Biometrika (1997), **84**, 3, *pp.* 567–577
Printed in Great Britain

A nonparametric estimation procedure for bivariate extreme value copulas

BY P. CAPÉRAÀ, A.-L. FOUGÈRES AND C. GENEST
Département de mathématiques et de statistique, Université Laval, Québec, Canada G1K 7P4
e-mail: caperaa@mat.ulaval.ca fougeres@mat.ulaval.ca genest@mat.ulaval.ca

SUMMARY

A bivariate extreme value distribution with fixed marginals is generated by a one-dimensional map called a dependence function. This paper proposes a new nonparametric estimator of this function. Its asymptotic properties are examined, and its small-sample behaviour is compared to that of other rank-based and likelihood-based procedures. The new estimator is shown to be uniformly, strongly convergent and asymptotically unbiased. Through simulations, it is also seen to perform reasonably well against the maximum likelihood estimator based on the correct model and to have smaller L_1, L_2 and L_∞ errors than any existing nonparametric alternative. The $n^{\frac{1}{2}}$ consistency of the proposed estimator leads to nonparametric estimation of Tawn's (1988) dependence measure that may be used to test independence in small samples.

Some key words: Asymptotic theory; Copula; Dependence function; Extreme value distribution; Nonparametric estimation.

1. INTRODUCTION

Management of environmental resources often requires the analysis of multivariate extreme values. Optimal reservoir management for water resource systems, for example, requires evaluation of probabilities of peak flow events from neighbouring basins. As a result of the regional nature of hydrological phenomena, extreme events typically exhibit some form of dependence that calls for multivariate data modelling. In applications of this type, the analyst may only have access to componentwise maxima or may prefer to restrict attention to such data to avoid dealing with the strong dependence that typically exists between successive measurements taken over time.

Concentrating on the bivariate case from now on, suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ form a random sample of pairs whose components represent the largest values of two characteristics observed over the same period of time. According to Pickands (1981), such data may be appropriately modelled by a distribution function $L(x, y) = C\{F(x), G(y)\}$ with marginals $F(x)$ and $G(y)$ and copula

$$C(u, v) = \mathrm{pr}\{F(X) \leqslant u, G(Y) \leqslant v\} = \exp\left[\log(uv) A\left\{\frac{\log(u)}{\log(uv)}\right\}\right], \tag{1}$$

defined for all $0 \leqslant u,\ v \leqslant 1$ in terms of a convex function A on $[0, 1]$ satisfying $\max(t, 1-t) \leqslant A(t) \leqslant 1$ for all $0 \leqslant t \leqslant 1$. Thus C depends only on the one-dimensional dependence function A. This generator is tied to many interesting aspects of the model, as described by Tawn (1988) and A. Khoudraji, who shows, in an unpublished 1995 Université Laval doctoral dissertation, that Kendall's and Spearman's nonparametric measures of dependence are respectively given by

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)}\, dA'(t), \quad \rho = 12 \int_0^1 \{A(t)+1\}^{-2}\, dt - 3.$$

Examples of parametric forms for A were given by Hutchinson & Lai (1990, Ch. 9). Tawn (1988) showed how estimation would proceed for some of these families, given a random sample of extreme value data. Smith, Tawn & Yuen (1990) also considered estimators of A' and A'', but the latter has the disadvantage of not being always well defined. Both of these papers suggest that nonparametric estimators of A giving the shape of the dependence function may be used as an exploratory tool for parametric model selection and assessment.

The classical nonparametric estimator of A is that of Pickands (1981). Its asymptotic properties were established by Deheuvels (1991), who proposed a variant that corrects its endpoint bias. Noting that Pickands' estimator is not convex, Tiago de Oliveira (1989) suggested another procedure that was largely ignored in the subsequent literature because of its obvious weaknesses. Finally, Smith et al. (1990) adapted the kernel method for density estimation to produce smooth estimates of A based on Pickands' proposal. However, the behaviour of their estimators is not fully understood, even in large samples.

This paper proposes a new nonparametric estimator of A, studies its asymptotic properties, and compares its small-sample behaviour with that of its major competitors. It is shown that this estimator, defined explicitly by (5), is asymptotically unbiased and uniformly, strongly consistent for A. The limiting distribution of the associated process on $[0, 1]$ is also exhibited. In addition, simulation results are reported which indicate that the new estimator has smaller L_1, L_2 and L_∞ errors than alternative estimators of Pickands (1981) and Deheuvels (1991) in different experimental conditions. Comparisons with the kernel based method of Smith et al. (1990) were also made, but ultimately had to be ignored, because of inherent numerical problems. Additional comparisons are presented between the nonparametric estimators and the method of maximum likelihood based on correct and incorrect parametric models exhibiting various degrees of asymmetry.

