AD-A120 208 NORTH CAROLINA UNIV AT CHAPEL HILL CENTER FOR STOCHAS--ETC F/G 12/1
BIVARIATE EXTREMES: MODELS AND STATISTICAL DECISION.(U)
JUN 82 J T OLIVEIRA F49620-82-C-0009
UNCLASSIFIED TR-14 AFOSR-TR-82-0853 NL

1 OF 1
AD A
120208

END
DATE
FILMED
11-82
DTIC

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

BIVARIATE EXTREMES: MODELS AND STATISTICAL DECISION

by

J. Tiago de Oliveira

TECHNICAL REPORT #14

June 1982

F49620-82-C-0009

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 82-0853	2. GOVT ACCESSION NO. AD-A120 208	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) BIVARIATE EXTREMES: MODELS AND STATISTICAL DECISION		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER TR#14
7. AUTHOR(s) J. Tiago de Oliveira		8. CONTRACT OR GRANT NUMBER(s) F49620-82-C-0009
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill NC 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE61102F; 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Directorate of Mathematical & Information Sciences Air Force Office of Scientific Research Bolling AFB DC 20332		12. REPORT DATE June 1982
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bivariate extremes; asymptotic distributions; dependence function; differentiable and non-differentiable models for (asymptotic) bivariate maxima; statistical decision; estimation and testing of the dependence parameters; test of independence; regression.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

After obtaining the asymptotic distribution of bivariate maxima, a direct characterization of the asymptotic distribution is given; the 5 known models are described through their dependence functions and some properties obtained. Known statistical decision results for the models are described.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BIVARIATE EXTREMES:

MODELS AND STATISTICAL DECISION

by

J. Tiago de Oliveira
Department of Statistics, Faculty of Sciences of Lisbon
Center for Statistics and Applications (I.N.I.E.)
Academy of Sciences of Lisbon
and
University of North Carolina at Chapel Hill

Summary: After obtaining the asymptotic distribution of bivariate maxima, a direct characterization of the asymptotic distribution is given; the 5 known models are described through their dependence functions and some properties obtained. Known statistical decision results for the models are described.

Key words: Bivariate extremes; asymptotic distributions; dependence function; differentiable and non-differentiable models for (asymptotic) bivariate maxima; statistical decision: estimation and testing of the dependence parameters, test of independence; regression.

Introduction: Bivariate (and multivariate) asymptotic distributions of extremes are useful for dealing with many concrete problems as the largest ages of death for men and women, whose distribution naturally splits in the product of the margins, by independence; the floods (or droughts) at two different places of the same river, each year; bivariate (or multivariate) extreme meteorological data (pressure , temperature, wind velocity, etc.) each week; largest waves each week, etc.

Evidently, the target of a study of asymptotic distributions of bivariate extremes is to obtain asymptotic probabilistic behavior and also, to provide bivariate models of (asymptotic) extremes that fit observed data. It can be said that, although some problems are solved, the methods found until now cover much less area than the theory for extremes, which itself may be said to be in adolescence but not, yet, in adult age. Thus bivariate extremes may be, now, at the end of infancy and multivariate extremes are yet even younger! It will be seen that, in many cases, we do not have the best test, the best estimation procedure, etc. - although we have one - and in other cases nothing at all is known. An example: the separation of the bivariate extreme models - so important for applications - is only now being considered! In general, the few papers up to now choose one model from the beginning or compare two of them by the use of Kolmogoroff-Smirnov or another test, see, for instance, Gumbel and Goldstein (1964).

Let us briefly recall the basic ideas relating to univariate extremes. Let $\{X_n\}$ be a sequence of i.i.d. random variables with distribution function $F(x)$. Then $\mathrm{Prob}\{\max(X_1,\ldots, X_n) \le x\} = F^n(x)$. We can ask whether there exist sequences of attraction coefficients $\{(\lambda_n, \delta_n)\}$ $(\delta_n > 0)$ such that

$$\mathrm{Prob}\{\frac{\max(X_1,\ldots, X_n) - \lambda_n}{\delta_n} \le x\} = F^n(\lambda_n + \delta_n\, x)$$

has a non-degenerate (weak) limiting distribution function $L(x)$. It is well known that, by Khintchine's theorem on convergence of types, that if this happens the *sequence* $\{(\lambda_n, \delta_n)\}$ *is* not unique and an equivalent sequence, $(\bar{\lambda}_n, \bar{\delta}_n)$, i.e. leading to the same limiting distribution $L(x)$ is such that $(\bar{\lambda}_n - \lambda_n)/\delta_n \to 0$ and $\bar{\delta}_n/\delta_n \to 1$. With a convenient choice of $\{(\lambda_n, \delta_n)\}$ we have the reduced (or standard) forms

$$\Psi_\alpha(z) = e^{-(-z)^\alpha} \quad \text{if } -\infty < z \le 0$$

$$= 1 \quad \text{if } 0 \le z < -\infty \; ; \; \alpha > 0 \; ;$$

$$\Lambda(z) = \exp(-e^{-z}) \; , \; -\infty < z < +\infty \quad \text{and}$$

$$\Phi_\alpha(z) = 0 \quad \text{if } -\infty < z \le 0$$

$$= e^{-z^{-\alpha}} \quad \text{if } 0 < z < +\infty \; , \; \alpha > 0.$$

The basic paper is Gnedenko (1943), with previous ones of Fisher-Tippett (1928), Fréchet (1927), Gumbel (1935) and von Mises (1935). For more details see Gumbel (1958).

Ψ_α, Λ and Φ_α are called, respectively, Weibull, Gumbel and Frechet (standard) distributions; in practical applications we have to introduce location and dispersion parameters.

