EXTREME VALUE THEORY FOR CERTAIN NON-STATIONARY SEQUENCES

by

G.A. Feeney
Australian Bureau of Statistics
Belconnen, Australia

and

P.K. Sen
Department of Biostatistics
University of North Carolina, Chapel Hill, NC 27514

Institute of Statistics Mimeo Series No. 1421

December 1982

Extreme Value Theory for Certain Non-stationary Sequences

G.A. Feeney
Australian Bureau of Statistics, Belconnen, Australia

P.K. Sen*
University of North Carolina, Chapel Hill, NC, USA

Summary. Under a Markovian structure on a sequence of random variables which can be partitioned into $m(\geq 1)$ jointly dependent subsequences (where within each subsequence the random variables have a common marginal distribution which may vary between the subsequences), the asymptotic distribution theory of the sample extreme values is developed. The asymptotic independence of the subsequence extreme values is also studied.

*Work supported by the National Heart, Lung and Blood Institute, Contract NIH-NHLBI-71-2243L from the National Institutes of Health.

1. Introduction.

Classical extreme value theory, as developed by Gnedenko (1943), is based on the assumption that the sequence of observations $\{X_i, i \geq 1\}$ is independent and identically distributed (i.i.d.). Watson (1954) was able to show that m-dependent stationary sequences have the same limiting extreme value distribution as their associated i.i.d. sequence. A similar case holds for other stationary sequences too [See Leadbetter, Lindgren and Rootzen (1979)]. We will consider the case of some dependent sequence having a certain non-stationary structure. In particular, we will consider the sequence $\{X_n\}$ as being partitioned into $m(\geq 1)$ jointly dependent subsequences where within each subsequence the X_i have a common marginal distribution function, and we allow these to vary between the subsequences. It should be noted that the construction of the subsequence is quite flexible, and, for some relevance to some practical problems, we may refer to the last section.

Let $X_j^{(\ell)}$ denote the j-th observation in the ℓ-th subsequence, for $j = 1,\dots,n_\ell$ and $\ell = 1,\dots,m$. Since the total sample size is n, the sample size for the ℓ-th subsequence, n_ℓ is approximately equal to $n_0 = [n/m]$ and $\Sigma_{\ell=1}^{m} n_\ell = n$. Also, let

$$Z_{n_\ell}^{(\ell)} = \max\{X_j^{(\ell)};\ 1 \leq j \leq n_\ell\}\ ,\quad 1 \leq \ell \leq m. \tag{1.1}$$

We refer to $Z_{n_\ell}^{(\ell)}$ as the ℓ-th subsequence maximum, $1 \leq \ell \leq m$, and we denote the overall maximum by

$$Z_n = \max\{Z_{n_\ell}^{(\ell)};\ 1 \leq \ell \leq m\} = \max\{X_j^{(\ell)};\ 1 \leq j \leq n_\ell,\ 1 \leq \ell \leq m\} \quad . \tag{1.2}$$

[Actually, the sequence $\{X_i\}$ is reindexed to $\{X_j^{(\ell)}\}$ by letting $i \to (j-1)m+\ell$, $1 \leq \ell \leq m$, $j \geq 1$.] The main purpose of this investigation is to study suitable regularity conditions under which the $Z_{n_\ell}^{(\ell)}$ are asymptotically jointly independent. Prior research on multivariate extreme values of observations were independent. In our case, we can see that since the successive observations are not independent, on letting $\underset{\sim}{X}_j = (X_j^{(1)},\ldots,X_j^{(m)})$, $j\geq 1$, $\underset{\sim}{X}_j$ and $\underset{\sim}{X}_{j+1}$ (or $\underset{\sim}{X}_{j-1}$) are not independent vectors, even when the X_i form an m-dependent sequence. However, under a Markovian structure on the vectors $\{\underset{\sim}{X}_j,\ j\geq 1\}$, we will show that under conditions similar to those of Sibuya (1960), the $Z_{n_\ell}^{(\ell)}$ are asymptotically jointly independent.

