Statistical Assessment of Extreme Weather Phenomena Under Climate Change

NCAR Advanced Study Program Summer Colloquium 2011, 624 June

Practice Sets for the R tutorial on EVA in R

1 Fitting the GEV to data

1. Simulate a sample of size 500 from the GEV distribution with location 3, scale 1.5 and shape 0.1.
2. Fit the GEV to the simulated data from 1 above, and check the assumptions (i.e., analyze the qq-plot).
3. Do the assumptions for fitting the GEV to these simulated data appear reasonable?
4. Plot the profile likelihood for the shape parameter found from the fit in 2 above. What are the 95% CI for this parameter? Is this parameter significantly different from zero at the 5% significance level?
5. Load the `SEPTsp` data set from package `extRemes`.
6. See the help file for this data set to learn what each field represents, and other information. For example, what is the source of these data? Where might one learn more about them?
7. Make a line plot of the maximum temperature over a one month period against year for these data.
8. Make a scatter plot of the standard deviation of maximum temperature against the maximum temperature. Does there appear to be much correlation?
9. Fit the GEV to the maximum temperature field.
10. Make a QQ-Plot for this fit. Do the assumptions for using the GEV appear reasonable for these data?

11. Estimate 95% CI's for the shape parameter. What can you say about the behavior of maximum temperature for Sept-Iles, Québec based on these data?

12. Make a line plot of the year against minimum temperature.

13. Fit the GEV to the minimum temperature data, and check the QQ-Plot (**Hint**: remember to take the negative transformation of the variable first).

14. Estimate a 95% CI for the shape parameter. What can you say about minimum temperature for Sept-Iles, Québec based on these data?

15. Load the `Denmint` data set from `extRemes`.

16. Take the annual maximum of the negative of the minimum temperature. (**Hint**: use `DenmintAM <- aggregate( -Denmint$Min, by=list( Denmint$Year), FUN=max, na.rm=TRUE)$x`).

17. Make a line plot of year against negative minimum temperature. Does there appear to be any temporal trend in these data? (**Hint**: use `yr <- unique( Denmint$Year)` and `plot(yr, DenmintAM, type="l")`).

18. Fit a linear regression of year against negative minimum temperature (**Hint**: See the help file for `lm`). Is there a significant linear trend in these data (**Hint**: use the `summary` function on the `lm` fitted object)?

19. Use `cbind` to make a new matrix that has `yr` and `DenmintAM` as its columns, and give the columns names. Fit the negative minimum temperature to a GEV (without any trend).

20. Make the QQ-Plot for this fit. Do the model assumptions appear to be reasonable?

21. Test the Gumbel hypothesis for these data (i.e., the null hypothesis is that $\xi = 0$).

22. Make a return level plot for the negative minimum temperature with (delta method) 95% CI's. **Hint**: see the help file for `return.level`.

23. Interpret this return level plot for a gas/power company wanting to understand the risk of too much demand for gas in Denver in any given year (**Hint**: remember the return levels are for the negative of minimum temperature).

24. Fit the negative minimum temperature data to the GEV with a linear trend in the location parameter for $t = 1, 2, \ldots$.

25. Check the QQ-Plot for this fit. Do the assumptiions for the model fit appear to be reasonable?

26. Perform a likelihood ratio test for $\mu_1 = 0$ in the fit from above. Is the result consistent with the result from the regression fit?

2 Threshold Excess Models

1. Load the Phoenix minimum temperature data (`Tphap`) from `extRemes`.

2. Make a line plot of these data? What features do you notice?

3. Take the negative transform the variable `MinT` from `Tphap`.

4. Fit the GPD over a range of thresholds to the (negative) `MinT` variable. Does -73 degrees appear to be a reasonable threshold? (**Hint**: You may need to try different ranges and numbers of thresholds).

5. Fit the GPD using a threshold of -73 degrees to the (negative) `MinT` variable.

6. Decluster the (negative) `MinT` series for the threshold of -73 degrees (decluster by `Year`), and re-fit the GPD to the de-clustered series (**Hint**: see the help file for `dclust`). How does the fit compare to the previous one? How do the QQ-Plots for each compare?

7. What conclusions would you make about the appropriateness of a stationary GPD model for these data? Do the underlying assumptions appear to be valid based on the qq-plots?

