ESTIMATION WITH MULTIVARIATE EXTREME VALUE DISTRIBUTIONS, WITH APPLICATIONS TO ENVIRONMENTAL DATA

By

YOUSHI CHEN

B.S c., Beijing Institute of Technology, 1983

M.S c., Beijing Institute of Technology, 1986

A THESIS SUBMITED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1991

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date Oct. 10, 1991

Abstract

Several parametric families of multivariate extreme value distributions (Hüsler and Reiss 1989, Tawn 1990, Joe 1990a, 1990b) have been proposed recently. Applications to multivariate extreme value data sets are needed to assess the adequacy of the known families in their fit to data. Different families are compared in their range of multivariate dependence and their ease of use for maximum likelihood estimation. Some useful conclusions have been made from experience with several environmental data sets.

TABLES OF CONTENTS

List of tables

ACKNOWLEDGMENT

I would like to express my deep gratitude to my supervisor, Professor Harry Joe, for his constant encouragement, guidance and assistance in my study and in the producing of this thesis. Very special thanks are due Professor Mohan Delampady for reading the manuscripts and making constructive comments. Finally, I would like to thank the Department of Statistics, The University of British Columbia, for supporting me with financial aid and encouragement.

Chapter 1 Introduction

Extreme value theory has a great number of applications. We list two cases where the largest or the smallest "measurement" are of interest.

1. Air Pollution. Air pollution concentration is expressed in terms of proportion of a specific pollutant in the air. Concentrations are recorded at equal time intervals, and it is required by law to keep the largest annual concentrations below given limits.

2. Natural Disasters. Floods, heavy rains, extreme temperatures, extreme atmospheric pressures, winds and other phenomena can cause extensive human and material loss. Communities can take preventive action to minimize their effects even if such disasters cannot be completely avoided. In dams, dikes, canals, and other structures the choice of building materials and methods of architecture can take some of these potential disasters into account. Engineering decisions that confront such problems should be based on a very accurate theory, because inaccuracies can be very expensive. For example, a dam built at a huge expense may not last long before collapsing.

In example 1, at each time point, there may be a vector $\mathbf{X}$ of measurements, consisting of concentrations of a pollutant at several air quality monitoring stations in a network or consisting of concentrations of several pollutants at a given monitoring stations. In example 2, there may be a vector $\mathbf{X}$ of measurements at several positions in a spatial grid.

In the examples, we have over time a number n of random measurements $\mathbf{X}_1, \ldots, \mathbf{X}_n$, and the behaviour of either $\mathbf{Z}_n = \max(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ or $\mathbf{W}_n = \min(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ can be of interest. Here max and min refer to componentwise maxima and minima; see Chapter 2 for details.

Statistical inferences of the upper or lower tail of the multivariate distribution of $\mathbf{X}$ can be made from using extreme value theory to model the multivariate distribution of

$\mathbf{Z_n}$ or $\mathbf{W_n}$.

There has been much recent research in multivariate extreme value theory; see Galambos (1987, Chapter 5) and Resnick (1987, Chapter 5) for some general theory. The set of possible of multivariate extreme value distributions has been shown to be of infinite dimension. Smith, Tawn and Yuen (1990) study a density estimation based method for nonparametrically fitting a multivariate extreme value distribution. Except in the bivariate case, the method was not too promising. An alternative is to find a finite-dimensional parametric subfamily of the set of multivariate extreme value distributions. A good subfamily would be "dense" in the complete family. Parametric families have recently been proposed by Hüsler and Reiss (1989), Tawn (1990), Joe (1990a,b), Coles and Tawn (1991). Experience is now needed to assess the adequacy of fit of these families to multivariate extreme value data sets. Joe (1990b) did some work on this but did not include the Hüsler and Reiss family in his comparisons.

One goal of this thesis is to do maximum likelihood estimation with the Hüsler and Reiss family. This turns out be very difficult and details are given in Chapter 3, where comparisons are also made with results in Joe (1990b).

Another goal of the thesis is to deal with missing values in computing maxima. Joe (1990b) deleted cases where there were many missing values so that maxima could not be properly computed. These maxima were assumed to be missing at random. Here, in Chapter 4, we treat these cases as right-censored maxima, that is the actual maxima is known to be above the maxima of all non-missing observations.

The thesis proceeds next (in Chapter 2) with definitions and properties of general multivariate distributions and of multivariate extreme value distribution. The last chapter (Chapter 5) of this thesis makes some final conclusions and points out a direction for further research.

Chapter 2. Multivariate Extreme Value Distribution

2.1 Introduction

In this chapter we describe some general properties of multidimensional distributions and copulas. Then we introduce some univariate extreme value theory and the generalized extreme value distribution. Following this, we concentrate on multivariate extreme value distributions, the main topic of this chapter.

2.2 Multidimensional Distributions

Let us define the p-dimensional random variable $\mathbf{X}$ as the vector $\mathbf{X} = (X_1, \ldots, X_p)$ The distribution function $F(\mathbf{x}) = F(x_1, \ldots, x_p)$ is defined as

$$F(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = P(X_1 \leq x_1, \ldots, X_p \leq x_p),$$

where $\mathbf{X} \leq \mathbf{Y}$ means $X_i \leq Y_i, \quad$ for $1 \leq i \leq p$.

For future reference, we also define

$$\mathbf{X} + \mathbf{Y} = (X_1 + Y_1, \ldots, X_p + Y_p),$$

$$\mathbf{XY} = (X_1 Y_1, \ldots, X_p Y_p),$$

and

$$\mathbf{X}/\mathbf{Y} = (X_1/Y_1, \ldots, X_p/Y_p).$$

Random variables with the same distribution as $\mathbf{X}$ will be denoted by $\mathbf{X}_1, \ldots, \mathbf{X}_p$, and the components of $\mathbf{X}_j$ by X_{ij}, that is, X_{ij} is the ith component of $\mathbf{X}_j$, or $\mathbf{X}_j = (X_{1j}, \ldots, X_{pj})$.

The order statistics of the ith component $(X_{i1}, \ldots, X_{in})$ are

$$X_{in}^{(1)} \leq X_{in}^{(2)} \leq \ldots \leq {X_{in}}^{(n)},$$

and we denote

$$W_i = X_{in}^{(1)}, \qquad Z_i = X_{in}^{(n)}.$$

In extreme value theory, an objective is to investigate the existence of the asymptotic distribution of

$$\mathbf{W}_n = (W_1, \ldots, W_p), \qquad \mathbf{Z}_n = (Z_1, \ldots, Z_p).$$

For this, we need some theory of multidimensional distribution functions.

A multidimensional distribution function $F(\mathbf{x})$ of a random vector $\mathbf{X}$ has the following elementary properties:

P1 (Bounds): $0 \leq F(x_1, \ldots, x_p) \leq 1 \qquad \forall x_1, \ldots, x_p$.

P2 (Monotonicity): F is nondecreasing in each of its arguments x_i, $1 \leq i \leq p$.

P3 (Limits): $\lim_{x_k \to -\infty} F(x_1, \ldots, x_p) = 0$.

P4 (Marginal distributions): If $x_j \to +\infty$, then $F(\mathbf{x})$ tends to an $(p-1)$-dimensional distribution, which is the distribution of the vector obtained by removing its jth component. We can obtain the univariate marginal distribution $F_j(x_j)$ by letting each x_i, $i \neq j$ tend to $+\infty$.

P5 (Density): If F has derivatives of order p, then

$$f = \frac{\partial^{(p)} F}{\partial x_1 \cdots \partial x_p} \geq 0,$$

and f is the probability density function.

P6 (Rectangles):

$$\begin{aligned} P(a_1 < X_1 \leq b_1, \ldots, a_p < X_p \leq b_p) &= F(b_1, b_2, \ldots, b_p) - \sum_{i=1}^{p} F_i + \\ &\quad \sum_{i<j} F_{ij} + \cdots + (-1)^p F(a_1, \ldots, a_p) \geq 0, \end{aligned}$$

where for a subset I of $\{1, \ldots, p\}$, F_I is the value of $F(s_1, \ldots, s_k)$ with $s_i = a_i$ if $i \in I$, and $s_i = b_i$ if $i \notin I$.

P7 (the Fréchet Bounds):

$$\max(0, \sum_{i=1}^{p} F_i(x_i) - p + 1) \leq F(x_1, \ldots, x_p) \leq \min(F_1(x_1), \ldots, F_p(x_p)).$$

If $F_1, \ldots, F_p$ are all continuous, then the upper bound corresponds to distribution of $(X_1, F_2^{-1}F_1(X_1), \ldots, F_p^{-1}F_1(X_1))$. In general the lower bound is not a multivariate distribution function for $p \geq 3$. For example, suppose F is a three-dimensional distribution function with marginal functions satisfying $F_1(x_1) = 1/2$, $F_2(x_2) = 1/2$ and $F_3(x_3) = 1/2$. Then

$$\begin{aligned} & F(+\infty, +\infty, +\infty) - F_1(x_1) - F_2(x_2) - F_2(x_3) \\ & \qquad + \max(0, F_1(x_1) + F_2(x_2) - 1) + \max(0, F_1(x_1) + F_3(x_3) - 1) + \\ & \qquad \max(0, F_2(x_2) + F_3(x_3) - 1) - \max(0, F_1(x_1) + F_2(x_2) + F_3(x_3) - 2) \\ = & \quad 1 - 3/2 = -1/2 < 0, \end{aligned}$$

which contradicts Property 6.

P8 (Survival function): Let B_j be the event $(X_j > x_j)$ and $j(k) = (j_1, \ldots, j_k)$ be a subset of $\{1, \ldots, p\}$, then let

$$G_{j(k)}(x_{j_1}, \ldots, x_{j_k}) = P(B_{j_1} \cap \cdots \cap B_{j_k}), \qquad S_0(\mathbf{x}) = 1,$$

and

$$S_k(\mathbf{x}) = \sum_{1 \leq j_1 \leq j_2 \cdots j_k \leq p} G_{j_k}(x_{j_1}, \ldots, x_{j_k}), \quad 1 \leq k \leq p.$$

For $p \geq 2$, then

$$F(x_1, \ldots, x_p) = \sum_{k=0}^{p} (-1)^k S_k(x_1, \ldots, x_p).$$

In addition, for any integer $0 \leq q \leq (p-1)/2$,

$$\sum_{k=0}^{2q+1} (-1)^k S_k(\mathbf{x}) \leq F(\mathbf{x}) \leq \sum_{k=0}^{2q} (-1)^k S_k(\mathbf{x}).$$

The proof of most of the above results is straightforward; for the proof of properties P7 and P8, readers are referred to Galambos (1987).

2.3 Copulas

Assume that $F(\mathbf{x})$ is a p-dimensional distribution function with univariate marginals $F_i(x_i)$, for $1 \leq i \leq p$. Let $C(\mathbf{y})$ be a p-dimensional function over the unit cube $0 \leq y_i \leq 1$, $1 \leq i \leq p$, and such that it increases in each of its variables and

$$F(x_1, \ldots, x_p) = C[F_1(x_1), \ldots, F_p(x_p)].$$

Then the function $C(\mathbf{y})$ is called a copula of $F(\mathbf{x})$. If C is a copula, then

$$G(y_1, \ldots, y_p) = C[G_1(y_1), \ldots, G_p(y_p)]$$

is a multivariate distribution with univariate margins G_i, for $i = 1, ..., p$, where G_i is an arbitrary univariate distribution function. When needed, we write $C_F = C_F(\mathbf{y}) = C(\mathbf{y})$ for $F(\mathbf{x})$. Note that some people call $C(\mathbf{y})$ a dependence function.

If $F(x_1, \ldots, x_p)$ is a continuous p-variate cumulative distribution function with univariate margins $F_j(x_j), j = 1, \ldots, p$, then the associated copula is

$$C(u_1, \ldots, u_p) = F(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)), \qquad 0 \leq u_i \leq 1, i = 1, \ldots, p.$$

This is a multivariate distribution with uniform $(0, 1)$ margins.

Returning to extreme values, the ith marginal of $F^n(\mathbf{x})$ is $F_i^n(x_i)$, and thus

$$F^n(x_1, \ldots, x_p) = C_{F^n}[F_1^n(x_1, \ldots, F_p^n(x_p)].$$

On the other hand, we have

$$F^n(x_1, \ldots, x_p) = C_F^n[F(x_1), \ldots, F(x_p)].$$

A comparison of these last two equations leads to

$$C_{F^n}(\mathbf{y}) = C_F^n[y_1^{1/n}, \ldots, y_p^{1/n}].$$

2.4 Univariate Extremes and the Generalized Extreme Value Distribution

Let $X_1, \ldots, X_n$ denote independent and identically distributed (iid) random variables, with cumulative distribution function F. Let

$$Z_n = \max(X_1, \ldots, X_n), \qquad W_n = \min(X_1, \ldots, X_n).$$

By the iid assumption, we have

$$H_n(x) = P(Z_n \leq x) = F^n(x)$$

and

$$L_n(x) = P(W_n \leq x) = 1 - (1 - F(x))^n.$$

We seek conditions on $F(x)$ to guarantee the existence of sequences $\{a_n\}$, $\{b_n\}$, $\{c_n\}$, $\{d_n\}$ of constants such that, as $n \to \infty$

$$\lim_{n\to\infty} H_n(a_n + b_n x) = H(x)$$

and

$$\lim_{n\to\infty} L_n(c_n + d_n x) = L(x)$$

exist for all continuity points of $H(x)$ and $L(x)$ respectively, where $H(x)$ and $L(x)$ are nondegenerate distribution functions. Equivalently, $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ and $\{d_n\}$ are such that $(Z_n - a_n)/b_n$ and $(W_n - c_n)/d_n$ converge in distribution as $n \to \infty$.

A sufficient condition (see Galambos, 1987) is given next.

Assume that there are sequences a_n, c_n, $b_n > 0$, and $d_n > 0$, of real numbers such that, for all x and y,

$$\lim_{n\to+\infty} n[1 - F(a_n + b_n x)] = u(x),$$

$$\lim_{n\to+\infty} nF(c_n + d_n y) = w(y)$$

exist. Then

$$\lim_{n\to\infty} P(Z_n \leq a_n + b_n x) = \exp[-u(x)]$$

and

$$\lim_{n\to\infty} P(W_n \le c_n + d_n y) = 1 - \exp[-w(y)].$$

Example 1. (The Exponential Distribution). Let $X_1, \ldots, X_n$ be iid random variables with common distribution function

$$F(x) = 1 - e^{-x}, \qquad x \ge 0.$$

Then

$$1 - F(a_n + b_n x) = e^{-a_n} e^{-b_n x}.$$

In order to satisfy

$$\lim_{n\to+\infty} n[1 - F(a_n + b_n x)] = u(x),$$

we can choose $a_n = \log n$ and $b_n = 1$. Hence, $u(z) = e^{-z}$, and thus,

$$\lim_{n\to\infty} P(Z_n < \log n + z) = \exp(-e^{-z}).$$

On the other hand, we can choose $c_n = 0$ and $d_n = 1/n$, such that

$$\lim_{n\to+\infty} nF(c_n + d_n y) = \lim_{n\to+\infty} nF(y/n) = \lim_{n\to+\infty} n(1 - e^{-y/n}) = y.$$

Hence,

$$\lim_{n\to\infty} P(W_n \le c_n + d_n y) = \lim_{n\to+\infty} P(W_n \le y/n) = 1 - e^{-y}.$$

Example 2. Let $X_1, \ldots, X_n$ be iid random variables with common distribution function

$$F(x) = 1 - x^{-1}, \qquad x \ge 1.$$

To determine the limiting distribution of Z_n, we note that with $a_n = 0$ and $b_n = n$, we obtain

$$\lim_{n\to+\infty} n[1 - F(a_n + b_n y)] = \lim_{n\to+\infty} n\frac{1}{ny} = \frac{1}{y}, \qquad y > 0.$$

and thus

$$\lim_{n\to+\infty} P(Z_n < ny) = \exp(-\frac{1}{y}).$$

For maxima, it is well-known that there are only three types of nondegenerate distributions $H(x)$ that can be univariate limiting extreme value distributions. These results are contained in books devoted to extreme value theory, such as Galambos (1987). The three types are:

$$H_{1,\gamma}(x) = \begin{cases} \exp(-x^{\frac{-1}{\gamma}}) & x > 0, \gamma > 0, \\ 0 & \text{otherwise}, \end{cases}$$

$$H_{2,\gamma}(x) = \begin{cases} \exp(-(-x)^{\frac{-1}{\gamma}}) & x < 0, \gamma < 0, \\ 1 & x \geq 0, \end{cases}$$

$$H_{3,0}(x) = \exp(-e^{-x}) \qquad -\infty < x < +\infty, \gamma = 0.$$

After adding location and scale parameters the above three types of distributions can be summarized together as the generalized extreme value distribution (GEV):

$$H(x) = \exp\{-[\max(1 + \frac{\gamma(x-\mu)}{\sigma}, 0)]^{\frac{-1}{\gamma}}\} \qquad -\infty < \gamma < +\infty. \qquad (2-4-1)$$

The parameter γ can be interpreted as a tail index parameter of F. It measures the thickness of the tail of F; γ is larger if F has a thicker tail, and vice versa. The tail thickness is related to the rate of convergence of $1 - F$ to 0 as $x \to \infty$.

For example, let

$$1 - F(x) = [\max(1 + \gamma x, 0)]^{-1/\gamma}.$$

For $\gamma_2 < \gamma_1 < 0$ then $-1/\gamma_1 > -1/\gamma_2 > 0$ and

$$\lim_{x\to-1/\gamma_2} \frac{[\max(1+\gamma_1 x, 0)]^{-1/\gamma_1}}{[\max(1+\gamma_2 x, 0)]^{-1/\gamma_2}} = \infty.$$

For $\gamma_2 < 0$, $\gamma_1 = 0$, then

$$\lim_{x\to-1/\gamma_2} \frac{e^{-x}}{[\max(1+\gamma_2 x, 0)]^{-1/\gamma_2}} = \infty.$$

For $\gamma_2 = 0$, $\gamma_1 > 0$, then

$$\lim_{x\to\infty} \frac{[\max(1+\gamma_1 x, 0)]^{-1/\gamma_1}}{e^{-x}} = \lim_{x\to\infty} \frac{e^x}{[\max(1+\gamma_1 x, 0)]^{1/\gamma_1}} \geq \lim_{x\to\infty} \frac{e^x}{(x\gamma_1)^{1/\gamma_1}} = \infty.$$

For $\gamma_1 > \gamma_2 > 0$, then $0 > -1/\gamma_1 > -1/\gamma_2$, $1/\gamma_2 - 1/\gamma_1 > 0$ and

$$\lim_{x\to\infty} \frac{[\max(1+\gamma_1 x, 0)]^{-1/\gamma_1}}{[\max(1+\gamma_2 x, 0)]^{-1/\gamma_2}} = \lim_{x\to\infty} \frac{(1+\gamma_1 x)^{-1/\gamma_1}}{(1+\gamma_2 x)^{-1/\gamma_2}}$$

$$\geq \lim_{x\to\infty} \frac{(1+\gamma_1 x)^{1/\gamma_2}}{(1+\gamma_1 x)^{1/\gamma_1}} = \lim_{x\to\infty} (1+\gamma_1 x)^{1/\gamma_2 - 1/\gamma_1} = \infty.$$

So, the tail is heavier as γ increases.

For $\gamma > 0$, from Theorem 2.1.1 of Galambos (1987), we find $a_n = 0$ and $b_n = \gamma^{-1}[(1/n)^{-\gamma} - 1]$ so that

$$\overline{F}(b_n x) = 1 - F(b_n x) = [1 + (n^\gamma - 1)x]^{-1/\gamma} \sim n^{-1} x^{-1/\gamma}.$$

and

$$F^n(b_n x) = (1 - \overline{F}(b_n x))^n = (1 - n^{-1} x^{-1/\gamma})^n \overset{n\to\infty}{\to} \exp(-x^{-1/\gamma}).$$

For $\gamma < 0$, from Theorem 2.1.2 of Galambos (1987), we find $a_n = -1/\gamma$, $b_n = -\gamma^{-1} n^\gamma$ so that

$$\overline{F}(x) = (1+\gamma x)^{-1/\gamma}, \qquad \overline{F}(a_n + b_n x) = (-n^\gamma x)^{-1/\gamma} = n^{-1}(-x)^{-1/\gamma}$$

and

$$F^n(a_n + b_n x) = [1 - n^{-1}(-x)^{-1/\gamma}]^n \to \exp(-(-x)^{-1/\gamma}), \qquad x < 0.$$

For $\gamma = 0$, from Theorem 2.1.3 of Galambos (1987), we find $a_n = \log n$, $b_n = 1$ so that

$$\overline{F}(x) = e^{-x} \qquad \overline{F}(a_n + b_n x) = n^{-1} e^{-x}$$

and

$$F^n(a_n + b_n x) = [1 - n^{-1} e^{-x}]^n = \exp(-e^{-x}).$$

The GEV distribution is useful for statistical inference when the distribution is unknown, and the tail index must be estimated.

Since

$$\min(X_1,\ldots,X_n) = -\max(-X_1,\ldots,-X_n),$$

then, similarly for minima there are also three types of nondegenerate distributions as the univariate limiting extreme value distribution which can be summarized together and can be expressed in terms of (2-4-1)

$$\begin{aligned} L(x) &= 1 - H(-x) \\ &= 1 - \exp\{-[\max(1 + \frac{\gamma(-x-\mu)}{\sigma}, 0)]^{-1/\gamma}\}. \end{aligned}$$

2.5 Multivariate Extreme Value Distributions

We call a p-dimensional distribution function $F(\mathbf{x})$ nondegenerate if all of its univariate marginals are nondegenerate.

One object of the multivariate extreme theory is to seek conditions on $F(\mathbf{x})$, under which there are sequences $\{\mathbf{a_n}\}$ and $\{\mathbf{b}_n\}$ of vectors such that each component of $\{\mathbf{b}_n\}$ is positive and

$$P(\mathbf{Z}_n < \mathbf{a}_n + \mathbf{b}_n\mathbf{z}_n) = H_n(\mathbf{a}_n + \mathbf{b}_n\mathbf{z}_n) \rightarrow H(\mathbf{z}) \qquad (2-5-1)$$

for a nondegenerate p-dimensional function $H(\mathbf{z})$.

Suppose $\mathbf{X_1},\ldots,\mathbf{X_n}$ are independent random vectors with common distribution function F, then

$$H_n(\mathbf{z}) = P(Z_{n1} \leq z_1,\ldots,Z_{np} \leq z_p) = F^n(\mathbf{z}),$$

$$L_n(\mathbf{w}) = P(W_{n1} > w_1,\ldots,W_{np} > w_p) = G^n(\mathbf{w}),$$

where

$$G(\mathbf{w}) = P(X_1 > w_1,\ldots,X_p > w_p).$$

Any problem on $\mathbf{W}$ is equivalent to one on $\mathbf{Z}$ by changing the basic vector $\mathbf{X}$ to $(-\mathbf{X})$, therefore we concentrate on the vector $\mathbf{Z}$ of maxima.

Next we list and discuss various properties which $F(\mathbf{x})$ or $H(\mathbf{x})$ may possess.

Property 1. If $F_n(\mathbf{x})$ is a sequence of p-dimensional distributions, let the ith univariate marginal of $F_n(\mathbf{x})$ be $F_n^{(i)}(x_i)$. If $F_n(\mathbf{x})$ converges to a nondegenerate continuous distribution function $F(\mathbf{x})$, then for each i with $1 \le i \le p$, $F_n^{(i)}(x_i)$ converges to the ith marginal $F^{(i)}(x_i)$ of $F(\mathbf{x})$.

Later we shall see that all limiting distribution functions of multivariate extremes are continuous. Hence, Property 1 tells us that we can appeal to univariate case for determining the components of $\mathbf{a_n}$ and $\mathbf{b_n}$ whenever (2-5-1) holds.

The following several properties are important in multivariate extreme value theory.

Property 2. Let $\mathbf{X_1}, \ldots, \mathbf{X_n}$ be iid p-dimensional vectors with common distribution function $F(\mathbf{x})$. Then there are vectors $\mathbf{a_n}$ and $\mathbf{b_n} > \mathbf{0}$ such that $(\mathbf{Z_n} - \mathbf{a_n})/\mathbf{b_n}$ converges to nondegenerate distribution function $H(\mathbf{x})$, if and only if, each marginal belongs to the GEV family, and if

$$C_F^n[y_1^{1/n}, \ldots, y_p^{1/n}] \to C_H(y_1, \ldots, y_p), \qquad n \to \infty.$$

Property 2 tells us that given a p-variate distribution function $F_n(\mathbf{x})$, we can check if its marginals belong to the GEV family, if so, then we use the methods of the univariate case to determine the components of the vectors $\mathbf{a_n}$ and $\mathbf{b_n}$; furthermore we can determine $C_F(\mathbf{y})$ by its definition and check if $C_F^n(\mathbf{y}^{1/n})$ converges.

Example 3. Let (X, Y) have a bivariate exponential distribution

$$F(x, y) = 1 - e^{-x} - e^{-y} + G(x, y), \qquad (2-5-2)$$

where

$$G(x, y) = P(X > x, Y > y).$$

If $(\mathbf{Z_n} - \mathbf{a_n})/\mathbf{b_n}$ converges to a distribution $H(z_1, z_2)$, then we can choose $\mathbf{a_n} = (\log n, \log n)$, and $\mathbf{b_n} = (1,1)$, and obtain

$$F(\log n + z_1, \log n + z_2) = 1 - \frac{e^{-z_1} + e^{-z_2}}{n} + G(\log n + z_1, \log n + z_2).$$

For the Marshall-Olkin distribution

$$G(x,y) = \exp[-x - y - \lambda \max(x,y)], \qquad \lambda > 0.$$

The last term of (2-5-2) is

$$\begin{aligned}\exp\{-(2+\lambda)\log n - z_1 - z_2 - \lambda \max(z_1, z_2)\} &= \frac{1}{n^{2+\lambda}} \exp\{-z_1 - z_2 - \lambda \max(z_1, z_2)\} \\ &= O(n^{-2-\lambda}),\end{aligned}$$

and

$$\begin{aligned}\lim_{n\to\infty} F^n(\log n + z_1, \log n + z_2) &= \lim_{n\to\infty} \left(1 - \frac{e^{-z_1} + e^{-z_2}}{n}\right)^n \\ &= \exp(-e^{-z_1} - e^{-z_2}).\end{aligned}$$

For another example, let

$$G(x,y) = (e^{\theta x} + e^{\theta y} - 1)^{-1/\theta}, \qquad \theta > 0.$$

Then

$$\begin{aligned}G(\log n + z_1, \log n + z_2) &= (n^\theta e^{\theta z_1} + n^\theta e^{\theta z_2} - 1)^{-1/\theta} \\ &\sim \frac{1}{n}(e^{\theta z_1} + e^{\theta z_2})^{-1/\theta},\end{aligned}$$

and

$$\begin{aligned}\lim_{n\to\infty} F^n(\log n + z_1, \log n + z_2) &= \lim_{n\to\infty} (1 - \frac{e^{z_1} + e^{z_2} - (e^{\theta z_1} + e^{\theta z_2})^{-1/\theta}}{n})^n \\ &= \{\exp[-(e^{z_1} + e^{z_2})]\}\{\exp[(e^{\theta z_1} + e^{\theta z_2})^{-1/\theta}]\}.\end{aligned}$$

In the second example, there is dependence in the limiting extremes.

Property 3. A p-variate continuous distribution function $H(\mathbf{x})$ is a limit distribution in (2-5-1) if, and only if, its univariate marginals belong to the GEV family and if its copula C_H satisfies

$$C_H^k(y_1^{1/k}, \ldots, y_p^{1/k}) = C_H(y_1, \ldots, y_p), k = 1, 2, \cdots. \qquad (2-5-3)$$

Property 3 tells us that if the limit $H(\mathbf{x})$ exists, then we check condition (2-5-3). If it holds and if C_H is a copula, then we have got the actual limit distribution.

Example 4. Let $H_1(x), \ldots, H_p(x)$ belong to the GEV family, then

$$H(\mathbf{x}) = H_1(x) \cdots H_p(x)$$

is a possible limit of (2-5-1). This limit consists of independent univariate marginals. By assumption, the condition of Property 2 on the marginals is satisfied. Furthermore, by definition,

$$C_H(\mathbf{y}) = y_1 \cdots y_p$$

for which the condition of (2-5-3) is evident. An appeal to Property 3 yields the claim.

Example 5. The distribution function

$$H(x_1, \ldots, x_p) = \exp\{\exp[-\min(x_1, \ldots, x_p)]\}$$

is a limit in (2-5-1).

The marginal distributions are $H_i(x_i) = \exp(-e^{-x_i}) = H_{3,0}(x_i)$. Therefore it remains to check the validity of the condition (2-5-3):

$$\exp\{-\exp[-\min(x_1, \ldots, x_p)]\} = \min\{\exp[-\exp(-x_i)] \quad 1 \leq i \leq p\},$$

and

$$C_H(y_1, \ldots, y_p) = \min(y_1, \ldots, y_p).$$

Hence for $k \geq 1$,

$$C_H^k(y_1^{1/k}, \ldots, y_p^{1/k}) = [\min(y_1^{1/k}, \ldots, y_p^{1/k})]^k = \min(y_1, \ldots, y_p).$$

Example 6. The distribution

$$H(x_1, x_2) = H_{3,0}(x_1)H_{3,0}(x_2)[1 + \frac{1}{2}(1 - H_{3,0}(x_1))(1 - H_{3,0}(x_2))]$$

does not occur as a limit in (2-5-1). (See Galambos(1987).) Even though the marginals $H_1(x_1) = H_{3,0}(x_1)$ and $H_2(x_2) = H_{3,0}(x_2)$ the copula

$$C_H(y_1, y_2) = y_1 y_2[1 + \frac{1}{2}(1 - y_1)(1 - y_2)]$$

fails to satisfy (2-5-3).

Let $\mathbf{X} = (X_1, \ldots, X_p)$ be a vector with distribution function $F(\mathbf{x})$. Let $j(k) = (j_1, \ldots, j_k)$, $1 \leq k \leq p$ be a vector with components $1 \leq j_1 \leq j_2 \cdots \leq j_k \leq p$. The distribution function $F_{j(k)}(x_{j_1}, \ldots, x_{j_k})$ is a k-dimensional marginal distribution, which is obtained from $F(\mathbf{x})$ by letting $x_i \to +\infty$ for all $i \notin \{j_1, \ldots, j_k\}$.

Let

$$G_{j(k)}(x_{j_1}, \ldots, x_{j_k}) = P(X_{j_1} > x_{j_1}, \ldots, X_{j_k} > x_{j_k}).$$

Assume that $F(\mathbf{x})$ is such that each of its univariate marginals belongs to the GEV family. Then, there are a_{in} and $b_{in} > 0$ such that

$$\lim_{n\to\infty} F_i^n(a_{in} + b_{in}x) = H_i(x), \quad 1 \leq i \leq p, \qquad (2-5-4)$$

where $H_i(x)$ belongs to the GEV family. We assume that a_{in} and $b_{in} > 0$ have been determined and put $\mathbf{a_n} = (a_{1n}, \ldots, a_{pn})$ and $\mathbf{b_n} = (b_{1n}, \ldots, b_{pn})$.

Property 4. $(\mathbf{Z_n} - \mathbf{a_n})/\mathbf{b_n}$ converges to a nondegenerate distribution $H(\mathbf{x})$ if and only if for each fixed vector j(k) and $\mathbf{x}$ for which $H_i(x_i)$, $1 \leq i \leq p$, of (2-5-4) are positive, the limit

$$\lim_{n\to\infty} nG_{j(k)}(a_{j_1n} + b_{j_1n}x_{j_1n}, \ldots, a_{j_kn} + b_{j_kn}x_{j_kn}) = h_{j_k}(x_{j_1}, \ldots, x_{j_k}) \qquad (2-5-5)$$

are finite, and the function

$$H(\mathbf{x}) = \exp\{\sum_{k=1}^{p}(-1)^k \sum_{1\leq j_1\leq j_2\cdots\leq j_k\leq p} h_{j_k}(x_{j_1},\ldots,x_{j_k})\} \tag{2-5-6}$$

is a nondegenerate distribution function. If the actual limit distribution of $(\mathbf{Z_n} - \mathbf{a_n})/\mathbf{b_n}$ is the one given in (2-5-6), then the following inequalities hold. Let $s \geq 0$ be an integer, then

$$H(\mathbf{x}; 2s+1) \leq H(\mathbf{x}) \leq H(\mathbf{x}; 2s), \tag{2-5-7}$$

where $H(\mathbf{x}, 0) = 1$ and

$$H(\mathbf{x}, r) = \exp\{\sum_{k=1}^{r}(-1)^k \sum_{1\leq j_1\leq j_2\cdots\leq j_k\leq p} h_{j_k}(x_{j_1},\ldots,x_{j_k})\}. \tag{2-5-8}$$

When $s = 0$ in (2-5-7), the inequality $H(\mathbf{x}, 1) \leq H(\mathbf{x}) \leq 1$ obtains. This means that $H(\mathbf{x})$ is never exceeded by the product of its univariate marginals, and $H(\mathbf{x})$ never exceeds 1.

Example 7. Let $F(x, y) = 1 - e^{-x} - e^{-y} + G(x, y)$ where

$$G(x, y) = (e^{\theta x} + e^{\theta y} - 1)^{-1/\theta}.$$

From the univariate case we know that

$$\lim_{n\to\infty} F_1^n(\log n + x) = \exp(-e^{-x}), \qquad x \geq 0,$$

$$\lim_{n\to\infty} F_2^n(\log n + y) = \exp(-e^{-y}), \qquad y \geq 0.$$

Since $G_1(x) = e^{-x}$ and $G_2(y) = e^{-y}$

$$h_1(x) = \lim_{n\to\infty} nG(\log n + x) = \lim_{n\to\infty} ne^{-(\log n + x)} = e^{-x}.$$

Similarly, $h_2(y) = e^{-y}$, and

$$h_{12}(x, y) \quad = \quad \lim_{n\to\infty} nG(\log n + x, \log n + y)$$

$$
\begin{aligned}
&= \lim_{n\to\infty} n(n^\theta e^{\theta x} + n^\theta e^{\theta y} - 1)^{-1/\theta} \\
&= \lim_{n\to\infty} (e^{\theta x} + e^{\theta y} - \frac{1}{n^\theta})^{-1/\theta} \\
&= (e^{\theta x} + e^{\theta y})^{-1/\theta}.
\end{aligned}
$$

Therefore, the limiting extreme value distribution is

$$
\begin{aligned}
H(x,y) &= \exp\{-h_1(x) - h_2(y) + h_{12}(x,y)\} \\
&= \exp\{-e^{-x} - e^{-y} + (e^{\theta x} + e^{\theta y})^{-1/\theta}\}.
\end{aligned}
$$

The following definition is needed for the next property.

Definition: The p-dimensional unit simplex S is the set of vectors $\mathbf{q}$ with nonnegative components q_i such that $\sum_{i=1}^{p} q_i = 1$.

Property 5 (The Pickands Representation for a min-stable exponential distribution). $G(x_1, \ldots, x_p)$, with univariate exponential margins, is a survival function satisfying

$$
-\log G(tx_1, \ldots, tx_p) = -t \log G(x_1, \ldots, x_p)
$$

for all $t > 0$ if and only if G has the representation

$$
-\log G(x_1, \ldots, x_p) = \int_{S_p} [\max_{1\le i\le p}(q_i x_i)] dU(q_1, \ldots, q_p), \ x_i \ge 0, i = 1, \ldots, p, \quad (2-5-9)
$$

where $S_p = \{(q_1, \ldots, q_p) : q_i \ge 0, i = 1, \ldots, p, \sum_i q_i = 1\}$ is the p-dimensional unit simplex and U is a finite measure on S_p.

Property 5 is a result of Pickands and an alternative statement of Theorem 5.4.5 of Galambos (1987). It is not an easy task to give the representation (2-5-9) for a given $H(\mathbf{x})$, but one usually does not aim at giving another form of $H(\mathbf{x})$ when it is already known. The value of (2-5-9) lies in its possibility of generating functions which are limits in (2-5-1). For applications of Property 5, see Joe (1990a). An example from there is:

$$
\exp\{-A(z_1, \ldots, z_p; \lambda_B, B \subset (1, \ldots, p), \delta)\} = \exp\{-\sum_{B\ne\emptyset} \lambda_B (\sum_{i\in B} x_i^\delta)^{1/\delta}\}, \delta \ge 1. \quad (2-5-10)
$$

For each integer $m \geq 2$, it can be directly shown that

$$G(x_1, \ldots, x_m) = \exp\{-(x_1^\delta + \cdots + x_m^\delta)^{1/\delta}\} \qquad (2-5-11)$$

is a survival function over $x_j \geq 0, j = 1, \ldots, m,$ for $1 \leq \delta \leq \infty$. (2-5-11) satisfies the homogeneity of degree one condition of Property 5. Hence by Property 5 , if $m \leq p$, there is a measure V_m on S_m and a measure U_m on S_p such that

$$\begin{aligned}(x_1^\delta + \cdots + x_m^\delta)^{1/\delta} &= \int_{S_m} [\max_{1 \leq i \leq m} (q_i x_i)] dV_m(q_1, \ldots, q_m) \\ &= \int_{S_p} [\max_{1 \leq i \leq m} (q_i x_i)] dU_m(q_1, \ldots, q_p),\end{aligned}$$

where U_m is a representation of V_m in p dimensions which puts all mass on $q_1, \ldots, q_m$. Therefore,

$$\sum_{B \neq \emptyset} \lambda_B (\sum_{i \in B} x_i^\delta)^{1/\delta}$$

has the representation

$$\int_{S_p} [\max_{1 \leq i \leq p} (q_i x_i)] dU(q_1, \ldots, q_p),$$

where dU has the form $\sum_B \lambda_B dU_B$. By Property 5, (2-5-10) is a distribution.

2.6 Some relation between extreme value theory and Central Limit theory

Since Central Limit theory is more familiar to statisticians, in this section we show some similarities and differences between Central Limit theory and Extreme Value theory.

Consider the univariate case first. Suppose that $X_1, \ldots, X_n$ are iid with distribution F. Let

$$S_n = X_1 + \cdots + X_n,$$

$$Z_n = \max\{X_1, \ldots, X_n\}, \qquad W_n = \min\{X_1, \ldots, X_n\}.$$

We would like to know whether there are any sequences $\{a_n\}$ and $\{b_n\}$ such that

$$\frac{S_n - a_n}{b_n}, \qquad \frac{Z_n - a_n}{b_n}, \qquad \text{or} \quad \frac{W_n - a_n}{b_n}$$

converge in distribution.

From Central Limit theory, if F has a finite second moment, then there exist $\{a_n\}$ and $\{b_n\}$ such that the limiting distribution $(S_n - a_n)/b_n$ has a normal distribution.

Definition. Let $H(x)$ be a nondegenerate distribution function, then $F(x)$ is in the domain of attraction of $H(x)$ if there are sequences $\{a_n\}$ and $\{b_n\} > 0$ such that

$$\lim_{n\to\infty} F^n(a_n + b_n x) = H(x).$$

From extreme value theory, we know that the limiting distribution $H(x)$ of $(Z_n - a_n)/b_n$ belongs to the GEV family if the distribution function $F(x)$ is in the domain of attraction of $H(x)$.

For the multivariate case, assume that $\mathbf{X_1}, \ldots, \mathbf{X_n}$ are iid with distribution F. Let

$$\mathbf{X_i} = (X_{i1}, \ldots, X_{ip}), \qquad \mathbf{S_n} = \mathbf{X_1} + \cdots + \mathbf{X_n},$$

$$\mathbf{Z_n} = (\max(X_{11}, \ldots, X_{n1}), \ldots, \max(X_{1p}, \ldots, X_{np})).$$

Are there any $\{\mathbf{a_n}\}$ and $\{\mathbf{b_n}\} > \mathbf{0}$ such that

$$\frac{\mathbf{S_n} - \mathbf{a_n}}{\mathbf{b_n}} = \left(\frac{S_{n1} - a_{n1}}{b_{n1}}, \ldots, \frac{S_{np} - a_{np}}{b_{np}}\right)$$

or

$$\frac{\mathbf{Z_n} - \mathbf{a_n}}{\mathbf{b_n}}$$

converge in distribution? If the limit of $(\mathbf{S_n} - \mathbf{a_n})/\mathbf{b_n}$ and the second order moments of F exist, then from Central Limit theory the limit of $(\mathbf{S_n} - \mathbf{a_n})/\mathbf{b_n}$ has a multinormal distribution.

If the limit of $(\mathbf{Z_n} - \mathbf{a_n})/\mathbf{b_n}$ exists, from extreme value theory, each univariate margin must be in the GEV family. An analogy between multinormal and multivariate extreme value distribution is:

(1) Suppose $(Z_1, \ldots, Z_p)$ has a multinormal distribution, then linear combinations of univariate normal, i.e. $a_1 Z_1 + \cdots + a_p Z_p$ has a univariate normal distribution.

(2) Suppose $X_j \sim H(\cdot, \gamma_j, \mu_j, \sigma_j)$. After transformation to an exponential survival function

$$Z_j = -\log(H(X_j; \gamma_j, \mu_j, \sigma_j))$$

have an exponential distribution. If $(Z_1, \ldots, Z_p)$ has a min-stable multivariate exponential distribution, then from Property 5,

$$G(z_1, \ldots, z_p) = \exp\{-A(z_1, \ldots, z_p)\} \qquad (2-6-1)$$

and $V = \min(\frac{Z_1}{w_1}, \ldots, \frac{Z_p}{w_p})$ has an exponential distribution $e^{-vA(w_1, \ldots, w_p)}$, where $w_j > 0$. That is, weighted minima are univariate exponential.

Note also the main difference: The multinormal distributions form a finite-dimensional parametric family, the multivariate extreme value distributions form an infinite-dimensional family (as given by Pickand's representation).

Since parametric inference is easier than nonparametric inference, a "goal" is to find good parametric (finite dimensional) subfamilies of the infinite-dimensional family (see the next section).

2.7 Parametric families of multivariate extreme value distributions.

A multivariate extreme value distribution with margins transformed to a survival function with exponential survival functions as univariate margin is a min-stable exponential distribution. Let G be a p-dimensional min-stable exponential distribution, from extreme value theory $A = -\log(G)$ satisfies

$$A(tz_1, \ldots, tz_p) = tA(z_1, \ldots, z_p), \qquad \forall t > 0.$$

Let G_s be the marginal survival function of G, then $G_s(\mathbf{z}_s) = \exp\{-A_s(\mathbf{z}_s)\}$ where A_s is obtained from A by setting $z_j = 0$ for $j \notin s$ and s is a subset of $\{1, \ldots, p\}$.

Here we will list some parametric families of multivariate min-stable exponential distributions in terms of A. These families can be found in Joe (1990a,b), Tawn (1989), Coles and Tawn (1991), Hüsler and Reiss (1989). The interpretations of the parameters are given after all the families are listed.

The families are given here for $p = 2, 3, 4$, from which the general multidimensional form can be seen.

$$A(z_1, z_2; \delta) = (z_1^{\delta} + z_2^{\delta})^{1/\delta}, \qquad \delta \geq 1, \qquad (2-7-1)$$

$$A(z_1, z_2; \tau) = z_1 + z_2 - (z_1^{-\tau} + z_2^{-\tau})^{-1/\tau}, \qquad \tau \geq 0, \qquad (2-7-2)$$

$$A(z_1, z_2, z_3; \delta, \delta_3) = ((z_1^{\delta} + z_2^{\delta})^{\delta_3/\delta} + z_3^{\delta_3})^{1/\delta_3}, \quad \delta \geq \delta_3 \geq 1, \qquad (2-7-3)$$

$$\begin{aligned} &A(z_1, z_2, z_3; \tau, \tau_3) \\ = \quad & z_1 + z_2 + z_3 - (z_1^{-\tau} + z_2^{-\tau})^{-1/\tau} - (z_1^{-\tau_3} + z_3^{-\tau_3})^{-1/\tau_3} - (z_2^{-\tau_3} + z_3^{-\tau_3})^{-1/\tau_3} + \\ & \left((z_1^{-\tau} + z_2^{-\tau})^{\tau_3/\tau} + z_3^{-\tau_3}\right)^{-1/\tau_3}, \quad \tau \geq \tau_3 \geq 0, \qquad (2-7-4) \end{aligned}$$

$$A(z_1, z_2, z_3, z_4; \delta, \delta_3, \delta_4) = ([(z_1^{\delta} + z_2^{\delta})^{\delta_3/\delta} + z_3^{\delta_3}]^{\delta_4/\delta_3} + z_4^{\delta_4})^{1/\delta_4}, \qquad 1 \leq \delta_4 \leq \delta_3 \leq \delta, \qquad (2-7-5)$$

$$A(z_1, z_2, z_3, z_4; \delta, \delta_2, \delta_4) = ((z_1^{\delta} + z_2^{\delta})^{\delta_4/\delta} + (z_3^{\delta_2} + z_4^{\delta_2})^{\delta_4/\delta_2})^{1/\delta_4}, \qquad 1 \leq \delta_2, \delta_4 \leq \delta, \qquad (2-7-6)$$

$A(z_1, z_2, z_3, z_4; \tau, \tau_3, \tau_4)$

$$
\begin{aligned}
= \quad & z_1 + z_2 + z_3 + z_4 - (z_1^{-\tau} + z_2^{-\tau})^{-1/\tau} - (z_1^{-\tau_3} + z_3^{-\tau_3})^{-1/\tau_3} - (z_2^{-\tau_3} + z_3^{-\tau_3})^{-1/\tau_3} - \\
& (z_1^{-\tau_4} + z_4^{-\tau_4})^{-1/\tau_4} - (z_2^{-\tau_4} + z_4^{-\tau_4})^{-1/\tau_4} - (z_3^{-\tau_4} + z_4^{-\tau_4})^{-1/\tau_4} + \\
& \left((z_1^{-\tau} + z_2^{-\tau})^{\tau_3/\tau} + z_3^{-\tau_3}\right)^{-1/\tau_3} + \left((z_1^{-\tau} + z_2^{-\tau})^{\tau_4/\tau} + z_4^{-\tau_4}\right)^{-1/\tau_4} + \\
& \left((z_1^{-\tau_3} + z_3^{-\tau_3})^{\tau_4/\tau_3} + z_4^{-\tau_4}\right)^{-1/\tau_4} + \left((z_2^{-\tau_3} + z_3^{-\tau_3})^{\tau_4/\tau_3} + z_4^{-\tau_4}\right)^{-1/\tau_4} - \\
& \left(((z_1^{-\tau} + z_2^{-\tau})^{\tau_3/\tau} + z_3^{-\tau_3})^{\tau_4/\tau_3} + z_4^{-\tau_4}\right)^{-1/\tau_4}, \quad \tau \geq \tau_2 \geq \tau_4 \geq 0. \quad (2-7-7)
\end{aligned}
$$

$$
\begin{aligned}
& A(z_1, z_2, z_3, z_4; \tau, \tau_2, \tau_4) \\
= \quad & z_1 + z_2 + z_3 + z_4 - (z_1^{-\tau} + z_2^{-\tau})^{-1/\tau} - (z_1^{-\tau_4} + z_3^{-\tau_4})^{-1/\tau_4} - (z_2^{-\tau_4} + z_3^{-\tau_4})^{-1/\tau_4} - \\
& (z_1^{-\tau_4} + z_4^{-\tau_4})^{-1/\tau_4} - (z_2^{-\tau_4} + z_4^{-\tau_4})^{-1/\tau_4} - (z_3^{-\tau_2} + z_4^{-\tau_2})^{-1/\tau_2} + \\
& \left((z_1^{-\tau} + z_2^{-\tau})^{\tau_4/\tau} + z_3^{-\tau_4}\right)^{-1/\tau_4} + \left((z_1^{-\tau} + z_2^{-\tau})^{\tau_4/\tau} + z_4^{-\tau_4}\right)^{-1/\tau_4} + \\
& \left(z_1^{-\tau_4} + (z_3^{-\tau_2} + z_4^{-\tau_2})^{\tau_4/\tau_2}\right)^{-1/\tau_4} + \left(z_2^{-\tau_4} + (z_3^{-\tau_2} + z_4^{-\tau_2})^{\tau_4/\tau_2}\right)^{-1/\tau_4} - \\
& \left((z_1^{-\tau} + z_2^{-\tau})^{\tau_4/\tau} + (z_3^{-\tau_2} + z_4^{-\tau_2})^{\tau_4/\tau_2}\right)^{-1/\tau_4}, \quad 1 \leq \tau_2, \tau_4 \leq \tau. \quad (2-7-8)
\end{aligned}
$$

Finally, the Hüsler and Reiss(1989) model with $p = 2$ is

$$
A(z_1, z_2; \lambda) = \Phi(\lambda + \frac{\log z_1 - \log z_2}{2\lambda})z_2 + \Phi(\lambda + \frac{\log z_2 - \log z_1}{2\lambda})z_1, \quad \lambda \geq 0. \quad (2-7-9)
$$

For $p \geq 2$, let

$$
\lambda_{i,j} \in (0, \infty) \quad \forall 1 \leq i, j \leq p \quad \text{with} \quad i \neq j.
$$

Put

$$
\Lambda = (\lambda_{i,j})_{i,j \leq p} \qquad \lambda_{i,i} = 0 \qquad \forall i.
$$

Moreover , for $2 \leq k \leq p$ and $\mathbf{m} = (m_1, \ldots, m_k)$ with $1 \leq m_1 < m_2 < \ldots < m_k \leq p$ define

$$
\Gamma_{k,\mathbf{m}} = \left[2(\lambda^2_{m_i,m_k} + \lambda^2_{m_j,m_k} - \lambda^2_{m_i,m_j})\right]_{i,j \leq k-1}. \qquad (2-7-10)
$$

Furthermore , let $S(\cdot|\Gamma)$ denote the survivor function of a normal random vector with mean vector $\mathbf{0}$ and covariance matrix Γ. The extension of (2-7-4) is

$$A(z_1,\ldots,z_p;\lambda_{ij},1\le i<j\le p)=\sum_{k=1}^{p}(-1)^k\sum^{k} h_{k,\mathbf{m}}(-\log z_{m_1},\ldots,-\log z_{m_k}) \qquad (2-7-11)$$

with

$$h_{k,\mathbf{m}}(\mathbf{y})=\int_{y_k}^{\infty} S\left[(y_i-z+2\lambda^2_{m_i,m_k})_{i=1}^{k-1}|\Gamma_{k,\mathbf{m}}\right]e^{-z}dz,$$

for $2\le k\le p$, where $\sum^k$ means summation over all p-vectors $\mathbf{m}=(m_1,\ldots,m_k)$ with $1\le m_1<m_2<\ldots<m_k\le p$, and $h_{1,m}(y)=e^{-y}$ for $m=1,\ldots,p$. The cases $p=3,4$ are given explicitly in Chapter 3.

For comparison, note that the bivariate models (2-7-1), (2-7-2) and (2-7-9) each have a single dependence parameter. For (2-7-1) and (2-7-2) dependence increases as the parameter increases; for (2-7-9) dependence decreases as the parameter increases. The trivariate models (2-7-3) and (2-7-4) have respectively models (2-7-1) and (2-7-2) for each of the three bivariate margins. For the trivariate model (2-7-3) (respectively (2-7-4)), δ (respectively τ) is the dependence parameter for the (1,2) bivariate margin; δ_3 (respectively, τ_3) is the dependence parameter for both the (1,3) and (2,3) bivariate margins. We have something similar for the 4-variate models (2-7-5) to (2-7-8). The 4-variate models (2-7-5) to (2-7-8) also have respectively models (2-7-1) and (2-7-2) for each of the six bivariate margins. For (2-7-5), δ is the dependence parameters for the (1,2) bivariate margin; δ_3 is the dependence parameter for the (1,3) and (2,3) bivariate margins; δ_4 is the dependence parameter for the (1,4), (2,4) and (3,4) margins. For (2-7-6), δ is the dependence parameter for the (1,2) bivariate margin; δ_2 is the dependence parameter for the (3,4) bivariate margin; δ_4 is the dependence parameter for the (1,3), (1,4), (2,3) and (2,4) margins. The model (2-7-11) has a dependence parameter for each of the $p(p-1)/2$ bivariate margins. For (2-7-11), the dependence parameter of the (i,j)

bivariate margin, with $i < j$, is λ_{ij}, and the (i, j) margin has form (2-7-9).

Models (2-7-3) to (2-7-8) and their extensions have $p - 1$ parameters in total; as above, some of the bivariate margins have the same dependence parameter. Hence these models do not have as much flexibilty in the dependence pattern but they have a simpler form than (2-7-11).

2.8 Estimation for a parametric family of multivariate extreme value distribution.

Joe (1990b) suggests the following procedures for fitting of a parametric multivariate distribution to iid $\mathbf{X_1}, \ldots, \mathbf{X_n}$.

(1) Fit the p univariate margins separately by maximum likelihood using the GEV family.

(2) Transform each margin so that each transformed variable has exponential survival function $G_i(z_i) = e^{-z_i}$.

(3) Suppose that the transformed data are iid p-vectors, then compare fit of different families of min-stable multivariate exponential distributions. Check for parameter estimation consistency with bivariate and higher order margins.

(4) Go back to the original data and estimate both multivariate parameters and parameters of the univariate margins simultaneously, using maximum likelihood if a copula has been decided on as being acceptable.

In (3), to check the fit of a model, we compare parametric estimates of A with a nonparametric estimate of A. Besides the consistency checking from we know that if $(Z_1, \ldots, Z_p)$ is a random vector with the min-stable exponential survival function $G = e^{-A}$, then

$$E(\min[\frac{Z_1}{w_1}, \ldots, \frac{Z_p}{w_p}]) = [A(w_1, \ldots, w_p)]^{-1}$$

for $w_j > 0, j = 1, \ldots, p$. So if $\mathbf{Y_1}, \ldots, \mathbf{Y_p}$ are a transformed random sample, exponential probability plots of $V_i = \min_j Y_{ij}/w_j$ can be used to check the assumption of min-stability for multivariate exponential distribution. If the plots are adequate, a non-parametric estimate of $A(w_1, \ldots, w_p)$ is given by $n / \sum_{i=1}^{n} V_i$ which can be compared with parametric estimates of A from different models.

Chapter 3 Maximum Likelihood Estimation With The Hüsler-Reiss Model

Hüsler and Reiss (1989) obtained a new class of multivariate extreme value distributions by taking a non-standard extreme-value limit.

Let $\mathbf{X}$ and $\mathbf{M}_n$ be as defined in Chapter 2. For a mutivariate normal random vector, having all correlation coefficients smaller than one, Sibuya (1960) proved that the marginal maxima $\mathbf{M}_{n,1}, \ldots, \mathbf{M}_{n,p}$ are asymptotically independent. The rate of convergence to asymptotic independence is slower as the correlation coefficients increase. Since in practice the sample size n never goes to infinity but is a fixed positve integer, is there another asymptotic formulation which may provide a better approximation? This motivated the work by Hüsler and Reiss.

For the bivariate case, Hüsler and Reiss let the correlation coefficient $\rho = \rho_n$ increase towards 1 as the sample size increases. It is shown that the marginal maxima are neither asymptotically independent nor completely dependent if $(1 - \rho_n)\log n$ converges to a positive constant as $n \to \infty$. This was extended to dimensions $p > 2$.

Hüsler and Reiss define their limiting distribution by the c.d.f. or survival function. To get a maximum likelihood estimation, the density is needed. To obtain the density, the following result for the conditional distribution of a multidimensional normal distribution is needed.

Result 1:

Let $\mathbf{U} = (U_1, \ldots, U_a)$ and $\mathbf{V} = (V_1, \ldots, V_b)$ be random vectors such that

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim MVN\left(\mathbf{0}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

where $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}$ are $a \times a, a \times b, b \times a, b \times b$ matrices respectively, and suppose Σ_{11} and Σ_{22} are nonsingular. Then

$$\mathbf{U}|\mathbf{V} = \mathbf{v} \sim MVN\left(\Sigma_{12}\Sigma_{22}^{-1}\mathbf{v}; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right).$$

Let $F_{\mathbf{X}}$ and $\overline{F}_{\mathbf{X}}$ denote the distribution function and survival function for a random vector $\mathbf{X}$, and let $f_{\mathbf{X}}$ and $f_{\mathbf{X}|\mathbf{Y}}$ denote the density function of $\mathbf{X}$ and the conditional density function of $\mathbf{X}$ when $\mathbf{Y}$ is fixed. Then the joint distribution of $\mathbf{X}$ and $\mathbf{Y}$ is

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})f_{\mathbf{Y}}(\mathbf{y}).$$

Let $\phi(\mathbf{x};\Sigma)$, $\overline{\Phi}(\mathbf{x};\Sigma)$ denote the multivariate normal density and survival function with zero mean vector and covariance matrix Σ. Then

$$f_{\mathbf{U},\mathbf{V}}(\mathbf{u},\mathbf{v}) = \phi(\mathbf{u}-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{v};\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\phi(\mathbf{v};\Sigma_{22})$$

and

$$\begin{aligned}
\overline{F}_{\mathbf{U},\mathbf{V}}(\mathbf{u},\mathbf{v}) &= \overline{\Phi}((\mathbf{u},\mathbf{v});\Sigma) \\
&= \int_{v_1}^{\infty}\cdots\int_{v_b}^{\infty}\int_{u_1}^{\infty}\cdots\int_{u_a}^{\infty}\phi(\mathbf{s}-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{t};\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\phi(\mathbf{t};\Sigma_{22})d\mathbf{s}d\mathbf{t} \\
&= \int_{v_1}^{\infty}\cdots\int_{v_b}^{\infty}\overline{\Phi}(\mathbf{u}-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{t};\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\phi(\mathbf{t};\Sigma_{22})d\mathbf{t}.
\end{aligned}$$

Therefore,

$$\frac{\partial^b\overline{\Phi}((\mathbf{u},\mathbf{v});\Sigma)}{\partial v_1\cdots\partial v_b} = (-1)^b\Phi(\mathbf{u}-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{v};\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\phi(\mathbf{v};\Sigma_{22}).$$

Now we are ready to derive the density for the Hüsler and Reiss survival function when the univariate margins exponential have been transformed to distributions with mean 1.

We provide details for dimensions $p = 2,3,4$ from which the general case will be apparent. The notation gets increasingly difficult as p increases.

3.1 Bivariate Case

Frome Section 2.7,

$$G(w_1,w_2,\lambda) = \exp\{-A_2(w_1,w_2,\lambda)\},$$

where

$$A_2(w_1, w_2, \lambda) = \Phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2))w_2 + \Phi(\lambda + \frac{1}{2\lambda}\log(w_2/w_1))w_1, \qquad \lambda \geq 0.$$

Since $\frac{\partial G}{\partial w_1} = -\frac{\partial A_2}{\partial w_1}G$ and $\frac{\partial G}{\partial w_2} = -\frac{\partial A_2}{\partial w_2}G$, we obtained the density function of $G(w_1, w_2, \lambda)$:

$$\frac{\partial^2 G}{\partial w_1 \partial w_2} = \frac{\partial}{\partial w_1}(\frac{\partial G}{\partial w_2}) = \frac{\partial}{\partial w_1}(-\frac{\partial A_2}{\partial w_2}G) = G(\frac{\partial A_2}{\partial w_1}\frac{\partial A_2}{\partial w_2}G - \frac{\partial^2 A_2}{\partial w_1 \partial w_2}).$$

From $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ we have $\phi'(x) = x\phi(x)$. Also

$$\begin{aligned}
&-\log[\phi(\lambda + \frac{1}{2\lambda}\log(w_2/w_1))/w_1] \\
&= \lambda^2/2 + (\log w_2 - \log w_1)/2 + \frac{1}{8\lambda^2}((\log w_2 - \log w_1)^2 + \log w_1 + \log\sqrt{2\pi} \\
&= \lambda^2/2 + (\log w_1 - \log w_2)/2 + \frac{1}{8\lambda^2}((\log w_1 - \log w_2)^2 + \log w_2 + \log\sqrt{2\pi} \\
&= -\log[\phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2))/w_2].
\end{aligned}$$

Hence

$$\phi(\lambda + \frac{1}{2\lambda}\log(w_2/w_1))/w_1 = \phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2))/w_2.$$

Therefore

$$\begin{aligned}
A_{1,2} &:= \frac{\partial A_2}{\partial w_1} \\
&= \frac{\partial}{\partial w_1}(\Phi(\lambda + \frac{1}{2\lambda}\log(w_2/w_1))w_2 + \Phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2))w_1) \\
&= -\frac{1}{2\lambda}\phi(\lambda + \frac{1}{2\lambda}\log(w_2/w_1))(w_2/w_1) + \frac{1}{2\lambda}\phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2))(w_1/w_2) \\
&\quad + \Phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2)) \\
&= \Phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2)).
\end{aligned}$$

Similarly,

$$A_{2,1} = \frac{\partial A_2}{\partial w_2} = \Phi(\lambda + \frac{1}{2\lambda}\log(w_2/w_1)).$$

Finally,

$$A_{12} := \frac{\partial^2 A_2}{\partial w_1 \partial w_2} = \frac{\partial}{\partial w_2}(\frac{\partial A_2}{\partial w_1})$$

$$= \frac{\partial}{\partial w_2}(\Phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2))$$
$$= \frac{-1}{2\lambda w_2}\phi(\lambda + \frac{1}{2\lambda}\log(w_1/w_2)).$$

From $A_{1,2}$, $A_{2,1}$ and A_{12} we can get the density function $g(w_1, w_2, \lambda)$.

3.2 Trivariate Case

Let

$$\Sigma = 2\begin{pmatrix} 2\lambda_{13}^2 & \lambda_{13}^2 + \lambda_{23}^2 - \lambda_{12}^2 \\ \lambda_{13}^2 + \lambda_{23}^2 - \lambda_{12}^2 & 2\lambda_{23}^2 \end{pmatrix}$$

be the matrix in (2-7-10) and let

$$\rho_{12} = \rho_{21} = \frac{\lambda_{13}^2 + \lambda_{23}^2 - \lambda_{12}^2}{2\lambda_{13}\lambda_{23}},$$

$$\rho_{13} = \rho_{31} = \frac{\lambda_{12}^2 + \lambda_{23}^2 - \lambda_{23}^2}{2\lambda_{12}\lambda_{23}},$$

$$\rho_{23} = \rho_{32} = \frac{\lambda_{12}^2 + \lambda_{13}^2 - \lambda_{23}^2}{2\lambda_{12}\lambda_{13}},$$

i.e.,

$$\rho_{ij} = \frac{\lambda_{ik}^2 + \lambda_{jk}^2 - \lambda_{ij}^2}{2\lambda_{ik}\lambda_{jk}}$$

for i, j, k distinct.

Then $h_{3,123}$ in (2-7-11) can be written as

$$C(w_1, w_2, w_3, \lambda_{12}, \lambda_{13}, \lambda_{23}) = \int_0^{w_3} \overline{\Phi}(\lambda_{13}+\frac{1}{2\lambda_{13}}(\log q - \log w_1), \lambda_{23}+\frac{1}{2\lambda_{23}}(\log q - \log w_2); \rho_{12})dq$$

where with some abuse of notation

$$\overline{\Phi}(\cdot;\rho) = \overline{\Phi}(\cdot; \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}).$$

From Section 2.7, we have

$$G(w_1, w_2, w_3) = \exp\{-A_3(w_1, w_2, w_3, \lambda_{12}, \lambda_{13}, \lambda_{23})\}$$

where with simplification of (2-7-11),

$$
\begin{aligned}
A_3(w_1, w_2, w_3, \lambda_{12}, \lambda_{13}, \lambda_{23}) \quad &= \quad -w_1 - w_2 - w_3 + A_2(w_1, w_2, \lambda_{12}) + \\
& \qquad A_2(w_1, w_3, \lambda_{13}) + A_2(w_2, w_3, \lambda_{23}) + C(w_1, w_2, w_3).
\end{aligned}
$$

Constraints are $0 \leq \lambda_{ij} \leq \infty$ and $0 \leq \rho_{ij} \leq 1$.

By symmetry, equivalent forms for C can be obtained by permutations of subscript indices:

$$
\begin{aligned}
& C(w_1, w_2, w_3, \lambda_{12}, \lambda_{13}, \lambda_{23}) \\
& = \int_0^{w_1} \overline{\Phi}(\lambda_{13} + \frac{1}{2\lambda_{13}}(\log q - \log w_3), \lambda_{12} + \frac{1}{2\lambda_{12}}(\log q - \log w_2); \rho_{23}) dq \\
& = \int_0^{w_2} \overline{\Phi}(\lambda_{12} + \frac{1}{2\lambda_{12}}(\log q - \log w_1), \lambda_{23} + \frac{1}{2\lambda_{23}}(\log q - \log w_3); \rho_{13}) dq.
\end{aligned}
$$

The density function for $G(w_1, w_2, w_3)$ is:

$$
\begin{aligned}
& g(w_1, w_2, w_3) = (-1)^3 \frac{\partial^3}{\partial w_1 \partial w_2 \partial w_3} \exp\{-A_3(w_1, w_2, w_3, \lambda_{12}, \lambda_{13}, \lambda_{23})\} \\
& = G(\frac{\partial A_3}{\partial w_1}\frac{\partial A_3}{\partial w_2}\frac{\partial A_3}{\partial w_2} - \frac{\partial A_3}{\partial w_1}\frac{\partial^2 A_3}{\partial w_2 \partial w_3} - \frac{\partial A_3}{\partial w_2}\frac{\partial^2 A_3}{\partial w_1 \partial w_3} - \frac{\partial A_3}{\partial w_3}\frac{\partial^2 A_3}{\partial w_1 \partial w_2} + \frac{\partial^3 A_3}{\partial w_1 \partial w_2 \partial w_3}).
\end{aligned}
$$

From

$$
\phi(x_1, x_2; \rho) = \phi(x_1)\phi(\frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}})\frac{1}{\sqrt{1-\rho^2}},
$$

$$
\begin{aligned}
\overline{\Phi}(x_1, x_2; \rho) \quad &= \quad \int_{x_1}^{\infty}\int_{x_2}^{\infty} \phi(\frac{z_2 - \rho z_1}{\sqrt{1-\rho^2}})\frac{1}{\sqrt{1-\rho^2}} dz_2 \phi(z_1) dz_1 \\
&= \quad \int_{x_1}^{\infty} \overline{\Phi}(\frac{x_2 - \rho z_1}{\sqrt{1-\rho^2}})\phi(z_1) dz_1
\end{aligned}
$$

and

$$
\frac{\partial \overline{\Phi}}{\partial x_1}(x_1, x_2; \rho) = \overline{\Phi}(\frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}})\phi(x_1).
$$

Let

$$
C_{1,23} := \frac{\partial C}{\partial w_1} = \overline{\Phi}(\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_1/w_3), \lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2); \rho_{23}).
$$

By symmetry, we have $C_{2,13} := \frac{\partial C}{\partial w_2}$ and $C_{3,12} := \frac{\partial C}{\partial w_3}$. Similarly,

$$\begin{aligned} C_{12,3} := \frac{\partial^2 C}{\partial w_1 \partial w_2} = \\ \overline{\Phi}[((\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_1/w_3)) - \rho_{23}(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2)))/\sqrt{1-\rho_{23}^2}] \times \\ \phi(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2))\frac{1}{2\lambda_{12}w_2} \end{aligned}$$

By symmetry, we also have $C_{13,2} := \frac{\partial^2 C}{\partial w_1 \partial w_3}$ and $C_{23,1} := \frac{\partial^2 C}{\partial w_2 \partial w_3}$. Finally,

$$\begin{aligned} C_{123} &:= \frac{\partial^3 C}{\partial w_1 \partial w_2 \partial w_3} \\ &= \phi(\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_3/w_1), \lambda_{23} + \frac{1}{2\lambda_{23}}\log(w_3/w_2); \rho_{12})\frac{1}{4\lambda_{13}\lambda_{23}w_1w_2} \\ &= \phi(\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_3/w_1))\phi[((\lambda_{23} + \frac{1}{2\lambda_{23}}\log(w_2/w_3)) - \\ &\quad \rho_{12}(\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_1/w_3)))/\sqrt{1-\rho_{12}^2}] \times \frac{1}{4\lambda_{13}\lambda_{23}w_1w_2\sqrt{1-\rho_{12}^2}}. \end{aligned}$$

Combining $C_{1,23}$, $C_{2,13}$, $C_{3,12}$, $C_{12,3}$, $C_{13,2}$, $C_{23,1}$, C_{123},and $\frac{\partial A_2}{\partial w_1}$, $\frac{\partial A_2}{\partial w_2}, \frac{\partial A_2}{\partial w_3}, \frac{\partial^2 A_2}{\partial w_1 \partial w_2}$, $\frac{\partial^2 A_2}{\partial w_1 \partial w_3}$, $\frac{\partial^2 A_2}{\partial w_2 \partial w_3}$, and $\frac{\partial^3 A_2}{\partial w_1 \partial w_2 \partial w_3}$ we have an expression for the density function $g(w_1, w_2, w_3)$.

3.3 Four-Dimensional Case

Let

$$\Sigma = 2\begin{pmatrix} 2\lambda_{14}^2 & \lambda_{14}^2 + \lambda_{24}^2 - \lambda_{12}^2 & \lambda_{14}^2 + \lambda_{34}^2 - \lambda_{13}^2 \\ \lambda_{14}^2 + \lambda_{24}^2 - \lambda_{12}^2 & 2\lambda_{24}^2 & \lambda_{24}^2 + \lambda_{34}^2 - \lambda_{23}^2 \\ \lambda_{14}^2 + \lambda_{34}^2 - \lambda_{13}^2 & \lambda_{24}^2 + \lambda_{34}^2 - \lambda_{23}^2 & 2\lambda_{34}^2 \end{pmatrix}$$

be the matrix in (2-7-10) and let

$$\rho_{ijk} = \frac{\lambda_{ij}^2 + \lambda_{ik}^2 - \lambda_{jk}^2}{2\lambda_{ij}^2\lambda_{ik}^2},$$

where $\lambda_{ij} = \lambda_{ji}$ and, i, j, k are distinct and between 1 and 4. Then $h_{4,1234}$ in (2-7-11) can be written as:

$$
\begin{aligned}
D(w_1, w_2, w_3, w_4) = \\
\int_0^{w_4} \overline{\Phi}(\lambda_{14} + \frac{1}{2\lambda_{14}}\log(q/w_4), \lambda_{24} + \frac{1}{2\lambda_{24}}\log(q/w_4), \\
\lambda_{34} + \frac{1}{2\lambda_{34}}\log(q/w_4), R(\rho_{412}, \rho_{413}, \rho_{423}))dq,
\end{aligned}
$$

where

$$
R(a, b, c) = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}.
$$

By symmetry, equivalent forms for D are

$$
\begin{aligned}
& D(w_1, w_2, w_3, w_4) \\
= & \int_0^{w_1} \overline{\Phi}(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(q/w_2), \lambda_{13} + \frac{1}{2\lambda_{13}}\log(q/w_3), \\
& \lambda_{14} + \frac{1}{2\lambda_{14}}\log(q/w_4), R(\rho_{123}, \rho_{124}, \rho_{134}))dq \\
= & \int_0^{w_2} \overline{\Phi}(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(q/w_1), \lambda_{23} + \frac{1}{2\lambda_{23}}\log(q/w_3), \\
& \lambda_{24} + \frac{1}{2\lambda_{24}}\log(q/w_4), R(\rho_{213}, \rho_{214}, \rho_{234}))dq \\
= & \int_0^{w_3} \overline{\Phi}(\lambda_{13} + \frac{1}{2\lambda_{13}}\log(q/w_1), \lambda_{23} + \frac{1}{2\lambda_{23}}\log(q/w_2), \\
& \lambda_{34} + \frac{1}{2\lambda_{34}}\log(q/w_4), R(\rho_{312}, \rho_{314}, \rho_{324}))dq.
\end{aligned}
$$

Constraints are:

$$
0 \le \lambda_{ij} \le \infty, \qquad i \ne j, 0 \le \rho_{ijk} \le 1,
$$

i, j, k are distinct and

$$
R(\rho_{412}, \rho_{413}, \rho_{423}), R(\rho_{123}, \rho_{124}, \rho_{134}), R(\rho_{213}, \rho_{214}, \rho_{234}), R(\rho_{312}, \rho_{314}, \rho_{324})
$$

are positive definite.

Taking derivatives, we have

$$D_{1,234} := \frac{\partial D}{\partial w_1} = \overline{\Phi}(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2),$$
$$\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_1/w_3), \lambda_{14} + \frac{1}{2\lambda_{14}}\log(w_1/w_4), R(\rho_{123}, \rho_{124}, \rho_{134}))$$

By interchange of 1 and 2, 1 and 3, 1 and 4 in $D_{1,234}$ we can obtain $D_{2,134} := \frac{\partial D}{\partial w_2}$, $D_{3,124} := \frac{\partial D}{\partial w_3}$, $D_{4,123} := \frac{\partial D}{\partial w_4}$. By making use of Result 1 with $\Sigma_{11} = \begin{pmatrix} 1 & \rho_{134} \\ \rho_{134} & 1 \end{pmatrix}$, $\Sigma_{12} = \begin{pmatrix} \rho_{123} \\ \rho_{124} \end{pmatrix}$, $\Sigma_{22} = 1$ we have

$$D_{12,34} := \frac{\partial^2 D}{\partial w_1 \partial w_2} = \frac{\partial D_{1,234}}{\partial w_2} =$$
$$\overline{\Phi}(\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_1/w_3) - \rho_{123}(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2)), \lambda_{14} + \frac{1}{2\lambda_{14}}\log(w_1/w_4) -$$
$$\rho_{124}(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2)), R_{34,12})\phi(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2))\frac{1}{2\lambda_{12}w_2},$$

where

$$\begin{aligned} R_{34,12} &= \begin{pmatrix} 1 & \rho_{134} \\ \rho_{134} & 1 \end{pmatrix} - \begin{pmatrix} \rho_{123} \\ \rho_{124} \end{pmatrix} \begin{pmatrix} \rho_{123} & \rho_{124} \end{pmatrix} \\ &= \begin{pmatrix} 1 - \rho_{123}^2 & \rho_{134} - \rho_{123}\rho_{124} \\ \rho_{134} - \rho_{123}\rho_{124} & 1 - \rho_{124}^2 \end{pmatrix}. \end{aligned}$$

By permutation of indices we can easily get

$D_{13,24} := \frac{\partial^2 D}{\partial w_1 \partial w_3}$, $D_{14,23} := \frac{\partial^2 D}{\partial w_1 \partial w_4}$, $D_{23,14} := \frac{\partial^2 D}{\partial w_2 \partial w_3}$, $D_{24,13} := \frac{\partial^2 D}{\partial w_2 \partial w_4}$, $D_{34,12} := \frac{\partial^2 D}{\partial w_3 \partial w_4}$.

Again using Result 1 with $\Sigma_{11} = 1$,

$$\Sigma_{12} = \begin{pmatrix} \rho_{123} \\ \rho_{124} \end{pmatrix}, \Sigma_{22} = \begin{pmatrix} 1 & \rho_{134} \\ \rho_{134} & 1 \end{pmatrix}$$

then

$$\Sigma_{22}^{-1} = \frac{1}{1-\rho_{134}^2}\begin{pmatrix} 1 & -\rho_{134} \\ -\rho_{134} & 1 \end{pmatrix},$$

$$R_{123,4} = 1 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 1 - (\rho_{123}^2 + \rho_{124}^2 - 2\rho_{123}\rho_{124}\rho_{134})/(1-\rho_{134}^2)$$

and

$$D_{123,4} := \frac{\partial^3 D}{\partial w_1 \partial w_2 \partial w_3} =$$
$$\overline{\Phi}\{(\lambda_{14} + \frac{1}{2\lambda_{14}}\log(w_1/w_4) - [(\rho_{123} - \rho_{124}\rho_{134})(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2)) +$$
$$(\rho_{124} - \rho_{123}\rho_{134})(\lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_1/w_3))]/(1-\rho_{134}^2))/\sqrt{R_{123,4}}\}$$
$$\phi(\lambda_{12} + \frac{1}{2\lambda_{12}}\log(w_1/w_2), \lambda_{13} + \frac{1}{2\lambda_{13}}\log(w_1/w_3); \rho_{123})\frac{1}{4\lambda_{12}\lambda_{13}w_2w_3}.$$

By symmetry and permution of indices we can easily get $D_{124,3} := \frac{\partial^3 D}{\partial w_1 \partial w_2 \partial w_4}$, $D_{134,2} := \frac{\partial^3 D}{\partial w_1 \partial w_3 \partial w_4}$, $D_{234,1} := \frac{\partial^3 D}{\partial w_2 \partial w_3 \partial w_4}$. Finally,

$$D_{1234} := \frac{\partial^4 D}{\partial w_1 \partial w_2 \partial w_3 \partial w_4} =$$
$$\phi(\lambda_{14} + \frac{1}{2\lambda_{14}}\log(w_1/w_4), \lambda_{24} + \frac{1}{2\lambda_{24}}\log(w_2/w_4), \lambda_{34} + \frac{1}{2\lambda_{34}}\log(w_3/w_4),)$$
$$R(\rho_{412}, \rho_{413}, \rho_{423}))\frac{1}{8\lambda_{14}\lambda_{24}\lambda_{34}w_1w_2w_3}$$

From Sections 2.7 we have

$$G(w_1, w_2, w_3, w_4) = \exp[-A_4(w_1, w_2, w_3, w_4, \lambda_{12}, \lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{23}, \lambda_{24}, \lambda_{34})],$$

where

$$A_4(w_1, w_2, w_3, w_4, \lambda_{12}, \lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{23}, \lambda_{24}, \lambda_{34}) =$$
$$h_1(w_1) + h_1(w_2) + h_1(w_3) + h_1(w_4) - h_2(w_1, w_2) - h_2(w_1, w_3) - h_2(w_1, w_4) -$$
$$h_2(w_2, w_3) - h_2(w_2, w_4) - h_2(w_3, w_4) + h_3(w_1, w_2, w_3) + h_3(w_1, w_2, w_4) +$$
$$h_3(w_1, w_3, w_4) + h_3(w_2, w_3, w_4) + h_4(w_1, w_2, w_3, w_4)$$

$$
\begin{aligned}
&= w_1 + w_2 + w_3 + w_4 - [w_1 + w_2 - A_2(w_1, w_2, \lambda_{12})] - \\
&\quad [w_1, w_3 - A_2(w_1, w_3, \lambda_{13})] - [w_1 + w_4 - A_2(w_1, w_4, \lambda_{14})] - \\
&\quad [w_2 + w_3 - A_2(w_2, w_3, \lambda_{23})] - [w_2 + w_4 - A_2(w_2, w_4, \lambda_{24})] - \\
&\quad [w_3 + w_4 - A_2(w_3, w_4, \lambda_{34})] + C(w_1, w_2, w_3, \lambda_{12}, \lambda_{13}, \lambda_{23}) + C(w_1, w_2, w_4, \lambda_{12}, \lambda_{14}, \lambda_{24}) + \\
&\quad C(w_1, w_3, w_4, \lambda_{13}, \lambda_{14}, \lambda_{34}) + C(w_2, w_3, w_4, \lambda_{23}, \lambda_{24}, \lambda_{34}) - D(w_1, w_2, w_3, w_4) \\
&= -2(w_1 + w_2 + w_3 + w_4) + A_2(w_1, w_2, \lambda_{12})] + A_2(w_1, w_3, \lambda_{13})] + \\
&\quad A_2(w_1, w_4, \lambda_{14})] + A_2(w_2, w_3, \lambda_{23})] + A_2(w_2, w_4, \lambda_{24})] + \\
&\quad A_2(w_3, w_4, \lambda_{34})] + C(w_1, w_2, w_3, \lambda_{12}, \lambda_{13}, \lambda_{23}) + C(w_1, w_2, w_4, \lambda_{12}, \lambda_{14}, \lambda_{24}) + \\
&\quad C(w_1, w_3, w_4, \lambda_{13}, \lambda_{14}, \lambda_{34}) + C(w_2, w_3, w_4, \lambda_{23}, \lambda_{24}, \lambda_{34}) - D(w_1, w_2, w_3, w_4)
\end{aligned}
$$

and

$$
\begin{aligned}
g(w_1, w_2, w_3, w_4) &= (-1)^4 \frac{\partial^4}{\partial w_1 \partial w_2 \partial w_3 \partial w_4} G(w_1, w_2, w_3, w_4) \\
&= G\Big(\frac{\partial A_4}{\partial w_1}\frac{\partial A_4}{\partial w_2}\frac{\partial A_4}{\partial w_3}\frac{\partial A_4}{\partial w_4} - \frac{\partial^2 A_4}{\partial w_1 \partial w_2}\frac{\partial A_4}{\partial w_3}\frac{\partial A_4}{\partial w_4} \\
&\quad - \frac{\partial^2 A_4}{\partial w_1 \partial w_3}\frac{\partial A_4}{\partial w_2}\frac{\partial A_4}{\partial w_4} - \frac{\partial^2 A_4}{\partial w_1 \partial w_4}\frac{\partial A_4}{\partial w_2}\frac{\partial A_4}{\partial w_3} - \frac{\partial^2 A_4}{\partial w_2 \partial w_3}\frac{\partial A_4}{\partial w_1}\frac{\partial A_4}{\partial w_4} \\
&\quad - \frac{\partial^2 A_4}{\partial w_2 \partial w_4}\frac{\partial A_4}{\partial w_1}\frac{\partial A_4}{\partial w_3} - \frac{\partial^2 A_4}{\partial w_3 \partial w_4}\frac{\partial A_4}{\partial w_1}\frac{\partial A_4}{\partial w_2} + \frac{\partial^2 A_4}{\partial w_1 \partial w_2}\frac{\partial^2 A_4}{\partial w_3 \partial w_4} + \\
&\quad \frac{\partial^2 A_4}{\partial w_1 \partial w_3}\frac{\partial^2 A_4}{\partial w_2 \partial w_4} + \frac{\partial^2 A_4}{\partial w_1 \partial w_4}\frac{\partial^2 A_4}{\partial w_2 \partial w_3} + \frac{\partial^3 A_4}{\partial w_1 \partial w_2 \partial w_3}\frac{\partial A_4}{\partial w_4} + \\
&\quad \frac{\partial^3 A_4}{\partial w_1 \partial w_2 \partial w_4}\frac{\partial A_4}{\partial w_3} + \frac{\partial^3 A_4}{\partial w_1 \partial w_3 \partial w_4}\frac{\partial A_4}{\partial w_2} + \frac{\partial^3 A_4}{\partial w_2 \partial w_3 \partial w_4}\frac{\partial A_4}{\partial w_1} - \\
&\quad \frac{\partial^4 A_4}{\partial w_1 \partial w_2 \partial w_3 \partial w_4}\Big).
\end{aligned}
$$

Upon differentiation,

$$
\frac{\partial A_4}{\partial w_1} = -2 + A_{1,2} + A_{1,3} + A_{1,4} + C_{1,23} + C_{1,24} + C_{1,34} - D_{1,234}.
$$

By symmetry, $\frac{\partial A_4}{\partial w_2}, \frac{\partial A_4}{\partial w_3}, \frac{\partial A_4}{\partial w_4}$ are like $\frac{\partial A_4}{\partial w_1}$ with indices 1 and 2, 1 and 3, 1 and 4 interchanged. Also,

$$
\frac{\partial^2 A_4}{\partial w_1 \partial w_2} = A_{12} + C_{12,3} + C_{12,4} - D_{12,34}
$$

and

$$\frac{\partial^3 A_4}{\partial w_1 \partial w_2 \partial w_3} = C_{123} - D_{123,4}$$

By the same reason as the above we can easily get

$$\frac{\partial^2 A_4}{\partial w_1 \partial w_3}, \frac{\partial^2 A_4}{\partial w_1 \partial w_4}, \frac{\partial^2 A_4}{\partial w_2 \partial w_3}, \frac{\partial^2 A_4}{\partial w_2 \partial w_4}, \frac{\partial^2 A_4}{\partial w_3 \partial w_4},$$

$$\frac{\partial^3 A_4}{\partial w_1 \partial w_2 \partial w_4}, \frac{\partial^3 A_4}{\partial w_1 \partial w_3 \partial w_4}, \frac{\partial^3 A_4}{\partial w_2 \partial w_3 \partial w_4}.$$

Finally,

$$\frac{\partial^4 A_4}{\partial w_1 \partial w_2 \partial w_3 \partial w_4} = -D_{1234}.$$

3.4 Computer implementation of the Hüsler-Reiss model

The parameter λ_{ij}'s are estimated with a quasi-Newton routine. The multivariate normal survival function is computed with the routine of Schervish (1984). Other routines that are needed are a numerical integration routines and routines for the functions: $C_{1,23}, C_{12,3}, C_{123}, D_{1,234}, D_{12,34}, D_{123,4}, D_{1234}$. Then $C_{1,24}, C_{13,4}$ etc., can be computed from these routines by changing the parameters of the routines.

