Statistical Theory of Extremes
Homepage: http://www.gathacognition.com/book/gcb14
http://dx.doi.org/10.21523/gcb1

Part 3
Multivariate extremes

Chapter 11

Bivariate Extremes: Models and Statistical Decision

J. Tiago de Oliveira
Academia das Ciências de Lisboa (Lisbon Academy of Sciences), Lisbon, Portugal.

Abstract

Large samples, $F^n(x,y)$ is replaced by an asymptotic approximation for statistical decision of extremes. The discussions decide the fitness and accuracy of those approximations. The differentiable models, non-differentiable models of statistical decision, intrinsic estimation of the dependence functions, ready-to-wear model and estimations of the association between the margins with some examples are discussed in this chapter. Differentiable models, the logistic one gives a good fit but shows the logistic and the natural models can be very close and difficult to separate for moderate samples.

Published Online

23 June 2017

Keywords

Statistical decision;
Asymptotic approximation;
Samples;
Differentiable models;
Logistic modes.

Editor(s)

J.C. Tiago de Oliveira

11.1 Introduction

As regards statistical decision for bivariate extremes, we will use an approach similar to the one used for univariate extremes: we will replace, for large samples, $F^n(x,y)$ by an asymptotic approximation

$$L\left(\frac{x-\lambda}{\delta},\frac{x-\lambda'}{\delta'}\right),$$

where, $L(x,y)$ will have the expressions given before, thus reducing the difficulty of not knowing $F^n(x,y)$ to the simpler (!) difficulties of not knowing the dependence structure ($k(.)$ or $A(.)$ for instance) and four

Originally published in 'Statistical Analysis of Extremes', 1997, 2016
http://dx.doi.org/10.21523/gcb1.1718
© 2017 GATHA COGNITION® All rights reserved.

independent margin parameters. Recall that we have the results similar to the ones used to justify the substitution of $F^n(x)$ by $L(\frac{x-\lambda}{\delta})$ in the univariate case:

The stability and the uniformity of convergence are valid as well as results similar to those of Watson and others, showing that even in the case of some dependence between the successive observations the limiting distributions are still the same, are also true.

In the asymptotic approximation of $L(\frac{x-\lambda}{\delta}, \frac{y-\lambda'}{\delta'})$ to the actual and unknown $F^n(x, y)$, we will use, in principle, a dependence function, although Geffroy's sufficient condition could suggest the independence model. The existence of a strong correlation between the maxima (or minima), a correlation not completely destroyed in some applications by the relatively large number of observations from which the maxima (or minima) are taken, suggests the use of an asymptotic bivariate extremes distribution function as a better approximation to the actual (unknown) distribution function of bivariate extremes. This stand is also justified by the fact that the discrepancy between the independence and the dependence cases may be very large in a probabilistic sense but not expressed by the usual indicators. As an example, we can cite the fact that the uniform distance between the biextremal model with standard Gumbel margins, where the distribution function is $\Lambda(x, y|\theta) = \exp\{-e^{-x} + e^{-y} - \min(e^{-x}, \theta\, e^{-y})\}$ (note that $\Lambda(x, y|0) = \Lambda(x).\Lambda(y)$ — independence — and that $\Lambda(x, y|1) = \Lambda(\min(x, y))$, — the diagonal case —) , and the independence case is

$$\sup|_{x,y} \Lambda(x, y|\theta) - \Lambda(x, y|0)| = \theta(1+\theta)^{-1-1/\theta};$$

consequently if we take the independence model as an approximation to a biextremal model with θ close to 1, i.e., close to the diagonal case, we could obtain probability evaluations for quadrants with an error around $1/4$, the value of the distance for $\theta = 1$, but that can differ by almost 1 if the set used is a strip containing the first diagonal!

Experience will finally decide the fitness and accuracy of those approximations. But, for the moment, as the (m, N) natural model shows, we can have any finite number of parameters, forgetting the four of the margins. And it also can be shown that for no one-parameter models (with standard margins), do there exists sufficient statistics. An alternative is the intrinsic estimation of the dependence function $(k \text{ or } A)$, as will be seen.

For some models we do not have the best test, the best estimation procedure, etc. (we may have one, but not necessarily the best) although in other situations nothing at all is known. An example: the statistical choice of

bivariate extreme models — so important for applications — is now beginning to be considered! In general, the few papers up to now have chosen one model from the beginning or compare two of them by the use of the Kolmogoroff-Smirnov statistic or another test; see, for instance, Gumbel and Golstein (1964).

As a general rule in statistical decision for bivariate maxima with Gumbel margins, the papers Tiago de Oliveira (1970), (1975a), (1975b), (1980), (1984) and (1989) contain results that are improved here.

As regards modelling, we have to deal with some models which have a planar density and with others, which appear naturally, that have a singular component.

Finally, let us recall that as the margins in both cases (Gumbel and exponential) have all moments, we also have crossed moments, etc., a fact that will be used.

Note that for all models to be considered, we only give the correlation coefficients and index of dependence when they have a compact form. The correlation coefficients can then easily be used for statistical decision. The non-parametric correlation coefficients and the index of dependence are the same for the corresponding $\Lambda(x,y|\theta)$ and $S(x',y'|\theta)$; the classical correlation coefficient is different.

It is important to note that separation between models is difficult for moderate samples. A graphical explanation is the essence of Exercise 3.23.

11.2 The differentiable models: statistical decision

Until now only two models of bivariate extremes with a planar density (with a point exception in one case) have been considered and studied in some depth in the literature: the logistic model and the mixed model, both appearing in Gumbel (1962) and (1966). Both are symmetrical but nothing precludes the existence of asymmetric models, which is the case in the majority of non-differentiable models to be considered in the next section. Here, as well as for non-differentiable models, we will always study them under the form of bivariate maxima with standard Gumbel margins, i.e., with the distribution function $\Lambda(x,y) = (\Lambda(x)\Lambda(y))^{k(y-x)}$ with the dependence function $k(w)$; at the end of each case we will consider the corresponding forms for bivariate minima with standard exponential margins, i.e., with the survival function $S(x',y') = \exp\{-(x'+y')\, A(\frac{y'}{x'+y'})\}$ $(x'+y' \geq 0)$ obtained by the transformation $X' = e^{-X}, Y' = e^{-Y}$ and with $A(u) = k(\log(\frac{1-u}{u}))$, as known.

The *logistic model* has the distribution function $\Lambda(x,y|\theta) = \exp\{-e^{-x/(1-\theta)} + e^{-y/(1-\theta)})^{1-\theta}\}, 0 \leq \theta < 1$ (for $\theta = 0$ we have independence and for $\theta = 1^-$ we get the diagonal case, which is the only situation without density). It is so called because

$$\text{Prob}\{Y - X \leq w|\theta\} = D(w|\theta) = (1 + e^{-w/(1-\theta)})^{-1}$$

corresponding to the dependence function

$$k(w|\theta) = \frac{(1+e^{-w/1-\theta})^{-1-\theta}}{1+e^{-w}};$$

note that we must have $0 \leq \theta \leq 1$, as comes from the inequality $\frac{\max(1+e^w)}{1+e^w} \leq k(w) \leq 1$.

The margins are exchangeable as shown by the expression of $\Lambda(x,y|\theta)$, by the fact that $k(-w|\theta) = k(w|\theta)$ or even because $D(w|\theta) + D(-w|\theta) = 1$.

The equidistribution curves are very easy to draw. The median curve has the equation

$$e^{-x/(1-\theta)} + e^{-y/(1-\theta)} = (\log 2)^{1/(1-\theta)}$$

which is similar, with similarity ratio $1/(1-\theta)$, to the equidistribution curve for the independence case.

The classical correlation coefficient is $\rho(\theta) = \theta(2-\theta)$, increasing from $\rho(0) = 0$ to $\rho(1) = 1$, the difference-sign correlation coefficient is $\tau(\theta) = \theta$, increasing from $\tau(0) = 0$ to $\tau(1) = 1$, the medial correlation coefficient is $v(0) = 4^{(1-2^{-\theta})} - 1$, increasing from $v(0) = 0$ to $v(1) = 1$. Linear regression of Y in X is, evidently, $L_y(x|\theta) = \gamma + \theta(2-\theta)(x-\gamma)$. It can be shown that the index of dependence

$$\sup|_{x,y} \Lambda(x,y|\theta) - \Lambda(x,y|0)| = (1-2^{-\theta})2^{-\theta/(2^{-\theta}-1)};$$

increases from $0 (\text{at } \theta = 0)$ to $1/4 = .25$ $(\text{at } \theta = 1)$. It is, thus intuitive that the distance between the independence $(\theta = 0)$ and the assumed model for $\theta > 0$ is small in general, and for small samples it will naturally be difficult to separate models.

Owing to the importance of independence, as a reference because in general we can expect association, we can formulate as confirmation of dependence an L.M.P. test of $\theta = 0$ vs $\theta > 0$. As for the sample $\{(x_1,y_1),\ldots,(x_n,y_n)\}$ the likelihood is

$$L(x_i, y_i|\theta) = \prod_1^n \frac{\partial^2 \Lambda(x_i,y_i|\theta)}{\partial x_i \, \partial y_i},$$

the rejection region has the form

$$\frac{\partial}{\partial \theta} \left(\sum_1^n \frac{\partial^2 \Lambda(x_i, y_i|\theta)}{\partial x_i \ \partial y_i}\right)|_{\theta=0} > c_n$$

or

$$\sum_1^n v(x_i, y_i) \geq c_n$$

where

$$v(x,y) = x + y + x\, e^{-x} + y\, e^{-y} + (e^{-x} + e^{-y} - 2)\log(e^{-x} + e^{-y}) + \frac{1}{e^{-x}+e^{-y}}.$$

In the case of independence, $v(X, Y)$ has mean value zero but infinite variance; evidently $1/n \,.\sum_1^n v(x_i, y_i) \xrightarrow{a.s.} 0$, in the case of independence, by Khintchine's theorem. Although then the usual Central Limit Theorem is not applicable, it was shown by Tawn (1988) that $\sum_1^n v(x_i, y_i)/\sqrt{(n \log n)/2}$ is asymptotically standard normal. Estimation by the maximum likelihood is possible and Cramér — regular for $\theta \neq 0$, having for $\theta = 0$ the difficulties arising from the infinite variance of $v(X, Y)$. Thus we have $c_n \sim \sqrt{(n \log n)/2}\ N^{-1}(1-\alpha)$, where $N(.)$ denotes the standard normal distribution function and α the significance level of the one-sided tests.

As θ is a dependence parameter, it is natural to use the correlation coefficients whose estimators are, as is well known, independent of the location and dispersion parameters of the margins, to test $\theta = 0$ vs. $\theta > 0$. It was shown, in Tiago de Oliveira (1964) and (1975), that the most efficient is the classical correlation coefficient. Recall that, in the case of independence, $\sqrt{n}\, r$ is asymptotically standard normal. As $\rho(\theta) = \theta(2-\theta)$, a point estimator of θ is given by $r = \theta^*(2-\theta^*)$ or $\theta^* = 1 - \sqrt{1-r}$, which leads to no truncation of the estimator if $r \geq 0$; otherwise we will take $\theta^* = 0$ (independence).

A quick approach may be the quadrants method (see Annex 5); as the medians of the margins, in reduced form, are $\tilde{\mu} = -\log\log 2 = 0.3665129$ we have $p(\theta) = \text{Prob}\{X \leq \tilde{\mu}, Y \leq \tilde{\mu}\} = \exp\left(-2^{1-\theta}\ \log 2\right)$, so the probability that both components are at the same side of the medians,

i.e., $\{X \leq \tilde{\mu}, Y \leq \tilde{\mu}\}$ or $\{X > \tilde{\mu}, Y > \tilde{\mu}\}$ is $2p(\theta)$. If N denotes the number of sample pairs with components both smaller or both larger than the margin medians $\tilde{\mu}$, we know that $\sqrt{n}\,\frac{N/n-2\,p(\theta)}{\sqrt{2\,p(\theta)\,(1-2\,p(\theta))}}$ is asymptotically standard normal. The point estimator of θ is then given by

$$\theta^{**} = 1 - \frac{\log\,(\log(2n/N)/\log 2)}{\log 2}$$

and the condition $0 \leq \theta^{*} \leq 1$ is equivalent to $1/2 \leq N/n \leq 1$; as the RHS inequality is trivial, if the LHS one is violated (i.e., $n > 2N$) we get $\theta^{**} < 0$ and we must assume independence (if the model is valid).

