Chapter 2

Background Theory & Methods I – Independent Sequences

2.0 Introduction

A precept of extreme-value theory is to employ as statistical models distributions that arise as limiting distributions for random variables associated with stochastic processes. These limits correspond to levels that are asymptotically extreme in some sense and, as such, justify their use as models for extreme values. The asymptotic distributions of two constructions are reviewed in this chapter for sequences of independent and identically distributed random variables: the maximum of a block of consecutive variables and the point process of consecutive variables. Univariate sequences are covered in Section 2.1; multivariate sequences are covered in Section 2.2. Statistical applications of the results are also demonstrated.

More complete introductions to extreme-value theory and methods are available in Leadbetter *et al.* (1983), Resnick (1987), Embrechts *et al.* (1997), Coles (2001) and Beirlant *et al.* (2003).

2.1 Univariate Sequences

Let $\{X_i\}_{i\geq 1}$ be a sequence of independent and identically distributed random variables with marginal distribution function F, that is each X_i is distributed according to F. Denote the maximum of n consecutive elements of the sequence by

$$M_n = \max\{X_1, \ldots, X_n\}.$$

As n increases, M_n approaches the upper end-point, $\omega = \sup\{x : F(x) < 1\}$, of F and the limiting distribution of M_n is a point mass at ω. A normalisation is required to obtain a non-degenerate limit. As in the central limit theorem, a linear normalisation is traditional and the limit, as n approaches infinity, is sought for the distribution function

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n)$$

for sequences of constants $a_n > 0$ and b_n. The following theorem, due to Fisher and Tippett (1928), characterises all of the possible limit distributions. Denote weak convergence by $\xrightarrow{w}$.

Theorem 2.1. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad \text{as } n \to \infty$$

for a non-degenerate distribution function G, then G is a generalised extreme-value (GEV) distribution function,

$$G(x) = \exp\left[-\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-1/\xi}\right], \tag{2.1}$$

defined on $\{x : 1 + \xi(x-\mu)/\sigma > 0\}$ for parameters $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. For $\xi = 0$, G is defined by the limit as $\xi \to 0$.

The generalised extreme-value distribution function (2.1) comprises three sub-classes:

$$\begin{array}{lll} \text{I} & \exp\left[-\exp\left\{-\left(\frac{x-\beta}{\alpha}\right)\right\}\right] & \text{for } x \in \mathbb{R}, \\ \text{II} & \exp\left\{-\left(\frac{x-\beta}{\alpha}\right)^{-\gamma}\right\} & \text{for } x > \beta, \\ \text{III} & \exp\left[-\left\{-\left(\frac{x-\beta}{\alpha}\right)\right\}^{\gamma}\right] & \text{for } x < \beta, \end{array}$$

where $\alpha > 0$ and $\gamma > 0$. These sub-classes correspond to the generalised extreme-value distribution with $\xi = 0$, $\xi > 0$ and $\xi < 0$, and are known by the names I – Gumbel, II – Fréchet and III – Weibull. 'Standard' versions of these distributions refer to the special case $\alpha = 1$, $\beta = 0$ and $\gamma = 1$.

The approximation $P(M_n \leq x) \approx G((x - b_n)/a_n)$ for large n motivates the generalised extreme-value distribution (2.1) as a model for maxima of blocks with large but finite lengths since the normalising constants can be assimilated into the location and scale parameters μ and σ. For example, the model could be fitted to a sample of annual maximum temperatures by maximising the likelihood. Although the extremes of non-stationary processes will not be treated in this thesis, any non-stationarity evident in the data can be accounted for by incorporating covariate information into the models. For example, μ could be replaced by $\alpha+\beta t$ to allow a linear trend in time t.

