NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2012/2013

MAS8304

Environmental Extremes: Mid–semester test

Time allowed: 50 minutes

Candidates should attempt all questions. Marks for each question are indicated.

There are THREE questions on this paper.

Answers to questions should be entered directly on this question paper in the spaces provided. This question paper must be handed in at the end of the test.

Name:..

Exponential distribution:

If $X \sim Exp(\lambda)$, then it has distribution function

$$F_X(x;\lambda) = 1 - e^{-\lambda x},$$

$x > 0, \lambda > 0$.

Generalised Extreme Value distribution:

If $X \sim GEV(\mu, \sigma, \xi)$ then it has distribution function

$$G_X(x;\mu,\sigma,\xi) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, & \xi \neq 0; \\ \exp\left\{-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right\}, & \xi = 0, \end{cases}$$

$-\infty < \mu < \infty, -\infty < \xi < \infty, \sigma > 0, a_+ = \max(0, a)$.

Generalised Pareto distribution:

If $X \sim GPD(\tilde{\sigma}, \xi)$ then it has distribution function

$$H_X(x;\tilde{\sigma},\xi) = \begin{cases} 1 - \left(1+\dfrac{\xi x}{\tilde{\sigma}}\right)_+^{-1/\xi}, & \xi \neq 0; \\ 1 - \exp\left(-\dfrac{x}{\tilde{\sigma}}\right) & \xi = 0, \end{cases}$$

$-\infty < \xi < \infty, \tilde{\sigma} > 0, a_+ = \max(0, a)$.

1. Suppose $X_1, \ldots, X_n$ is a sequence of independent Weibull random variables, that is

$$F(x) = \exp\left\{-(-x)^{\alpha}\right\}.$$

Suppose $\alpha = 1$. Show that the limit distribution of $(M_n - b_n)/a_n$, where

$$M_n = \max(X_1, \ldots, X_n),$$

is of extreme value type, and identify the distribution. Also clearly identify the values of the normalising constants a_n and b_n.

Answer:

$$\begin{aligned}
\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} &= \Pr\left\{M_n \leq a_n z + b_n\right\} \\
&= \Pr(X_1 \leq a_n z + b_n) \times \ldots \times \Pr(X_n \leq a_n z + b_n) \\
&= \left[\Pr(X \leq a_n z + b_n)\right]^n,
\end{aligned}$$

as the X_i are independent. Now, since $\alpha = 1$,

$$F_X(x) = e^x,$$

giving

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} = \left[e^{a_n z + b_n}\right]^n = e^{n(a_n z + b_n)}.$$

This looks like a Weibull random variable (i.e. type III extreme value distribution) with $\alpha = 1$ (see above); thus, we need

$$\begin{aligned}
n(a_n z + b_n) &= z \qquad \text{that is} \\
a_n z + b_n &= \frac{z}{n}.
\end{aligned}$$

This would hold if $b_n = 0$ and $a_n = \dfrac{1}{n}$.

[Total Q1: 10 marks]

2. Annual maximum flood discharges, in units of 1000 cubed feet per second, are available for two locations on the Fox River in Wisconsin, USA: Berlin (upstream) and Wrightstown (downstream). The data span a period of 33 years, from 1980 to 2012 (inclusive). Two analysts work separately on the data collected at Berlin and Wrightstown.

Lucy, the analyst working on the flood data at Berlin, assumes a GEV with $\xi = 0$, as is often the case for extreme river discharges at upstream locations. She finds the following maximum likelihood estimates of the remaining GEV location and scale parameters:

$$\hat{\mu} = 3.38 \quad \text{and} \quad \hat{\sigma} = 1.45,$$

with associated hessian of the corresponding log–likelihood function:

$$H = \begin{pmatrix} -11.771 & 4.176 \\ 4.176 & -19.864 \end{pmatrix}.$$

Duncan, the analyst working on the flood data for Wrightstown, assumes a GEV with $\xi \neq 0$, and uses the ismev package in R. After storing the data in `foxflood`, he obtains the following output:

```
A=gev.fit(foxflood)
$conv
[1] 0

$nllh
[1] 98.01564

$mle
[1] 6.0169817  5.1352751 -0.4485899

$se
[1] 1.0287297 0.8266614 0.1744666
```

(a) Why does the R output above support Duncan's decision to assume that $\xi \neq 0$ for Wrightstown? Comment on the tail behaviour of the flood discharges at Wrightstown, as suggested by Alan's results.

[**4 marks**]

Answer:

We have a negative estimate for the shape parameter ξ; in fact, the 95% confidence interval for ξ is

$$-0.4486 \pm 1.96 \times 0.1745 \longrightarrow (-0.791, -0.107)$$

which is wholly negative (does not pass through zero). This suggests a bounded upper tail for flood discharges at Wrightstown.

Question 2 continued on next page

(b) Use Lucy's results to obtain estimated standard errors for $\hat{\mu}$ and $\hat{\sigma}$ at Berlin.