The problem considered here must be distinguished from situations considered by Coles & Tawn (1991, 1994) and by Joe, Smith & Weissman (1992), in which one attempts to determine the dependence structure of extreme events, based on a sample of data that are merely in the domain of attraction of a bivariate extreme value distribution. In that context, a point process representation theorem of de Haan (1985) and Resnick (1987, Ch. 5) leads to a threshold method which, though not directly related to model (1), nonetheless depends on the estimation of A or its derivatives.

Motivation for the introduction of the new estimator is given in § 2. Asymptotic considerations described in § 3 then lead to the statement of the paper's main result, Proposition 4·1, and to an operational definition of the proposed estimator in § 4. Next, simulation results comparing the small-sample behaviour of this new estimator with alternatives are reported in § 5. Finally, § 6 exploits the speed of convergence of the proposed estimator to construct a test of independence based on an estimation of coefficient $2\{1-A(\frac{1}{2})\}$ of Tawn (1988). Strategies for circumventing the hypothesis of known marginals are mentioned in the discussion. Mathematical developments are relegated to appendices.

2. Motivation

Since the marginal distributions of L are assumed known, the estimation of A can be based on the pairs $(U_i, V_i) \equiv \{F(X_i), G(Y_i)\}$ $(1 \leqslant i \leqslant n)$, which form a random sample from copula (1). The starting point of this investigation is the following proposition, which implies that A can be expressed as a function of $H(z) = \mathrm{pr}(Z_i \leqslant z)$, the distribution function of the pseudo-observations $Z_i = \log(U_i)/\log(U_i V_i)$ $(1 \leqslant i \leqslant n)$.

Proposition 2·1. *Let $C(u, v)$ be an extreme value copula with generator A. The distribution function of the random variable $Z = \log(U)/\log(UV)$ is given by*

$$H(z) = z + z(1-z)D(z),$$

where $D(z) = A'(z)/A(z)$ and $A'(z)$ denotes the right derivative of A for all $0 \leqslant z < 1$.

The proof of this proposition stems from the easily checked formula

$$\mathrm{pr}(U \leqslant u, Z \leqslant z) = H(z)u^{A(z)/z}.$$

As a consequence of this result, one gets

$$\frac{A(t)}{A(s)} = \exp\left\{\int_s^t \frac{H(z)-z}{z(1-z)}\,dz\right\}$$

for arbitrary choices of $0 \leqslant s \leqslant t \leqslant 1$. Since $A(0) = A(1) = 1$, one may write

$$A(t) = \exp\left\{\int_0^t \frac{H(z)-z}{z(1-z)}\,dz\right\} = \exp\left\{-\int_t^1 \frac{H(z)-z}{z(1-z)}\,dz\right\}.$$

Let H_n be the empirical distribution function of $Z_1, \ldots, Z_n$. Replacing H by H_n in the above expressions yields two possible nonparametric estimators for A, denoted by

$$A_n^0(t) = \exp\left\{\int_0^t \frac{H_n(z)-z}{z(1-z)}\,dz\right\}, \quad A_n^1(t) = \exp\left\{-\int_t^1 \frac{H_n(z)-z}{z(1-z)}\,dz\right\}.$$

The asymptotic properties of these two preliminary estimators are studied in the following section. This will lead us, in § 4, to estimate $\log A$ by

$$\log A_n(t) = p(t)\log A_n^0(t) + \{1-p(t)\}\log A_n^1(t), \tag{2}$$

where $p(t)$ is an appropriate weight function. The performance of A_n as an estimator of A will then be compared to that of its competitors in § 5.

3. Asymptotic properties of the two preliminary estimators

Before looking at the asymptotic behaviour of $\log A_n^i$ for $i = 0, 1$, one should note that they are both unbiased estimators of $\log A$ for all sample sizes, since H_n is itself unbiased for H. This fact is formally stated below for $\log A_n^0$, together with its limiting behaviour as an estimator of $\log A$ and as a process on $[0, 1]$.