8. Make a return level graph for the fit to the non-declustered data. From the R prompt, type `X11()` to open a new plot device, and then make a return level graph for the fit to the de-clustered data. Compare the two graphs side-by-side. Do they differ much? (**Hint**: you can use `gpd.diag` or `return.level` to do this, the former will also give the qq-plot among others).

9. Load the data set called `Denversp` from the package `extRemes`.

10. See the help file for this data set to learn what it contains.

11. Make a scatter plot of precipitation against hour. What do you notice?

12. Make a scatter plot of precipitation against day and then year. Any patterns or trends?

13. Choose a threshold for fitting the GPD to these data (**Hint**: use 0.1 and 0.8 as the lower and upper limits in `gpd.fitrange`). Does 0.395 mm appear to be a reasonable choice for a threshold?

14. Make another scatter plot of precipitation against hour, and add a red horizontal dashed line at 0.395 (**Hint**: `abline( h=0.395, col="red", lwd=2)`). Do the data appear to be independent over the threshold?

15. Fit a GPD to the Denver precipitation data. Plot the diagnostics for the fit. Is this model reasonable for these data?

16. Estimate a 95% CI for the shape parameter. Is the shape parameter significantly different from zero at the 5% level for either fit?

17. Decluster the precipitation field from `Denversp` using runs de-clustering with `r=1`.

18. Re-fit the newly de-clustered field to the GPD. Is this fit any different from the previous ones?

19. Estimate the Poisson rate parameter associated to a threshold of 0.395 mm for the de-clustered precipitation data.

20. Estimate the Poisson rate parameter. Is it nearly the same as the estimate obtained above? (**Hint**: use the relation $\hat{\lambda} = \left[1 + \frac{\hat{\xi}}{\hat{\sigma}}(u - \hat{\mu})\right]^{-1/\hat{\xi}}$).

21. Make a QQ-Plot for the point process model fit (if you haven't already). Do the assumptions for the model appear to be reasonable?

22. Now, let's fit the Poisson rate parameter including an annual cycle using the `glm` function. First, make a binary vector where 1 indicates an excess over 0.395 mm (**Hint**: use `ind <- Denversp$Prec > 0.395`). Next, make a data frame containing the cyclic trends over time (i.e., containing $\sin(2\pi t/365.25)$ and $\cos(2\pi t/365.25)$, where $t =$ `Denversp$Hour`). Now use `glm` with `family=poisson()` to fit the model $\hat{\lambda}(t) = \hat{\lambda}_0 + \hat{\lambda}_1 \sin(2\pi t/365.25) + \hat{\lambda}_2 \cos(2\pi t/365.25)$ (where $\hat{\lambda}(t)$ is the indicator vector), and use `summary` to see the results. Is there a significant (say at the 5% level) annual cycle in the Poisson rate parameter?

23. Fit the Denver precipitation data to a point process model with no parameter covariates, and threshold of 0.395 mm.

24. Check the QQ-Plot. Are the model assumptions reasonable?

25. Fit the Denver precipitation data to a point process model for a threshold of 0.395 mm, and with a cyclic variation in the location parameter as $\hat{\mu}(t) = \hat{\mu}_0 + \hat{\mu}_1 \sin(2\pi t/24)$ for $t =$`Denversp$Hour`.

26. Perform a likelihood ratio test for $\mu_1 = 0$ in the above model. Is the fit significant? Are the model assumptions reasonable?

27. Now do the same but use both the sine and cosine fields. You might need to use `Method = BFGS quasi-Newton`. Are the model assumptions reasonable here?

28. Fit the Denver precipitation data to a point process model with a cyclic trend in the scale parameter (i.e., $\log \sigma(t) = \sigma_0 + \sigma_1 \sin(2\pi t/24) + \sigma_2 \cos(2\pi t/24)$). Is the trend significant? Are the model assumptions reasonable based on the QQ-Plot?

29. Given the results here, and the results from declustering previously, which approach would you recommend for these data?

3 Beyond

1. List out the arguments for the function, `optim` (**Hint**: use the `args` function).

2. See the help file for the function, `optim`.

3. Analyze the `Peak` data set from `extRemes`. Is a block maxima or threshold excess model more appropriate here? Do there appear to be any trends in the data?

4. Analyze the maximum winter temperature for Sept-Iles. Do any of the other fields included with the data set make sense to try as covariates?