It is evident that by logarithmic transformations we can reduce Weibull and Frechet forms to the Gumbel one, so that, for theoretical study - although not for practical applications - we can concentrate on the Gumbel limiting form, as will be done in the next sections.

Note that as $\max(x_1, \ldots, x_n) = -\min(-x_1, \ldots, -x_n)$ the analogous limiting forms for minima are $1 - L(-z)$, i.e., $1 - \Phi_\alpha(-z)$, $1 - \Lambda(-z)$ and $1 - \Psi_\alpha(-z)$.

Let us recall, as it will be useful in the sequel, that for the Gumbel distribution $\Lambda(z)$ the mean value of $\mu_1' = \gamma = 0.57722$ (Euler constant), the variance

$\mu_2 = \sigma^2 = \pi^2/6$, the skewness coefficient $\mu_3/\sigma^3 = 1.1396$ and the kurtosis coefficient $\mu_4/\sigma^4 = 5.4$.

Asymptotic behavior of bivariate maxima: Consider now a sequence of i.i.d. random pairs $\{(X_n, Y_n)\}$ with distribution function $F(x,y)$. Analogously $\mathrm{Prob}\{\max(x_1,\ldots, x_n) \le x, \max(y_1,\ldots, y_n) \le y = F^n(x,y)$. We can seek a pair of sequences $\{(\lambda_n, \delta_n), (\lambda'_n, \delta'_n)\}$ such that

$$\mathrm{Prob}\{\frac{\max(x_1,\ldots, x_n)-\lambda_n}{\delta_n} \le x \quad , \quad \frac{\max(y_1,\ldots, y_n)-\lambda'_n}{\delta'_n} \le y\} = F^n(\lambda_n+\delta_n x, \lambda'_n+\delta'_n y)$$

do have a (weak) limiting non-degenerate distribution function $L(x,y)$. If this happens the Boole-Frechet inequality shows that the margins also have (weak) limiting distributions of the marginal maxima and, thus, are of the three forms previously given. In relation to what has been said before, we will, from now on, suppose that the limiting distributions of the margins are of Gumbel form:

$$L(x_1 +\infty) = \Lambda(x) \quad , \quad L(+\infty, y) = \Lambda(y) \ .$$

Using Khintchine's theorem, as is done for the univariate case and imposing Gumbel margins we can show that $L(x,y)$ must satisfy the (stability) relation

$$L^k(x,y) = L(x-\log k, y-\log k)$$

for an integer k positive. Passing from the positive integer k to rational $r(>0)$ and finally to real $t(>0)$ we get

$$L^t(x,y) = L(x-\log t, y-\log t)$$

Taking now $x = \log t$ we have $L(x,y) = L^{e^{-x}}(0,y-x)$. Putting now $L(0,w) = \exp(-(1+e^{-w})k(w))$ we have shown, finally, that the limiting (and stable) distribution of maxima with Gumbel margins are of the form

$$L(x,y) = \Lambda(x,y) = \exp(-(e^{-x}+e^{-y})k(y-x)) = \{\Lambda(x)\Lambda(y)\}^{k(y-x)}$$

It remains to study now the dependence function $k(w)$, obviously continuous and non-negative, for Λ to be a distribution function. Those results are well known. They can be found, with different forms of margins in Finkelshteyn (1953), Tiago de Oliveira (1958), Geffroy (1958/59) and Sibuya (1959) with a synthesis of the results in Tiago de Oliveira (1962/63). Subsequent results are in Tiago de Oliveira (1975) and (1980). Galambos (1968) contains a more recent account.

A characterization of the distribution function $\Lambda(x,y)$ can be made in the following way. It is immediate that a random pair (X,Y) with distribution function $\Lambda(x,y)$ is such that $V = \max(X+a,Y+b)$ has a Gumbel distribution function with a location parameter, i.e., $\max(X+a,Y+b) - \{\log(e^a+e^b) + \log k(a-b)\}$ has a standard Gumbel distribution function. In fact

$$\mathrm{Prob}\{\max(X+a,Y+b)-\lambda(a,b)\le z\} = F(z+\lambda(a,b)-a,z+\lambda(a,b)-b) = \Lambda(z)$$

or

$$F(z-a,z-b) = \Lambda(z-\lambda(a,b))$$

If we put $z-a = p$, $z-b = q$ we get

$$F(p,q) = \Lambda(z-\lambda(z-p,z-q))$$

and, thus, $z - \lambda(z-p,z-q)$ is independent of z. Taking now $z = q$ and $\lambda(q-p,0) = \log(1+e^{q-p}) + \log k(q-p)$ we obtain the desired form.

Let us now describe the dependence function $k(w)$. Although a continuous function we cannot show that it has a 2nd derivative and consequently we cannot expect a bivariate extreme random pair with distribution function $\Lambda(x,y) = [\Lambda(x)\Lambda(y)]^{k(y-x)}$ to have a planar density. In fact, from the Boole-Fréchet inequality

$$\max(0,\Lambda(x)+\Lambda(y)-1) \le \Lambda(x,y) \le \min(\Lambda(x),\Lambda(y))$$

we have, replacing x and y by $x + \log n$ and $y + \log n$, raising to the power n and letting $n \to \infty$, the limit inequality

$$\Lambda(x)\Lambda(y) \le \Lambda(x,y) \le \min(\Lambda(x),\Lambda(y))$$

or

$$\exp\{-(e^{-x}+e^{-y})\} \le \Lambda(x,y) \le \exp(-e^{-\min(x,y)})$$

Evidently the upper limit corresponds to the case where the reduced margins pair (X,Y) is concentrated, with probability 1, in the first diagonal, the so-called diagonal case, which is singular; the lower limit corresponds to independence. For the dependence function we have

$$(\text{diagonal})\frac{\max(1,e^{-w})}{1+e^{-w}} \le k(w) \le 1 \text{ (independence)}.$$

Note that $k(-\infty) = k(+\infty) = 1$. The behavior of the dependence function can be described through the behavior of the median line $\Lambda(x,y) = \frac{1}{2}$ or $(e^{-x}+e^{-y})k(y-x) = \log 2$; note that the median curve is always in the plane area defined by the curve for the diagonal case $\max(e^{-x},e^{-y}) = \log 2$ and the curve for independence $e^{-x}+e^{-y} = \log 2$.

