NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2015/2016

MAS8306

Environmental Extremes: Mid–semester test

Time allowed: 50 minutes

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

There are TWO 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. (a) After each of the following statements, delete as appropriate. **(4 marks)**

The 50 year return level for daily temperatures in Seville, Spain, is 44.7°C. Thus, in the next 100 years we will see temperatures of at least 44.7°C on two days. **TRUE/**

In the limit, the shape parameters in the Generalised Extreme Value distribution and the Generalised Pareto distribution are equivalent. **/FALSE**

Confidence intervals for return levels using the profile log-likelihood usually have a higher upper bound than those obtained assuming Normality. **/FALSE**

A Fréchet distribution might be appropriate for data on sunshine hours as the limiting distribution has a finite upper bound. **TRUE/**

(b) Analyses of extreme wind speeds (miles per hour) at a location on the US east coast are summarised in the table below. Values in parentheses are standard errors.

	Estimated parameters			Return level estimates $\hat{z}_r$		
	Location	Scale	Shape	$r = 10$	$r = 50$	$r = 200$
Annual maxima: GEV	67.2	7.3	–1.2	73.1	72.5	85.3
	(13.3)	(2.2)	(0.2)	(10.0)	(11.3)	(13.5)
Threshold excesses: GPD	—	6.8	0.1	73.6	78.3	86.2
		(1.9)	(0.1)	(15.2)	(7.1)	(6.9)

Identify **four** things in this table which you think are suspicious, and say *why* you think they are suspicious. **(8 marks)**

Answer:

[**Total = 12 marks**]

*[This page has been left blank for your solution to the last question, **should you need it**]*

2. At midday every day since January 1st 1965 the sea level at a location in the Venetian Lagoon, X centimetres, has been recorded. These sea level recordings are complete to the end of last year (2015).

(a) Analysts at the *European Climate Data Consortium* (ECDC) extract the set of annual maximum sea levels from records at this location and store them in the vector `anmax` in R. Use the ECDC's (edited) output below to help you answer the following questions.

```
> library(ismev)
> gev.fit(anmax)$cov
$conv
[1] 0

$nllh
[1] 10.7145

$mle
[1] 111.09925486  17.17548761  -0.07673265

            [,1]        [,2]         [,3]
[1,]  6.90642105  1.16790535 -0.057843529
[2,]  1.16790535  3.25213339 -0.036209952
[3,] -0.05784353 -0.03620995  0.005405396
```

(i) The generalised extreme value (GEV) distribution has been fitted to the set of annual maxima. Circle the value of the maximised GEV likelihood. **(1 mark)**

10.7145 **–10.7145** **227.5612**

(ii) Would you be satisfied that the function used has obtained the maximum likelihood estimates of the GEV? Briefly explain. **(2 marks)**

Answer:

(iii) Obtain 95% confidence intervals for each of the estimated GEV parameters. **(6 marks)**

Answer:

Question 2 continued on next page

[Part of this page has been left blank for your solution to the previous question]

(iv) Estimate the 100 year return level for this location in the Venetian Lagoon. **(3 marks)**

Answer:

(v) Current research suggests catastrophic flooding in Venice if sea levels at this location exceed two metres. Use the fitted GEV to estimate the probability of catastrophic flooding in Venice in any given future year. **(3 marks)**

Answer:

Question 2 continued on next page

(b) Researchers at *Ca' Foscari* University, in Venice, adopt a different modelling approach to the ECDC. Their investigations suggest an exponential distribution for the daily sea level observations at this location, that is,

$$X|\zeta \sim Exp(\zeta).$$

They decide to take a threshold-based approach to modelling sea level extremes.

(i) Show that, above some high threshold u, the distribution of excess over u is Generalised Pareto (GP) with scale $1/\zeta$ and shape 0. **(7 marks)**

Answer:

Question 2 continued on next page

(ii) Construct the log-likelihood function for the GP scale parameter in (i) and obtain the maximum likelihood estimator for this parameter. Confirm that your estimator does indeed maximise the log-likelihood. **(6 marks)**

Answer:

Question 2 continued on next page

(iii) The r-year return level for this model is given by

$$z_r = u + \frac{\log(r n_y \lambda_u)}{\zeta},$$

where $\lambda_u = \Pr(X > u)$ and n_y is the average number of observations per year. 765 sea level observations exceed a chosen threshold of $u = 145$cm, giving

$$\sum_{i=1}^{765}(x_i - u) = 3912.2 \text{ cm}.$$

Estimate the 100 year return level using the threshold-based approach for modelling extremes. **(5 marks)**
[Hint: 12 years are leap years over the data collection period 1965–2015]

Answer:

(iv) Complete the following variance-covariance matrix for $(\hat{\lambda}_u, \hat{\zeta})^\mathsf{T}$. You may use the space below to show any working, should you need to. **(2 marks)**

Answer:

$$V = \begin{pmatrix} \text{---} \quad \text{---} \times 10^{-6} & \text{----} \;\; \text{----} \\ & \\ 0 & 5.2307 \end{pmatrix}.$$

(c) Describe two potential problems with the threshold-based approach used throughout part (b). Why might these problems not be an issue in the analysis in part (a)? **(3 marks)**

Answer:

[Total = 38 marks]

THE END