3.5 Data Analysis

Joe (1990b) constructed some multivariate extreme value environmental data sets from Bay Area and from the Great Vancouver Regional District.

The first data set consists of weekly maxima of ozone concentrations (in parts per hundred million) for several monitoring stations in the Bay Area; the weeks were for the months of April to October in 1983-1987. For comparison with Joe (1990b), the same subset of 5 stations are used.

The second data set consists of weekly maxima of SO_2, NO_2 and ozone concentrations (in parts per hundred milion) for several monitoring stations in the Greater Vancouver Regional District: the weeks were for the months of April to September to October for July 1984 to October 1987. Joe (1990b) eliminated all weeks with too many missing values, and for comparison the same reduced data are used. In Chapter 4, weeks with missing values are used as censored data.

Details of exploratory data analysis of both data sets are given in Joe (1990b).

Joe (1990b) compares the fit of the models (2-7-1) to (2-7-8) and their multivariate extensions, as well as other models in Joe (1990a). The models in Joe (1990a) which have some appealing theoretical properties were found to fit worse even though the models had more parameters. A partial explanation is that the parameters in these models are not easily interpretable as in (2-7-1) to (2-7-11). For comparison with a baseline, the likelihood of the multivariate normal copula with exponential margins was computed.

The comparison of the Hüsler-Reiss model with the best fitting of (2-7-1) to (2-7-8) and with the multivariate normal copula is given in Tables 1 to 7. Only the parameter estimates for the Hüsler-Reiss model are given. Note that the likelihoods given are for the dependence parameters, after the univariate margins have been transformed to exponential survival functions (see Section 2.8). Also all likelihoods based on margins (or subsets of data from subsets of stations) are given. This is to check on consistency of the parameter estimates for multivariate margins of different orders. This check provides an indication of the goodness of fit of the multivariate models. A final note for the tables is that for 3 or more stations, labelled as $1, \ldots, p$, the parameter estimates are given in lexicographical order, that is, $\lambda_{12}, \lambda_{13}, \ldots, \lambda_{1p}, \lambda_{23}, \ldots, \lambda_{2p}, \ldots, \lambda_{p-1,p}$. For example, consider the pairs of stations PT, SJ in Table 1, where the other stations are abbreviated CC, VA,ST. In the trivariate subset, (CC, PT, SJ), the λ parameter for (PT,SJ) is the third value (0.588); in the trivariate subset, (PT, SJ, VA), it is the first value (0.595). in the 4-variate subset, (CC,PT, SJ, VA), it is the fourth value (0.567), etc. Recall from Chapter 2, that a smaller value of λ means more dependence.

Some overall conclusions from the tables are:

(1) The parameter estimates for a particular λ are similar over different likelihoods.

(2) Of the 37 bivariate likelihoods, the likelihood of the Hüsler-Reiss model is higher than those of both models (2-7-1) and (2-7-2) in 13 cases, and it is higher than that of the multivariate copula in 18 cases.

(3) Of the 25 trivariate likelihoods, the likelihood of the Hüsler-Reiss model is higher than those of both models (2-7-3) and (2-7-4) in 11 cases. With the adjustment for the number of parameters ((2-7-3) and (2-7-4) have 2 parameters, the Hüsler-Reiss trivariate model has 3 parameters), through a penalty of half the number of parameters for the log likelihood(the Akaike information criterion), the Hüsler-Reiss model is only better for the first data set in Table 1.

(4) Of the 9 four-variable likelihoods, the likelihood of the Hüsler-Reiss model is higher than those of the models (2-7-5) to (2-7-8) in 6 cases. This is true even with adjustment for the number of parameters (6 for the Hüsler-Reiss model and 3 for models (2-7-5) to (2-7-8)).

Overall, it appears that the Hüsler-Reiss model provides a better fit for Tables 1 and 3, and maybe Table 5. Therefore it is a useful model in addition to those in Joe (1990b).

An indication of whether the Hüsler-Reiss model is better can be seen from the estimates for the bivariate likelihoods. If the pattern of the λ's does not fit the bivariate patterns that are possible with models (2-7-3) to (2-7-8), then the Hüsler-Reiss model should be better. (The restrictions of the bivariate patterns for models (2-7-3) to (2-7-8) can be seen from taking the bivariate margins in these models.)

Table 1. Bay Area - Ozone(n=120)

Notation: Concord(CC),Pittsburg(PT),Vallejo(VA), San Jose(SJ) Santa RosT).					
Likelihood for dependence parameters.					
	estimate	negative log likelihood			
subset	H-R	δ's	τ's	normal	H-R model
CC, PT	0.353	154.28	154.08	153.50	154.75
CC, SJ	0.451	176.17	175.96	179.40	176.91
CC, VA	0.532	190.70	190.38	192.41	190.30
CC, ST	0.708	211.60	211.03	214.24	210.47
PT, SJ	0.595	199.85	199.34	200.48	199.06
PT, VA	0.581	196.07	196.05	198.52	197.15
PT, ST	0.749	214.80	214.37	214.54	213.92
SJ, VA	0.579	197.68	197.40	202.49	197.15
SJ, ST	0.740	214.56	213.96	218.74	213.52
VA, ST	0.599	199.86	199.62	203.12	199.54
CC,PT,SJ	0.354,0.453 0.588	221.32	216.41	212.94	212.23
CC,PT,VA	0.353,0.532 0.580	225.26	221.36	223.87	222.92
CC,PT,ST	0.352,0.704 0.750	244.94	245.20	245.70	243.86
CC,SJ,VA	0.451,0.536 0.576	245.38	243.37	248.67	242.80
CC,SJ,ST	0.450,0.716 0.741	268.11	265.32	272.20	265.69
CC,VA,ST	0.533,0.708 0.606	271.62	271.40	273.54	268.28
PT,SJ,VA	0.595,0.583 0.579	267.09	265.27	269.90	264.27
PT,SJ,ST	0.594,0.755 0.738	290.80	288.08	291.56	287.04
PT,VA,ST	0.582,0.752 0.606	279.04	277.97	278.94	275.45
SJ,VA,ST	0.580,0.747 0.607	282.20	278.01	284.13	275.85
CC,PT,SJ VA	0.357,0.441 0.539,0.567 0.581,0.574	288.42	280.47	280.01	275.72
CC,PT,SJ ST	0.355,0.441 0.70, 0.567 0.738,0.741	311.54	305.04	303.61	299.69
CC,PT,VA ST	0.353,0.534 0.706,0.581 0.754,0.607	307.17	303.67	304.20	300.66
CC,SJ,VA ST	0.450,0.538 0.714,0.577 0.750,0.609	329.94	324.70	329.57	321.01
PT,SJ,VA ST	0.594,0.585 0.755,0.581 0.746,0.610	352.88	347.34	349.94	342.36

Table 2. GVRD-Sulphur Dioxide(n=70)

Notation: Confederation Park(CP), Second Narrows(SN), Anmore(AN),Rocky Point Park(RP).					
Likelihood for dependence parameters.					
	estimate	negative log likelihood			
subset	H-R	δ's	τ's	normal	H-R model
CP, SN	0.845	129.55	129.54	129.37	129.58
CP, AN	1.086	136.11	135.96	134.82	135.94
CP, RP	0.905	131.45	131.67	129.61	132.00
SN, AN	1.068	135.36	135.54	133.97	135.62
SN, RP	1.229	137.04	137.40	136.48	137.70
AN, RP	0.756	123.39	124.54	123.63	126.08
CP,SN,AN	0.840,1.068 1.062	193.83	193.82	192.16	193.44
CP,SN,RP	0.846,0.891 1.157	190.28	192.97	189.26	190.89
CP,AN,RP	1.045,0.909 0.750	183.89	188.01	182.94	187.05
SN,AN,RP	1.064,1.172 0.750	187.61	190.59	187.30	190.93
CP,SN,AN RP	0.840,1.060 0.900,1.032 1.155,0.748	241.86	247.15	240.07	244.05

Table 3. GVRD-Nitrogen Dioxide(n=84)

Notation: Kensington Park(KP), Confederation Park(CP), Anmore(AN), Rocky Point Park(RP).					
Likelihood for dependence parameters.					
	estimate	negative log likelihood			
subset	H-R	δ's	τ's	normal	H-R model
KP, CP	0.378	112.97	113.10	120.15	114.06
KP, AN	0.415	119.26	119.31	124.01	119.59
KP, RP	0.362	110.64	110.64	119.83	111.53
CP, AN	0.466	127.84	127.30	128.81	126.55
CP, RP	0.346	108.53	108.27	107.87	107.71
AN, RP	0.356	110.48	110.22	112.64	109.91
KP,CP,AN	0.379,0.417 0.463	147.96	146.20	154.95	146.27
KP,CP,RP	0.382,0.363 0.343	133.11	126.51	137.25	128.55
KP,AN,RP	0.416,0.363 0.353	136.99	131.88	143.35	132.88
CP,AN,RP	0.463,0.345 0.356	139.29	137.49	135.81	132.40
KP,CP,AN RP	0.382,0.418 0.363,0.457 0.342,0.352	161.42	151.95	160.77	149.77

Table 4. GVRD-Ozone(n=84)

Notation: Marpole(MA), Confederation Park(CP), Anmore(AN), Rocky Point Park(RP).					
Likelihood for dependence parameters.					
	estimate	negative log likelihood			
subset	H-R	δ's	τ's	normal	H-R model
MA, CP	0.521	89.17	89.23	88.00	89.69
MA, AN	0.829	101.63	102.47	101.80	104.54
MA, RP	0.763	99.94	100.50	100.88	102.21
CP, AN	0.747	98.78	99.60	95.96	101.93
CP, RP	0.680	97.23	97.72	93.19	98.95
AN, RP	0.284	61.55	61.73	72.02	64.66
MA,CP,AN	0.526,0.803 0.741	129.02	131.08	126.47	133.31
MA,CP,RP	0.526,0.738 0.676	125.73	129.42	123.81	130.07
MA,AN,RP	0.816,0.747 0.284	103.97	105.08	115.26	108.96
CP,AN,RP	0.734,0.670 0.285	101.13	101.70	107.48	105.76
MA,CP,AN RP	0.505,0.763 0.705,0.700 0.642,0.291	130.01	133.28	137.97	136.71

Table 5. GVRD-Station 5 (Confederation Park). (n=64)

Likelihood for dependence parameters.					
	estimate	negative log likelihood			
subset	H-R	δ's	τ's	normal	H-R model
SO_2, NO_2	1.359	127.33	126.92	125.49	126.77
SO_2, O_3	1.033	122.83	122.88	122.10	123.01
NO_2, O_3	0.932	120.31	120.56	118.39	120.76
SO_2, NO_2, O_3	1.312,1.023	181.73	180.37	176.24	179.48
	0.930				

Table 6. GVRD-Station 7 (Anmore). (n=76)

Likelihood for dependence parameters.					
	estimate	negative log likelihood			
subset	H-R	δ's	τ's	normal	H-R model
SO_2, NO_2	1.167	148.68	148.81	146.10	148.84
SO_2, O_3	1.134	148.21	148.22	145.17	148.25
NO_2, O_3	0.526	118.65	119.11	124.25	120.64
SO_2, NO_2, O_3	1.132,1.112	189.37	190.11	192.69	191.96
	0.526				

Table 7. GVRD-Station 9 (Rocky Point Park). (n=63)

Likelihood for dependence parameters.					
	estimate	negative log likelihood			
subset	H-R	δ's	τ's	normal	H-R model
SO_2, NO_2	1.065	120.87	121.48	118.51	121.99
SO_2, O_3	1.063	120.38	121.10	118.57	122.28
NO_2, O_3	0.622	107.55	106.99	106.00	106.30
SO_2, NO_2, O_3	1.019,1.050	162.60	163.61	159.92	164.04
	0.625				

Chapter 4. Analysis With Missing Data

In statistical analysis, it may happen by accident or some other reason that some of the observations are missing, cannot be collected, or in some other way are not obtainable. In such a situation it is obvious that the routine method is not appropriate.

Now we would like to reanalyze the data from Greater Vancouver Regional District with missing values among the hourly measurements. If there are missing hourly measurements, we treat the daily or weekly maxima as right-censored.

For right censored data, let x_i be either the observed value or the right censored value, and let

$$\delta_i = \begin{cases} 1 & \text{if } x_i \text{ is right-censored} \\ 0 & \text{if } x_i \text{ is not right-censored} \end{cases}$$

If $f(t;\theta)$ is a parametric family of models for the iid data and if $S(t;\theta)$ is the survival function, then the likelihood function with right-censored data is

$$L = \Pi_{i=1}^{n} f(x_i, \theta)^{1-\delta_i} S(t_i, \theta)^{\delta_i},$$

see, for example, Lawless(1980) for details.

We will use the procedure in Section 2 proposed by Joe (1990b) for fitting of a parametric multivariate distribution to independent and identically distributed p-vectors $\mathbf{X_1}, \ldots, \mathbf{X_n}$. The first step is to fit generalized extreme value distribution to the univariate margin separatly by maximum likelihood. That is, the jth univariate margin is:

$$F_j(x, \gamma_j, \mu_j, \sigma_j) = \exp\{-(1 + \gamma_j[(x - \mu_j)/\sigma_j])_+^{-1/\gamma_j}\},$$

where $(z)_+ = \max(z, 0)$. Assuming that the measurements are iid, the likelihood is

$$\Pi_{i=1}^{n}[f_j(x_{ij}; \gamma_j, \mu_j, \sigma_j)]^{1-\delta_{ij}}[\overline{F}_j(x_{ij})]^{\delta_{ij}}$$

where $\overline{F}(x;\gamma,\mu,\sigma) = 1 - F(x;\gamma,\mu,\sigma)$ and $f(x;\gamma,\mu,\sigma) = \frac{dF(x;\gamma,\mu,\sigma)}{dx}$, and subscript j has been added for the jth component.

After fitting the univariate margins, make the transformation

$$y_{ij} = (1 + \gamma_j[(x_{ij} - \mu_j)/\sigma_j])^{-1/\gamma_j}$$

so that original random variables become exponential random variables. The y_{ij}'s are not treated as left censored data. Now for multivariate parameters, we take the univariate margins to be exponential.