If there is a planar density, i.e., $k''(w)$ exists, then as it is easily obtained by derivation, $k(w)$ must satisfy the relations:

$$k(-\infty) = k(+\infty) = 1 \quad ,$$

$$[(1+e^{w})k(w)]' \ge 0 \quad ,$$

$$[(1+e^{-w})k(w)]' \le 0$$

and

$$(1+e^{-w})k''(w) + (1-e^{-w})k'(w) \ge 0 \, ,$$

the corresponding conditions for the general case being, as $\Delta^2_{x,y}\Lambda(x,y) \ge 0$,

$$k(-\infty) = k(+\infty) = 1 \quad ,$$

$$(1+e^{w})k(w) \quad \text{a non-decreasing function} \, ,$$

$$(1+e^{-w})k(w) \quad \text{a non-increasing function} \quad ,$$

and $$\Delta^2_{x,y}\ [(e^{-k}+e^{-y}k(y-x)] \le 0\ .$$

Some other properties can be ascribed to $k(w)$. The first one is the symmetry condition, i.e., if $k(w)$ is a dependence function, then $k(-w)$ is also a dependence function. The proof is immediate if we consider the conditions in the differentiable case (where a planar density does exist) and slightly longer in the general case. If $k(w) = k(-w)$ then (X,Y) is an exchangeable pair and $\Lambda(x,y) = \Lambda(y,x)$.

Also it is immediate that if $k_1(w)$ and $k_2(w)$ are dependence functions, any mixture $\theta k_1(w) + (1-\theta)\ k_1(w)$, $0 \le \theta \le 1$, is also a dependence function . The set of dependence functions is, then, convex. And this convexity property

$$\Lambda(x,y) = \Lambda^{\theta}{}_1(x,y) \cdot \Lambda_2^{1-\theta}(x,y)$$

is very useful in obtaining models: the mixed model as well as the Gumbel model, are such examples.

Another method of generating models is the following. Let (X,Y) be an extreme random pair, with dependence function $k(w)$ and standard Gumbel margins and consider the new random pair $(\tilde{X},\tilde{Y})$ with $\tilde{X} = \max(X+a,Y+b)$, $\tilde{Y} = \max(X+c,Y+d)$. To have standard Gumbel margins we must have $(e^a+e^b)k(a-b) = 1$ and $e^c+e^d)k(c-d) = 1$. Then we have

$$\tilde{k}(w) = \frac{[e^{\max(a+w,c)} + e^{\max(b+w,d)}]k[\max(a+w,c) - \max(b+w,d)]}{1+e^w}$$

with (a,b) and (c,d) satisfying the conditions written above. This max-technique will be used towards the end of the paper to generate the biextremal and natural models.

Let us stress that independence has a very important position on a limiting situation. If we denote by $P(a,b)$ the function defined by $\text{Prob}\{X>x,Y>y\} = P(F(x,+\infty), F(+\infty,y))$ Sibuya (1960) has shown that the necessary and sufficient condition for having limiting independence is that $P(1-s,1-s)/s \to 0$ as $s \to 0$. He also showed that the necessary and sufficient condition for having the diagonal case as a limit situation is that $P(1-s,1-s)/s \to 1$ as $s \to 0$. With the first result we can show, easily, that the maxima of the binormal distribution has independence as a limiting distribution if $|\rho| < 1$.

Also Geffroy (1958/59) showed that a sufficient condition for limiting independence is that

$$\frac{1 + F(x,y) - F(x,w_y) - F(w_x,y)}{1 - F(x,y)} \to 0 \text{ as } x \to w_x \text{ and } y \to w_y \ ,$$

w_x and w_y being the right end points of the support of X and Y.

Sibuya conditions (and Geffroy sufficient conditions) are easy to interpret: we have limiting independence if $\text{Prob}\{X>x,Y>y\}$ is a vanishing summand of $\text{Prob}\{X>x \text{ or } Y>y\}$ and the diagonal case as limit if $\text{Prob}\{X>x,Y>y\}$ is the leading summand of $\text{Prob}\{X>x \text{ or } Y>y\}$.

It is known that a random pair (X,Y) with distribution function F(x,y) has positive association if

$$\text{Prob}\{X\le x,Y\le y\} + \{\text{Prob } X>x,Y>y\}$$

is larger than or equal to the corresponding probabilities in the case of independence; intuitively this means that large (small) values of one of the variables are associated with large (small) values of the other. It is immediate that this reduces to

$$F(x,y) \ge F(x,+\infty)F(+\infty,y) .$$

The inequality, obtained from Boole-Frechet inequality,

$$\Lambda(x)\Lambda(y) \le \Lambda(x,y) \le \min(\Lambda(x),\Lambda(y))$$

shows that this is the case for bivariate extreme pairs, as could be anticipated. This result is due to Sibuya (1960).

The results on correlation that follow and some regression results to be given later continue to illuminate the situation.