2. The Extreme Value Distribution of the Maximum of the Subsequence Maxima

Under the usual conditions, given in Gnedenko (1943), de Haan (1970) and Galambos (1978), it is clear that, whenever it exists

$$H_\ell(x) = \lim_{n_\ell\to\infty} \Pr\{Z_{n_\ell}^{(\ell)} \leq a_{n_\ell}^{(\ell)} x + b_{n_\ell}^{(\ell)}\} \quad , \tag{2.1}$$

where H_ℓ is a non-degenerate distribution function for $\ell = 1,\ldots,m$. The form of H_ℓ is determined by the type of parent distribution, F_ℓ, where $F_\ell(x) = \Pr\{X_j^{(\ell)} \leq x\}$. However, H_ℓ must be one of the three possible types of extreme value distributions, see Galambos (1978). In addition $a_{n_\ell}^{(\ell)}$, positive, and $b_{n_\ell}^{(\ell)}$ are normalizing constants for $\ell = 1,\ldots,m$. Therefore

$$\Pr\{Z_n \leq a_n x + b_n\} = \Pr\{Z_n^{(\ell)} \leq a_n x + b_n;\ \ell = 1,\ldots,m\} \quad , \tag{2.2}$$

where

$$Z_n = \max\{X_i;\ i = 1,\ldots,n\} \quad ,$$

and a_n, positive, and b_n are real constants. We can then re-express (2.2) as

$$\Pr\{Z_n \le a_n x + b_n\} = \Pr\left\{ \frac{Z_{n_\ell}^{(\ell)} - b_{n_\ell}^{(\ell)}}{a_{n_\ell}^{(\ell)}} \le \left(\frac{a_n}{a_{n_\ell}^{(\ell)}}\right) x + \frac{b_n - b_{n_\ell}^{(\ell)}}{a_{n_\ell}^{(\ell)}};\quad \ell = 1,\ldots,m \right\} . \quad (2.3)$$

If it could be shown that the $Z_{n_\ell}^{(\ell)}$ are asymptotically jointly independent then (2.3) becomes, in the limit;

$$\lim_{n\to\infty} \Pr\{Z_n \le a_n x + b_n\} = \prod_{\ell=1}^{m} \lim_{n\to\infty} \Pr\left\{ \frac{Z_{n_\ell}^{(\ell)} - b_{n_\ell}^{(\ell)}}{a_{n_\ell}^{(\ell)}} \le \left(\frac{a_n}{a_{n_\ell}^{(\ell)}}\right) x + \frac{b_n - b_{n_\ell}^{(\ell)}}{a_{n_\ell}^{(\ell)}} \right\} , \quad (2.4)$$

since n_ℓ; $\ell = 1,\ldots,m$; tends to infinity if and only if n tends to infinity. Therefore by equation (2.1) and of Gnedenko (1943)

$$\lim_{n\to\infty} \Pr\{Z_n \le a_n x + b_n\} = \prod_{\ell=1}^{m} H_\ell(A_\ell x + B_\ell) , \quad (2.5)$$

where a_n, positive, and b_n are such that

$$A_\ell = \lim_{n_\ell\to\infty} \frac{a_n}{a_{n_\ell}^{(\ell)}} \quad (2.6)$$

and

$$B_\ell = \lim_{n_\ell\to\infty} \frac{b_n - b_{n_\ell}^{(\ell)}}{a_{n_\ell}^{(\ell)}} , \quad (2.7)$$

for $\ell = 1,\ldots,m$.

The choice of the form of a_n and b_n, like $a_{n_\ell}^{(\ell)}$ and $b_{n_\ell}^{(\ell)}$, will depend on the type of marginal distribution functions, F_ℓ, $\ell = 1,\ldots,m$. It can be shown that the only sensible choice for a_n and b_n are such that $A_\ell x + B_\ell$ equals infinity for all but the dominant sub-sequence(s), for which it equals x. Denote the corresponding normalizing constants by $a_{n_t}^{(t)}$ and $b_{n_t}^{(t)}$. Then we can see that

$a_n = a^{(t)}_{n_t}$, $b_n = b^{(t)}_{n_t}$, and

$$\lim_{n\to\infty} \Pr\{Z_n \le a^{(t)}_{n_t} x + b^{(t)}_{n_t}\} = H_t(x) \quad , \tag{2.8}$$

where τ ($\le m$) is the multiplicity of t; however with probability one, for continuous distributions $F_1,\ldots,F_m$ not agreeing in the tails, $\tau = 1$. From (2.8) we see that the extreme value distribution for the maximum of the entire sequence reduces to the product of the extreme value distributions for the maximum of the dominant sub-sequences, if we can assume the asymptotic joint independence of the sub-sequence maxima.

Some preliminary results need to be established before determining the conditions under which asymptotic joint independence holds.

Let $G(\underset{\sim}{x}) = \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{x}\}$, $x \in E^m$, $G_{[\ell]}(x) = \Pr\{X_j^{(\ell)} \le x\}$, $1 \le \ell \le m$ and $G_{[\ell\ell']}(x,y) = \Pr\{X_j^{(\ell)} \le x,\ X_j^{(\ell')} \le y\}$, $1 \le \ell \ne \ell' \le m$. Note that the $G_{[\ell]}$ are not assumed to be identical, so are the $G_{[\ell\ell']}$. Further let $G^*(\underset{\sim}{x},\underset{\sim}{y}) = \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{x},\ \underset{\sim}{X}_{j+1} \le \underset{\sim}{y}\}$ and let $G^*_{[\ell\ell']}(x,y) = \Pr\{X_j^{(\ell)} \le x,\ X_{j+1}^{(\ell')} \le y\}$. Also, let

$$\Omega(G) = G(\underset{\sim}{x}) / \prod_{\ell=1}^{m} G_{[\ell]}(x_\ell) \tag{2.9}$$

$$\Omega^*(G^*) = G^*(\underset{\sim}{x},\underset{\sim}{y}) / [G(\underset{\sim}{x})G(\underset{\sim}{y})] \quad . \tag{2.10}$$

Our Basic Assumptions are:

(I) For every $\ell \ne \ell' = 1,\ldots,m$, as $y_\ell, y_{\ell'} \to +\infty$,

$$\frac{\Pr\{X_j^{(\ell)} > y_\ell,\ X_j^{(\ell')} > y_{\ell'}\}}{\min(\Pr\{X_j^{(\ell)} > y_\ell\},\ \Pr\{X_j^{(\ell')} > y_{\ell'}\})} \to 0 \tag{2.11}$$

and for every ℓ, $\ell' = 1,\ldots,m$,

$$\frac{\Pr\{X_j^{(\ell)}>y_\ell,\ X_{j+1}^{(\ell')}>y_{\ell'}\}}{\min(\Pr\{X_j^{(\ell)}>y_\ell\},\ \Pr\{X_j^{(\ell')}>y_{\ell'}\})} \to 0 \tag{2.12}$$

(II) The vectors $\underset{\sim}{X}_j$ have the Markovian property

$$\begin{aligned}&\Pr\{\underset{\sim}{X}_j \epsilon A_j | \underset{\sim}{X}_r \epsilon A_r,\ r\leq j-1\}\\ &= \Pr\{\underset{\sim}{X}_j \epsilon A_j | \underset{\sim}{X}_{j-1} \epsilon A_{j-1}\} \quad ,\end{aligned} \tag{2.13}$$

for all $\{A_j,\ j\geq 1\} \in \{A_j,\ j\geq 1\}$, $j\geq 1$.

In defense of this assumption, we may remark that if the $X_i^* = X_i - EX_i$, $i\geq 1$, form an autoregressive sequence of order k, where $k\leq m$, then we may write

$$\underset{\sim}{X}_j = \underset{\sim}{\mu} + \underset{\sim}{B}\underset{\sim}{X}_{j-1} + \underset{\sim}{e}_j \ , \quad \text{for all } j\geq 1 \quad ,$$

where $\underset{\sim}{B}$ is an $m\times m$ matrix of constants and the $\underset{\sim}{e}_j$ are independent of the $\underset{\sim}{X}_{j-1}$, so that assumption (II) holds. Assumption (II) holds also for independent (non-stationary) processes.