Notice that $p(\theta) = \exp\{(-2^{1-\theta}\log 2)\}$ increases from $p(0) = 1/4$ to $p(1) = 1/2$.

It is easy, by the use of the δ-method, to show that $\log 2\,.\sqrt{\frac{n\,N}{n-N}}\,\log\frac{2\,n}{N}\,(\theta^{**} - \theta)$ is asymptotically standard normal.

The logistic model for bivariate minima with standard exponential margins is immediately obtained by the transformation $X' = e^{-X}, Y' = e^{-Y}$ $(A(u) = k(\log(\frac{1-u}{u}))$ has the following survival function:

$$S(x', y'|\theta) = P(X' \geq x', Y' \geq y') = \exp\{-[x'^{1/(1-\theta)} + y'^{1/(1-\theta)}]^{1-\theta}\}$$

for $x', y' \geq 0$; also we get $A(u|\theta) = (u^{1/(1-\theta)} + (1-u)^{1/(1-\theta)})^{1-\theta}$ and the equisurvival median curve is $x'^{1/(1-\theta)} + y'^{1/(1-\theta)} = (\log 2)^{1/(1-\theta)}$; for $\theta = 1$ we obtain a straight lines. We also have $B(u|\theta) = \text{Prob}\left\{\frac{Y'}{X'+Y'} \leq u\right\} = 1 - D\left(\log\frac{1-u}{u}\middle|\theta\right) = \frac{u^{1/(1-\theta)}}{u^{1/(1-\theta)}+(1-u)^{1/(1-\theta)}}$. The function $v(x', y')$ is $v(x', y') = -(\log x' + x'\log x' + \log y' + y'\log y') + (x' + y' - 2)\log(x' + y') + 1/(x' + y')$.

The classical correlation coefficient $\rho = \int_0^1 A^{-2}\,(u|\theta)\,d\,u - 1$, by the transformation $u = (1 + (\frac{1-t}{t})^{1-\theta})^{-1}$, takes the form $\rho(\theta) = (1-\theta)\int_0^1 t^{-\theta}\,(1-t)^{-\theta}\,d\,t - 1 = (1-\theta)B(1-\theta, 1-\theta) - 1 = (1-\theta)\frac{\Gamma^2(1-\theta)}{\Gamma(2(1-\theta))} - 1$ (as $B(.,.)$ is the beta function), which increases from $\rho(0) = 0$ to $\rho(1^-) = 1$ as $(1-\theta)\,\Gamma(1-\theta) \to 1$ for $\theta \to 1^-$.

Let us now consider the *mixed model*. Its distribution function is $\Lambda(x, y|\theta) = \exp\,(-e^{-x} - e^{y} + \theta/(e^{x} + e^{y}))\,, 0 \leq \theta \leq 1$. When $\theta = 0$ we

have independence but for $\theta = 1$ it does not degenerate, in the diagonal case, as happens with the logistic model.

The model has the dependence function $k(w|\theta) = 1 - \theta \frac{w}{(1+e^w)^2} = k(-w|\theta)$, and so the model is exchangeable, as also follows from

$$D(w|\theta) = \frac{e^w}{1+e^w} \cdot \frac{(1+e^w)^2 - \theta}{(1+e^w)^2 - \theta\, e^w} = 1 - D(-w|\theta).$$

The dependence function $k(w|\theta)$ is the $(1-\theta, \theta)$ mixture of the dependence functions 1 and $1 - \frac{e^w}{(1+e^w)^2}$; the inequality $\frac{\max(1,e^w)}{1+e^w} \leq k(w) \leq 1$ shows that the domain $[0,1]$, of cannot be enlarged.

The equidistribution curves are drawn as usual. The median curve takes the form $e^{-x} + e^{-y} - \theta/(e^x + e^y) = \log 2$.

The classical correlation coefficient has the expression $\rho(\theta) = \frac{6}{\pi^2} \cdot (\arccos(1-\theta/2))^2$ which increases from $\rho(0) = 0$ to $\rho(1) = 2/3$, the medial correlation coefficient is $v(\theta) = 2^{\theta/2} - 1$ increasing from $v(0) = 0$ to $v(1) = 1$; the model is, thus, not fit to express a strong dependence, as confirmed below by the index of dependence. Linear regression of Y is X is evidently $L_y(x|\theta) = \gamma + \frac{6}{\pi^2} (\arccos(1-\theta/2))^2 (x - \gamma)$. We have also

$$\sup_{x,y} |\Lambda(x,y|\theta) - \Lambda(x)\, y|0)| = \frac{\theta}{4-\theta} (1-\theta/4)^{4/\theta}$$

increasing from $0 (\text{at } \theta = 0)$ to $3^3/4^4 = .105$ at $= 1$. The smaller variation of the correlation coefficient and of the index of dependence shows that the deviation from independence is smaller and more difficult to detect.

The LMP test of $\theta = 0$ vs. $\theta > 0$ is given by the rejection region

$$\sum_1^n v(x_i, y_i) \geq c_n$$

with

$$v(x,y) = 2\frac{e^{2x} \cdot e^{2y}}{(e^x + e^y)^3} - \frac{e^{2x} + e^{2y}}{(e^x + e^y)^2} + \frac{1}{e^x + e^y}.$$

As before, the situation is marred by the fact that for independence, although $v(X,Y)$ has mean value zero, the variance is infinite. The usual Central Limit Theorem cannot be applied but, as before, $1/n \cdot \sum_1^n v(x_i, y_i) \xrightarrow{a.s.} 0$. But as shown also by Tawn (1988), $\sum_1^n v(x_i, y_i) / \sqrt{(n \log n)/15}$ is also asymptotically standard normal; then $c_n \sim \sqrt{(n \log n)/15}\ N^{-1}(1-\alpha)$.

Estimation by the maximum likelihood is also possible and Cramér-regular for $\theta \neq 0$ with analogous difficulties for $\theta = 0$, as before.

Correlation coefficients, independent of the margin parameters, are then to be used to test $\theta = 0$ vs. $\theta > 0$. Once more the usual correlation coefficient is the best. The point estimator of θ, if $0 \leq r \leq 2/3$, is given by $\theta^* = 2(1 - \cos\sqrt{\frac{\pi^2}{6}}\, r)$ with the natural truncation otherwise.

A quick approach may also use the quadrants method. As the medians of the margins are $\tilde{\mu} = -\log\log 2$ we have $p(\theta) = \text{Prob}\{X \leq \tilde{\mu}, Y \leq \tilde{\mu}\} = 2^{\theta/2}/4$ which varies from $p(\theta) = 1/4$ to $p(1) = \sqrt{2}/4$. If N/n denotes the fraction of sample pairs both smaller or both larger than the margin medians, we know that $\sqrt{n}\frac{N/n - 2\,p(\theta)}{\sqrt{2\,p(\theta)(1-2\,p(\theta))}}$ is asymptotically standard normal. If $1/2 \leq N/n \leq \sqrt{2}/4$ the estimator of θ is then given by $\theta^{**} = 2(1 + \frac{\log N/n}{\log 2})$ being truncated as before when the double inequality does not apply and, once more by the δ-method, we see that $\frac{\log 2}{2}\sqrt{\frac{n\,N}{n-N}}\,(\theta^{**} - \theta)$ is asymptotically standard normal.

The mixed model for bivariate minima with standard exponential margins immediately has the survival function

$$S(x', y'|\theta) = \exp\{-(x' + y') + \frac{\theta\, x'+y'}{x'+y'}\}, x' + y' \geq 0;$$

we have $A(u|\theta) = 1 - \theta\, u(1-u)$ and the equisurvival median curve is $x' + y' - \theta\frac{x'\,y'}{x'+y'} = \log 2$. We have $B(u|\theta) = 1 - D(\log\frac{1-u}{u}|\theta) = \frac{u - \theta\, u(1-u)^2}{1-\theta\, u(1-u)}$.The function $v(x', y')$ is $v(x', y') = 2\,\frac{x'\cdot y'}{(x'+y')^3} - \frac{x'^2+y'^2}{(x'+y')^2} + \frac{x'\cdot y'}{x'+y'}$. The classical correlation coefficient is simply

$$\rho(\theta) = \frac{8}{(4-\theta)\sqrt{4\,\theta - \theta^2}}\,\text{tg}^{-1}\sqrt{\frac{\theta}{4-\theta}} - \frac{2-\theta}{4-\theta}$$

increasing from $\rho(0) = 0$ to $\rho(1) = \frac{1}{3}(\frac{4\,\pi}{3\sqrt{3}} - 1) = .4727997$.

Finally, we discuss the *statistical choice* between the two non-separated models (logistic and mixed), non-separated because they have in common the independence case ($\theta = 0$ in both cases). As developed in Tiago de Oliveira (1984) and generalized in Tiago de Oliveira (1985), the decision rule (with v_L

and v_M denoting the v functions for the logistic and mixed models, given before) is:

decide for independence if $\sum_1^n v_L(x_i, y_i) \le a_n, \sum_1^n v_M(x_i, y_i) \le b_n$;

decide for the logistic model if $\sum_1^n v_L(x_i, y_i) > a_n$

and $\sum_1^n \frac{v_L(x_i,y_i)}{a_n} > \sum_1^n \frac{v_M(x_i,y_i)}{b_n}$;

decide for the mixed model if $\sum_1^n v_M(x_i, y_i) > b_n$

and $\sum_1^n \frac{v_M(x_i, y_i)}{b_n} > \sum_1^n \frac{v_L(x_i, y_i)}{a_n}$.

It was shown, in Tawn (1988), that the asymptotic distribution of the pair $(\sum_1^n v_L(x_i, y_i)/\sqrt{(n \log n)/2}, \sum_1^n v_M(x_i, y_i)/\sqrt{(n \log n)/15})$ is asymptotically binormal independent with standard margins, and so we should take, with the usual meanings, $a_n \sim \sqrt{(n \log n)/2}\, N^{-1}(1 - \alpha/2)$ and $b_n \sim \sqrt{(n \log n)/15}\, N^{-1}(1 - \alpha/2)$.

Let us now make some comments.

It must be strongly emphasized that the tests and estimators based on the correlation coefficients are independent of the values of the margin parameters. However, in the other cases we have to "estimate" reduced values by using $\hat{x}_i = (x_i - \hat{\lambda}_x)/\hat{\delta}_x$ and $\hat{y}_i = (y_i - \hat{\lambda}_y)/\hat{\delta}_y$ in the Gumbel margins case, the effect of which is not known.

It is natural to consider, as a first step in the statistical analysis of random pairs of maxima or minima, a test for independence, although in many cases we may expect its rejection, maxima or minima being almost always dependent, as observed above.

We can now summarize the results for a test of independence given in Tiago de Oliveira (1965).

