Lund Institute of Technology
Mathematical Statistics

Centre for Mathematical Sciences

Statistical Modeling of Extreme Values,
FMSN55/MASM15

Computer Assignment 1

This assignment is a compulsory part of the course. At the end of the session each group's results will be reviewed and graded as *pass* or *fail*. Note that in order to be able to finish the assignment in the specified time above, you have to read through the whole assignment in addition to the necessary parts in the book and manuals described below **before** you attend the computer session.

Block maxima

Sea level Data

The main objective of this computer assignment is to get familiar with the library `extRemes` and its graphical user interface `in2extRemes` in `R` which contain functions for extreme value analysis discussed in the book by S. Coles.
Read through the whole analysis in Section 3.4.1 of the book and during the computer session check that you get the same results as those reported in the book. You should also read Appendix A in Coles' book regarding computational aspects of this analysis in `S-Plus` which also holds for `R`.

Getting started

Log in at one of the PCs in the computer room MH:230 or MH:231 using your *STIL*-account. Click on the icon "MClogin" on the desktop and login again with the same user name and password. This will attach the hard drive "`L:`" where your working directory will be saved. Note that you need to do this **before** you start up the software package `R`. Choose the latest version of `R` from the `Start` menu. If you have problems either logging in or starting `R`, ask for help.
Some tips and hints when typing in `R` code:

- `R` is case sensitive (so `LM()` and `lm()` are not equivalent).
- `R` is tolerant to the use of spaces, so `x <- 1` and `x<-1` are equivalent; though, the former being considered to be more readable.
- You can use the arrow keys to speed things up. The 'up' arrow gives you the previous command that you typed.
- The usual prompt sign for `R` is `>`. If you get a `+` prompt sign instead, it means that `R` is awaiting the completion of the previous command that you typed in. This can happen because you have forgotten to close parentheses, for instance. Just type in the remainder of the command.

We will use the data on annual maximum sea levels at Port Pirie, South Australia. The dataset was introduced first in Example 1.1, page 4 of the book and the extreme value analysis of data has been discussed in Section 3.4.1, pages 59-64. To repeat this analysis, start `R` and write `library(in2extRemes)` to attach the related libraries and get access to the dataset and `R` functions. To start the graphical user interface run the command `in2extRemes()` in `R` window.
The dataset on maximum sea levels is stored in `R` dataset `portpirie.R` and can be accessed in the following locations

- `/usr/common/extremvarde/R/datasets/` (from Linux computers in the labs)
- `P:\Rdata` (from Windows computers in the labs)
- `http://www.maths.lth.se/matstat/kurser/fms155mas231/datasetsR.html`.

To read the dataset into `R` choose `File`→`Read Data` in `in2extRemes` window. Note that the dataset has been saved as `R` source so you should choose `R source` under `File Type` in the window which pops up. You do not need to change any other options but do not forget to assign a name of your choice to the dataset under `Save As(in R)`. In the follwing we will assume that you have saved the dataset with the name `portpirie` in `R`. Otherwise you just need to change the following commands accordingly.

Statistical analysis

It is important to understand how you can access different results of your analysis in `extRemes` package. After reading the dataset you can run the

command `names(portpirie)` to see the names of different parts in the list which has been created. To access each part of the list you need to add $ and the corresponding name to the object. For instance, `portpirie$data` contains the data.
Throughout this assignment assume all confidence intervals have confidence level 95%.