As the covariance between X and Y can be written, as it is well known

$$\operatorname{cov}(X,Y) = \int_{-\infty}\int^{+\infty} [F(x,y) - F(x,+\infty)F(+\infty,y)]\, dxdy$$

we have in our case

$$\operatorname{cov}(X,Y) = -\int_{-\infty}^{+\infty} \log k(w)\, dw$$

and the correlation coefficient

$$\rho_\Lambda = -\frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log k_\Lambda(w)\, dw \ .$$

As $k(w) \le 1$ we have $0 \le \rho$, as could be expected from the positive association It is very easy to show that for the diagonal case

$$k_1(w) = \frac{\max(1,e^w)}{1 + e^w}$$

we have $\rho = 1$. Evidently the value of ρ does not identify the dependence function (or the distribution): ρ is the same for $k(w)$ and $k(-w)$. But $\rho = 0$, as $k(w) \le 1$ implies $k(w) = 1$, or independence. Writing now ρ under the form

$$\rho = 1 - \frac{6}{\pi^2}\int_{-\infty}^{+\infty} \log\frac{k(w)}{k_1(w)}\, dw$$

we see, analogously, that $\rho = 1$ or $k(w) \ge k_1$ implies $k(w) = k_1(w)_1$. That is the diagonal case.

Other common correlation coefficients are

grade-correlation coefficient $\quad \chi = 12\int_{-\infty}^{+\infty} \frac{e^w}{(1+e^w)^2(1+k(w))^2}\, dw - 3$

difference-sign correlation coefficient $\quad \tau = 1 = \int_{-\infty}^{+\infty} D(w)(1-D(w))\, dw$

where

$$D(w) = \operatorname{Prob}(Y-X\le w) = \frac{1}{1+e^{-w}} + \frac{k'(w)}{k(w)} \ ;$$

the inverse relation between D(w) and k(w) is

$$k(w) = \frac{\exp(\int_{-\infty} D(t)\ dt)}{1 + e^{w}} .$$

Recall that k'(w) exists almost everywhere.

In the case of existence of a planar density, which means the existence of D'(w), the conditions on k(w) are equivalent to the conditions that D(w) is a distribution function with mean value zero

$$(\int_{-\infty}^{0} D(w)\ dw = \int_{0}^{+\infty} (1-D(w))\ dw)$$

and such that

$$D'(w) \geq D(w)(1-D(w)) .$$

Note that in the independence case (k(w) = 1) we have

$$D(w) = \frac{1}{1 + e^{-w}} \qquad \text{(the logistic distribution)}$$

and

$$D'(w) = D(w)(1-D(w)) .$$

The differentiable models; statistical decisions: In this section we will suppose that the assumed models have standard Gumbel margins (or, equivalently, that the location and dispersion margin parameters are known).

Up to now only two differentiable models - i.e., models with planar density, with a point exception in one case - are known. They appear in Gumbel (1961). One is the logistic model so called because its distribution function of the reduced difference W = Y - X is

$$D_{\theta}(w) = (1+e^{-w/(1-\theta)})^{-1}$$

corresponding to the dependence function

$$k_{\theta}(w) = \frac{(1+e^{-w/(1-\theta)})^{1-\theta}}{1 + e^{-w}} .$$

For $\theta = 0$ we have independence ($k_0(w) = 1$) and for $\theta = 1$ we have the diagonal case

$$k_1(w) = \frac{\max(1, e^{-w})}{1 + e^{-w}} \quad ,$$

which is the only case where we do not have a planar density.

As $k_\theta(w) = k_\theta(-w)$ the margins are exchangeable, as also can be shown by the form of the distribution function

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

The correlation coefficient has the expression $\rho(\theta) = \theta(2-\theta)$ which increases from $\rho(0) = 0$ to $\rho(1) = 1$; Kendall's τ has the expression $\tau(\theta) = \theta$.

It can be shown that

$$\sup_{x,y} |\Lambda_\theta(x,y) - \Lambda_0(x,y)| = (1-2^{-\theta})\ 2^{-\frac{\theta}{2^\theta - 1}}$$

which increases from 0 (at $\theta = 0$) to ¼ (at $\theta = 1$). It is, then, 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 probably be impossible to distinguish them. It is thus natural to use a one-sided test of $\theta = o$ vs. $\theta > 0$. Denoting by $p_\theta(x,y)$ the density $\partial^2\Lambda_\theta/\partial x\, dy$ ($p_0(x,y) = \Lambda'(x)\Lambda'(y)$) it is well known that the locally most powerful test of $\theta = 0$ vs. $\theta > 0$ is given by the critical region

$$\textstyle\sum_1^u v(x_i, y_i) \geq a_n$$

where

$$v(x,y) = \frac{\partial}{\partial\theta} \log p_\theta(x,y)|_{\theta = 0} \ .$$

In our case we get

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

whose mean value is zero but has infinite variance at $\theta = 0$. Thus the usual central limit theorem is not applicable and we have to resort to simulation.

A similar situation occurs for the <u>mixed model</u> whose dependence function is the $(1-\theta,\theta)$ mixture of the dependence function $k_0(w) = 1$ (independence) and $k_1(w) = 1 - \frac{e^w}{(1+e^w)^2}$ for which $k_\theta(w) = 1 - \theta\frac{e^w}{(1+e^w)^2}$. In this case we always have a planar density. The distribution function of the difference $W = Y - X$ is

$$D_\theta(w) = \frac{e^w}{1+e^w}\,\frac{(1+e^w)^2-\theta}{(1+e^w)^2-\theta e^w}$$

For $\theta = 0$ we have independence, as noted, but for $\theta = 1$ we have dependence but not the diagonal case. The Boole-Fréchet inequality shows that the domain $[0,1]$ for θ cannot be enlarged.