The comparison of the usual estimators of the classical, difference-sign, grade, and medial correlation coefficients for one-sided tests, owing to the non-negativity of the values of those coefficients for the bivariate extremal distributions, shows, if we are in the model with bivariate margins, that:

a) the difference-sign and grade correlation are asymptotically equivalent as tests for dependence with respect to any (reasonable) system of alternatives considered, and as the Spearman correlation is always asymptotically equivalent to the grade correlation, the three are equivalent in this case; note that as the grade correlation can be optimal for testing for independence — see Tiago de Oliveira (1985) — with $F(x, y|\theta)$ such that

$F(x,+\infty|\theta) = F_1(x), F(+\infty,y|\theta) = F_2(y), F(x,y|0) = F_1(x)\,F_2(x)$ only if $\frac{\partial\, F(x,y|\theta)}{\partial\,\theta}|_{\theta=0} = F_1(x)\,(1-F_1(x))\,F_2(y)(1-F_2(y))$ which is not the case; no one of the correlation coefficients tests above can be optimal;

b) for the two differentiable models it was shown that the classical correlation coefficient

$$r = \sum (x_i - \bar{x})\,(y_i - \bar{y})/\{\sum (x_i - \bar{x})^2\,(y_i - \bar{y})^2\}^{1/2}$$

gives the best test of the ones above based on correlation coefficients; as $\sqrt{n}\,r$ is asymptotically standard normal we will accept the hypothesis of independence if $\sqrt{n}\,r \le \lambda_\alpha$ where $N(\lambda_\alpha) = 1-\alpha$;

c) the Gumbel and Mustafi (1967) test is equivalent to use of medial correlation;

d) the strip method devised by Posner, Rodemich, Ashlock and Lurie (1969) also has an efficiency smaller than that of the use of the classical product-moment method.

Evidently, some of those results — the non-parametric ones — are immediately extended.

There exist asymmetric extensions of the logistic model; see Tawn (1988). Also there exists another asymmetric model under study, presented in Tiago de Oliveira (1989c), with high correlations $(\rho > .77)$; it has a non-differentiable extension such that $0 \le \rho \le 1$. It gave a good fit to Fox River.

11.3 The non-differentiable models; statistical decision

We know essentially three forms of the non-differentiable models — i.e., with a singular component, thus not having a planar density: the Gumbel model, the biextremal model, and the natural model, the last one being recently generalized by the author (see the next chapter for some details); but no statistical results have appeared so far except when it degenerates into the natural model.

The *Gumbel model* has the distribution function is $\Lambda(x,y|\theta) = \exp\{-(e^{-x} - e^{y} - \theta\min(e^{-x}, e^{-y})\}, 0 \le \theta \le 1$; its dependence function is $k(w|\theta) = 1 - \theta\frac{\min(1,e^{-w})}{1+e^{-w}}$ which is the $(1-\theta,\theta)$ mixture of $k_1(w) = 1$ (independence) and $k_D(w) = \frac{\max(1,e^{-w})}{1+e^{-w}}$ (diagonal case). The model is exchangeable as $k(w|\theta) = k(-w|\theta)$ and we have $D(w|\theta) = \frac{1-\theta}{1-\theta+e^{-w}}$ if

$w < 0$, $D(w|\theta) = \frac{e^w}{1-\theta+e^w}$ if $w > 0$, with a jump of $\frac{\theta}{2-\theta}$, at $w = 0$, which is probability $P(Y = X)$; if $w \neq 0$ we have $D(w|\theta) + D(-w|\theta) = 1$. This model was introduced in Tiago de Oliveira (1971) and is a conversion, to Gumbel margins, of a bivariate model with exponential margins, due to Matshall and Olkin (1967).

The median equidistribution curve is, evidently, $e^{-x} + e^{-y} - \theta \min(e^{-x}, e^{-y}) = \log 2$. The classical correlation coefficient can take the form $\rho(\theta) = \frac{12}{\pi^2} \int_0^\theta \frac{\log(2-t)}{1-t} d\,t$, the difference-sign correlation coefficient is $\tau(\theta) = \theta/(2-\theta)$, the grade correlation coefficient is $\chi(\theta) = \frac{3\,\theta}{4-\theta}$, and the medial correlation coefficient is $v(\theta) = 2^\theta - 1$, all the coefficients increasing from 0 to 1 as θ varies from 0 to 1. The index of dependence is

$$\sup_{x,y} |\Lambda(x,y|\theta) - \Lambda(x)\,y|0) = \frac{\theta}{2-\theta}\,(1-\theta/2)^{2/\theta}$$

increasing also from 0 to $1/4$ as θ increases from 0 to 1.

The use of formulae given for curvilinear regression leads to

$$\bar{y}(x|\theta) = \gamma + \log(1-\theta) + \int_{(1-\theta)e^{-x}}^{+\infty} e^{-t}/t \cdot d\,t - (1-\theta)\exp(\theta\, e^{-x}) \int_{e^{-x}}^{+\infty} e^{-t}/t \cdot d\,t$$

$$= x + \int_0^{(1-\theta)e^{-x}} \frac{1-e^{-t}}{t} d\,t - (1-\theta)\exp(\theta\, e^{-x}) \int_{e^{-x}}^{+\infty} \frac{e^{-t}}{t} d\,t.$$

The form of $\bar{y}(x|\theta)$ is very different from the linear regression $\gamma + \rho(\theta)\,(x-\gamma)$, but the non-linearity index NL oscillates between 0 (from $\theta = 0$ to $\theta = 1$) and .006; the improvement is very small in the passage from the (simple) linear regression to the (more complex) curvilinear regression: in fact, a large difference between the two regressions only happens for very small or very large values of X, whose probability is very small: see Tiago de Oliveira (1974).

Let us now consider statistical decision for this model.

The estimation of the margin parameters is avoided using the correlation coefficients. The test for independence can be done, as before, based on the fact that if $\theta = 0$ (independence) $\sqrt{n}\,r$ is asymptotically standard normal.

The method of quadrants can be used, once more with $p(\theta) = \text{Prob}\{X \leq \tilde{\mu}, Y \leq \tilde{\mu}\} = 2^\theta/4 = 2^{\theta-2}$, increasing from $p(0) = 1/4$ to $p(1) =$

$1/2$. Then $\sqrt{n}\frac{N/n-2\,p(\theta)}{\sqrt{2\,p(\theta)(1-2\,p(\theta))}}$ is asymptotically standard normal and $\theta^{**}=\frac{\log N/n}{\log 2}+1$ (supposing that $N/n\geq 1/2$) is an estimator of θ, with truncation if necessary. Also it is immediate to show that $\sqrt{n}\frac{\log 2}{\sqrt{2^{1-\theta}-1}}(\theta^{**}-\theta)$ and $(\log 2)\sqrt{\frac{n\,N}{n-N}}(\theta^{**}-\theta)$ are asymptotically standard normal.

The Gumbel model for bivariate minima with standard exponential margins was presented, in a different approach, as said before, by Marshall and Olkin (1967). We have $S(x',y'|\theta)=e^{-x'-y'+\theta\min(x',y')}, 0\leq\theta\leq 1$; we have $A(u|\theta)=1-\theta\min(u,1-u)$ a mixture $(\theta,1-\theta)$ of independence $(A(u)=1)$ and diagonal case $(A(u)=\max(u,1-u))$ as before, naturally; and the equisurvival median curve is $x'+y'-\theta\min(x',y')=\log 2$.

We have $B(u|\theta)=1-D(\log\frac{1-u}{u}|\theta)=\frac{(1-\theta)u}{1-\theta\,u}$ if $u<\frac{1}{2}, B(u|\theta)=\frac{u}{1-\theta(1-u)}$ if $u\geq\frac{1}{2}$ with a jump of $\theta/(2-\theta)$ at $u=1/2$, as could be expected from the result for $D(w|\theta)$.

The classical correlation coefficient is $\rho(\theta)=\theta/(2-\theta)$ increasing from $\rho(0)=0$ to $\rho(1)=1$.

Let us now consider the *biextremal model*. It appears naturally in the study of extremal processes; see Tiago de Oliveira (1968).

It can defined directly as follows: consider (X,Z) independent reduced Gumbel random variables and form the new pair (X,Y) with $Y=\max(X-a,Z-b)$, a and b such that $\text{Prob}\{Y\leq y\}=\text{Prob}\{\max(X-a,Z-b)\leq y\}=\Lambda(y)$, which implies $e^{-a}+e^{-b}=1$. Putting $e^{-a}=\theta, e^{-b}=1-\theta(0\leq\theta\leq 1)$, we have $Y=\max(X+\log\theta\,,Z+\log(1-\theta))$. This was the initial use of the max-technique to generate new models and is a particular case of a natural model, to be defined below, with $a=0, b=+\infty, c=-\log\theta$ and $d=-\log(1-\theta)$. The distribution function is $\Lambda(x,y|\theta)=\exp\{-\max(e^{-x}+(1-\theta)e^{-y},e^{-y})\}, (0\leq\theta\leq 1)$ and the dependence function is immediately $k(w|\theta)=1-\min(\theta,e^{w})/(1+e^{w})=(1-\theta+\max(\theta,e^{w}))/(1+e^{w})$ and we have $D(w|\theta)=0$ if $w<\log\theta$ and $(1+(1-\theta)e^{-w})^{-1}$ if $w\geq\log\theta$ with a jump of θ at $w=\log\theta$, which is equal to the $\text{Prob}\{Y=X+\log\theta\}$, and so the singular part is concentrated at the line $y=x+\log\theta$, with probability θ .Evidently $Y\geq X+\log\theta$ with probability one. For $\theta=0$ and $\theta=1$ we obtain independence and the diagonal case. Notice that $k(w|\theta)\neq k(-w|\theta)$ and so the model is not exchangeable.

The classical correlation coefficient of the biextremal model is

$$\rho(\theta) = \frac{6}{\pi^2}\int_0^{\theta} \frac{\log t}{1-t}\, d\, t = 1 + \frac{6}{\pi^2}\, L(\theta),$$

$L(\theta) = \int_1^{\theta} \frac{\log w}{w-1} d\, w$ being the Spence integral; $p(\theta)$ increases from $p(0) = 0$ to $p(1) = 1$. Linear regression is thus $L_y(x) = \gamma + \rho(\theta)\,(x - \gamma)$.

The grade correlation, difference-sign correlation, and medial correlation coefficients are easily seen to be $\chi(\theta) = 3\theta/(2+\theta), \tau(\theta) = \theta$ and $v(\theta) = 2^{\theta} - 1$, all increasing from 0 to 1 as θ varies from 0 to 1.

The index of dependence is

$$\sup_{x,y} |\Lambda(x, y|\theta) - \Lambda(x, y|0)| = \theta/(1+\theta)^{(1+\theta)/\theta},$$

increasing from 0 to $1/4$ as θ increases from 0 to 1.

The curvilinear regressions — here we do not have exchangeability — are

$$\bar{y}(x|\theta) = x + \log\theta + \int_0^{(1-\theta)/\theta \cdot e^{-x}} \frac{1 - e^{-t}}{t}\, d\, t =$$

$$= x + \log(1-\theta) + \int_{\frac{1-\theta}{\theta} e^{-x}}^{+\infty} \frac{e^{-t}}{t} d\, t$$

and

$$\bar{x}(y|\theta) = y - \log\theta - (1-\theta)\exp(\theta\, e^{-\bar{x}}) \int_{\theta\, e^{-x}}^{+\infty} \frac{e^{-t}}{t} d\, t.$$

The comparison with the linear regressions $L_y(x|\theta) = \gamma + \rho(\theta)\,(x - y)$ and $L_x(y|\theta) = \gamma + \rho(\theta)\,(x - \gamma)$ shows that the non-linearity index varies from 0 to 0.068 in the first curvilinear regression and from 0 to 0.007 in the second one: see Tiago de Oliveira (1974). We do not, therefore, have a substantial improvement in using curvilinear regression, although the comparison between $\bar{y}(x|\theta)$ and $L_y(x|\theta)$ and between $\bar{x}(y|\theta)$ and $L_x(y|\theta)$ shows a great difference. The explanation is as before: the small probability mass attached to the regions where the regressions (linear and curvilinear) differ strongly.

The test of independence ($\theta = 0$ vs. $\theta > 0$) can be made by using the classical correlation coefficient as usual.

The quadrants method can be applied as $p(\theta) = \text{Prob}\{X \le \tilde{\mu}, Y \le \tilde{\mu}\} = 2^{\theta}/4$, equal to the expression for the Gumbel model; we can use the techniques described before.