1. Plot the sea level data against year (`Plot`→`Scatter Plot`).

2. Find maximum likelihood estimates of the parameters for generalized extreme value distribution (`Analyze`→`Extreme Value Distributions`). Run the command `names(portpirie)` again. Note that a new component with the name `models` has been added to the object. Writing `names(portpirie$models)` will show you the models which you have fitted so far. If this is your first analysis you will see only `fit1`. Write `names(portpirie$models$fit1)` to see which parts are included as the result of the fit. For instance `portpirie$models$fit1$results` will print the parameter estimates, negative log likelihood, hessian matrix and so on.
   Answer:

3. Find covariance matrix of ML-estimates (`Analyze`→`Fit Summary`). You can also find the covariance matrix by using either `summary(portpirie$models$fit1)` or `parcov.fevd(portpirie$models$fit1)`.
   Answer:

4. Note that the covariance matrix in `in2extRemes` is calculated based on the inverse of observed Fisher information. To see this calculate the inverse of the hessian matrix by `solve(portpirie$models$fit1$results$hessian)`.

Does the result agree with your calculations in the previous part?
Answer:

5. Confidence intervals for the parameters in GEV based on asymptotic Normal distribution of MLEs (`Analyze→Parameter Confidence Intervals`).
Answer:

6. Check that your results for confidence intervals agree with what you get from this command: `ci(portpirie$models$fit1,type="parameter")`.
Answer:

7. Confidence interval for the shape parameter in GEV based on profile likelihood using the following command[1]:

[1]Note that according to the `in2extRemes` tutorial (page 35) `in2extRemes` only allows for changing the search limits for return levels (for which this approach is arguably more

`ci(portpirie$models$fit1,type="parameter",method="proflik", which.par=3,xrange= c(-0.5,0.5))`.
Answer:

8. Maximum likelihood estimates of 10- and 100-year return levels and their variances. Note that for matrix multiplication in `R` you have to use "`%*%`" . (see page 56 in the book)
Answer:

9. Confidence intervals for 10- and 100-year return levels based on delta method (`Analyze`→`Parameter Confidence Intervals`).
Answer:

10. Confidence intervals for 10- and 100-year return levels based on profile likelihood method (`Analyze`→`Parameter Confidence Intervals`). Change

necessary) and not individual parameter estimates. The software will attempt to find an appropriate search range, but will not always succeed; leading to intervals that are clearly incorrect (e.g., the estimated parameter does not fall inside the limits). In such a case, the search range can be changed in the command-line code of `extRemes`, but not from the `in2extRemes` window.

the "Number of points at which to calculate profile likelihood" to 100 and the "Profile Search Range" to 4 to 6. Alternatively, you can find the profile likelihood interval for 10-year return level with this command:

```
ci(portpirie$models$fit1,type="return.level",return.period=10,
method="proflik",nint=100,xrange=c(4,6)).
```

Answer:

11. Maximum likelihood estimate of upper endpoint of distribution (see pages 56-57 in the book).
Answer:

12. Confidence intervals for upper endpoint of distribution based on delta method (see pages 56-57 in the book). Comment on the result.
Answer:

13. Diagnostic plots for the GEV fit with discussion of different plots (`Plot→Fit Diagnostics`).
Answer:

14. Note the the diagnostic plots obtained above are different from what you see in the book. The reason for this is that the default plots in `in2extRemes` package have been changed since the book was published. To get similar plots as in the book you can use the following commands (assuming you have saved the data with the name `portpirie` and you want to plot for `fit1`):

```
par(mfrow=c(2,2))
plot(portpirie$models$fit1,type="probprob",main="Probability plot")
plot(portpirie$models$fit1,type="qq",main="Quantile plot")
plot(portpirie$models$fit1,type="rl",main="Return level plot")
plot(portpirie$models$fit1,type="hist",main="Histogram")
```

15. Maximum likelihood estimates of the parameters in Gumbel distribution (`Analyze`→`Extreme Value Distributions`). What are the parameters estimates in this model?
Answer:

16. Covariance matrix of ML-estimates in Gumbel distribution.
Answer:

17. Confidence intervals for the parameters in Gumbel distribution based on observed information matrix.
Answer:

18. Diagnostic plots for the Gumbel fit with discussion of different plots.
Answer:

19. Likelihood ratio test for hypothesis $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$ (`Analyze→Likelihood-ratio test`). Does the Gumbel distribution provides an acceptable model for the data?
Answer: