

MAS3918/8918: Extremes

Problems sheet 2: Solutions to assessed questions

1. (15 marks in total)

- (a) Assuming a Gumbel-type tail implies that $\xi = 0$ when using the GPD for the set of threshold exceedances. From question 10 in Chapter 7, we know that the log-likelihood is therefore

$$\ell(\sigma; \mathbf{y}) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^k y_i;$$

we now maximise $\ell(\sigma; \mathbf{y})$. Now

$$\frac{\partial \ell}{\partial \sigma} = -k/\sigma + \sum_{i=1}^k y_i \sigma^{-2}. \quad (2 \text{ marks})$$

Equating to zero, and replacing σ with $\hat{\sigma}$, gives

$$\begin{aligned} \frac{k}{\hat{\sigma}} &= \frac{\sum_{i=1}^k y_i}{\hat{\sigma}^2} \\ k\hat{\sigma}^2 &= \hat{\sigma} \sum_{i=1}^k y_i \\ \hat{\sigma} &= k^{-1} \sum_{i=1}^k y_i = \bar{y}. \quad (1 \text{ mark}) \end{aligned}$$

Subtracting $u = 40$ from all observations, and then discarding any resulting values which are less than zero, gives the set of threshold exceedances $\mathbf{y}$:

$$8 \quad 1 \quad 6 \quad 12 \quad 17 \quad 11 \quad 3$$

Thus, for the pollution concentration extremes, we have

$$\hat{\sigma} = \frac{8 + 1 + \dots + 3}{7} = 8.286. \quad (1 \text{ mark})$$

- (b) We have

$$\frac{\partial^2 \ell}{\partial \sigma^2} = k\sigma^{-2} - 2 \sum_{i=1}^k y_i \sigma^{-3}; \quad (1 \text{ mark})$$

evaluating at $\sigma = \hat{\sigma} = 8.286$ gives

$$I_O = - \left(\frac{7}{8.286^2} - 2 \times \frac{58}{8.286^3} \right) = 0.1019; \quad (1 \text{ mark})$$

thus, $\text{var}(\hat{\sigma}) = I_O^{-1} = 9.8135$, giving $\text{e.s.e.}(\hat{\sigma}) = \sqrt{9.8135} = 3.1327$. (2 marks)

(c) For the t -observation return level, we need to solve

$$\Pr(X > x) \approx \lambda_u \cdot S(x) = \frac{1}{t} \quad (*)$$

for $x = z_t$, where S is the survival function ($1 - \text{CDF}$) for the Generalised Pareto Distribution.

In the case $\xi \neq 0$ we have

$$\begin{aligned} S(x) &= 1 - H(x) \\ &= 1 - \left(1 - \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]_+^{-1/\xi} \right) \\ &= \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]_+^{-1/\xi}, \end{aligned}$$

as in Eq. (19) (page 36) of the lecture notes; we then solve (*) for $x = \hat{z}_t$ after replacing σ , ξ and λ_u with their maximum likelihood estimates. This part is not needed for the solution to this question.

For the case $\xi = 0$, we replace $S(x)$ with

$$\begin{aligned} S(x) &= 1 - H(x) \\ &= 1 - \left(1 - \exp \left\{ -\frac{x - u}{\sigma} \right\} \right) \\ &= \exp \left\{ -\frac{x - u}{\sigma} \right\} \quad \text{(1 mark)} \end{aligned}$$

as provided on page 31 of the lecture notes (since for the case $\xi = 0$ we assume the limit as $\xi \rightarrow 0$, which gives the exponential distribution with rate $1/\sigma$). Again, we solve (*) for $x = \hat{z}_t$ after replacing σ and λ_u with their maximum likelihood estimates, giving:

$$\begin{aligned} \hat{\lambda}_u \cdot \exp \left\{ -\frac{\hat{z}_t - u}{\hat{\sigma}} \right\} &= \frac{1}{t} \\ -\frac{(\hat{z}_t - u)}{\hat{\sigma}} &= \log \left(t \hat{\lambda}_u \right)^{-1} \\ \hat{z}_t - u &= \hat{\sigma} \log(t \hat{\lambda}_u) \\ \hat{z}_t &= u + \hat{\sigma} \log(t \hat{\lambda}_u). \quad \text{(2 marks)} \end{aligned}$$

Now, putting onto an annual scale to get the r -year return level requires a replacement of t with $r \times n_y$ (as we did in the lecture notes for the case $\xi \neq 0$), where n_y is the number of observation per year **(1 mark)**. This gives

$$\hat{z}_r = u + \hat{\sigma} \log(r n_y \hat{\lambda}_u),$$

as required.

For our data, we have $\hat{\sigma} = 8.286$. Also,

$$\hat{\lambda} = 7/21 = 1/3, \quad \text{(1 mark)}$$

and $n_y = 365.25$ as we have daily data (thus an average of 365.25 observations per year to account for leap years). This gives

$$\hat{z}_1 = 40 + 8.286 \log(1 \times 365.25/3) = 79.789 \mu\text{g per } 10^4 \text{m}^3. \quad \text{(2 marks)}$$

2. (10 marks in total)

Below are outline solutions for one site – Twickenham. Note that students have been allocated to one of 13 such sites, so answers will vary by student; students must also choose an appropriate threshold after a subjective assessment of a mean residual life plot, meaning that even students allocated to the same site might have slightly different results. Provided students' solutions look sensible, full marks can be obtained. For reference, indicative results for each site are provided in the table below.