Choose a sequence $\underset{\sim}{y} = \underset{\sim}{y}^{(n)} = (y_1^{(n)},\dots,y_m^{(n)})$, such that

$$G(\underset{\sim}{y}^{(n)}) = 1 - r_n \tag{2.14}$$

where $-n\log G(\underset{\sim}{y}^{(n)})$ $(\sim -nr_n) \to -r$; $r\epsilon(0,\infty)$. Then, let

$$G_{[\ell]}(y_\ell^{(n)}) = r_{n_\ell} \ , \quad 1\leq\ell\leq m;\ r_n^* = r_{n_1}+\dots+r_{n_m} \quad . \tag{2.15}$$

Note that

$$G(\underset{\sim}{y}^{(n)}) \leq G_{[\ell]}(y_\ell^{(n)}) \ , \quad \text{for all } \ell\epsilon[1,m] \quad , \tag{2.16}$$

so that $r_{n_\ell} \leq r_n$, for all $\ell = 1,\dots,m$.

Lemma 2.1

$$[\Omega(G(\underset{\sim}{y}^{(n)}))]^{n_0} \to 1 \quad \text{as} \quad n(\sim mn_0)\to\infty \quad .$$

Proof:

$$n_o \log\Omega(G(\underset{\sim}{y}^{(n)}))$$

$$= n_0 \log G(\underset{\sim}{y}^{(n)}) - n_0 \sum_{\ell=1}^{m} \log G_{[\ell]}(y_\ell^{(n)})$$

$$= n_0 \log(1-r_n) - n_0 \sum_{\ell=1}^{m} \log(1-r_{n_\ell}) \quad . \tag{2.17}$$

Now

$$r_n = 1 - G(\underset{\sim}{y}^{(n)})$$

$$= \Pr\{X_j^{(\ell)} > y_\ell^{(n)} \quad \text{for at least one} \quad \ell: \quad 1 \le \ell \le m\}$$

$$= \sum_{\ell=1}^{m} \Pr\{X_j^{(\ell)} > y_\ell^{(n)}\} - \sum_{\ell \ne \ell'=1}^{m} \Pr\{X_j^{(\ell)} > y_\ell^{(n)},\ X_j^{(\ell')} > y_{\ell'}^{(n)}\} + \ldots$$

Therefore by (2.11),

$$r_n = r_n^* - 0(1) r_n \quad ,$$

so that by (2.17)

$$n_0 \log\Omega(G(\underset{\sim}{y}^{(n)})) = n_0 \log(1-r_n^*+0(1)r_n) - n_0 \sum_{\ell=1}^{m} \log(1-r_{n_\ell})$$

$$= n_0 r_n^* + 0(1)(n_0 r_n) - n_0 \sum_{\ell=1}^{m} r_{n_\ell} + 0(1)$$

$$= 0(1)$$

as

$$r_n^* = \sum_{\ell=1}^{m} r_{n_\ell} = 0(\frac{1}{n}) = 0(\frac{1}{n_0}) \tag{2.18}$$

Q.E.D.

Lemma 2.2

$$n_0 \log \Omega^*(G^*(\underset{\sim}{y}^{(n)}, \underset{\sim}{y}^{(n)})) \to 0 \quad \text{as} \quad n_0 \to \infty \quad .$$

Proof:

Note that

$$1 - G(\underset{\sim}{y}^{(n)}, \underset{\sim}{y}^{(n)}) = \Pr\{\text{at least one of } \underset{\sim}{X}_j, \underset{\sim}{X}_{j+1} \le \underset{\sim}{y}^{(n)}\}$$

$$= 2[1-\Pr\{\underset{\sim}{X}_j \le \underset{\sim}{y}^{(n)}\}] - \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{y}^{(n)}, \underset{\sim}{X}_{j+1} \le \underset{\sim}{y}^{(n)}\}$$

$$= 2r_n - \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{y}^{(n)}, \underset{\sim}{X}_{j+1} \le \underset{\sim}{y}^{(n)}\} \qquad (2.19)$$

where by (2.12) and the fact that

$$\Pr\{\underset{\sim}{X}_j \nleq \underset{\sim}{y}^{(n)}, \underset{\sim}{X}_{j+1} \nleq \underset{\sim}{y}^{(n)}\}$$