The distribution function is

$$\Lambda_\theta(x,y) = \exp-\{e^{-x}+e^{-y}-\frac{\theta}{e^x+e^y}\} = \Lambda(x)\Lambda(y)\exp(\frac{\theta}{e^x+e^y})$$

and the pair is exchangeable as can be seen, also, because $k_\theta(w) = k_\theta(-w)$. The correlation coefficient is

$$\rho(\theta) = \frac{6}{\pi^2}(\text{arc}\cos(1-\theta/2))^2 \quad ,$$

increasing from $\rho(0) = 0$ to $\rho(1) = 2/3$. We have also

$$\sup_{x,y}|\Lambda_\theta(x,y) - \Lambda_0(x,y)| = \frac{\theta}{4-\theta}\,(1-\frac{\theta}{4})^{\frac{4}{\theta}}$$

which increases from 0 (at $\theta = 0$) to $3^3/4^4 = 0.106$. The smaller variation of the correlation coefficient and of the distance shows that the deviation from independence is smaller and most difficult to detect.

Once more, the locally most powerful test of $\theta = 0$ vs. $\theta > 0$ leads to the critical region

$$\sum_1^u v(x_i,y_i) \geq a_n$$

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

The mean value is zero but the variance is also infinite. Those difficulties show that we must apply the usual methods of data analysis, although inefficient.

The combination of the use of correlation coefficients (product-moment, difference-sign and grade) and also the step and quadrants method described briefly in Tiago de Oliveira (1975, 1980) show that the most efficient of all is the (product-moment) correlation coefficient, for testing $\theta = 0$ vs. $\theta > 0$ in both models. Naturally, until further advances appear, it seems natural to use correlation coefficients to estimate θ. Note that all those methods are independent of the margin parameters. For confidence intervals, owing to the difficulty of getting the variance of ρ as a function of θ (in both models) it may, even, be useful to use the quadrants method which estimates the probability of the components of the random pair to be both larger or both smaller than the medians of the margins, by its observed frequency. As the margin medians, in reduced form, are $\tilde{\mu} = -\log\log 2$, the probability already referred to has the expression $2\Lambda_\theta(\tilde{\mu},\tilde{\mu})$ which amounts to $p(\theta) = \exp(\log 2\times(1-2^{1-\theta}))$ in the logistic model and to $p(\theta) = \frac{1}{2}\times 2^{\theta/2}$ for the mixed model. θ is estimated by $p(\theta^*) = N/n$ where N is the number of observed pairs whose components are both smaller or both larger than the sample margin medians and it is known that

$$\sqrt{n}\,\frac{N/n-p(\theta)}{\sqrt{p(\theta)(1-p(\theta))}}$$

is asymptotically normal.

No other statistical decision problems (such as regression analysis, discrimination and forecasting) have been dealt with for both models; as said separation of the two models is now under study.

The non-differentiable models; statistical decisions: The biextremal and the Gumbel models were the only ones considered until recently; now there is a third model, the natural, which, in some way, generalizes the biextremal model (see Tiago de Oliveira (1970, 1971, 1974, 1975, 1975', 1980)).

The biextremal model appears naturally in extremal processes (see Tiago de Oliveira (1968) and references therein). One way of introducing it directly through the max-technique is to consider a standard Gumbel independent pair (X,Z) and take the new pair (X,Y) with

$$Y = \max(X+\log\theta, Z+\log(1-\theta)) \quad , \ 0 \le \theta \le 1.$$

It has the distribution function

$$\Lambda_\theta(x,y) = \exp\{-\max(e^{-x}+(1-\theta)e^{-y}, e^{-y})\}$$

and the dependence function

$$k_\theta(w) = \frac{1-\theta+\max(\theta,e^w)}{1+e^w} = 1 - \frac{\min(\theta,e^w)}{1+e^w}$$

As $k_\theta(w) \neq k_\theta(-w)$ the random variables are not exchangeable.

The distribution function of the (reduced) difference is $D_\theta(w) = 0$ if $w<\log\theta$ and $D_\theta(w) = (1+(1-\theta)e^{-w})^{-1}$ if $w\ge\log\theta$ with a jump of θ at $\log\theta$. It is immediate that

$$\mathrm{Prob}\{Y\ge X+\log\theta\} = 1 ,$$

and, so, a singular part is concentrated at the line $y = x + \log\theta$, with probability θ. For $\theta = 0$ and $\theta = 1$ we obtain independence and the diagonal case. We have

$$\sup|\Lambda_\theta(x,y) - \Lambda_0(x,y)| = \theta(1+\theta)^{-\frac{1+\theta}{\theta}}$$

which, as for the logistic model, increases from 0 (at $\theta = 0$) to 1 (at $\theta = 1$).

As $$\mathrm{Prob}\{Y-X\ge\log\theta\}$$

and so $$\mathrm{Prob}\{\min(Y_i-X_i)\ge\log\theta\} = 1$$

it is natural to estimate θ by

$$\theta^* = \min(e^{y_i - x_i}, 1)$$

whose distribution function is $\mathrm{Prob}\{\theta^* \le z\} = 0$ if $z < 0$, $= 1 - (\frac{1-\theta}{z- \; -\theta})^n$ if $\theta \le z < 1$, $= 1$ if $z > 1$, the variance of θ being asymptotic to $2(1-\theta)^n/n^2$ if $\theta > 0$ and asymptotic to $1/n^2$ if $\theta = 0$. A natural test of independence, at significance level α, using θ^* is to accept $\theta = 0$ vs. $\theta > 0$ if $\theta^* \le \alpha^{-1/n} - 1$. The correlation coefficient of the biextremal model is

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

increasing from $\rho(0) = 0$ to $\rho(1) = 1$.

