NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2015/2016

MAS8306: Solutions (Q1–Q4)

Topics in Statistics Specimen Paper

Time allowed: 2 hours

Candidates should attempt all questions. Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are SIX questions on this paper.

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

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

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

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) If there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$,

$$\Pr\left\{(M_n - b_n)/a_n \le x\right\} \to G(x)$$

for some non–degenerate distribution G, then G is of the same type as one of the following distributions:

$$I : G(x) = \exp\{-\exp(-x)\} \qquad -\infty < x < \infty;$$

$$II : G(x) = \begin{cases} 0 & x \le 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases}$$

$$III : G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0, \alpha > 0 \\ 1 & x \ge 0. \end{cases}$$

[**4 marks**]

(b) We know that

$$\Pr\{M_n \le x\} = F^n(x) \approx G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

for some parameters μ, σ and ξ. Hence

$$n\log F(x) \approx -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi} \qquad (\ddagger)$$

Substitution of the result given, into (‡), gives

$$1 - F(u) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}$$

for large u. Similarly, for $y > 0$,

$$1 - F(u+y) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}.$$

Then

$$\begin{aligned}
\Pr\{X > u + y | X > u\} &= \frac{1 - F(u+y)}{1 - F(u)} \\
&\approx \frac{\frac{1}{n}\left[1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} \\
&= \left[\frac{1 + \xi\left(\frac{u-\mu}{\sigma}\right) + \xi\frac{y}{\sigma}}{1 + \xi\left(\frac{u-\mu}{\sigma}\right)}\right]_+^{-1/\xi} \\
&= \left[1 + \frac{\xi y}{\tilde{\sigma}}\right]_+^{-1/\xi}, \qquad \text{as required,}
\end{aligned}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

[**8 marks**]

Question 1 continued on next page

(c) (i) The r year return level, z_r, is the value we can expect to be exceeded once, on average, every r years.

[2 marks]

(ii)

$$\Pr(X > u + y | X > u) \approx \left[1 + \frac{\hat{\xi} y}{\hat{\sigma}}\right]_{+}^{-1/\hat{\xi}},$$

Now

$$\Pr(X > u + y | X > u) = \frac{\Pr(X > u + y)}{\Pr(X > u)},$$

giving

$$\Pr(X > u + y) = \lambda_u \Pr(X > u + y | X > u),$$

where $\lambda_u = \Pr(X > u)$. Thus,

$$\begin{aligned} \Pr(X > u + y) &\approx \lambda_u \left[1 + \frac{\xi y}{\sigma}\right]_{+}^{-1/\xi} \qquad \text{i.e.} \\ \Pr(X > x) &\approx \lambda_u \left[1 + \xi\left(\frac{x - u}{\sigma}\right)\right]_{+}^{-1/\xi}. \end{aligned}$$

Equating the RHS of the above expression to $1/t$ and solving for $x = z_t$ gives the t–<u>observation</u> return level:

$$z_t \approx u + \frac{\sigma}{\xi}\left[(t\lambda_u)^{\xi} - 1\right];$$

On an annual scale, we now replace t with $r \times n_y$ to get the r–<u>year</u> return level z_r:

$$z_r = u + \frac{\sigma}{\xi}\left[(r n_y \lambda_u)^{\xi} - 1\right],$$

as required.

[6 marks]

[Total Q1: 20 marks]

2. (a) Model `M2` allows for a linear trend in the location parameter, that is

$$\mu(t) = \beta_0 + \beta_1 t,$$

where $t = 1, 2, \ldots, 51$. In fact, the slope term β_1 corresponds to the annual rate of change in annual maximum sea surges at Grand Isle. It is positive, suggesting increasing levels of sea surge extremes over time.

Model `M1` assumes the location parameter is constant.

[3 marks]

(b) We have

$$D = 2\{-41.74466 - (-48.40894)\} = 13.32856,$$

which is large relative to $\chi_1^2(0.05) = 3.841$. Thus, the linear trend component explains a substantial amount of the variation in the data, and is likely to be a genuine effect in the sea surge process.