$$= \Pr\{\text{at least for one } 1 \le \ell, \ell' \le m, X_j^{(\ell)} > y_\ell^{(n)}, X_{j+1}^{(\ell')} > y_{\ell'}^{(n)}\} \quad , \qquad (2.20)$$

we conclude that (2.20) = $0(r_n)$, so that (2.19) is $2r_n + 0(r_n)$. Hence

$$n_0 \log \ ^*(G^*(\underset{\sim}{y}^{(n)}, \underset{\sim}{y}^{(n)})) = n_0 \log(1-2r_n+0(r_n)) - n_0 2 \log(1-r_n)$$

$$= -2n_0 r_n + 0(1) + 2n_0 r_n + 0(1)$$

$$= 0(1) \qquad (2.21)$$

Q.E.D.

Now consider the expression

$$\Pr\{Z_{n_\ell}^{(\ell)} \le y_n^{(\ell)}, 1 \le \ell \le m\} \quad (n_1 = \ldots = n_m = n_0 = [\tfrac{n}{m}])$$

$$= \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{y}^{(n)'}, \forall 1 \le j \le n_0\}$$

$$= \Pr\{\underset{\sim}{X}_1 \le \underset{\sim}{y}^{(n)}\} \prod_{j=2}^{n_0} \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{y}^{(n)} | \underset{\sim}{X}_s \le \underset{\sim}{y}^{(n)}, \forall s \le j-1\}$$

$$= \Pr\{\underset{\sim}{X}_1 \le \underset{\sim}{y}^{(n)}\} \prod_{j=2}^{n_0} \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{y}^{(n)} | \underset{\sim}{X}_{j-1} \le \underset{\sim}{y}^{(n)}\} \qquad \text{(by (2.22))}$$

$$= \Pr\{\underset{\sim}{X}_1 \le \underset{\sim}{y}^{(n)}\} \prod_{j=2}^{n_0} \Pr\{\underset{\sim}{X}_j \le \underset{\sim}{y}^{(n)}, \underset{\sim}{X}_{j-1} \le \underset{\sim}{y}^{(n)}\} / \Pr\{\underset{\sim}{X}_{j-1} \le \underset{\sim}{y}^{(n)}\}$$

$$= [G(\underset{\sim}{y}^{(n)})]^{n_0} [\Omega^*(G^*(\underset{\sim}{y}^{(n)}, \underset{\sim}{y}^{(n)}))]^{n_0 - 1}$$

$$= \prod_{\ell=1}^{m} \{[G_{[\ell]}(y_\ell^{(n)})]^{n_0}\} [\Omega(G(\underset{\sim}{y}^{(n)})]^{n_0} [\Omega^*(G^*(\underset{\sim}{y}^{(n)}, \underset{\sim}{y}^{(n)}))]^{n_0 - 1} \qquad (2.23)$$

Note that by our choice, for every $\ell = 1, \ldots, m$,

$$[G_{[\ell]}(y_\ell^{(n)})]^{n_0} \to H_\ell(x_\ell) \quad , \qquad (2.24)$$

whenever H_ℓ exists, where

$$x_\ell: \; y_\ell^{(n)} = a_{n_\ell} x + b_{n_\ell} \quad .$$

Thus $H_\ell(x_\ell)$ is the limiting distribution function of the standardized form of $Z_{n_\ell}^{(\ell)}$. Combining (2.23) with Lemmas 2.1 and 2.2 we conclude that

$$\lim_{n_0 \to \infty} \Pr\{Z_{n_\ell}^{(\ell)} \le y_\ell^{(n)}, \; 1 \le \ell \le m\} = \prod_{\ell=1}^{m} \lim_{n_0 \to \infty} \Pr\{Z_{n_\ell}^{(\ell)} \le y_\ell^{(n)}\}\}$$

$$= \prod_{\ell=1}^{m} H_\ell(x_\ell) \quad , \qquad (2.25)$$

which proves the desired result; i.e., the sub-sequence maxima are asymptotically jointly independent.

We now prove that the asymptotic pairwise independence of maxima implies the asymptotic joint independence of maxima, in the context of the sub-sequences.

