**1.** Researchers at the École Polytechnique Fédérale de Lausanne, in Switzerland, are analysing snow depth recordings as part of a research project looking at factors contributing to the occurrence of avalanches. To investigate, for a set of daily snow depth measurements (in cm) at a location in the Swiss Alps, a researcher extracts the maximum value from each year, over the period 1961–2012 (inclusive). This set of annual maxima is stored in the vector `snow` in R.

A function called `gev.loglik` is written, in R, to return the *negative* log–likelihood of the generalised extreme value (GEV) distribution; this is a function of the parameter vector for the GEV (`theta`) and the annual maximum snow depths (`snow`). The non–linear minimisation routine `nlm` is then applied to `gev.loglik`:

```
> theta=c(mean(snow),sd(snow),0.1)                #LINE 1
> A=nlm(gev.loglik,theta,hessian=TRUE)
> A
$minimum
[1] 264.815

$estimate
[1] 75.81065681 34.54750759 -0.05934882

$gradient
[1]  4.326391e-07 -3.751443e-07  6.878054e-06

$hessian                                          #LINE 13
            [,1]        [,2]       [,3]
[1,]  0.04182309 -0.01241791  0.6181536
[2,] -0.01241791  0.07925069  0.9518501
[3,]  0.61815359  0.95185010 90.8240111

$code
[1] 1

$iterations
[1] 27
```

(a) Using the R output above,

(i) *briefly* explain `#LINE 1` of the code;

**Answer:**

Initialises the GPD parameter vector $\theta=(\mu,\sigma,\xi)$ at $\mu = \text{mean}$, $\sigma = \text{s.d.}$ and $\xi = 0.1$. Necessary to implement the nlm routine.

1

(ii) Write down the maximum likelihood estimates of the GEV parameters $\mu$, $\sigma$ and $\xi$, to three decimal places, and use these to estimate the 50 year return level snow depth at this location;

**Answer:**

(1) $\hat{\mu} = 75.811$

$\hat{\sigma} = 34.548$

$\hat{\xi} = -0.059$

(1) $$\hat{z}_{50} = 75.811 - \frac{34.548}{0.059}\left[\left(-\log\left(1-\frac{1}{50}\right)\right)^{0.059} - 1\right]$$

$$= 196.224 \text{ cm}$$

(iii) write down the value of the log–likelihood function at the estimates you declared in part (ii);

**Answer:**

(1) $-264.815$

(iv) write down the observed information matrix $I_O$, with reference to #LINE 13 of the code;

**Answer:**

(1)

$$I_O = \begin{pmatrix} 0.048 & -0.012 & 0.618 \\ -0.012 & 0.079 & 0.952 \\ 0.618 & 0.952 & 90.824 \end{pmatrix}.$$

(v) write down a line of R code that would give us the variance–covariance matrix for the GEV parameters.

**Answer:**

(1)

```
> solve(A$hessian)
```

*Question 1 continued on next page*

(b) Applying the code you gave in part (a)(v) (provided your code is correct!), gives the following variance–covariance matrix $V$ for $\boldsymbol{\theta} = (\mu, \sigma, \xi)^T$:

$$V = \begin{pmatrix} 30.806 & & \\ 8.403 & 16.272 & \\ -0.298 & -0.232 & 0.015 \end{pmatrix}.$$

In the past, a Gumbel distribution has been used to model extreme snow depths at this site. Comment on the suitability of the simpler Gumbel model, with reference to the matrix $V$ above.

**Answer:**

The 95% CI for $\xi$ is $-0.059 \pm 1.96 \times \sqrt{0.015}$, giving

1 $(-0.299, 0.181)$.

1 Since this includes zero, the simpler Gumbel model may well be suitable.

(c) The researchers at École Polytechnique Fédérale de Lausanne believe that, in any year, an avalanche is likely to occur if the annual maximum snow depth exceeds 1.7 metres. For this site in the Swiss Alps, estimate the number of years, in the next 200 years, in which we can expect to see an avalanche.

**Answer:**

2 $$P(\text{snow depth} > 170\text{cm}) = 1 - \exp\left\{-\left[1 + \hat{\xi}\left(\frac{170 - \hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\}$$

$$= 1 - \exp\left\{-\left[1 - 0.059\left(\frac{170 - 75.811}{34.548}\right)\right]^{-1/-0.059}\right\}$$

1 $$= 0.0499.$$

So we expect $200 \times 0.0499 \approx 10$ years of avalanches!

(d) Researchers also believe there could be relationship between the annual maximum snow depths year–on–year. How would such a relationship affect the analysis performed in this question?

**Answer:**

2 Not really affected – different $\mu$ and $\sigma$, but we estimate these parameters anyway. Of course, this is provided there is no <u>long-range</u> dependence – <u>ie</u> "Leadbetter's $D(u_n)$ condition" holds.

[Total Q1: 13 marks]

**2.** Given the recent increase in frequency and severity of summer heatwaves across Mediterranean Europe, the European Drought Centre (EDC) was established in 2011 to analyse historical temperature records across several locations off the coast of Spain, France and Italy.