As regards the estimation of θ, as $y_i - x_i \geq \log\theta$ and $0 \leq \theta \leq 1$, the best estimator of the left-end point $\log\theta$ is $\log\tilde{\theta} = \min(y_i - x_i, 0)$ and so $\tilde{\theta} = e^{\min(y_i - x_i, 0)} = \min(e^{y_i - x_i}, 1)$. We have, for $z < 1$, $\text{Prob}\{\tilde{\theta} \leq z|\theta\} = \text{Prob}\{\min_1^n(e^{y_i - x_i}) \leq z|\theta\} = 1 - \text{Prob}\{\min_1^n(y_i - x_i) > \log z|\theta\} = 1 - (1 - D(\log z|\theta))^n$ and so have $\text{Prob}\{\tilde{\theta} \leq z|\theta\} = 0$ if $z < \theta$, $\text{Prob}\{\tilde{\theta} \leq z|\theta\} = 1 - (\frac{z}{1-\theta+z})^n$ if $\theta \leq z < 1$ and $\text{Prob}\{\tilde{\theta} \leq z|\theta\} = 1$ if $z \geq 1$; the jumps have the values $1 - (1-\theta)^n$ at $z = 0$ and $(\frac{1-\theta}{2-\theta})^n$ at $z = 1$; if $\theta \neq 1$ we see that $\tilde{\theta} \xrightarrow{P} \theta$; see Tiago de Oliveira (1970) and (1975b) for other details.

The biextremal model for minima has the survival function $S(x', y'|\theta) = e^{-\max(x' + (1-\theta)y', y')}$, $0 \leq \theta \leq 1, x', y' \geq 0$; we obtain $A(u|\theta) = \max(u, 1 - \theta\, u)$ and $B(u|\theta) = 1 - D\,(\log\frac{1-u}{u}|\theta) = \frac{(1-\theta)\,u}{1-\theta\,u}$ if $u < \frac{1}{1+\theta}$, $B(u|\theta) = 1$ if $u \geq \frac{1}{1+\theta}$ with a jump of θ at $u = 1/(1+\theta)$. The equisurvival median curve is $\max(x' + (1-\theta)\,y', y') = y' + \max(x' - \theta\, y', 0) = \log 2$. The classical correlation coefficient takes the value $\rho(\theta) = \theta$.

Let us now consider the *natural model*. Using the max-technique, from a pair $(\overline{X}, \overline{Y})$ of independent reduced Gumbel random variables we get a new (dependent) random pair $X = \max(\overline{X} - a, \overline{Y} - b)$, $Y = \max(\overline{X} - c, \overline{Y} - d)$ with Gumbel margins if and only if $e^{-a} + e^{-b} = 1$ and $e^{-c} + e^{-d} = 1$. The dependence function is $k(w|a, b, c, d) = (e^{-\min(a, c+w)} + e^{-\min(b, d+w)})(1 + e^w)$, whose tails coincide with those of the diagonal case $k_D(w) = \frac{\max(1, e^w)}{1+e^w}$. It is easy to see that $W = Y - X = \min(a + \overline{W}, b) - \min(c + \overline{W}, d)$ is such that $a - c \leq W \leq b - d$ if $a + d < b + c$ and $b - d \leq W \leq a - c$ if $b + c < a + d$. The case $a + b = b + c$ is irrelevant, because W is then an almost sure random variable with $W = a - c = b - d$, and as W must be zero (its mean value is zero) we have $a = c, b = d, X = Y$.

From now on we will suppose $a + d < b + c$, the case $b + c < a + d$ being dealt with by exchange of X and Y. Let us put $\alpha = a - c, b - d = \beta$; by the fact that $\overline{W}$ has a mean value of zero we have $\alpha < 0 < \beta$. Using the relations $e^{-a} + e^{-b} = 1$ and $e^{-c} + e^{-d} = 1$, with $a - c = \alpha$ and $b - d = \beta$ we get the final expression

$$k(w|\alpha, \beta) = \frac{(e^{\beta} - 1)\max(1, e^{\alpha - w}) + (1 - e^{\alpha})\max(1, e^{\beta - w})}{(e^{\beta} - e^{\alpha})(1 + e^{-w})}$$

or, in more detail,

$k(w|\alpha,\beta) = \frac{1}{1+e^{w}}$ if $w < \alpha$,

$= \frac{e^{\beta}-1+(1-e^{\alpha})\ e^{\beta-w}}{(e^{\beta}-e^{\alpha})(1+e^{-w})}$ if $\alpha \le w \le \beta$,

$= \frac{1}{1+e^{w}}$ if $\beta < w$;

as said before for $w < \alpha$ and for $w > \beta$ we have $k(w|\alpha,\beta) = k_D(w)$.

Note that $k(w|-\beta,-\alpha) = k(-w|\alpha,\beta)$; we get an exchangeable model if $\alpha = -\beta$. As said before note also that for $\alpha = \log\theta, \beta = +\infty$, we get the biextremal model and $\alpha = -\infty, \beta = -\log\theta$ its dual.

Using now the parameters (α,β) ($\alpha < 0 < \beta, \alpha \le W \le \beta$ with probability one), we see that $\Lambda(x,y|\alpha,\beta) = \exp\{-[(e^{\beta}-1)\max(e^{-x}, e^{\alpha-y}) + (1-e^{\alpha})\max(e^{-x}, e^{\beta-y})] /(e^{\beta}-e^{\alpha})\}, -\infty < \alpha \le 0 \le \beta < +\infty$. $D(w|\alpha,\beta)$ is given by:

$D(w|\alpha,\beta) = 0$ if $w < \alpha$

$= \frac{e^{\beta}-1}{e^{\beta}-1+(1-e^{\alpha})\ e^{\beta-w}}$ if $\alpha \le w \le \beta$,

$= 1$ if $\beta \le w$;

with jumps of $\frac{(e^{\beta}-1)e^{\alpha}}{e^{\beta}-e^{\alpha}}$ at $w = \alpha$ and of $\frac{(1-e^{\alpha})}{e^{\beta}-e^{\alpha}}$ at $w = \beta$; $[\alpha,\beta]$ is, thus, the support of $W = Y - X, R = \beta - \alpha$ the range and so when we have $-\infty < \alpha \le 0 \le \beta < +\infty$ there exists, as $\alpha \le W = Y - X \le \beta$, a strong stochastic connection between X and Y; this model thus has flexibility to express various situations.

The classical correlation coefficient can be written is

$$\rho(\alpha,\beta) = 1 - \frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log\frac{k(w|\alpha,\beta)}{k_D(w)}\,dw =$$

$$= 1 - \frac{6}{\pi^2}\int_{\alpha}^{\beta} \log\frac{e^{\beta}-1+(1-e^{\alpha})\,e^{\beta-w}}{(e^{\beta}-e^{\alpha})\max(1+e^{-w})}\,dw;$$

it is immediate that $\rho(\alpha,\beta) \to 1$ as $\alpha,\beta \to 0$ and that $\rho(\alpha,\beta) \to 0$ as $-\alpha,\beta \to +\infty$.

The difference-sign correlation coefficient is

$$\tau(\alpha,\beta) = 1 - \int_{\alpha}^{\beta} D(w|\alpha,\beta)\,(1 - D(w|\alpha,\beta))\,dw =$$

$$= 1 - \frac{(e^{\beta} - 1)(1 - e^{\alpha})}{(e^{\beta} - e^{\alpha})} = \frac{e^{\beta+\alpha} - 2\,e^{\alpha} + 1}{e^{\beta} - e^{\alpha}};$$

the medial correlation coefficient takes the form

$$v(\alpha, \beta) = 2^{(e^{\beta+\alpha} - 2\,e^{\alpha} + 2)/(e^{\beta} - e^{\alpha})} - 1.$$

The equidistribution median curve is

$$(e^{\beta} - 1)\max(e^{-x}, e^{\alpha - y}) + (1 - e^{\alpha})\max(e^{-x}, e^{\beta - y}) = \log 2 \cdot (e^{\beta} - e^{\alpha}).$$

The index of dependence is

$$(1 - k_0)\exp\{\frac{k_0}{1 - k_0}\log k_0\} \text{ where } k_0 = \min(\frac{1}{e^{\alpha}+1}, \frac{e^{\beta}}{e^{\beta}+1}).$$

The linear regressions with reduced margins are given by the straight lines $L_y(x) = \gamma + \rho(\alpha, \beta)(x - \gamma)$ and $L_x(y) = \gamma + \rho(\alpha, \beta)(y - \gamma)$. For the curvilinear regression, as the exchange of X and Y is equivalent to the exchange of $-\alpha$ and β, it is sufficient to compute the curvilinear regression

$$\bar{y}(x|\alpha, \beta) = x + \beta - \frac{(e^{\beta} - 1)}{e^{\beta} - e^{\alpha}}\exp(\frac{1 - e^{\alpha}}{e^{\beta} - e^{\alpha}}e^{-x}) \cdot \int_{\frac{1 - e^{\alpha}}{e^{\beta} - e^{\alpha}}e^{-x}}^{\frac{e^{\beta}}{e^{\alpha}}\frac{1 - e^{\alpha}}{e^{\beta} - e^{\alpha}}e^{-x}} \frac{e^{-t}}{t}\,dt.$$

Analogously to the situation for the biextremal model, we may expect that regression curves are a good approximation, also in the same sense, to the curvilinear regression.

As $\alpha \le w_i \le \beta$, with $\alpha < 0 < \beta$, we can take as estimators of α and β the statistics $\alpha^* = \min(0, y_i - x_i)$ and $\beta^* = \min(0, y_i - x_i)$. We have exchangeability ($k(w|\alpha, \beta) = k(-w|\alpha, \beta)$) iff

$$\begin{aligned} k(w|\alpha, \beta) &= \frac{1}{1+e^{w}} && \text{if} \quad w < -\beta \\ &= \frac{e^{\beta}}{1+e^{\beta}} && \text{if} \quad -\beta \le w \le \beta \\ &= \frac{1}{1+e^{-w}} && \text{if} \quad \beta \le w \end{aligned}$$

and, consequently

$$\begin{aligned} D(w|\beta) &= 0 && \text{if} \quad w < \beta \\ &= \frac{1}{1+e^{-w}} && \text{if} \quad -\beta \le w \le \beta \\ &= 1 && \text{if} \quad \beta \le \alpha, \end{aligned}$$

with jumps of $(1 + e^{\beta})^{-1}$ at $-\beta$ and β.

Also we have

$$\rho(\beta) = 1 - \frac{12}{\pi^2}\int_0^\beta \frac{w}{1+e^{-w}}\, d\, w.$$

$$\tau(\beta) = \frac{2}{1+e^{\beta}}.$$

$$v(\beta) = 2^{2/(1+e^{\beta})} - 1,$$

$$\bar{y}(x|\beta) = x + \beta - \frac{e^{\beta}}{1+e^{\beta}} \exp\frac{e^{-x}}{1+e^{\beta}} \int_{\frac{e^{-x}}{1+e^{\beta}}}^{\frac{e^{\beta-x}}{1+e^{\beta}}} \frac{e^{-t}}{t}\, d\, t\,.$$

and $\beta^* = \max|x_i|$ as estimator of β.