Block maxima contain only partial information about the extremes of a process; more detail is provided by other large observations, for which point processes yield useful limiting models. Let $\{X_i : i \in \mathcal{I}\}$ represent the locations of points, indexed by a set $\mathcal{I}$, occurring randomly in a space S. A point process Q counts the number of points in regions of S:

$$Q(A) = \sum_{i \in \mathcal{I}} \delta_{X_i}(A) \qquad A \subseteq S,$$

where $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \notin A$. The expected number of points in a set A is given by the intensity measure $\Lambda(A)$.

For $\{X_i\}_{i\geq 1}$ a sequence of independent and identically distributed random variables, define

$$Q_n(\cdot) = \sum_{i=1}^{n} \delta_{(i/n\,,\,(X_i-b_n)/a_n)}(\cdot).$$

This two-dimensional point process counts points on $(0,1] \times \mathbb{R}$. The following theorem, due to Pickands (1971), characterises the limit of Q_n, where the linear normalisation is used again to avoid degeneracy. Denote convergence in distribution by $\xrightarrow{d}$.

Theorem 2.2. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad \textit{as } n \to \infty$$

for a non-degenerate distribution function G with lower and upper end-points $\omega_0 = \inf\{x : G(x) > 0\}$ and $\omega_1 = \sup\{x : G(x) < 1\}$, then $Q_n \xrightarrow{d} Q$, where Q is a non-homogeneous Poisson process on $(0,1] \times (\omega_0, \omega_1)$ with intensity measure

$$\Lambda(A) = -(b-a)\log G(x) \tag{2.2}$$

on $A = (a,b) \times [x, \omega_1)$ when $0 \leq a < b \leq 1$ and $\omega_0 < x < \omega_1$.

The theorem motivates the Poisson process Q as a model for those points $(i/n, X_i)$ for which X_i exceeds a high threshold u. The model can be fitted by maximising the likelihood

$$L(\mu, \sigma, \xi) = \exp\{-\Lambda(A)\} \prod_{i:X_i>u} \lambda(i/n, X_i),$$

where $A = (0,1] \times [u, \omega_1)$ and

$$\lambda(t,x) = \frac{1}{\sigma}\left\{1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-1/\xi-1}$$

is the intensity associated with Λ at a point (t,x) in A.

Another useful model for threshold exceedances can be derived from Theorem 2.2

as follows.

$$
\begin{aligned}
P\left(\frac{X_1-b_n}{a_n}>u+x \,\middle|\, \frac{X_1-b_n}{a_n}>u\right) &\to \frac{\Lambda\{(0,1]\times(u+x,\omega_1)\}}{\Lambda\{(0,1]\times(u,\omega_1)\}}\\
&= \frac{\log G(u+x)}{\log G(u)}\\
&= \left(1+\frac{\xi x}{\sigma_u}\right)^{-1/\xi},
\end{aligned}
$$

where $\sigma_u = \sigma + \xi(u-\mu)$. This limit is called the generalised Pareto (GP) distribution (Pickands, 1975) and provides a model for the distribution of excesses over high thresholds:

$$
P(X_1 - u > x \mid X_1 > u) \approx \left(1+\frac{\xi x}{\sigma_u}\right)^{-1/\xi} \tag{2.3}
$$

for large u on $\{x : x \geq 0 \text{ and } 1+\xi x/\sigma_u > 0\}$.

2.2 Multivariate Sequences

Let $\{X_i\}_{i\geq 1}$ be a sequence of independent and identically distributed vector random variables $X_i = (X_{i1}, \ldots, X_{iD})$ with dimension D and marginal distribution function F. Many definitions are possible for the maximum of n consecutive elements of a multivariate sequence. The classical line in multivariate extremes (Sibuya, 1960) corresponds to the marginal ordering described by Barnett (1976). Another, conditional approach has been proposed recently by Heffernan and Tawn (2003). This thesis follows the classical line for which the maximum of n consecutive elements is defined to be the componentwise maximum,

$$M_n = (M_{n1}, \dots, M_{nD}),$$

where $M_{nd} = \max\{X_{1d}, \dots, X_{nd}\}$ for each $1 \le d \le D$. Non-degenerate limits are obtained for the distribution function

$$P\left(\frac{M_n - b_n}{a_n} \le x\right) = F^n(a_n x + b_n),$$

where $x = (x_1, \dots, x_D)$, $a_n = (a_{n1}, \dots, a_{nD}) > 0$ and $b_n = (b_{n1}, \dots, b_{nD})$. All operations are performed componentwise, so $a_n > 0$ is translated as $a_{n1} > 0, \dots, a_{nD} > 0$. The following theorem, due to de Haan and Resnick (1977), characterises all of the possible limit distributions.

