\sigma}^2 \frac{n}{(n-1)^2(n-2)}$. The 95% confidence interval for the scale parameter under the hypothesis given in the instruction was (7.640062, 10.528359).

19

The likelihood ratio test between the two models was conducted with the result $LR = 2(logL(m_2) - logL(m_1)) = 4.6$, where m_1 is the more restrictive, exponential model, and m_2 is the GPD model provided by the interface's functions. The difference in degrees of freedom was 1, and so the LR was compared to the 95%quantile of the $\chi^2(1)$ distribution, which was 3.841459. LR clearly exceeds this, and the null hypothesis that $\xi = 0$ can therefore be rejected at a 95% confidence level.

20

In 5 the QQ-plot is shown comparing the quantiles of the model to the empirically obtained quantiles. The quantiles seem reasonably linear in the beginning but for larger values they deviate. The QQ-plot for setting the shape parameter to zero looks a bit worse than the QQ-plot obtained for the model where the shape parameter is non-zero. The units differ between the two plots but this does not influence the look of the plot.

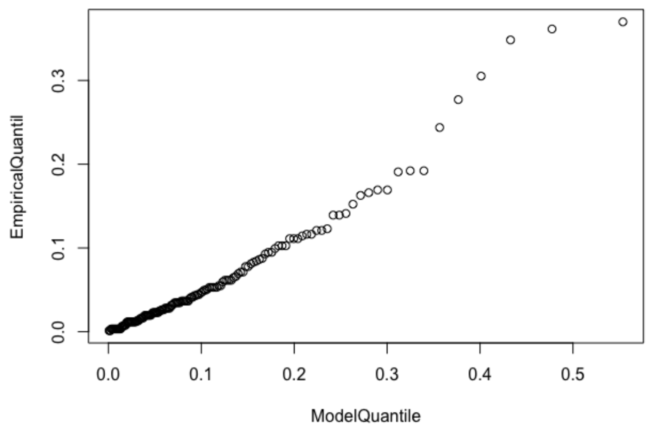

Figure 5: The QQ-plot comparing the quantiles of the model to the empirically obtained quantiles.

3.1

	$u=-10$		$u=-20$	
	$r=2$	$r=4$	$r=2$	$r=4$
n_c	26	17	40	28
$\hat{\sigma}$	14.0(3.9)	14.9(5.5)	14.2(3.1)	19.8(5.1)
$\hat{\xi}$	−0.4(0.2)	−0.4(0.3)	−0.24(0.16)	−0.43(0.19)
$\hat{x}_{100}$	23.7	24.3	30.0	24.0
$\hat{\theta}$	0.35	0.23	0.21	0.15

Table 4: Table acc. to analysis in section 5.3.3 in the course book

3.2

Looking at Table 4, it becomes apparent that setting $u = -20$ yields larger n_c for both values of r. This means more stability in the parameter estimations, which can be seen in the parentheses containing the standard error of the parameter estimations. The high value of $\hat{x}_{100}$ for $u = -20$ and $r = 2$ can possibly be due to a too small value of r which would mean that dependent samples are treated as independent. It is also clear that value of the shape parameter differs for these values. On a final note it should be noted that all confidence intervals for the estimations above are fairly large, which is likely to do with the relatively small data set.