Passing to standard exponential margins and introducing the new parameters $\alpha' = e^{-\alpha} \geq 1 \geq \beta' = e^{-\beta}$, we get

$$S(x', y'|\alpha', \beta') = \exp\{-\frac{(1-\beta')\max(\alpha' x', y') + (\alpha'-1)\max(\beta' x', y')}{\alpha'-\beta'}\};$$

the dependence function

$$A(u|\alpha', \beta') = \frac{(1-\beta')\max(\alpha'(1-u), u) + (\alpha'-1)\max(\beta'(1-u), u)}{\alpha'-\beta'}\ , \text{ i.e.,}$$

$$A(u|\alpha', \beta') = 1 - u \ \text{ if } \ 0 \leq u \leq \frac{\beta'}{\beta'+1}$$

$$= \frac{\alpha'(1-\beta') + (\alpha'\beta'-1)u}{\alpha'-\beta'} \qquad \text{if } \ \frac{\beta'}{\beta'+1} \leq u \leq \frac{\alpha'}{\alpha'+1}$$

$$= u \ \text{ for } \ \frac{\alpha'}{\alpha'+1} \leq u \leq 1.$$

$B(u|\alpha', \beta')$ is equal to 0 if $u < \beta'/(1+\beta')$, to $\frac{(\alpha'-1)u}{(1-\beta')\alpha' - (1-\alpha'\beta')u}$ if $\beta'/(1+\beta') \leq u < \alpha'/(1+\alpha')$ and to 1 when $\alpha'/(1+\alpha') \leq u \leq 1$; the jumps at $u = \beta'/(1+\beta')$ and $u = \alpha'/(1+\alpha')$ are respectively $(\alpha'-1)\beta'/(\alpha'-\beta')$ and $(1-\beta')/(\alpha'-\beta')$. The support for $\frac{Y'}{X'+Y'}$ is thus $[\frac{\beta'}{\beta'+1}, \frac{\alpha'}{\alpha'+1}]$ or equivalently the angle $\beta' X' \leq Y' \leq \alpha' X'$ and not all the first quadrant as in the other examples.

Let us make a quick reference to a generalization of the natural model, the $(2, N)$ *natural model*, a particular case of the (m, N) *natural model* (for $m = 2$) described briefly in the next Chapter, still under study.

Let $Z_1, \ldots, Z_n$ be N independent reduced Gumbel random variables and define $X = \max_1^n (Z_j - p_j), Y = \max_1^n (Z_j - p_j)$.

The random pair (X, Y) has reduced Gumbel margins iff $\sum_1^N e^{-p_j} = 1$ and $\sum_1^N e^{-q_j} = 1$; then we have the distribution function

$$\Lambda(x, y|p, q) = \text{Prob}\{X \leq x, Y \leq y\} = \exp\{\sum_{j=1}^N \max(e^{-x-p_j}, e^{-y-q_j})\},$$

with the dependence function

$$k(w|p, q) = \frac{\sum_1^N \max(e^{-p_j}, e^{-q_j-w})}{1+e^{-w}}.$$

The model thus has $2(N-1)$ parameters. The corresponding model for bivariate minima with standard exponential margins has the survival function $S(x', y'|p', q') = \exp\{-\sum_1^N \max(x'p_j', y'q_j')\}$, where $p_j' = e^{-p_j}, q_j' = e^{-q_j}$, $\sum_1^N p_j' = 1, \sum_1^N q_j' = 1$, with the dependence function

$$A(u|p, q) = \sum_1^N \max(p_j'(1-u), q_j' u).$$

11.4 Intrinsic estimation of the dependence functions

We have given the necessary and sufficient conditions on $k(w)$ and $A(u)$ for $\Lambda(x, y) = \exp\{-(e^{-x} + e^{-y})\, k(y-x)\}$ to be a distribution function of bivariate maxima with reduced Gumbel margins and for $S(x', y') = \exp\{-(x'+y')\, A(\frac{y'}{x'+y'})\}, x', y' \geq 0$ to be the survival function of bivariate minima with standard exponential margins. Let us recall that, for $A(u), u \in [0,1]$, we have $A(0) = A(1) = 1, 0 \leq -A'(0), A'(1) \leq 1$ and $A(u)$ convex.

We will call an *intrinsic estimator*, sometimes also called *non-parametric*, of $A(u)$ $(k(w))$ an estimator (i.e., a statistical function) converging in probability to the function $A(u)$ $(k(w))$, for every $u \in [0,1]$ $(-\infty \leq w \leq +\infty)$ that satisfies the necessary and sufficient conditions to be an $A(u)$ $(k(w))$. In Tiago de Oliveira (1989b) and in Deheuvels and Tiago de Oliveira (1989) a first, and heavy, approach was made. Later, in Tiago de Oliveira (1991), a new and simpler approach was obtained. The procedure presented in Tiago de Oliveira (1989b), whose study was continued in Deheuvels and Tiago de Oliveira (1989), is more complicated than this one; also the estimator given there does not have all moments for any size sample, as opposed to what happens with the new one.

Let us consider, for simplicity, the intrinsic estimation of $A(u)$ from an i.i.d. sample of n pairs $\{(x_i, y_i)\}$ with survival function $S(x, y)$, satisfying *all* necessary and sufficient conditions given above; the conversion to the estimation of $k(w)$ is immediate.

For convenience of the exposé we will use $\Omega(u) = 1 - A(u)$; we have the necessary and sufficient conditions: $\Omega(0) = \Omega(1) = 0, 0 \leq \Omega'(0), -\Omega'(1) \leq 1$ and $\Omega(u)$ concave and also from $\max(u, 1-u) \leq A(u) \leq 1$ we obtain $0 \leq \Omega(u) \leq \min(u, 1-u) = l(u)$.

Let us consider to estimate $\Omega(u)$, the statistical function

$$\overline{\Omega}_n(u) = \overline{\Omega}_n\big(u|x_i', y_i'; \alpha\big) = \frac{c(\alpha)}{n} \sum_1^n \min\left(\frac{1-u}{\alpha + x_i'}, \frac{u}{\alpha + y_i'}\right).$$

where we will take $\alpha(= \alpha_n) \to 0$.

As $\min\left(\frac{1-u}{\alpha+x_i'}, \frac{u}{\alpha+y_i'}\right) \leq \frac{l(u)}{\alpha}$, we see that each summand has all the moments. We will seek function $(s)c = c(\alpha)$ and $\alpha = \alpha_n$ sufficient to guarantee that $\overline{\Omega}_n(u) \xrightarrow{\text{m.s.}} \Omega(u)$. From the properties of each summand we can prove that $\overline{\Omega}_n(0) = \overline{\Omega}_n(1) = 0$ and $\overline{\Omega}_n(u)$ is concave, but the two other conditions on $\Omega'(u)$ are not necessarily verified (the conditions on the derivatives are valid for *each* summand only if $0 < c(\alpha) \leq \alpha$, which will not be the case). As a consequence we will use the truncation $\Omega_n^*(u) = \min\big(\overline{\Omega}_n(u), l(u)\big)$, which is a simple intrinsic estimator of $\Omega(u)$.

Let us begin by considering $\overline{\Omega}_n(u)$ in the complete dependence case:

In the diagonal case, where $X' = Y'$ with probability one, $X'(= Y')$ has the standard exponential distribution.

Then we have

$$M(\overline{\Omega}_n(u)) = M\left(c(\alpha) \min\left(\frac{1-u}{\alpha + x'}, \frac{u}{\alpha + y'}\right)\right) = \left(c(\alpha) \int_0^{+\infty} \frac{e^{-x'}}{\alpha + x'} \, dx\right) l(u) =$$

$$= \left(c(\alpha) e^{\alpha} \int_{\alpha}^{+\infty} \frac{e^{-t}}{t} \, dt\right) \cdot l(u) = c(\alpha) \, J_1(\alpha) \, l(u) =$$

and

$$V\big(\overline{\Omega}_n(u)\big) = \frac{1}{n} V\left(c(\alpha) \min\left(\frac{1-u}{\alpha + x'}, \frac{u}{\alpha + y'}\right)\right) = \frac{c^2(\alpha)}{n} [\, J_2(\alpha) - J_1^2(\alpha)] \, l^2(u)$$

where $J_1(\alpha) = e^{\alpha} \int_{\alpha}^{+\infty} \frac{e^{-t}}{t}\, dt$ and $J_2(\alpha) = e^{\alpha} \int_{\alpha}^{+\infty} \frac{e^{-t}}{t^2}\, dt$; note that $J_1'(\alpha) = J_1(\alpha) - 1/\alpha$ and $J_2'(\alpha) = J_2(\alpha) - 1/\alpha^2$. It is immediate that $J_1(\alpha) \sim -\log \alpha$ and $J_2(\alpha) \sim -1/\alpha$ as $\alpha \to 0$.

Clearly the unbiased estimator of $l(u)$ leads to the choice $c(\alpha)\, J_1(\alpha) = 1$ and the condition that $V\big(\overline{\Omega}_n(u)\big) \to 0$ takes the form $\frac{1}{n}\big(\frac{J_2(\alpha)}{J_1^2(\alpha)} - 1\big) \to 0$ or, as $\frac{J_1^2(\alpha)}{J_2(\alpha)} \to 0$ if $\alpha \to 0$, $\frac{1}{n}\frac{J_2(\alpha)}{J_1^2(\alpha)} \sim \frac{1}{n\,\alpha\,(\log \alpha)^2} \to 0$ as $\alpha = \alpha_n \to 0$.

Let us analyse other choices of $c(\alpha)$ which may be simpler and more convenient. Evidently we must have $c(\alpha)\, J_1(\alpha) \to 1$ and $\frac{c^2(\alpha)}{n}(J_2(\alpha) - J_1^2(\alpha)) \to 0$ or $c(\alpha)\, J_1(\alpha) \to 1$ and, also, $\frac{1}{n\,\alpha\,(\log \alpha)^2} \to 0$. The mean-square error of the estimator is thus

$$\mathrm{MSE} = (c(\alpha)\, J_1(\alpha) - 1))^2\, l^2(u) + \frac{c^2(\alpha)}{n}\Big(J_2(\alpha) - J_1^2(\alpha)\Big)\, l^2(u),$$

which converges to zero under the conditions stated above.

We could seek the function $c(\alpha)$ minimizing the MSE but this has no practical interest.

As $\int_0^{\alpha} \frac{1-e^{-t}}{t}\, dt - \int_{\alpha}^{+\infty} \frac{e^{-t}}{t}\, dt = \gamma + \log \alpha$ we see immediately that $J_1(\alpha) + \log \alpha \to -\gamma$ as $\alpha \to 0$. But the choice $c(\alpha) = 1/(-\gamma - \log \alpha)$ $c(\alpha)\, J_1(\alpha) \to 1)$ has $c(\alpha) \leq 0$ for values of $\alpha \approx 1$. Let us take then $c(\alpha) = \frac{1}{a - \log \alpha}$. The bias is then $(c(\alpha) J_1(\alpha) - 1)\, l(u) = (\frac{J_1(\alpha)}{a - \log \alpha} - 1)\, l(u) \sim \frac{\gamma + a}{\log \alpha}\, l(u)$; a must be chosen such that $a - \log \alpha > 0$ and as we are going to deal with small values of α, we can suppose $\alpha \leq 1$ and then take $a > 0$; note that this is equivalent to using $\alpha' = e^{-a}\, \alpha$ instead of α.

We can then compute asymptotically the mean-square error of the estimator. We have

$$\mathrm{MSE} = \{[(\frac{a+\gamma}{a - \log \alpha})^2 + 0(\frac{1}{(a - \log \alpha)^2})] + [\frac{1}{n\,\alpha (a - \log \alpha)^2} + 0(\frac{1}{n\,\alpha (a - \log \alpha)^2})]\}\, l^2(u)$$

where the first term is the square of the bias and the second corresponds to the asymptotic evaluation of the variance $\frac{J_2(\alpha) - J_1^2(\alpha)}{n\,(a - \log \alpha)^2}$ because $J_2(\alpha) \sim 1/\alpha$ and $J_1(\alpha) \sim -\log \alpha$, as stated previously.

As $\alpha \to 0$, and so $-\log \alpha + \infty$, the condition on $\alpha = \alpha_n$ is to be such that $n\,\alpha (a - \log \alpha)^2$ or $n\,\alpha (\log \alpha)^2 \to +\infty$.

The discussion of the dominant terms in the MSE shows immediately that the squared bias is dominant if $n\,\alpha_n \to \infty$, is of the same order the variance if $n\,\alpha_n \to H(0 < H < +\infty)$, and the variance term is dominant if $n\,\alpha_n \to 0$.