[2 marks]

(c) (i) The 95% confidence interval is $-0.028 \pm 1.96 \times 0.083 = (-0.191, 0.135)$ which suggests a Gumbel–type tail (unbounded).

[2 marks]

(ii) For the GEV, we have

$$z_r = \mu + \frac{\sigma}{\xi}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\xi} - 1\right].$$

Here, we have $t = 59$ for the year 2020, giving:

$$\begin{aligned}
\hat{\mu} &= 3.182 + 0.0185 \times 59 = 4.2735, \\
\hat{\sigma} &= 0.4780, \\
\hat{\xi} &= -0.0278.
\end{aligned}$$

Substitution in the return level equation (with $r = 100$)

$$\hat{z}_{100} = 4.2735 - \frac{0.4780}{0.0278}\left[(-\log(1 - 1/100))^{0.0278} - 1\right] = 6.34 \text{ feet}.$$

[2 marks]

[Total Q2: 9 marks]

3. (a) Now the variance–covariance matrix for $(\hat{\sigma}, \hat{\xi})$ is given by

$$\begin{aligned}(-H)^{-1} &= \frac{1}{38.21 \times 2027.11 - 256.7^2}\begin{pmatrix} 2027.11 & -256.7 \\ -256.7 & 38.21 \end{pmatrix} \\ &= \begin{pmatrix} 0.175 & -0.022 \\ -0.022 & 0.003 \end{pmatrix}.\end{aligned}$$

Also,

$$\mathsf{var}(\hat{\lambda}_u) = \hat{\lambda}_u(1-\hat{\lambda}_u)/N = \frac{235}{30581}\left(1 - \frac{235}{30581}\right)/30581 = 2.49 \times 10^{-7},$$

giving

$$V = \begin{pmatrix} 2.49 \times 10^{-7} & 0 & 0 \\ 0 & 0.175 & -0.022 \\ 0 & -0.022 & 0.003 \end{pmatrix}.$$

[5 marks]

(b) We have

$$\hat{z}_{50} = 30 - \frac{5.10}{0.44}\left[(50 \times 365.25 \times 0.00768)^{-0.44} - 1\right] = 40.27^\circ\text{C}.$$

[2 marks]

(c) From the plot, using a cut-off of $0.5 \times \chi_1^2(0.05) = 1.921$, gives

$$(39.75^\circ\text{C}, 41.4^\circ\text{C}).$$

[2 marks]

(d) Problems:

- How do we choose κ?
- κ too large $\longrightarrow$ too few cluster peaks
- κ too small $\longrightarrow$ independence assumption violated
- Sensitivity of estimates of σ and ξ (and possible return levels) on the choice of κ (e.g. Fawcett and Walshaw (2012)!)

[2 marks]

[Total Q3: 11 marks]

4. (a) Example: Often, it is the combination of extremes of many environmental variables that cause destruction/devastation. For example, during a hurricane, it might be the combination of (i) extremely strong winds, and (ii) extremely heavy rain, that results in sea–surge and mass flooding.

[2 marks]

(b) (i) The assumptions are that X_i and Y_i, $i = 1, 2, \ldots, n$ are unit Fréchet distributed. In practice, this is achieved by marginally fitting the GPD to excesses of u_x for X and excesses of u_y for Y, and then transforming to unit Fréchet using the fitted GPD parameters $(\hat{\sigma}_x, \hat{\xi}_x)$ and $(\hat{\sigma}_y, \hat{\xi}_y)$.

[2 marks]

(ii) The parameter α controls the degree of extremal dependence between X and Y. For independent extremes, $\alpha = 1$; as $\alpha \to 1$ we have increasing levels of extremal dependence.

[2 marks]

(iii) If $x_i < u_x$ and $y_i > u_y$, then we have

$$\left.\frac{\partial H}{\partial y}\right|_{(u_x, y_i)} = y_i^{-(1/\alpha+1)} \left(u_x^{-1/\alpha} + y_i^{-1/\alpha}\right)^{\alpha-1} \exp\left\{-\left(u_x^{-1/\alpha} + y_i^{-1/\alpha}\right)^{\alpha}\right\}$$

[4 marks]

[Total Q4: 10 marks]

THE END