An analyst at the EDC uses a threshold–based approach to analyse maximum daily temperatures at Ibiza, in the Balearic Islands, for a period of 43 years (11 of which were leap years). Data for June 1972 are completely missing.

A mean residual life plot suggests using a threshold of 30.5°C to identify temperatures as extreme. To address the problem of short term dependence between consecutive extremes, the analyst employs runs declustering, arbitrarily choosing a separation interval of $\kappa = 10$ hours; this gives 97 'independent' cluster peak excesses to which the generalised Pareto distribution (GPD) can be fitted.

On fitting the GPD, the analyst computes the following 95% confidence intervals for the scale and shape parameters $\sigma$ and $\xi$ (respectively):

$$\sigma \;:\; (3.313, 5.828)$$
$$\xi \;:\; (-0.971, -0.492)$$

(a) Comment on the tail of the fitted distribution. Do your observations here seem sensible given the type of data being analysed? **[2 marks]**

**Answer:**

Weibull-type tail, since the CI for $\xi$ is wholly negative. Thus, the tail of the fitted distribution is bounded above—seems sensible for temperature extremes.

(b) Using this model, the $r$–*year return level*, $z_r$, can be estimated as

$$\hat{z}_r \;=\; u + \frac{\hat{\sigma}}{\hat{\xi}}\left[(r n_y \hat{\lambda}_u)^{\hat{\xi}} - 1\right], \tag{1}$$

where $r$ is the return period, $n_y$ is the average number of observations per year, and, generically, $\hat{\theta}$ is the maximum likelihood estimate of $\theta$.

(i) In plain English, explain what is meant by the $r$–*year return level.* **[1 mark]**

**Answer:**

The value that is exceeded once, on average, every $r$ years.

*Question 2 continued on next page*

(ii) Briefly explain the role of $\hat{\lambda}_u$ in Equation (1), and find its value in this analysis of temperature extremes at Ibiza. **[2 marks]**

**Answer:**

(1) Gives us the unconditional distribution of a threshold exceedance Y, and represents the threshold exceedance rate.

We have

(1)
$$\hat{\lambda}_u = \frac{97}{(32 \times 365) + (11 \times 366) - 30}$$

$$= 0.00619.$$

(iii) Use your answer to part (ii) above, and the confidence intervals for $\sigma$ and $\xi$ given at the start of the question, to estimate the 100–year return level temperature on the island of Ibiza. **[4 marks]**

**Answer:**

(2)
$$\left\{\begin{array}{l} \sigma : (3.313, 5.828) \longrightarrow \hat{\sigma} = 4.5705 \\ \xi : (-0.971, -0.492) \longrightarrow \hat{\xi} = -0.7315 \end{array}\right.$$

so

$$\hat{z}_{100} = u + \hat{\sigma}/\hat{\xi}\left[(r n_y \hat{\lambda}_u)^{\hat{\xi}} - 1\right]$$

(2)
$$= 30.5 - \frac{4.5705}{0.7315}\left[(100 \times 365.25 \times 0.00619)^{-0.7315} - 1\right]$$

$$= 36.630^\circ C$$

Question 2 continued on next page

(c) The analyst working for the EDC requires a standard error for the 100–year return level estimate obtained in part (b)(iii). He is going to use the *delta method* to do this, giving:

$$\mathsf{var}(\hat{z}_{100}) \approx \nabla z_{100}^T V \nabla z_{100},$$

where

$$\nabla z_{100}^T = \left[\frac{\partial z_{100}}{\partial \sigma}, \frac{\partial z_{100}}{\partial \xi}\right] = [1.341, 7.737],$$

evaluated at $\hat{\lambda}_u$, $\hat{\sigma}$ and $\hat{\xi}$, and

$$V = \begin{pmatrix} \mathsf{var}(\hat{\sigma}) & \mathsf{cov}(\hat{\sigma}, \hat{\xi}) \\ \mathsf{cov}(\hat{\sigma}, \hat{\xi}) & \mathsf{var}(\hat{\xi}) \end{pmatrix}.$$

(i) *Briefly* explain what the analyst will be doing wrong here. What effect would this have on the standard error? **[2 marks]**

**Answer:**

He will not be accounting for our uncertainty in $\hat{\lambda}_u$. This will result in an under-estimated standard error for $\hat{z}_{100}$.

(ii) Obtain the correct standard error for the 100–year return level. *[Hint:* $\mathsf{cov}(\hat{\sigma}, \hat{\xi}) = -0.078$*]* **[14 marks]**

**Answer:**

We need to use

$$\nabla z_{100}^T = \left[\frac{\partial z_{100}}{\partial \lambda_u}, \frac{\partial z_{100}}{\partial \sigma}, \frac{\partial z_{100}}{\partial \xi}\right], \text{ and}$$

$$V = \begin{pmatrix} var(\hat{\lambda}_u) & 0 & 0 \\ 0 & Var(\hat{\sigma}) & Cov(\hat{\sigma}, \hat{\xi}) \\ 0 & cov(\hat{\sigma}, \hat{\xi}) & Var(\hat{\xi}) \end{pmatrix},$$

where $Var(\hat{\lambda}_u) = \dfrac{0.00619 \times 0.99381}{(32 \times 365) + (11 \times 366) - 30} = 3.924 \times 10^{-7}$.