The simplest form for $\alpha = \alpha_n$ is to take $\alpha_n = Rn^{-p}\ (p > 0)$. The condition $n\,\alpha_n(\log \alpha_n)^2 \to +\infty$ immediately imposes $0 < p \leq 1$ and the square of the bias and variance are asymptotic to, respectively, $\frac{(a+\gamma)^2}{p^2\,(\log n)^2}$ and $\frac{1}{R\,p^2\,n^{1-p}\,(\log n)^2}$. Their ratio is asymptotic to $(a+\gamma)^2\,R\,n^{1-p}$ having the squared bias (asymptotic) dominance if $p < 1$ and (asymptotic) equivalence if $p = 1$. In both cases the MSE is of the order of $(\log n)^{-2}$, and it is simpler to take $p = 1$ and so $\alpha_n = R/n$. Then $c(\alpha_n) = \frac{1}{a-\log R-\log n}$ and the condition $c(\alpha_n) > 0$ implies $a - \log R > 0$. We can take $R = 1(\alpha_1 = 1)$ and dispose of a. Taking for simplicity $a = 1$ we get the simple expression $\alpha_n = 1/n$ and $c(\alpha_n) = \frac{1}{1-\log \alpha_n} = \frac{1}{1+\log n}$, and then the squared bias is $\frac{(1+\gamma)^2}{(1+\log n)^2} + 0(\frac{1}{(1+\log n)^2})$, the variance is $\frac{1}{(1+\log n)^2} + 0(\frac{1}{(1+\log n)^2})$ and the MSE is $\frac{1+(1+\gamma)^2}{(1+\log n)^2} + 0(\frac{1}{(1+\log n)^2})$. The convergence, being of the order of $1/(\log n)^2$, is thus very slow. Notice that the condition $c(\alpha) < \alpha$ which was needed for the properties of $\Omega'(0)$ and $\Omega'(1)$ is not verified.

Clearly as $\overline{\Omega}_n(u) \xrightarrow{m.s.} \Omega(u) = l(u)$ then $\overline{\Omega}_n(u) \xrightarrow{P} l(u)$ and so $\Omega_n^*(u) = \min(\overline{\Omega}_n(u), l(u)) \xrightarrow{P} l(u)$.

As a further note, consider two choices: $\alpha_n = R\,n^{-p}$ and $\alpha_n^{'} = R'\,n^{-p'}(0 < p, p' \leq 1)$. The ratio of the variances corresponding to α_n and $\alpha_n^{'}$ is

$$\sim \frac{1/R\,p^2 n^{1-p}(\log n)^2}{1/R'p'^2\,n^{1-p'}(\log n)^2} = \frac{R'p'^2}{R\,p^2}\,n^{p-p'},$$

which converges to zero (variance of smaller asymptotic order) if $n^{p-p'} \to 0$ or $p < p'$, showing that although the variance would be of smaller order with smaller p this effect is outweighed by the bias effect and so the choice of $\alpha_n = \frac{1}{1+\log n}$ is reasonable.

Consider now the general case

$$\bar{\Omega}_n(u) = \frac{c(\alpha)}{n}\sum_1^n \min\left(\frac{1-u}{\alpha + x_i'}, \frac{u}{\alpha + y_i'}\right)$$

where, as seen before, we will take $\alpha_n = 1/n$ and $c(\alpha_n) = \frac{1}{1-\log \alpha_n} = \frac{1}{1+\log n}$.

We have

$$Q_{\alpha(z)} = \text{Prob}\{c(\alpha)\min\left(\frac{1-u}{\alpha + X'}, \frac{u}{\alpha + Y'}\right) \geq z\} =$$

$$= \text{Prob}\{X' \leq c(\alpha)\frac{(1-u)}{z} - \alpha, Y' \leq c(\alpha)\frac{u}{z} - \alpha\}$$

$$= 1 - S\left(\frac{c(1-u)}{z} - \alpha, 0\right) - S\left(0, \frac{c,u}{z} - \alpha\right) + S\left(\frac{c(1-u)}{z} - \alpha, \frac{c\cdot u}{z} - \alpha\right).$$

Then the mean value of $c(\alpha)\min\left(\frac{1-u}{\alpha+X'}, \frac{u}{\alpha+Y'}\right)$ is given by $\int_0^{+\infty} z\, d(1 - Q_\alpha(z)) =$

$\int_0^{+\infty} z\, d\, S\left(\frac{c(1-u)}{z} - \alpha, 0\right) + \int_0^{+\infty} z\, d\, S\left(\frac{c\, u}{z} - \alpha, 0\right) - \int_0^{+\infty} z\, d\, S\left(\frac{c(1-u)}{z} - \alpha, \frac{c\ u}{z} - \alpha\right) = I_1(\alpha) + I_2(\alpha) - I_3(\alpha).$

Let us then compute the summands. We have

$$I_1(\alpha) = \int_0^{+\infty} z\, d\, S\left(\frac{c(1-u)}{z} - \alpha, 0\right) = \int_0^{+\infty} z\, d\left(e^{-((c(1-u)/z)-\alpha)_+}\right) =$$

$$\int_{(c(1-u)/z)-\alpha>0}^{+\infty} z\, d\, \left(e^{-\left((c(1-u)/z)-\alpha\right)}\right) = e^{\alpha}\, c(1-u) \int_\alpha^{+\infty} \frac{e^{-\xi}}{\xi}\, d\, \xi$$

and

$$I_2(\alpha) = \int_0^{+\infty} z\, d\, S\left(0, \frac{c\, u}{z} - \alpha\right) = \int_0^{+\infty} z\, d\left(e^{-(((c\, u)/z)-\alpha)_+}\right) =$$

$$\int_{((c\, u)/z)-\alpha>0}^{+\infty} z\, d\left(e^{-\left(\left((c(1-u)/z)-\alpha\right)\right)}\right) = e^{\alpha}\, c\, u \int_\alpha^{+\infty} \frac{e^{-\xi}}{\xi}\, e\, \xi$$

and so $I_1(\alpha) + I_2(\alpha) = e^{\alpha}\, c(\alpha)\, J_1(\alpha) \to 1$.

Now compute $I_3(\alpha)$. We will show that $I_3(\alpha) \to A(u)$ and so $M(\bar{\Omega}_n(u)) = I_1(\alpha) + I_2(\alpha) - I_3(\alpha) \to 1 - A(u) = \Omega(u)$.

We have

$$I_3(\alpha) = \int_0^{+\infty} z\, d[\exp\{-[(\frac{c(1-u)}{z} - \alpha)_+ + (\frac{c\,u}{z} - \alpha)_+]\, A(\frac{(\frac{c\,u}{z}-\alpha)_+}{(\frac{c(1-u)}{z}-\alpha)_+ + (\frac{c\,u}{z}-\alpha)_+})\}].$$

Let us suppose, for convenience, that $0 \le u \le 1-u \le 1$. Then $(\frac{c(1-u)}{z} - \alpha)_+$ is different from zero when $c(1-u) > \alpha\, z$ and $(\frac{c\,u}{z} - \alpha)_+$ is different from zero when $c\,u > \alpha\, z$. Then $I_3(\alpha)$ can be written as

$$I_3(\alpha) = \int_0^{cu/\alpha} z\, d[\, e^{-(c/z-2\alpha)\, A((cu/z)-\alpha)/((c/z)-2\alpha))}] + \int_{cu/\alpha}^{c(1-u)/\alpha} z\, d(e^{-((c(1-u)/z)-\alpha)}).$$

The second integral is

$$c(\alpha)\, e^{\alpha}(1-u) \int_{\alpha}^{((1-u)/u)\alpha} \frac{e^{-\xi}}{\xi}\, d\,\xi = \frac{e^{\alpha}}{1-\log\,\alpha}(1-u) \int_{\alpha}^{((1-u)/u)\alpha} \frac{e^{-\xi}}{\xi}\, d\,\xi \to 0$$

as $\alpha \to 0$.

Let us consider the first integral. It can be written as

$$-\int_0^{c\,u/\alpha} z\, d(1 - e^{-(c/z-2\,\alpha)\, A((cu/z-\alpha)/(c/z-2\,\alpha))}),$$

which, by integration by parts, takes the form

$$-[z(1 - e^{-(c/z-2\,\alpha)\, A((c\,u/z-\alpha)/(c/z-2\,\alpha))})|_0^{c\,u/\alpha}] +$$

$$+ \int_0^{c\,u/\alpha} (1 - e^{-(c/z-2\,\alpha)\, A((c\,u/z-\alpha)/(c/z-2\,\alpha))}) d\, z$$

$$=$$

$$-\frac{c(\alpha)u}{\alpha}(1 - e^{-(1/u-2)\alpha}) + \int_0^{c\,u/\alpha}(1 - e^{-(c/z-2\,\alpha)\, A((c\,u/z-\alpha)/(c/z-2\,\alpha))})\, d\, z.$$

The first summand, as $c(\alpha) = \frac{1}{1-\log\,\alpha}$, converges to zero. Let us, then, study the integral. Now put $\xi = c/z$. The integral takes the form

$$= \int_{\alpha/u}^{+\infty} c(\alpha)(1 - e^{-(\xi-2\,\alpha)\, A((\xi u-\alpha)/(\xi-2\,\alpha))}\, \frac{d\,\xi}{\xi^2}.$$

But as $u \le 1-u$ $(u \le 1/2)$ we have $\frac{\xi u-\alpha}{\xi-2\,\alpha} < u$ and as $A(u)/u$ is non-increasing we obtain

$$\frac{A(\frac{\xi u-\alpha}{\xi-2\,\alpha})}{\frac{\xi u-\alpha}{\xi-2\,\alpha}} \ge \frac{A(u)}{u} \text{ or } (\xi - \alpha/u)\, A(u) \le (\xi - 2\,\alpha)\, A(\frac{\xi u-\alpha}{\xi-2\,\alpha});$$

and from $\frac{A(u)}{1-u}$ non-decreasing we get $(\xi - 2\,\alpha)\, A(\frac{\xi u-\alpha}{\xi-2\,\alpha}) \le (\xi - \frac{\alpha}{1-u})\, A(u)$ and, thus, the integral is bounded above and below by

$\int_{\alpha/u}^{+\infty} c(\alpha)(1-e^{-(\xi-\alpha/(1-u))\,A(u)})\,\frac{d\,\xi}{\xi^2}$ and $\int_{\alpha/u}^{+\infty} c(\alpha)(1-e^{-(\xi-(\alpha/u))\,A\,(u)})\,\frac{d\,\xi}{\xi^2}$.

Putting $\xi\,A(u)=t$ we get, respectively, the following bounds above and below:

$A(u)\,c(\alpha)\int_{\alpha A/u}^{+\infty}(1-e^{\alpha A/(1-u)-t})\,\frac{d\,t}{t^2}$ and $A(u)\,c(\alpha)\int_{\alpha A/u}^{+\infty}(1-e^{(\alpha A/u)-t})\,\frac{d\,t}{t^2}$.

But $\int_{p\alpha}^{+\infty}(1-e^{q\alpha-t})\frac{d\,t}{t^2}$, with $p,q>0$, by integration by parts, has the expression $\frac{1-e^{(q-p)\alpha}}{p^\alpha}+e^{q\alpha}\int_{p\alpha}^{+\infty}\frac{e^{-t}}{t}\,d\,t$ and with $c(\alpha)=\frac{1}{1-\log\,\alpha}$ we see immediately that $c(\alpha)\int_{p\alpha}^{+\infty}(1-e^{q\alpha-t})\frac{d\,t}{t^2}\to 1$, and so both bounds converge to $A(u)$. Thus, as said initially, $I_3(\alpha)\to A(u)$, and so the mean value $I_1(\alpha)+I_2(\alpha)-I_3(\alpha)\to 1-A(u)=\Omega(u)$. The proof when $1-u\le u$ leads to the same result.