In Tiago de Oliveira (1974) we gave the expression of the general regression of Y in X and X on Y. It was shown already, in regard to mean square error, that linear regression is a good approximation to the general regression. Although linear regression and general regression curves behave very differently for very large and very small values of x (or y), this can be explained because of the positive association and of the fact that the half-lines where they are very distinct have a very low probability and, thus, a very small weight in the mean square error.

Gumbel model has the distribution function

$$\Lambda_\theta(x,y) = \exp\{-[e^{-x}+e^{-y}-\theta\min(e^{-x},e^{-y})]\}$$

with the dependence function

$$k_\theta(w) = 1-\theta\,\frac{\min(1,e^w)}{1+e^w}$$

and as $k_\theta(w) = k_\theta(-w)$ the random variables X and Y are exchangeable. The distribution function of the (reduced) difference is

$$D_\theta(w) = \frac{1-\theta}{1-\theta+e^w} \quad \text{if } w < 0 \quad , \quad = \frac{e^w}{1-\theta+e^w} \quad \text{if } w > 0$$

with a jump of $\theta/(2-\theta)$ at $w = 0$, giving thus the probability $P(Y=X)$. The dependence function is the mixture $(1-\theta,\theta)$ of independence $(k_0(w) = 1)$ and the diagonal case $(k(w) = \frac{\max(1,e^w)}{1+e^w})$. For this

$$\sup_{x,y}|\Lambda_\theta(x,y) - \Lambda_0(x,y)| = \frac{\theta}{2-\theta}(1-\theta/2)^{2/\theta}$$

which, also, increases from 0 (at $\theta = 0$) to ¼ (at $\theta = 1$). ___

This model is a transformation for Gumbel margins of the Marshall and Olkin (1967) bivariate exponential model. If we denote by nf_n the number of points (x_i,y_i) with $x_i = y_i$ and by $T_n = 1/n\Sigma\max(e^{-x_i},e^{-y_i})$ the maximum likelihood estimator, it is then given by

$$\hat{\theta} = (T_n-1+\sqrt{(T_n-1)^2+4f_nT_n})/2T_n$$

taking $\theta = 0$ if the expression is negative. θ^*, not truncated, is asymptotically normal with mean value θ and variance

$$\frac{\theta(1-\theta)(2-\theta)}{n(1+\theta)}$$

Note that f_n and T_n are asymptotically independent and that

$$(\sqrt{n}\,\frac{(2-\theta)f_n-\theta}{\sqrt{2\theta(1-\theta)}}),\quad \sqrt{n}((2-\theta/T_n-1))$$

is asymptotically a binormal pair with standard margins. In particular we see that

$$\text{var}\,(f_n) = \frac{2\,(1-\theta)(2-\theta)^2}{n}$$

is zero at $\theta = 0$ as (X,Y), being independent form an absolutely continuous pair and at 1 as $P(X=Y) = 1$. The estimator

$$\hat{\theta} = (T_n-1+\sqrt{(T_n-1)^2+4f_nT_n})/2T_n$$

is asymptotically normal with mean value θ and variance $\theta(1-\theta)(2-\theta)/(1+\theta)n$;

evidently if $\theta^* < 0$ or $\theta^* > 1$ we must truncate.

To test independence ($\theta = 0$), as $f_n = 0$ and the variance of $\hat{\theta}$ is null, we can use T_n, $\sqrt{n}(2 \cdot T_n - 1)$ being asymptotically standard normal.

The correlation coefficient is

$$\rho(\theta) = \frac{12}{\pi^2} \int_0^\theta \frac{\log(2-t)}{1-t}\, dt$$

increasing from $\rho(0) = 0$ to $\rho(1) = 1$. As for the biextremal model the linear regression is a very good approximation to the general one, in the same sense as before (see Tiago de Oliveira (1974)).

Let us consider the natural model described in Tiago de Oliveira (1982). If we take independent random (reduced) Gumbel variables Z and T and consider a new random pair (X,Y) with $X = \max(Z-a, T-b)$, $Y = \max(Z-c, Y-d)$ with $a, b, c, d \geq 0$ such that the margins are standard we get $e^{-a}+e^{-b} = e^{-c}+e^{-d} = 1$. Then we have

$$\Lambda(x,y) = P(X \leq x, Y \leq y) = \exp\{-(e^{-x}+e^{-y})k(y-x)\}$$

$$k(w) = (\max(e^{-a}, e^{-c-w}) - \max(e^{-b}, e^{-d-w}))/(1+e^{-w}) \;,$$

if $a - c \leq b - d$, using all the introduced parameters. The random points as $a - c \leq y - x \leq b - d$, are contained in a strip parallel to the first diagonal, imposing thus a strong stochastic relation between X and Y if the bounds $a - c$ and $b - d$ are not infinite. As $a - c \leq 0 \leq b - d$ this strip contains the origin. Let us now introduce the parameters $\alpha, \beta \geq 0$, such that $a - c = a + \log(1-e^{-d}) = -\alpha$ and $b - d = -d - \log(1-e^{-a}) = \beta$. Note that $\alpha \geq 0$ or $\beta \geq 0$ imply $e^{-a}+e^{-d} \geq 1$. The final expression of the dependence function is

$$k_{\alpha\beta}(w) = \frac{1}{1+e^{w}} \qquad \text{if } w \leq -\alpha$$

$$= \frac{1-e^{-\beta}+(1-e^{-\alpha})e^{-w}}{(1-e^{-\alpha-\beta})(1+e^{-w})} \qquad \text{if } -\alpha \leq w \leq \beta$$

$$= \frac{1}{1 + e^{-w}} \qquad \text{if } w \geq \beta$$

Note that the left and right tails of $k_{\alpha,\beta}(w)$ coincide with the ones of the diagonal case. As said before the case $\alpha = \beta = +\infty$ corresponds to independence and $\alpha = \beta = 0$ to the diagonal case. The exchange of α and β corresponds to the exchange of X and Y. For $\alpha = -\log\theta$ and $\beta = +\infty$ $(0 \leq \theta \leq 1)$ we get the biextremal model and its dual (exchange of X and Y) is obtained for $\alpha = +\infty$, $\beta = -\log\theta$.