<u>Theorem 2.2</u>

The maxima $Z_{n_1}^{(1)}, \ldots, Z_{n_m}^{(m)}$ are asymptotically jointly independent if each pair of maxima $Z_{n_i}^{(i)}$, $Z_{n_j}^{(j)}$ are asymptotically independent, for $i<j = 1,\ldots,m$.

Proof:

$$\Pr\{Z_{n_1}^{(1)}\le x_1,\ldots,Z_{n_m}^{(m)}\le x_m\}$$

$$= \Pr\{\bigcap_{i=1}^{n_0}[X_i^{(1)}\le x_1,\ldots,X_i^{(m)}\le x_m]\}$$

$$= \prod_{i=1}^{n_0}\Pr\{X_i^{(1)}\le x_1,\ldots,X_i^{(m)}\le x_m \mid \bigcap_{j=1}^{i-1}[X_j^{(1)}\le x_1,\ldots,X_j^{(m)}\le x_m]\} \quad , \qquad (2.26)$$

using Bayes' rule. For notational convenience, let

$$X_{i-k}^{(j)} = Y_i^{(j-km)} \quad ,$$

for $k = 1,\ldots,i-1$ and $i = 1,\ldots,n_0$. So (2.26) becomes

$$\Pr\{Z_{n_1}^{(1)}\le x_1,\ldots,Z_{n_m}^{(m)}\le x_m\}$$

$$= \prod_{i=1}^{n_0}\Pr\{Y_i^{(1)}\le x_1,\ldots,Y_i^{(m)}\le x_m \mid \bigcap_{j=1}^{i-1}\bigcap_{\ell=1}^{m}[Y_i^{(\ell-(i-j)m)}\le x_\ell]\}$$

$$= \prod_{i=1}^{n_0}\left\{\prod_{h=1}^{m}\Pr\{Y_i^{(h)}\le x_h \mid \bigcap_{k=1}^{h-1}[Y_i^{(k)}\le x_k]\bigcap_{j=1}^{i-1}\bigcap_{\ell=1}^{m}[Y_i^{(\ell-(i-j)m)}\le x_\ell]\}\right\}$$

$$= \prod_{i=1}^{n_0}\left\{\prod_{h=1}^{m}\frac{F_{hi}(x_1,\ldots,x_h,x_1,\ldots,x_m,\ldots,x_1,\ldots,x_m)}{F_{(h-1)i}(x_1,\ldots,x_{h-1},x_1,\ldots,x_m,\ldots,x_1,\ldots,x_m)}\right\} \quad , \qquad (2.28)$$

where $F_{hi}(x_1,\ldots,x_h,x_1,\ldots,x_m,\ldots,x_1,\ldots,x_m)$ is a $(h+(i-1)m)$ variate distribution function. Further consideration of (2.28) yields

$$\Pr\{Z_{n_1}^{(1)}\le x_1,\ldots,Z_{n_m}^{(m)}\le x_m\}$$

$$= \left\{\prod_{h=1}^{m}\prod_{i=1}^{n_0}\frac{F_{hi}(x_1,\ldots,x_h,x_1,\ldots,x_m,\ldots,x_1,\ldots,x_m)}{F_h(x_h)F_{(h-1)i}(x_1,\ldots,x_{h-1},x_1,\ldots,x_m,\ldots,x_1,\ldots,x_m)}\right\}\prod_{h=1}^{m}F_h^n(x_h) \quad . \qquad (2.29)$$

Let

$$\Omega_{hi} = \frac{F_{hi}(x_1,\ldots,x_h,x_1,\ldots,x_m,\ldots,x_1,\ldots,x_m)}{F_1(x_1)\ldots F_h(x_h)[F_1(x_1)\ldots F_m(x_m)]^{i-1}} \quad . \tag{2.30}$$

Then,

$$\Pr\{Z_{n_1}^{(1)} \leq x_1,\ldots,Z_{n_m}^{(m)} \leq x_m\} = \prod_{h=1}^{m}\left\{\left\{\prod_{i=1}^{n} \frac{\Omega_{hi}}{\Omega_{(h-1)i}}\right\} F_h^n(x_h)\right\} \quad . \tag{2.31}$$