By the same technique we can show that

$$V(c(\alpha)\min(\tfrac{1-u}{\alpha+X'},\tfrac{u}{\alpha+Y'}))\sim\frac{c^2(\alpha)}{\alpha}C(u),$$

when $C(u)$ is a bounded function of u.

The condition of convergence is, once more, $V(\overline{\Omega}_n(u))\sim\frac{C(u)}{n\,\alpha(\log\,\alpha)^2}\to 0$ or $n\,\alpha(\log\alpha)^2\to 0$; as before for the diagonal case the previous analysis applies.

The final estimator is, as before, $\Omega_n^*(u)=\min\big(\overline{\Omega}_n(u),l(u)\big)\xrightarrow{P}\Omega(u)=1-A(u)$.

We have, then, for bivariate minima with standard exponential margins, as $\alpha_n=1/n$,

$$\overline{A}_n(u)=1-\frac{1}{1+\log n}\sum_1^n\min(\frac{1-u}{1+n\,x_i'},\frac{u}{1+n\,y_i'}),$$

and, more generally,

$$A_n^*(u)=1-\min(l(u),\frac{1}{1+\log n}\sum_1^n\min(\frac{1-u}{1+n\,x_i'},\frac{u}{1+n\,y_i'})),$$

Recall that $\overline{A}_n(u)$ can be used if $0\le-\overline{A}_n'(0),\overline{A}_n'(1)\le 1$ that is iff

$$0\le\sum_1^n\frac{1}{1+n\,x_i'},\sum_1^n\frac{u}{1+n\,y_i'}\le 1+\log n,$$

and, in that case, we get

$\bar{S}_n(x,y) = e^{-(x+y)\,\bar{A}_n(y/(x+y))} = e^{-(x+y)+1(1+\log n)} \cdot \sum_1^n \min(x/(1+n\,x_i^{'}), y/(1+n\,y_i^{'}))$.

As $\bar{A}_n(u) \xrightarrow{P} A(u)$ and $A_n^*(u) = \max(\bar{A}_n(u), 1-l(u)) \xrightarrow{P} A(u)$, we see that $\bar{S}_n(x',y') = \exp\{-(x'+y')\,\bar{A}_n(\frac{y'}{x'+y'})\} \xrightarrow{P} S(x',y')$ and also $S_n^*(x',y') = \exp\{-(x'+y')\,A_n^*(\frac{y'}{x'+y'})\} \xrightarrow{P} S(x',y')$.

For bivariate maxima with reduced Gumbel margins we have $\bar{k}_n(w) = \bar{A}_n(\frac{1}{1+e^w})$ and $k_n^*(w) = A_n^*(\frac{1}{1+e^w})$. Denoting by $\{(x_i,y_i)\}$ the sample we have (with the transformation (x,y) to $(x',y') = (e^{-x}, e^{-y})$ as stated,

$\bar{k}_n(w) = 1 - \frac{1}{(1+\log n)(1+e^w)} \sum_1^n \min(\frac{e^w}{1+e^{-n\,x_i}}, \frac{1}{1+e^{-n\,y_i}})$ and

$\bar{A}_n(x,y) = \exp\{-(e^{-x}+e^{-y}) + \frac{1}{1+\log n} \sum_1^n \min(\frac{e^{-x}}{1+n\,e^{-x_i}}, \frac{1}{1+n\,e^{-y_i}})\}$.

Pickands (1981) gave a much more complex estimator, converging with probability one but with a very tedious calculation.

11.5 Miscellaneous statistical results:

A ready-to-wear model:

In Ramberg, Dudewicz, Tadikamalle and Mytyka (1979) a family of quantile functions $Q(p) = F^{-1}(p)$ is proposed that the authors say fit the data well; this family, of the form (where $0^0 = 1$)

$$Q(p) = \lambda + \delta(p^\alpha - (1-p)^\beta),$$

extends Tukey's λ-functions and so contains the logistic distribution. As the latter plays a central role for $D(w)$ functions — corresponding to independence — it seems natural to study the quantile functions $Q(p)$ that verify the conditions to be a $D(w)$ function. First, as $D(w)$ has all moments $\mu_k^{'} = \int_0^1 Q^k(p)\,dp, k \geq 0$, we see immediately that we must have $\alpha \geq 0$ and $\beta \geq 0$; in the contrary case for k integer, such that $k\alpha + 1 < 0 \leq (k-1)\alpha + 1$ or $k\beta + 1 < 0 \leq (k-1)\beta + 1$, the integrals $\int_0^1 p^{k\alpha}\,dp$ or $\int_0^1 (1-p)^{k\beta}\,dp$ would diverge. Then for $Q(p)$ to be non-decreasing we must have $\delta > 0$ (the case $\delta = 0$ leads to the almost sure random variable at λ, a case that can also be obtained with $\alpha = 0$ and $\beta = 0$). It is immediate that the conditions $\int_{-\infty}^{+\infty} w\,d\,D(w) = 0$ and $D'(w) \geq D(w)\big(1 - D(w)\big)$ are equivalent

to $\int_0^1 Q(p)\,dp = 0$ and $(0 \le)\ p(1-p)\,Q'(p) \le 1$. The condition $\int_0^1 Q(p)\,dp = 0$ gives

$$\lambda + \delta\left(\frac{1}{\alpha+1} - \frac{1}{\beta+1}\right) = 0,$$

and so

$$Q(p) = \delta\left\{\left(p^{\alpha} - \frac{1}{\alpha+1}\right) + \left((1-p)^{\beta} - \frac{1}{\beta+1}\right)\right\},$$

$$Q(0) = \delta\left\{\left(0^{\alpha} - \frac{1}{\beta+1}\right) - \left(1 - \frac{1}{\beta+1}\right)\right\},$$

$$Q(1) = \delta\left\{\left(1 - \frac{1}{\alpha+1}\right) - \left(0^{\beta} - \frac{1}{\beta+1}\right)\right\}.$$

The support is $[Q(0), Q(1)]$ and the range $R = \delta\left(2 - 0^{\alpha} - 0^{\beta}\right)$, i. e., $R = 2\,\delta$ if $\alpha > 0$ and $\beta > 0$, $R = \delta$ if $\alpha > 0$ and $\beta = 0$ or $\alpha = 0$ and $\beta = 1$ and $R = 0$ in the degenerated case where $W(= Y - X) = 0$ with probability 1.

Let us now consider the condition $p(1-p)\,Q'(p) \le 1$ for each $p \in [0,1]$ which is equivalent to $\delta\ \sup_{0 \le p \le 1}[\alpha\,(p^{\alpha}(1-p) + \beta\,(1-p)^{\beta}] \le 1$ relating the three parameters $\alpha \ge 0, \beta \ge 0$ and $\delta > 0$.

We will consider only a limiting case when $\delta \to +\infty$ that already includes the logistic distribution; note that, as we have shown with the natural model, we can have cases where the range of $W = Y - X$ is a finite one. As $\frac{p^{\alpha} - 1/(\alpha+1)}{\alpha}$ has a limit $(1 + \log p)$ if $\alpha \to 0$, then $\delta\left(p^{\alpha} - \frac{1}{\alpha+1}\right)$ has a limit $A(1 + \log p)$ iff $\delta\alpha \to A (0 \le A < +\infty)$; analogously we shall have $\delta\beta \to B (0 \le A < +\infty)$ and the limiting quantile function is $A(1 + \log p) - B\left(1 + \log(1-p)\right), A \ge 0, B \ge 0$.

Note that $A = 0$ gives the exponential $p = 1 - e^{-(1+w/B)}$ if $w \ge -B$ and $p = 0$ for $w \le -B$; for $B = 0$ we obtain $p = e^{(w/A)-1}$ if $w \le A$ and $p = 1$ if $w \ge A$; for $A = B$ we get the logistic distribution $p = (1 + e^{-w/A})^{-1}$.

The condition $0 \le p(1-p)\,Q'(p) \le 1$ takes the form $0 \le A(1-p) + Bp \le 1$ and so we get the conditions on A and B, $0 \le A, B \le 1$.

The estimation of A and B can easily be made by least squares, i.e., associating the ordered statistics, $w'_{i,n}$ to the plotting positions $p_{i,n}$, and thus minimizing $\sum_1^n [w'_{i,n} - (A(1 + \log p_{i,n}) - B(1 + \log(1 - p_{i,n}))]^2$ and obtaining systematic (i.e., linear on the ordered statistics) estimators. There is

no special reason to use this method, which also would introduce the question of plotting positions.

Let us compute some correlation coefficients. We have

$$\rho = 1 - \frac{3}{\pi^2}\int_{-\infty}^{+\infty} w^2 \, d\, D(w) = 1 - \frac{3}{\pi^2}\int_0^1 Q^2(p) d\, p$$

$$= 1 - \frac{3}{\pi^2}\{A^2 + B^2 - 2(1 - \frac{\pi^2}{6})\, A\, B\}$$

as

$$\int_0^1 \log\, p\, \, d\, p = \int_0^1 \log(1-p)\, d\, p = -1, \int_0^1 (1 + \log p)^2\, d\, p$$

$$= \int_0^1 (1 + \log(1-p))^2\, d\, p = 1$$

and

$$\int_0^1 \log p \log(1-p)\, d\, p = \frac{\partial^2 B(a,b)}{\partial\, a\, \partial\, b}|_{a=b=1} = 2 - \pi^2/6.$$

We have then $\rho(A,B) = 1 - AB - \frac{3}{\pi^2}(A-B)^2 = 1 + (\frac{12}{\pi^2} - 1)\, AB - \frac{3}{\pi^2}\,(A+B)^2$.The logistic model $(A = B)$ has then $\rho = 1 - A^2 (\theta = 1 - A$ was the parameter previously used) oscillating between $0 (A = 1)$ and $1 (A = 0)$.

Also $\tau = 1 - \int_{-\infty}^{+\infty} w\, d\, D^2(w) = 1 - \int_0^1 Q(p) d\, p^2 = 1 - \frac{A+B}{2}$. Then taking $\tilde{\rho}^* = \max(0, \rho^*)$ and $\tilde{\tau}^* = \max(0, \tau^*)$ where ρ^* and τ^* are the usual estimators of ρ and τ, and the truncation by zero results from the fact that $\rho \geq 0$ and $\tau \geq 0$, the equations $\tilde{\rho}^* = \rho(A^*, B^*)$ and $\tilde{\tau}^* = \tau(A^*, B^*)$ do not define uniquely A^* and B^* because of the symmetry of $\rho(A,B)$ and $\tau(A,B)$ in A and B, but give two alternatives for (A^*, B^*) corresponding to the choice for A^* of the smallest or the largest of a second degree equation and dually for B^*.

A test of the logistic model as $\rho = 1 - A^2$ and $\tau = 1 - A$ is a test of $\rho = \tau(2 - \tau)$.

The use of quantiles can estimate A and B in an unique way.

As A and B play symmetrical roles it is natural to use quantiles of probabilities p and $q = 1 - p (> p \Leftrightarrow p \leq 1/2)$ and solve the equations

$$Q^*(p) = A^*(1 + \log p) - B^*(1 + \log q)$$

$$Q^*(q) = A^*(1 + \log q) - B^*(1 + \log p).$$

We get

$$A^* = \frac{(1 + \log q)\, Q^*(q) - (1 + \log p) Q^*(p)}{(1 + \log q)^2 - (1 + \log p)^2}$$

$$B^* = \frac{(1 + \log q)\, Q^*(p) - (1 + \log p) Q^*(q)}{(1 + \log q)^2 - (1 + \log p)^2}$$

(A^*, B^*) as linear combinations of the quantiles $Q^*(p)$ and $Q^*(q)$ have an asymptotically binomial distribution which can be obtained by the δ-method. It does not seem worthwhile to advance this line of thought until it is felt that this "ready-to-wear" approach is a good one.