It is immediately shown that

$$D_{\alpha,\beta}(w) = \text{Prob}\{Y-X \leq w\} = 0 \qquad \text{if } w < -\alpha$$

$$= \frac{1 - e^{-\beta}}{1-e^{-\beta}+(1-e^{-\alpha})e^{-w}} \qquad \text{if } -\alpha \leq w < \beta$$

$$= 1 \qquad \text{if } w \geq \beta$$

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

The correlation coefficient has the expression

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

and as

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

we get by subtracting and simple algebra

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

$$= 1 - \frac{6}{\pi^2}\Big\{-\frac{\alpha^2}{2} + (\alpha+\beta)\log\frac{1 - e^{-\beta}}{1-e^{-\alpha-\beta}} +$$

$$\int_{(1-e^{-\alpha})/(e^{\beta}-1)}^{(e^{\alpha}-1)/(1-e^{-\beta})} \frac{\log(1+t)}{t}\, dt\Big\}\,.$$

It is evident $\rho(\alpha,\beta) \to 1$ as $\alpha,\beta \to 0$ and $\rho(\alpha,\beta) \to 0$ as $\alpha,\beta \to +\infty$.

The linear regression with reduced margins is given by the straight lines

$$y - \gamma = \rho(\alpha,\beta)(x-\gamma) \quad \text{and} \quad x - \gamma = \rho(\alpha,\beta)(y-\gamma) \ .$$

For the general regression, as the interchange of X and Y is equivalent to the interchange of α and β it is sufficient to compute the regression curve

$$\overline{y}_{\alpha,\beta}(x) = x + \beta - \frac{1 - e^{-\beta}}{1-e^{-\alpha-\beta}} \exp\left(\frac{1-e^{-\alpha}}{e^{\beta}-e^{-\alpha}} e^{-x}\right) \times \int_{[(1-e^{-\alpha})/(e^{\beta}-e^{-\alpha})]e^{-x}}^{[e^{\beta}(e^{\alpha}-1)/(e^{\beta}-e^{-\alpha})]e^{-x}} \frac{e^{-t}}{t}\, dt \ ;$$

Analogously to the situation for the biextremal model we can expect that regression lines are a good approximation - in the same sense - to the general regression curve. When the margins are reduced, as $-\alpha$ and β are the bounds of the support of $D_{\alpha,\beta}(w)$ we can naturally estimate α and β from the $w_i = y_i - x_i$.

As $-\alpha \le \min(w_i) \le \max(w_i) \le \beta$ and $-\alpha \le 0 \le \beta$ the estimators of α and β - although biased (both in a one-sided manner) are

$$\alpha^* = - \min(0,\min(w_i)) \ (< \alpha)$$

$$\beta^* = \max(0, \min(w_i)) \ (< \beta)$$

As $\text{Prob}\{\alpha^* = 0\} = (1 - D_{\alpha,\beta}(0))^n$ and $\text{Prob}\{\beta^* = 0\} = D_{\alpha,\beta}(0)^n$

and

$$0 < D_{\alpha,\beta}(0) = \frac{2e^{\beta}-1-e^{\beta-\alpha}}{2(e^{\beta}-e^{-\alpha})} < 1 \quad \text{if} \quad \alpha,\beta > 0$$

we see that the probabilities converge to zero and, thus, the estimators are consistent. If $\alpha = \beta = 0$ we have the diagonal case and all the sample points are in the first diagonal.

Remarks on the non-parametric estimation of the dependence function: The fact that the set of dependence functions $\{k(w)\}$ is a convex set could suggest estimating the dependence function of the data under consideration by an average

$k^*(w) = \frac{1}{n}\sum_1^n k_s(w/w_i)$, where $k_s(w/w_i)$ is a special and convenient dependence function and w_i is the difference $y_i - x_i$ for the observed pair (x_i, y_i), evidently with standard margins. By Khintchine's theorem we know that $k^*(w)$ converges almost surely, for every w, to

$$\int_{-\infty}^{+\infty} k_s(w/t)\ dD(t)$$

which should be equal to

$$k(w) = \frac{\exp(\int_{-\infty}^{w} D(t)\ dt)}{1 + e^w} \quad .$$

The resulting integral equation

$$(1+e^w) \int_{-\infty}^{+\infty} k_s(w/t)\ dD(t) = \exp(\int_{-\infty}^{w} D(t)\ dt)$$

does not seem to be solvable for $k_s(w/v)$, for all admissible $\{D(w)\}$. This temptation must be discarded.

We can try to estimate directly $D(w)$ by the usual sample distribution function $\frac{1}{n}\sum_1^n H(w-w_i)$ where $H(w)$ is the Heavside jump function at $w = 0$ ($H(w) = 0$ if $w < 0$, $H(w) = 1$ if $w \geq 0$). It is easy to see that we get, then,

$$k^*(w) = \frac{\exp(\frac{1}{n}\sum_1^n (w-w_i)_+)}{1 + e^w}$$

where $(w)_+ = 0$ if $w < 0$ and $(w)_+ = w$ if $w \geq 0$. But we can see that although $k^*(-\infty) = 1$ we have $k^*(+\infty) = e^{-\bar{w}} \neq 1$. A possible modification is to take

$$k^*(w) = \frac{\exp(\frac{1}{n}\sum_1^n (w-w_i)_+)}{1 + e^{w-\bar{w}}}$$

which converges a.s. to

$$\frac{\exp(\int_{-\infty}^{w} D(t)\ dt)}{1 + e^w} = k(w)$$

as $n \to \infty$, because $\bar{w} \to 0$. We have, already, then $k^{**}(-\infty) = k^{**}(+\infty) = 1$

but k is not yet a dependence function.