Theorem 2.3. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{M_n - b_n}{a_n} \le x\right) \xrightarrow{w} G(x) \quad \text{as } n \to \infty \tag{2.4}$$

for a non-degenerate distribution function G, then each of the D one-dimensional component distributions of G is a generalised extreme-value distribution function,

$$G_d(x_d) = \exp\left[-\left\{1 + \xi_d\left(\frac{x_d - \mu_d}{\sigma_d}\right)\right\}^{-1/\xi_d}\right] \quad \text{for } 1 \le d \le D, \tag{2.5}$$

and $G(x) = G_*\{\zeta^{-1}(x)\}$, *where*

$$\zeta_d(x_d) = \mu_d + \frac{\sigma_d}{\xi_d}\left(x_d^{\xi_d} - 1\right) \quad \textit{for } 1 \leq d \leq D \tag{2.6}$$

and G_* *is a multivariate extreme-value distribution function with standard Fréchet components, that is*

$$G_*(x) = \exp\{-V(x)\} \tag{2.7}$$

for $x \in \mathbb{R}_+^D$, *where*

$$V(x) = \int_{S_D} \max_{1 \leq d \leq D} \left\{\frac{z_d}{x_d}\right\} \mathrm{d}H(z)$$

and H *is a non-negative, finite measure on the simplex*

$$S_D = \{z = (z_1, \ldots, z_D) \in \mathbb{R}_+^D \,:\, z_1 + \ldots + z_D = 1\}$$

for which

$$\int_{S_D} z_d \, \mathrm{d}H(z) = 1 \quad \textit{for } 1 \leq d \leq D. \tag{2.8}$$

Unlike the univariate case, the limiting distribution (2.4) is not determined by a finite number of parameters; rather, it is characterised by the $3D$ parameters of the component distributions (2.5) and the infinite-dimensional measure H. For the remainder of the thesis, it is assumed that the derivative, h, of H exists on the interior of S_D and on the interiors of all of the lower-dimensional boundaries, in which case the density of the multivariate extreme-value distribution (2.4) exists everywhere.

Estimation of multivariate extreme-value distributions, in particular when $D = 2$, will be discussed in Chapter 5. One option is to choose a parametric model for the measure H, or equivalently V, such that the conditions (2.8) are satisfied, and then fit the model to a sample of componentwise maxima by maximising the likelihood. An example is the logistic model (Gumbel, 1960), which in the bivariate case is

$$V(x_1, x_2) = \left(x_1^{-1/\alpha} + x_2^{-1/\alpha}\right)^{\alpha} \tag{2.9}$$

for dependence parameter $\alpha \in (0, 1]$. Non-parametric estimators also exist; a semi-parametric estimator is introduced in Chapter 5.

It is convenient to describe multivariate extreme-value distributions by separating the dependence structure (2.7) from the one-dimensional components (2.5). Assume, therefore, that the component distributions of F are standard Fréchet, that is $F_d(x_d) = \exp(-1/x_d)$ when $x_d > 0$ for $1 \leq d \leq D$. It follows that G_d is also standard Fréchet for $1 \leq d \leq D$; results for generalised extreme-value component distributions can be obtained by applying the change of variables (2.6). It is helpful also to introduce another coordinate system. For $x = (x_1, \ldots, x_D)$, define its radial component to be $r = x_1 + \ldots + x_D > 0$ and its angular component to be $w = (w_1, \ldots, w_D)$, where $w_d = x_d/r \in (0, 1)$.