11.6 The estimation of the association between the margins

The fact that we may use an intrinsic estimator of $A(u)$ or $k(w)$, even supposing that we are using $\overline{A}_n(w)$ or $\overline{k}_n(w)$, and not $A_n^*(u)$ or $k_n^*(w)$, does not facilitate the evaluation of the association. It is sufficient to recall that $\rho(k) = -\frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log k(w)\, dw$ or $\rho(A) = \int_0^1 A^{-2}(w) du - 1$ to see that they are awkward to obtain except for numerical computation.

As $\max(u, 1-u) \le A(u) \le 1$ and $\frac{\max(1, e^w)}{1+e^w} \le k(w) \le 1$, we can use as association indicators the ratios of areas between $A(u)$ or $k(u)$ and independence suitably normalized:

$$\xi(A) = \frac{\int_0^1 (1 - A(u))\, d\, u}{\int_0^1 (1 - \max(u, 1-u))\, d\, u} = 4 \int_0^1 (1 - A(u))\ d\, u$$

and

$$\xi(k) = \frac{\int_{-\infty}^{+\infty} (1 - k(w))\, d\, w}{\int_{-\infty}^{+\infty} (1 - \frac{\max(1, e^w)}{1+e^w}) d\, w} = \frac{1}{2 \log 2} \int_{-\infty}^{+\infty} \big(1 - k(w)\big)\ d\, w\ ;$$

note that the last coefficient could be suggested by the use of the approximation $1 - k(w)$ to $-\log k(w)$ which, as $1/2 \le k(w) \le 1$, has the following bounds

$$0 \le -\log k(w) - (1 - k(w)) \le -\log(1/2) - (1/2) = .1931471,$$

and so the renormalization substitutes the coefficient $\frac{6}{\pi^2} = .6079271$ by the coefficient $\frac{1}{2 \log 2} = .7213475$; it is immediate, by integrating the relation above, that $0 \le 2 \log 2 \cdot \xi(k) \le \frac{6}{\pi^2} \rho(k)$.

We get the estimators

$$\bar{\xi}_n(A) = 4\int_{-\infty}^{+\infty}(1-\bar{A}_n(w)(u))d\,u$$

$$= \frac{4}{1+\log n}\sum_1^n \int_0^1 \min\left(\frac{1-u}{1+n\,x_i'}, \frac{u}{1+n\,y_i'}\right) d\,u =$$

$$= \frac{2}{1+\log n}\sum_1^n \frac{1}{2+n(x_i'+y_i')}$$

and

$$\bar{\xi}_n(k) = \frac{1}{2\,\log 2}\int_{-\infty}^{+\infty}(1-\bar{k}_n(w))d\,w$$

$$= \frac{1}{2\,\log 2\cdot(1+\log n)}\sum_1^n \int_{-\infty}^{+\infty}\frac{1}{1+e^w}\min\left(\frac{e^w}{1+e^{-x_i}}, \frac{1}{1+e^{-y_i}}\right) d\,w$$

$$= \frac{1}{2\,\log 2\cdot(1+\log\,n)}\sum_1^n\left[\frac{1}{1+n\,e^{-x_i}}\log\left(1+\frac{1+n\,e^{-x_i}}{1+n\,e^{-y_i}}\right)+\right.$$

$$\left.\frac{1}{1+n\,e^{-y_i}}\log\left(1+\frac{1+n\,e^{-y_i}}{1+n\,e^{-x_i}}\right)\right]$$

11.7 A worked example

Consider the data for the flood discharges of the Ocmulgee River in Georgia, USA at Hawkinsville (upstream) and Macon (dowstream), for the 40 years 1910/1949, contained in Table 1 taken from Gumbel and Goldstein (1964).

The margins data can be accepted to have a Gumbel distribution and the ML estimators of the parameters are $\hat{\lambda}_H = 23.709, \hat{\delta}_H = 15.057, \hat{\lambda}_M = 26.378$ and $\hat{\delta}_M = 17.042$. The empirical Kolmogoroff-Smirnov statistics for the margins are $KS_H = .09442$ and $KS_M = .06228$ with $\sqrt{n}\,KS_H = .597$ and $\sqrt{n}\,KS_M = .394$ accepting thus the Gumbel model for the margins. The better analysis is to use the $\hat{V}_n$ statistics for each margin. The medians are $\tilde{\mu}_H = 30.15$ and $\tilde{\mu}_M = 31.80$; note that the years ranked 20 and 21 in both stations are not the same. The number of concordant pairs (for the quadrants method) is $N = 34$; the value given in Gumbel and Goldstein (1964) is $N = 33$ because they estimate the medians through an inefficient estimation of the margin parameters. The correlation coefficient is $\rho^* = .942$.

Table 1.

Year	Hawkinsville (1000 ft^3/s)	Macon (1000 ft^3/s)	Year	Hawkinsville (1000 ft^3/s)	Macon (1000 $ft^{3/}s$)
1910	18.8	28.8	1930	50.0	64.4
1911	5.9	8.5	1931	12.2	10.7
1912	44.4	44.8	1932	16.2	19.6
1913	52.0	51.0	1933	19.9	19.0
1914	5.9	4.8	1934	17.4	16.9
1915	20.1	19.1	1935	13.5	22.7
1916	40.4	47.8	1936	61.0	65.3
1917	27.0	25.4	1937	25.8	33.3
1918	14.3	14.3	1938	33.0	31.0
1919	40.0	31.0	1939	37.9	33.9
1920	45.2	66.2	1940	13.3	14.2
1921	30.0	37.0	1941	6.9	7.3
1922	44.0	48.6	1942	57.0	73.4
1923	30.3	28.3	1943	41.6	44.8
1924	15.2	21.0	1944	46.8	50.2
1925	79.0	72.5	1945	26.2	40.4
1926	19.3	28.3	1946	35.4	57.6
1927	7.6	7.9	1947	34.8	32.6
1928	42.4	47.1	1948	28.2	24.0
1929	70.5	73.4	1949	68.0	84.0

The estimation equation $2p^{**} = \frac{N}{n} = \frac{N_1 + N_2}{n}$ imposes, as $\frac{1}{4} \leq p^{**} \leq \frac{1}{2}, \frac{1}{2} \leq \frac{N}{n} \leq 1$ which is acceptable as $N/n = 34/40 = .85$ and $p^{**} = N/2n = .425$. Clearly this value of p^* rejects independence as can be seen.

As for the mixed model $p_M(\theta) = 2^{-2+\theta/2}, \max p_M(\theta) = p_M(1) = 2^{-3/2} = .354 < .425$, the mixed model should be rejected or, accepting it, take $\theta = 1$; the distance between $.354$ and $.425$ suggests avoiding the truncation. Although Posner, Rodemich, Ashlock and Lurie (1969) did find an application of the mixed model — see the data and the strip method in their paper — we may expect the model to be useful only for weak correlations.

Consider now the logistic model. As $p_L(\theta) = \exp\{-2^{1-\theta} \log 2\}$ has the maximum $p_L(1) = 1/2$, the model can be accepted. In that case we have $\theta^{**} = .696$. As $p_L(\theta) = \theta(2 - \theta)$ we get $\theta^* = 1 - \sqrt{1 - \rho^*} = .759$ close to θ^{**} as $\theta^* - \theta^{**} = .063$. But better estimators of θ must be found.

From the practical point of view it seems that, for the differentiable models, the logistic one seems to give a good fit in general. But it must be recalled that Tiago de Oliveira (1987) has shown that the logistic and the

natural models can be very close and difficult to separate for moderate samples. The latter is easily generalizable for asymmetric situations and even for more dimensions, which is not the case with the logistic model.

References

Deheuvels, P. and Tiago de Oliveira, J., 1989. On the non-parametric estimation of the bivariate extreme-value distributions. *Statist. and Prob. Letters*, 8, 315-323.

Gumbel, E. J. and Goldstein, N., 1964. Analysis of empirical bivariate extremal distributions. *J. Amer. Statist. Assoc*., 59, (307), 794-816.

Gumbel, E. J. and Mustafi, C. K., 1967. Some analytical properties of bivariate extremal distributions. *J. Amer.Statist . Assoc*., 62, 569-588.

Gumbel, E. J., 1962. Statistical theory of extreme values (main results). *Contributions to Order Statistics*, Ahmed R. Sarhan and Bernard G. Greenberg eds., 56-93, Wiley, New York.

Gumbel, E. J., 1966. Two systems of bivariate extremal distributions. *Bull. Int. Statist. Inst*., 41st. sess., Paris.

Marshall, A. W. and Olkin, I., 1967. A multivariate exponential distribution. *J. Amer.Statist . Assoc*., 62, (317), 30-44.

Pickands, J. III., 1981. Multivariate extreme value theory. *Bull. Int. Statist. Inst*., 43rd sess.,II, 859-878, Buenos Aires.

Posner, Edward C., Rodemich, Eugene R., Ashlock, John C. and Lurie,Sandra., 1969. Application of an estimator of high efficiency in bivariate extreme value theory. *J. Amer.Statist. Assoc*., 64, (328), 1403-1414.

Ramberg, J. S., Dudewicz, E. J., Tadiicamalla, P. R. and Mytytka, E. F., 1979. A probability distribution and its uses in fitting data. *Technometrics*, 21, (2), 201-214.

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

Tiago de Oliveira, J, 1964. L'indépendance des distributions extrêmales bivariées. *Publ. Inst . Statist . Univ. Paris*, XIII, 137-141.

Tiago de Oliveira, J, 1965. Statistical decision for bivariate extremes. *Portugaliae Math*, 24, 145-154.

Tiago de Oliveira, J, 1968. Extremal processes; definition and properties. *Publ. Inst. Statist. Univ. Paris*, XII, 25-36.

Tiago de Oliveira, J., 1970. Biextremal distributions: statistical decision. *Trab. Inv. Estad. Oper*., XXI, 1 07-117, Madrid.

Tiago de Oliveira, J., 1971. A new model of bivariate extremes; statistical decision. *Studi di Probabilitá, Statistica e Ricerca Operativa in onore di Giuseppe Pompilj*, Tip. Oderisi, Gubbio, 1-13.

Tiago de Oliveira, J, 1974. Regression in the nondifferentiable bivariate extreme models. *J. Amer. Statist. Assoc*., 69, 816-818.

Tiago de Oliveira, J, 1975a. Statistical decision for extremes. *Trab. Estad. Inv. Oper*., XXVI, 453-471, Madrid.

Tiago de Oliveira, J, 1975b. Bivariate and multivariate extreme distributions. *Statistical Distributions in Scientific Work*, G. P. Patil, S. Kotz and J. K. Ord eds., 1, 355-361, D.Reidel, Dordrecht.

Tiago de Oliveira, J, 1980. Bivariate extremes: foundations and statistics. *Multivariate Analysis -V*, P. R. Krishnaiah ed., 349-368, North-Holland, Amsterdam.

Tiago de Oliveira, J, 1984. Bivariate models for extremes: statistical decision. *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 131-153, D. Reidel, Dordrecht.

Tiago de Oliveira, J, 1985. Statistical choice of non-separated models. *Trab. Estad. Inv.Operat*., Madrid, XXX VI, 136-152.

Tiago de Oliveira, J, 1987. Comparaison entre les modèles bivariés logistique et naturel pour les maxima et extensions. *C. R. Acad. Sc. Paris*, 305, ser I, 481-484.

Tiago de Oliveira, J, 1989a. Intrinsic estimation of dependence structure for bivariate extremes. *Statist. and Prob. Letters*, 8, 213-218.

Tiago de Oliveira, J, 1989b. Statistical decision for bivariate extremes. *Extreme Value Theory*, Proceedings, Oberwolfach 1987, 246-261, Springer.

Tiago de Oliveira, J., 1989c. Bivariate models for floods of rivers. Presented at the 47th session of the I.S.I., Paris.

Tiago de Oliveira, J, 1991. Intrinsic estimation of the dependence functions in bivariate extremes; a new and simpler approach. To be publ. in *Commun. Statist.-Theor. Meth*.

*****