For bivariate data, suppose we have iid random pairs (Y_{11}, Y_{12}),...,(Y_{n1}, Y_{n2}),and observe $(Y_{11}, Y_{12}, \delta_{11}, \delta_{12})$,...,$(Y_{n1}, Y_{n2}, \delta_{n1}, \delta_{n2})$,where

$$\delta_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is left censored,} \\ 0 & \text{if } Y_{ij} \text{ is not censored,} \end{cases}$$

$i = 1, \ldots, n$ and $j = 1, 2$. The bivariate cdf is $F(y_1, y_2, \theta) = F = 1 - \overline{F}_1 - \overline{F}_2 + \overline{F}$, where $\overline{F}$ is the survival function and $\overline{F}_1$ and $\overline{F}_2$ are univariate margins of $\overline{F}$.

Therefore we can get the following contribution to the likelihood of the bivariate parameter θ:

$$\begin{cases} P(Y_1 = y_1, Y_2 = y_2) = \frac{\partial^2 F}{\partial y_1 \partial y_2} & \text{if } (\delta_{i1}, \delta_{i2}) = (0,0) \\ P(Y_1 \le y_1, Y_2 = y_2) = \frac{\partial F}{\partial y_2} = \frac{\partial \overline{F}}{\partial y_2} - \frac{\partial \overline{F}_2}{\partial y_2} & \text{if } (\delta_{i1}, \delta_{i2}) = (1,0) \\ P(Y_1 = y_1, Y_2 \le y_2) = \frac{\partial F}{\partial y_1} = \frac{\partial \overline{F}}{\partial y_1} - \frac{\partial \overline{F}_1}{\partial y_1} & \text{if } (\delta_{i1}, \delta_{i2}) = (0,1) \\ P(Y_1 \le y_1, Y_2 \le y_2) = F(y_1, y_2, \theta) & \text{if } (\delta_{i1}, \delta_{i2}) = (1,1) \ . \end{cases}$$

In the trivariate case, the generalization is straightforward and some details are given below. Suppose we have iid random variables (Y_{11}, Y_{12}, Y_{13}), ...,(Y_{n1}, Y_{n2}, Y_{n3}), and observe $(Y_{11}, Y_{12}, Y_{13}, \delta_{11}, \delta_{12}, \delta_{13})$, ..., $(Y_{n1}, Y_{n2}, Y_{n3}, \delta_{n1}, \delta_{n2}, \delta_{n3})$, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is left censored,} \\ 0 & \text{if } Y_{ij} \text{ is uncensored,} \end{cases}$$

$i = 1, \ldots, n$ and $j = 1, 2, 3$. The trivariate cdf is

$$F(y_1, y_2, y_2, \theta) = 1 - \overline{F}_1(y_1) - \overline{F}_2(y_1) - \overline{F}_3(y_1) + \overline{F}_{12}(y_1, y_2, \theta)$$
$$+\overline{F}_{13}(y_1, y_3, \theta) + \overline{F}_{23}(y_2, y_3, \theta) - \overline{F}(y_1, y_2, y_2, \theta)$$

where $\overline{F}$ is survival function and $\overline{F}_1$, $\overline{F}_2$, $\overline{F}_3$, $\overline{F}_{12}$, $\overline{F}_{13}$, and $\overline{F}_{23}$ are the univariate and bivariate margins of $\overline{F}$.

Let $C_{123} = (\delta_{i1}, \delta_{i2}, \delta_{i3})$, then we have

$$\begin{cases} P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = -\frac{\partial^3 F}{\partial y_1 \partial y_2 \partial y_2} & \text{if } C_{123} = (0,0,0) \\ P(Y_1 \le y_1, Y_2 = y_2, Y_3 = y_3) = \frac{\partial^2 F}{\partial y_2 \partial y_3} = \frac{\partial^2 \overline{F}_{23}}{\partial y_2 \partial y_3} - \frac{\partial^2 \overline{F}}{\partial y_2 \partial y_3} & \text{if } C_{123} = (1,0,0) \\ P(Y_1 = y_1, Y_2 \le y_2, Y_3 = y_3) = \frac{\partial^2 F}{\partial y_1 \partial y_3} = \frac{\partial^2 \overline{F}_{13}}{\partial y_1 \partial y_3} - \frac{\partial^2 \overline{F}}{\partial y_1 \partial y_3} & \text{if } C_{123} = (0,1,0) \\ P(Y_1 = y_1, Y_2 = y_2, Y_3 \le y_3) = \frac{\partial^2 F}{\partial y_1 \partial y_2} = \frac{\partial^2 \overline{F}_{12}}{\partial y_2 \partial y_3} - \frac{\partial^2 \overline{F}}{\partial y_1 \partial y_2} & \text{if } C_{123} = (0,0,1) \\ P(Y_1 \le y_1, Y_2 \le y_2, Y_3 = y_3) = \frac{\partial F}{\partial y_3} = -\frac{\partial \overline{F}_3}{\partial y_3} + \frac{\partial \overline{F}_{13}}{\partial y_3} + \frac{\partial \overline{F}_{23}}{\partial y_3} - \frac{\partial \overline{F}}{\partial y_3} & \text{if } C_{123} = (1,1,0) \\ P(Y_1 \le y_1, Y_2 = y_2, Y_3 \le y_3) = \frac{\partial F}{\partial y_2} = -\frac{\partial \overline{F}_2}{\partial y_2} + \frac{\partial \overline{F}_{12}}{\partial y_2} + \frac{\partial \overline{F}_{23}}{\partial y_2} - \frac{\partial \overline{F}}{\partial y_2} & \text{if } C_{123} = (1,0,1) \\ P(Y_1 = y_1, Y_2 \le y_2, Y_3 \le y_3) = \frac{\partial F}{\partial y_1} = -\frac{\partial \overline{F}_1}{\partial y_1} + \frac{\partial \overline{F}_{12}}{\partial y_1} + \frac{\partial \overline{F}_{13}}{\partial y_1} - \frac{\partial \overline{F}}{\partial y_1} & \text{if } C_{123} = (0,1,1) \\ P(Y_1 \le y_1, Y_2 \le y_2, Y_3 \le y_3) = F(y_1, y_2, y_2, \theta) & \text{if } C_{123} = (1,1,1) \,. \end{cases}$$

Clearly, this can be generalized to higher dimensions.

Computer implementation of likelihood

For models (2-7-1) to (2-7-8), the different contributions to the likelihood can easily be obtained using a symbolic manipulation software, such as Maple (Char et al., 1988, 5th edition), where derivatives can be output in Fortran format.

Data analysis

Here we reanalysize the Data Set 2 of Chapter 3 with the missing hourly values and censored weekly maxima.

The results of the multivariate extreme value analysis are summarized in Table 8−10. The values of the maximum likelihood estimates are given in the left hand side of each table, and the negative log likelihoods are in the right side of each table.

A larger value of μ means that the maxima for the station tend to be larger, a larger value of σ means more spread and a larger value of γ means that the hourly concentration has a heavier tail. Compared to the results of previous analysis, where weeks with many missing hourly measurements were treated as missing at random, we see that all parameter estimates of μ, σ and γ in three tables are mostly slightly larger (compare Tables 3 to 5 of Joe 1990b). This is not a surprising result.

After fitting the univariate margins, we obtain the transformed data which is left censored. The next step is to analyze the transformed data. The models for the multivariate exponential distribution were described before. The parameter estimates for the censored data are quite close to the estimates in Table 3 to 5 of Joe (1990b), but the relative values of the estimates are different in some cases. For trivariate case, the likelihoods for the δ's models and the τ's models diverge whereas they are close in the bivariate case. An explanation of the multivariate parts of Tables 8 to 10 is explained next for one case. For example, for (CP, SN, AN) in Table 8, the estimates of the parameters in model (2-7-3) are (1.567, 1.249), and the estimates of the parameters in (2-7-4) are (0.851, 0.507). For model (2-7-3) (respectively (2-7-4)), 1.567 (0.851) measures the dependence of the pair of stations (CP, SN), and 1.249 (0.507) measures the dependence of the pairs of (CP, AN) and (SN, AN). The model (2-7-3) is better in Table 10, and the model (2-7-4) is better in Table 8 and 9.

Conclusions are the same as in Joe (1990b) which means that there is no substantial difference with the missing values deleted.

Table 8. GVRD-Sulphur Dioxide(n=104)

Notation: Confederation Park(CP), Second Narrows(SN), Anmore(AN),Rocky Point Park(RP).

Likelihood for dependence parameters.

subset	estimate	estimate	negative log likelihood	
	δ's	τ's	δ's	τ's
CP, SN	1.560	0.843	185.581	185.664
CP, AN	1.220	0.468	205.635	205.345
CP, RP	1.462	0.729	200.587	200.556
SN, AN	1.292	0.535	191.869	192.223
SN, RP	1.447	0.720	190.815	190.669
AN, RP	1.647	0.905	195.814	197.452
Multivariate				
CP,SN,AN	1.567,1.249	0.851, 0.507	284.370	283.937
CP,SN,RP	1.550,1.456	0.823, 0.674	277.888	279.263
AN,RP,CP	1.650,1.348	0.875, 0.562	291.025	293.808
AN,RP,SN	1.675,1.395	0.868, 0.596	278.128	282.235
Estimate for univariate margin parameter				
CP	0.313	1.857	0.834	
SN	0.181	2.083	0.885	
AN	0.261	3.215	2.024	
RP	0.080	2.058	1.094	

Table 9. GVRD-Nitrogen Dioxide(n=84)

Notation: Kensington Park(KP), Confederation Park(CP), Anmore(AN),Rocky Point Park(RP).				
Likelihood for dependence parameters.				
subset	estimate	estimate	negative log likelihood	
	δ's	τ's	δ's	τ's
KP, CP	2.816	2.118	161.095	161.041
KP, AN	2.673	1.965	161.941	161.944
KP, RP	2.947	2.246	154.375	154.356
CP, AN	2.168	1.482	176.638	176.104
CP, RP	3.256	2.557	144.079	143.930
AN, RP	2.968	2.272	152.358	152.171
Multivariate				
KP,CP,AN	2.741, 2.259	2.061, 1.715	219.517	215.709
RP,KP,CP	2.945, 2.718	2.364, 2.069	194.823	189.065
RP,KP,AN	2.974, 2.597	2.226, 2.150	198.191	192.105
RP,CP,AN	2.961, 2.419	2.2.9, 1.934	204.392	199.517
Estimate for univariate margin parameter				
KP	0.082	5.701	2.186	
CP	0.179	6.093	2.442	
AN	0.086	3.750	1.760	
RP	0.078	5.228	1.658	

Table 10. GVRD-Ozone(n=84)

Notation: Marpole(MA), Confederation Park(CP), Anmore(AN),Rocky Point Park(RP).				
Likelihood for dependence parameters.				
subset	estimate	estimate	negative log likelihood	
	δ's	τ's	δ's	τ's
MA, CP	2.077	1.360	143.604	143.340
MA, AN	1.726	1.012	159.720	160.279
MA, RP	1.691	0.967	159.913	160.312
CP, AN	2.174	1.467	147.824	147.753
CP, RP	2.162	1.444	144.703	144.804
AN, RP	3.616	2.917	122.432	122.622
Multivariate				
MA,CP,AN	2.061, 1.835	1.322, 1.140	211.020	210.141
MA,CP,RP	1.948, 1.948	1.328, 1.073	207.476	210.365
AN,RP,MA	3.606, 1.759	2.868, 0.963	189.255	191.630
AN,RP,CP	3.586, 2.131	2.852, 1.439	176.798	177.085
Estimate for univariate margin parameter				
MA	-0.049	4.182	0.816	
CP	-0.019	5.054	1.384	
AN	0.047	5.250	1.595	
RP	0.091	5.054	1.761	

Chapter 5. Conclusion and Further Research

This thesis has applied maximum likelihood estimation of several parametric families of multivariate extreme value distribution, including the case where some maxima are right censored.

The Hüsler-Reiss family has more flexibility, with one parameter for each bivariate margin, than the parametric families used in Joe (1990b). However, its form is much more complicated and for diminsion $p > 3$, time-consuming integrations and computations of the multivariate normal cdf are needed. Also, unlike the other families, symbolic manipulation software cannot be used to obtain the density function from the survival function. The maximum likelihood estimation takes 10-20 hours on a Sun SPARCstation 330 for $p = 4$, and is expected to take weeks for $p = 5$ (the programming for this case has not been completed). If both the multivariate dependence parameters and the univariate marginal parameters are simultaneously estimated, the CPU time is much more than the preceding (this also has not been done). Fortunately, it seems possible from the maximum likelihood estimates of the bivariate likelihoods to check whether the simpler models in Joe (1990b) fit the data adequately, if so, the bivariate parameter estimates must follow certain patterns. If the simpler models appear not to be adequate, then the best model, among known ones, would likely be the Hüsler-Reiss model.

Further research is needed to derive families of multivariate extreme value distributions that have a dependence parameter for each bivariate margin and that have a simpler form than the Hüsler-Reiss model (simpler in the sense that maximum likelihood estimation becomes possible for $p \geq 5$). Some positive work in this direction has been initiated.

BIBLIOGRAPHY

Char, Bruce W., Geddes, K.O., Gonnet, G.H., Monagan, M.B., and Watt, S.M. (1989). Maple. 5th ed., WATCOM Publications Limited.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. J. R. Statist. Soc. 52, 377-392.

Galambos, J. (1987). The Asymptitic Theory of Extreme Order Statistics, second edition. Kreiger Publishing Co., Malabar, FL

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors between independence and complete dependence. Statistics & Probability Letters 7 283-286.

Joe H. (1990a). Families of min-stable multivariate exponential and multivariate extrem value distributions. Statistics & Probability Letters 9 (1990) 75-81.

Joe H. (1990b). On fitting multivariate extreme value distributions, with applications to environmental data. Technical report. Department of Statistics, University of British Columbia.

Lawless J. F. (1982). Statistical Models and Methods for Lifetime data. JOHN WILEY & SONS, Inc., New York.

Resnick, S.I. (1987). Extreme Values, Regular Variation and Point Processes. New York: Springer-verlag.

Schervish, J. (1984). Multivariate normal probabilities with error bound. Appl. Stat.33 81- 94.

Tawn J. A. (1988). Bivariate extreme value theory, models and estimation, Biometrika 75, 397-415.

Tawn J. A. (1990). Modelling multivariate extreme value distributions. Biometrika 77, 245-253.