1

[redacted]

*[This page is left blank for your solution to the last question]*

For $Var(\hat{\sigma})$:

$$\hat{\sigma} + 1.96 \times s.e.(\hat{\sigma}) = 5.828$$

$$\Rightarrow \quad s.e.(\hat{\sigma}) = \frac{5.828 - 4.5705}{1.96} = \underline{0.64158}.$$

① $\underline{\underline{\text{so}}}$ $Var(\hat{\sigma}) = 0.64158^2 = 0.4116.$

For $Var(\hat{\xi})$:

$$\hat{\xi} + 1.96 \times s.e.(\hat{\xi}) = -0.492$$

$$s.e.(\hat{\xi}) = \frac{-0.492 + 0.7315}{1.96} = \underline{0.1222}$$

① $\underline{\underline{\text{so}}}$ $Var(\hat{\xi}) = 0.1222^2 = 0.0149.$

So we have

② $$V = \begin{pmatrix} 3.924 \times 10^{-7} & 0 & 0 \\ 0 & 0.4166 & -0.078 \\ 0 & -0.078 & 0.0149 \end{pmatrix}$$

Also

$$\frac{\partial z_{100}}{\partial \lambda_u} = \frac{\partial}{\partial \lambda_u}\left[u + \left(\frac{\sigma}{\xi}(rn_y)^{\xi}\right)\lambda_u^{\xi} - \sigma/\xi\right]$$

② $$= \xi \times \sigma/\xi \, (rn_y)^{\xi} \lambda_u^{\xi - 1}$$

$$= \sigma (rn_y)^{\xi} \lambda_u^{\xi - 1}$$

*Question 2 continued on next page*

*[This page is left blank for your solution to the last question]*

At $\sigma = \hat{\sigma}$, $\xi = \hat{\xi}$ and $\lambda_u = \hat{\lambda}_u$, we have

(1) $$\frac{\partial z_{100}}{\partial \lambda_u} = 13.99963.$$

So

(5) $$\mathrm{Var}(\hat{z}_{100}) \approx [13.99963, 1.341, 7.737] \times \begin{pmatrix} 3.924\times10^{-7} & 0 & 0 \\ 0 & 0.4166 & -0.078 \\ 0 & -0.078 & 0.0149 \end{pmatrix} \times \begin{bmatrix} 13.99963 \\ 1.341 \\ 7.737 \end{bmatrix}$$

$$\approx (5.493\times10^{-6}, -0.0448, 0.0107) \times \begin{bmatrix} 13.99963 \\ 1.341 \\ 7.737 \end{bmatrix}$$

$$\approx 0.022786$$

(1) So, finally, we have

$$\text{s.e.}(\hat{z}_{100}) = \sqrt{0.022786}$$
$$= 0.151.$$

*Question 2 continued on next page*

(d) Why should the standard error you obtained in (c)(ii) not be used to construct a confidence interval for $z_{100}$? What procedure *should* be used? **[2 marks]**

**Answer:**

This would assume the likelihood surface for $\hat{z}_{100}$ is symmetric – it usually is not.

Profile log-likelihood should be used.

(e) The cluster separation interval $\kappa = 10$ was chosen arbitrarily. Briefly discuss how this could be problematic. **[2 marks]**

**Answer:**

If $\kappa$ is too small, successive cluster peaks will not be independent.

If $\kappa$ is too large, there will be too few cluster peaks to work with.

(f) Given the nature of the data being analysed, can you think of any other problems with this analysis of extremes as it stands? **[1 mark]**

**Answer:**

Temperature extremes will probably exhibit non-stationarity – most likely seasonal variability.

[**Total Q2: 30 marks**]

**3**. Suppose $X_1, X_2, \ldots, X_n$ is a sequence of independent $Exp(2)$ random variables. Show that the limiting distribution of excesses over a high threshold $u$ belongs to the Generalised Pareto family, and give the value of the scale and shape parameters.

**Answer:**

We want

4

$$\frac{1-F(u+y)}{1-F(u)} = \frac{1-\left\{1-e^{-2(u+y)}\right\}}{1-\left\{1-e^{-2u}\right\}}$$

$$= \frac{\cancel{e^{-2u}} \times e^{-2y}}{\cancel{e^{-2u}}} = e^{-2y}.$$

$\xi = 0$ 1 For a $GPD(\tilde{\sigma}, \xi = 0)$, we have

$$1 - H(y) = e^{-\frac{1}{\tilde{\sigma}}y},$$

$\tilde{\sigma} = \frac{1}{2}$ 1 so here $\tilde{\sigma} = \frac{1}{2}$ to give $1 - H(y) = e^{-2y}$.

1 Thus, the limiting distribution is $GPD(\frac{1}{2}, 0)$.

[Total Q3: 7 marks]

THE END