The following theorem, due to de Haan (1985), characterises the limit of the point process

$$Q_n(\cdot) = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}(\cdot).$$

Theorem 2.4. *If there exist sequences of constants $a_n > 0$ and b_n such that*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad \textit{as } n \to \infty \tag{2.10}$$

for a non-degenerate distribution function G with form (2.7), then $Q_n \xrightarrow{d} Q$, where Q is a non-homogeneous Poisson process on $\mathbb{R}_+^D \setminus \{0\}$ with intensity

$$\lambda(r, w) = r^{-2} h(w).$$

If the component distributions F_d are standard Fréchet then the normalising constants in the limit (2.10) can be set equal to $a_n = n$ and $b_n = 0$. In this case, the theorem provides a Poisson process model for points X_i/n on regions such as

$$A = \mathbb{R}_+^D \setminus \{[0, c_1) \times \ldots \times [0, c_D)\}$$

that are bounded away from the origin. If a parametric model such as the logistic (2.9) is chosen for the dependence structure then the Poisson process can be fitted by maximising the likelihood

$$\exp\{-\Lambda(A)\} \prod_{i: x_i/n \in A} \lambda(r_i, w_i),$$

where Λ is the intensity measure, $r_i = (x_{i1} + \ldots + x_{iD})/n$ and $w_{id} = x_{id}/(nr_i)$.

The Poisson process can also be fitted to points X_i on regions

$$A = \mathbb{R}_+^D \setminus \{[0, u_1) \times \ldots \times [0, u_D)\}, \tag{2.11}$$

where the u_d are high thresholds substituting for nc_d. Since $n^2\lambda(nr, w) = \lambda(r, w)$, the corresponding likelihood is proportional to

$$\exp\{-n\Lambda(A)\} \prod_{i:x_i \in A} \lambda(r_i, w_i),$$

where $r_i = x_{i1} + \ldots + x_{iD}$ and $w_{id} = x_{id}/r_i$.

This assumes that the component distributions F_d are standard Fréchet. When the F_d are unknown, the transformation $\tilde{x}_{id} = -1/\log F_d(x_{id})$ can be incorporated and the component distributions estimated simultaneously with the dependence structure. A model for each F_d is available if the generalised Pareto distribution (2.3) holds, that is

$$1 - F_d(x_d) = \{1 - F_d(u_d)\} \left\{1 + \xi_d \left(\frac{x_d - u_d}{\sigma_d}\right)\right\}^{-1/\xi_d}$$

when $x_d > u_d$. For $x_d \leq u_d$, F_d can be estimated by the empirical distribution function. Defining $\tilde{A} = \mathbb{R}_+^D \setminus \{[0, \tilde{u}_1) \times \cdots \times [0, \tilde{u}_D)\}$ and $\tilde{u}_d = -1/\log F_d(u_d)$, the likelihood becomes

$$\exp\{-n\Lambda(\tilde{A})\} \prod_{i:x_i \in A} \left\{\lambda(\tilde{r}_i, \tilde{w}_i) \prod_{d:x_{id} > u_d} \frac{\tilde{x}_{id}^2 (1 - e^{-1/\tilde{x}_{id}})^{1+\xi_d} e^{1/\tilde{x}_{id}}}{\sigma_d \{1 - F_d(u_d)\}^{\xi_d}}\right\}$$

since

$$\frac{\mathrm{d}\tilde{x}}{\mathrm{d}x} = \frac{\tilde{x}^2(1 - e^{-1/\tilde{x}})^{1+\xi} e^{1/\tilde{x}}}{\sigma\{1 - F(u)\}^{\xi}}.$$

The choice (2.11) of extreme region, A, is important here. For other choices the set of points that fall in the region can change with the parameter estimates of the component distributions. For further details on this model see Coles and Tawn (1991, 1994) and Joe *et al.* (1992).