[6 marks]

Answer:

We know that the variance–covariance matrix V can be found as

$$V = I_O^{-1}, \quad \text{where} \quad I_O = -H.$$

Thus,

$$\begin{aligned} V &= \begin{pmatrix} 11.771 & -4.176 \\ -4.176 & 19.864 \end{pmatrix}^{-1} \\ &= \frac{1}{11.771 \times 19.864 - 4.176^2} \begin{pmatrix} 19.864 & 4.176 \\ 4.176 & 11.771 \end{pmatrix} \\ &= \begin{pmatrix} 0.0918 & 0.0193 \\ 0.0193 & 0.0544 \end{pmatrix}. \end{aligned}$$

Thus, the standard errors are

$$\begin{aligned} \text{e.s.e.}(\hat{\mu}) &= \sqrt{0.0918} = 0.303 \\ \text{e.s.e.}(\hat{\sigma}) &= \sqrt{0.0544} = 0.233 \end{aligned}$$

(c) By constructing confidence intervals for the GEV location parameters in the usual way, comment on whether or not there is a significant difference between the annual maximum flood discharges at these two locations on the Fox river.

[6 marks]

Answer:

For Berlin, we have:

$$3.38 \pm 1.96 \times 0.303 \longrightarrow (2.786, 3.974) \times 1000 \text{ cubed feet per second.}$$

For Wrightstown, we have:

$$6.017 \pm 1.96 \times 1.029 \longrightarrow (4.000, 8.034) \times 1000 \text{ cubed feet per second.}$$

Yes, flood discharges *are* significantly different (at the 5% level of significance), as the two confidence intervals for μ do not overlap. But only just, as they nearly do! Discharges are significantly higher at Wrightstown than at Berlin.

Question 2 continued on next page

(d) For both locations, find the probability that, this year, the annual maximum flood discharge will exceed 5000 feet$^3/s$.

[6 marks]

Answer:

For Berlin, we have

$$\begin{aligned}\Pr(X > 5) &= 1 - G(5; \hat{\mu} = 3.38, \hat{\sigma} = 1.45)\\ &= 1 - \exp\left\{-\exp\left[-\left(\frac{5-3.38}{1.45}\right)\right]\right\}\\ &= 0.279.\end{aligned}$$

At Wrightstown, we have

$$\begin{aligned}\Pr(X > 5) &= 1 - G(5; \hat{\mu} = 6.017, \hat{\sigma} = 5.135, \hat{\xi} = -0.449)\\ &= 1 - \exp\left\{-\left[1 - 0.449\left(\frac{5-6.017}{5.135}\right)\right]_+^{1/0.449}\right\}\\ &= 0.701.\end{aligned}$$

(e) Over the next 100 years, in how many years can we expect the Fox river to have a flood discharge in excess of 5000feet$^3/s$ at *both* locations? *[Hint: Assume the river floods independently at Berlin and Wrightstown]*

[3 marks]

Answer:

$$\Pr(X > 5 \text{ at } \textit{both} \text{ locations}) = 0.279 \times 0.701 = 0.1956,$$

so we can this to happen in about $0.1956 \times 100 = 19.56 \approx 20$ of the next 100 years. Probably not appropriate to assume independence though!

[Total Q1: 25 marks]

3. You are working with a team of climatologists who are interested in temperature patterns across continental Europe. Daily minimum temperatures (°C) are recorded at Sarajevo, in Bosnia–Herzegovina, between 2006–2011 (inclusive). To work with the standard extreme value models for extremely *large* observations (as opposed to the extremely *small* observations as you have here), you decide to negate your data.

(a) In the context of the generalised extreme value distribution, what effect does working with the negated data have on estimates of the GEV parameters?

[2 marks]

Answer:

By negating our set of minima, we can apply the standard models to transformed data directly – the only affect this has is on the estimate of the GEV location parameter, which is changed by sign only.

(b) You decide to perform an analysis of threshold exceedances. With reference to Figure 1 overleaf, briefly explain why you might identify extremes as values which exceed a threshold of $u = -10$°C.

[1 mark]

Answer:

Above a point of about $u = -10$°C we observed (vague!) linearity in the mean residual life plot, suggesting the suitability of the generalised Pareto distribution for excesses above this level.

(c) You analyse the data in R. The negated temperatures are stored in the vector `neg.sarajevo`; you then use the ismev function `gpd.fit`, resulting in the (edited) output shown below. Look at it, and then answer the following questions.

```
B=gpd.fit(neg.sarajevo,-10)
$threshold
[1] -10

$nexc
[1] 233

$conv
[1] 0

$nllh
[1] 636.7434

$mle
[1]  7.4306027 -0.2728779

$se
[1] 0.59746049 0.04995889
```

Question 3 continued on next page

(i) Estimate the threshold exceedance rate λ.

[2 marks]

Answer:

We have six years of daily data, between 2006–2011. The only leap year is 2008, giving

$$5 \times 365 + 1 \times 366 = 2191 \text{ observations.}$$

Thus

$$\hat{\lambda} = 233/2191 \approx 0.106,$$

(ii) Given that $\mathsf{cov}(\hat{\sigma}, \hat{\xi}) = -0.025$, complete the variance–covariance matrix V for the parameter vector $\boldsymbol{\theta} = (\lambda, \sigma, \xi)^T$.