We could, owing to the central position of the logistic distribution, as associated with independence, and also due to its quite good behavior, try to estimate D(w) by

$$D^*(w) = \frac{1}{n}\sum_1^n \frac{1}{1+e^{-(w-w_i)/\delta_n}}$$

with $\delta_n \to 0$. In fact, we do not obtain a D(w) function and, so, the simpler estimator, up to now, is $k^{**}(w)$.

The area of non-parametric estimation of k(w) or D(w) by k- or D- functions seems, thus, still completely unexplored.

Remarks about the general situation: In general we do not have standard margins but margins with location and dispersion parameters. In that case, which seems natural is to estimate, independently, the margin parameters λ_x, δ_x (>0) and λ_y, δ_y (>0) by its ML-estimators $\hat{\lambda}_x, \hat{\delta}_x, \hat{\lambda}_y$ and $\hat{\delta}_y$, then obtain the "estimated" reduced values $\hat{x}_i = (x_i - \hat{\lambda}_x)/\hat{\delta}_x$ and $\hat{y}_i = (y_i - \hat{\lambda}_y)/\hat{\delta}_y$ and, finally, the "estimated" reduced difference $\hat{w}_i = \hat{y}_i - \hat{x}_i$ and proceed as before, with standard margins.

As a whole, in the differentiable cases we can expect good behavior - see the paper on δ-method by Tiago de Oliveira (1981) - but in the non-differentiable cases the situation can be more difficult, as is, especially, the case for the Gumbel model where, with probability one we do not have $\hat{x}_i = \hat{y}_i$. In this case the use of $\hat{T}$ is suggested, although it is much less efficient than the case of $\hat{f}_n$. This is, thus, another open area of study.

REFERENCES

B.V. Finkelshteyn (1953) - Limiting distribution of extremes of a variational series of a two-dimensional random variable, Dokl. Ak. Nauk. S.S.S.R., vol. 91 (in Russian).

R.A. Fisher and L.H.E. Tippett (1928) - Limiting forms of frequency distribution of largest and smallest members of a sample, Proc. Cambr. Phil. Soc., vol. 24, pt. 2.

M. Fréchet (1927) - Sur la loi de probabilité de l'e'cart maximum, Ann. Soc. Polon. Math. (Krakov), vol. 6.

J. Galambos (1968) - The asymptotic theory of m-th order extreme statistics, J. Wiley, NY.

J. Geffroy (1958-59) - Contributions à l'etude des valeurs extrêmes, Publ. Inst. Stat. Univ. Paris, vols. 7 and 8.

B.V. Gnedenko (1943) - Sur la distribution limite du terme maximum d'une série aléatoire, Ann. Math., vol. 44.

E.J. Gumbel (1935) - Les valeurs extrêmes des distributions statistiques, Ann. Inst. H. Poincaré, vol. V.

__________ (1958) - Statistics of Extremes, Columbia Univ. Press.

__________ (1961) - Multivariate extremal distributions, Bull. Int. Stat. Inst., 33e sess., 2e livr., Paris.

E.J. Gumbel and Neil Goldstein (1964) - Analysis of empirical bivariate extremal distributions, J. Amer. Stat. Assoc., vol. 62.

Albert W. Marshall and Ingram Olkin (1967) - A multivariate exponential distribution, J. Amer. Stat. Assoc., vol. 62.

R. von Mises (1935) - La distribution de la plus grande de n valeurs, Rev. Math. Union Interbalkanique, vol. I.

M. Sibuya (1960) - Bivariate extremal statistics, I, Ann. Inst. Stat. Math. vol. XI.

J. Tiago de Oliveira (1958) - Extremal distributions, Rev. Fac. Ciencias Lisbon, 2 ser., A, Mat., vol. VII.

__________ (1962-63) - Structure theory of bivariate extremes; extensions, Estudos Mat., Estat e Econom., vol. VII.

__________ (1968) - Extremal processes; definition and properties, Publ. Inst. Stat. Univ. Paris, vol. XVII.

_________ (1970) - Biextremal distributions: statistical decision, Trab. Estad. y Inv. Oper., vol. XXI, Madrid.

_________ (1971) - A new model of bivariate extremes: statistical decision, Studi di Probabilita', Statistica e Ricerca Operativa in onore di Giuseppe Pompilj, Tip. Oderisi, Gubbio, Italia.

_________ (1974) - Regression in the nondifferentiable bivariate extreme models, J. Amer. Stat. Assoc., vol. 69.

_________ (1975) - Bivariate and multivariate extreme distributions, Statistical Distributions in Scientific Work, G.P. Patil, S. Kotz and J.K. Ord, eds., vol. 1, D. Reidell Publ. Co.

_________ (1975') - Statistical decision for extremes, Trab. Estad. y Inv. Oper., vol. XXVI, Madrid.

_________ (1980) - Bivariate extremes: foundations and statistics, Proc. V Int. Symp. Multiv. Analysis, P.R. Krishnaiah, ed., North Holland.

_________ (1981) - The δ-method for the obtention of asymptotic distributions: applications, Publ. Inst. Stat. Univ. Paris, vol. XXVI.

_________ (1982) - Decision and modelling for extremes, Some recent advances in statistics, J. Tiago de Oliveira and B. Epstein eds., Academic Press, London.

MED

82