Therefore, by definition

$$\Omega_n(F_1^n(x_1),\ldots,F_m^n(x_m)) = \prod_{h=1}^{m}\prod_{i=1}^{n} \frac{\Omega_{hi}}{\Omega_{(h-1)i}} \tag{2.32}$$

If each pair of maxima are asymptotically independent, then by Lemma 2.1

$$\lim_{n\to\infty} \prod_{i=1}^{h} \Omega_{hi} = 1 \quad ,$$

for $h = 1,\ldots,m$. Therefore, from (2.32)

$$\lim_{n\to\infty} \Omega_n(F_1^n(x_1),\ldots,F_m^n(x_m)) = 1 \quad , \tag{2.33}$$

that is, the maxima $Z_{n_1}^{(1)},\ldots,Z_{n_m}^{(m)}$ are asymptotically jointly independent.

Using Theorem 2.2, and a result of Sibuya's, Sibuya (1960), we can state the following corollary.

Corollary 2.4

The maxima $Z_{n_1}^{(1)},\ldots,Z_{n_m}^{(m)}$ are asymptotically jointly independent if

$$P_{ij}(1-r,1-r) = o(r) \quad \text{for all} \quad i<j = 1,\ldots,m,$$

where

$$P_{ij}(F_i(x_i),\ F_j(x_j)) = \Pr\{X_i>x_i,X_j>x_j\} \quad . \tag{2.34}$$

This corollary then gives reasonable sufficient conditions under which

$$\lim_{n\to\infty} \Pr\{Z_n \le a_{n_t}^{(t)} x + b_{n_t}^{(t)}\} = H_t(x) \quad .$$

Let us, briefly, consider the condition, that is,

$$P(1-r,1-r) = o(r) \tag{2.35}$$

This is known as Sibuya's condition, given in Sibuya (1960). For example, let us suppose that the marginal distribution of X_1 is a Weibull (α_1,β_1). Furthermore, assume that the bivariate distribution is derived from the third Gumbel bivariate exponential distribution referred to in Gumbel (1960). That is

$$\Pr\{X_1>x_1,\ X_2>x_2\} = \exp\{-[\alpha_1^m x_1^{m\beta_1}+\alpha_2^m x_2^{m\beta_2}]^{1/m} \quad , \tag{2.36}$$

where m is a measure of the dependence between X_1 and X_2. In fact, we can show that

$$m = [1-\mathrm{Corr}(\log X_1,\ \log X_2)]^{-\frac{1}{2}}$$

Let $x_1 = F_1^{-1}(1-r)$ and $x_2 = F_2^{-1}(1-r)$. Thus,

$$\frac{P(1-r,1-r)}{r} = r^{2^{\frac{1}{m}}-1} \tag{2.37}$$

Therefore $P(1-r,1-r) = o(r)$ as long as m^{-1} does not equal zero; that is the correlation is not one. Similarly, it can be shown that if we assume that the marginals are log-normally distributed than Sibuya's condition reduces to the correlation not equalling one. The condition for asymptotic joint independence is then seen to be very general and quite reasonable.

3. The Distribution of the k-th Extreme

Let $X^{(\ell)}_{n_\ell-k+1:\,n_\ell}$ denote the k-th largest observation in the ℓ-th sub-sequence. Following Gnedenko (1943), de Haan (1970) and Galambos (1978) we see that, if it exists

$$H_\ell^{(k)}(x) = \lim_{n_\ell\to\infty} \Pr\{X^{(\ell)}_{n_\ell-k+1:\, n_\ell} \le a^{(\ell)}_{n_\ell} x + b^{(\ell)}_{n_\ell}\} \quad , \tag{3.1}$$

for $\ell = 1,\ldots,m$; k fixed, and

$$H_\ell^{(k)}(x) = H_\ell(x) \sum_{r=0}^{k-1} \frac{1}{r!}[-\log H_\ell(x)]^r \quad , \tag{3.2}$$

where $H_\ell(x)$ is as given by (2.1). In an analogous manner to the work in Section 2 we can show that