Twickenham

A mean residual life plot (not shown here) suggests a threshold of around 30mm might be appropriate. Thus: (1 mark) for R output

```
1 > gpd.fit(Twickenham, 30)
2 $threshold
3 [1] 30
4
5 $nexc
6 [1] 176
7
8 $conv
9 [1] 0
10
11 $nllh
12 [1] 527.05
13
14 $mle
15 [1] 7.69831847 -0.04647912
16
17 $rate
18 [1] 0.006601155
19
20 $se
21 [1] 0.87720226 0.08552665
```

Giving $\hat{\sigma} = 7.70(0.877)$ (1 mark) and $\hat{\xi} = -0.05(0.086)$ (1 mark), with standard errors in parentheses.

We also have $\hat{\lambda}_u = 0.0066(0.0005)$ (1 mark), as we know from the notes that

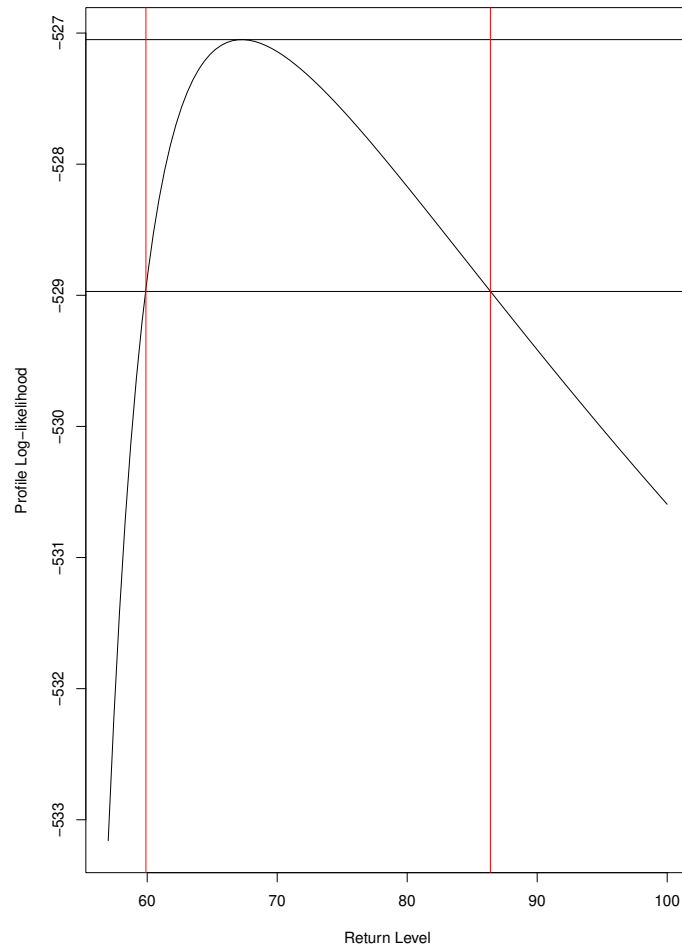
$$\text{s.e.}(\hat{\lambda}_u) = \sqrt{\hat{\lambda}_u(1 - \hat{\lambda}_u)/N} = \sqrt{0.0066(1 - 0.0066)/26662} = 0.0005.$$

For the 100-year return level, we use Eq. (21) in the lecture notes:

$$\begin{aligned}\hat{z}_{100} &= u + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(r n_y \hat{\lambda}_u \right)^{\hat{\xi}} - 1 \right] \\ &= 30 + \frac{7.70}{-0.05} \left[(100 \cdot 365.25 \cdot 0.0066)^{-0.05} - 1 \right] \\ &= 66.94\text{mm.} \quad (2 \text{ marks})\end{aligned}$$

For the profile log-likelihood plot, and after some experimentation/trial and improvement:

```
1 > gpd.prof(A, 100, xlow=50, xup=100, npy = 365.25, conf = 0.95, nint = 100)
2 > abline(v=86.4, col="red")
3 > abline(v=59.9, col="red")
```



Giving a 95% confidence interval of approximately (59.9, 86.4)mm. **(2 marks)** for plot, **(2 marks)** for interval.

Indicative results for other sites

	$\hat{\sigma}$	$\hat{\xi}$	$\hat{z}_{100}$
Twickenham	7.698	-0.046	67.274
Bishop Stortford	4.270	-0.090	35.079
Boston	3.860	0.009	39.260
Bradfield	6.335	-0.069	52.550
Cranwell	2.968	-0.006	29.046
Hebden Royd	4.887	0.022	49.834
Immingham	4.105	-0.016	39.534
Lancaster	15.882	-0.039	144.896
Leeming	3.242	0.039	34.862
Stratford	3.767	0.009	38.225
Lyme Regis	3.744	-0.024	33.929
Sheffield	7.358	-0.001	72.133
Stalybridge	6.075	0.039	61.750