[5 marks]

Answer:

$$V = \begin{pmatrix} 0.0004^2 & & \\ 0 & 0.597^2 & \\ 0 & -0.025 & 0.050^2 \end{pmatrix}$$

This is because we know that

$$\mathsf{var}(\hat{\lambda}) = \hat{\lambda}(1 - \hat{\lambda})/N = 0.106 \times 0.894/2191 = 0.0004.$$

(d) With reference to Figure 1 overleaf, find the estimated 50–year return level *minimum* daily temperature for Sarajevo, with it's associated 95% profile log–likelihood confidence interval. *[Hint: $\chi^2_1(0.05) = 3.841$]*

[5 marks]

Answer:

The profile log–likelihood is maximised at about 14°C, giving

$$\hat{z}_{50} = -14°\text{C}.$$

Similarly, we have $(-11.55, -19.75)$ for the confidence interval.

[Total Q3: 15 marks]

THE END

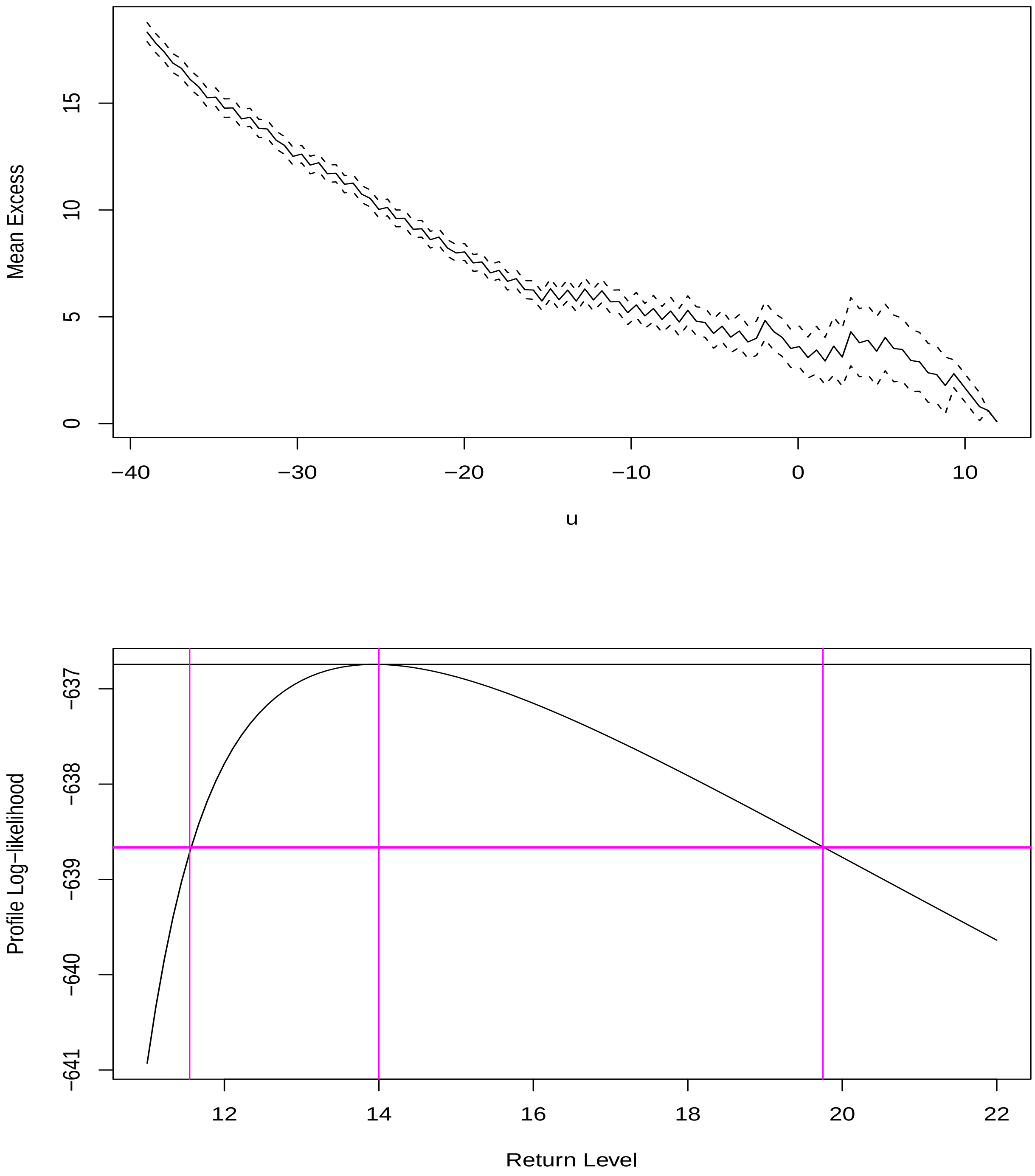

Figure 1: Mean residual life plot (top); profile log–likelihood curve for the 50–year return level.