$$\lim_{n_\ell\to\infty} \Pr\{X^{(\ell)}_{n_\ell-k+1:\, n_\ell} \le a_n x + b_n\} = H_\ell^{(k)}(A_\ell x + B_\ell) \quad ; \tag{3.3}$$

where A_ℓ and B_ℓ are as before. We noted earlier that A_ℓ and B_ℓ were such that

$$A_\ell x + B_\ell = \begin{cases} x & \text{for } \ell = t \\ \infty & \text{otherwise} \end{cases} \tag{3.4}$$

for $\ell = 1,\ldots,m$; where the t-th sub-sequence is the dominant one. The above results are true for the individual sub-sequences however, we are only interested in the properties of the k-th extreme of the entire sequence, denoted by $X_{n-k+1:\, n}$. Therefore

$$\lim_{n\to\infty} \Pr\{X_{n-k+1:\, n} \le a_n x + b_n\} = \lim_{\to\infty}\{1 - \Pr\{X_{n-k+1:\, n} > a_n x + b_n\}\} \tag{3.5}$$

The set $\{X_{n-k+1:\, n} > a_n x + b_n\}$ can be expressed as the union of a multitude of sets of the form

$$\{X^{(p)}_{n_p-j+1:\, n_p} > a_n x + b_n,\ X^{(q)}_{n_q-i+1:\, n_q} > a_n x + b_n, \ldots\} \quad . \tag{3.6}$$

However, due to (3.3) and (3.4) the probability of these sets is zero except for those which involve only the order statistics of the t-th sub-sequence. If we assume that each sub-sequence is unique then there is only such set, corresponding to the case where all the k observations greater than $a_n x+b_n$ belong to the t-th sub-sequence. Therefore, assuming the asymptotic joint independence of the sub-sequence b-th extremes, (3.5) reduces to

$$\lim_{n\to\infty} \Pr\{X_{n-k+1:\,n} \leq a_{n_t}^{(t)} x + b_{n_t}^{(t)}\} = H_t(x) \sum_{r=0}^{k-1} \frac{1}{r!} [-\log H_t(x)]^r \qquad (3.7)$$

We have therefore shown that for a certain m-dependent non-stationary sequence the limiting extreme value distribution can be obtained and is identical to that for the most extreme or dominant sub-sequence.

4. Applications

The above approach was motivated by the inappropriateness of the assumption of a stationary sequence adopted by Watson (1954) and more recently Leadbetter, Lindgren and Rootzén (1979) in relation to air pollution concentrations. These concentrations are such that consecutive days are strongly dependent and a weekly cyclical pattern is evident due to variations in traffic flows. In addition meteorological factors which typically have an effect for less than one week destroy the stationarity of the sequence of concentrations. It did, however, seem reasonable to assume that observations on different weeks follow a Markovian pattern.

Of major interest to the U.S. Environmental Protection Agency, for the setting of standards, is the distribution of the k-th largest daily maximum concentration over a long time period. This paper attempts to solve this problem by imposing some reasonable assumptions, detailed above, on the dependence structure of the observations.

References

1. Fréchet, M.: Les Probabilitiés Associées à un Système D'évènements Compatibles et Dépendants, Paris: Hermann 1940.

2. Galambos, J.: An Asymptotic Theory of Extreme Order Statistics, New York: John Wiley 1978.

3. Gnedenko, B.V.: Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math. 44, 423-453 (1943).

4. Gumbel, E.J.: Bivariate distributions. J. Amer. Statist. Assoc. 55, 698-707 (1960).

5. Haan, L. de: On Regular Variation and its Application ot the Weak Convergence of Sample Extremes, Mathematical Centre Tracts 32, Amsterdam: Mathematical Centre 1970.

6. Leadbetter, M.R., Lindgren, G. and Rootzén, H.: Extremal and Related Properties of Stationary Processes, Part 1: Extremes of Stationary Sequences, Sweden: University of Umea 1979.

7. Sibuya, M.: Bivariate extreme statistics. Ann. Inst. Stat. Math., 11, 195-210 (1960).

8. Watson, G.S.: Extreme values in samples from independent stationary stochastic processes. Ann. Math. Statist. 25, 798-800 (1954).