Proposition 3·1. *The statistic $\log A_n^0$ is an unbiased and uniformly, strongly consistent estimator of $\log A$. That is, one has $E(\log A_n^0) = \log A$ for all $n \geqslant 1$ and*

$$\sup_{t\in[0,1]} |\log A_n^0(t) - \log A(t)| \to 0$$

almost surely. In addition, the process $n^{\frac{1}{2}}(\log A_n^0 - \log A)$ is asymptotically Gaussian with *zero mean and covariance matrix*

$$\Gamma_0(s, t) = \int_0^s \int_0^t \frac{H(u \wedge v) - H(u)H(v)}{uv(1-u)(1-v)} \, dv \, du.$$

Since $H(u \wedge v) \geqslant H(u)H(v)$, it should be observed that $\Gamma_0(t, t)$ is monotone increasing in t, so that, in spite of its attractive properties, $\log A_n^0(t)$ is an increasingly unreliable estimator of $\log A(t)$ as $t \to 1$. A similar analysis shows that $n^{\frac{1}{2}}(\log A_n^1 - \log A)$ is asymptotically Gaussian with zero mean and covariance matrix

$$\Gamma_1(s, t) = \int_s^1 \int_t^1 \frac{H(u \wedge v) - H(u)H(v)}{uv(1-u)(1-v)} \, dv \, du,$$

and hence that the variance of $\log A_n^1(t)$ is a decreasing function of t. This phenomenon suggests that a combined estimator of the form (2) might be preferable to each of the $\log A_n^i$'s, so long as $p(t)$ gives comparatively more weight to $\log A_n^i$ in the neighbourhood of i. The following proposition delineates circumstances under which this combined estimator inherits the properties of the $\log A_n^i$'s.

PROPOSITION 3·2. *Suppose that p is a bounded function on $[0, 1]$. The statistic $\log A_n$ defined in (2) is then an unbiased and uniformly, strongly consistent estimator of $\log A$. In addition, the process $n^{\frac{1}{2}}(\log A_n - \log A)$ is asymptotically Gaussian with zero mean and variance function*

$$\Gamma(t) = p^2(t)\Gamma_0(t, t) + \{1 - p(t)\}^2\Gamma_1(t, t) + 2p(t)\{1 - p(t)\}C(t), \qquad (3)$$

where

$$C(t) = -\int_0^t \int_t^1 \frac{H(u)\{1 - H(v)\}}{uv(1-u)(1-v)} \, dv \, du \leqslant 0$$

is the asymptotic covariance of $n^{\frac{1}{2}} \log A_n^0(t)$ and $n^{\frac{1}{2}} \log A_n^1(t)$.

It would be natural to choose $p(t)$ so as to minimise the asymptotic variance $\Gamma(t)$, that is

$$p(t) = \frac{\Gamma_1(t, t) - C(t)}{\Gamma_0(t, t) + \Gamma_1(t, t) - 2C(t)} \quad (0 \leqslant t \leqslant 1). \qquad (4)$$

In this case, one has $0 \leqslant p(t) \leqslant 1$. However, as the terms on the right-hand side of (4) are unknown and may be inconvenient to estimate in practice, the simple choice $p(t) = 1 - t$ will be used for simulations reported in § 5.

4. DEFINITION OF A NEW ESTIMATOR

In view of Proposition 3·2, a natural estimator of A would be A_n, as defined implicitly by (2). If $Z_{(1)}, \ldots, Z_{(n)}$ stand for the ordered Z_i's, and if

$$Q_i = \left\{ \prod_{k=1}^{i} Z_{(k)}/(1 - Z_{(k)}) \right\}^{1/n} \quad (1 \leqslant i \leqslant n)$$

it is not difficult to see that A_n can be written in closed form as

$$A_n(t)=\begin{cases}(1-t)Q_n^{1-p(t)} & \text{if } 0\leqslant t\leqslant Z_{(1)},\\ t^{i/n}(1-t)^{1-i/n}Q_n^{1-p(t)}Q_i^{-1} & \text{if } Z_{(i)}\leqslant t\leqslant Z_{(i+1)}\quad (1\leqslant i\leqslant n-1),\\ tQ_n^{-p(t)} & \text{if } Z_{(n)}\leqslant t\leqslant 1,\end{cases} \tag{5}$$

provided that the Z_i's are distinct.

This estimator has the property that $A_n(0)=A_n(1)=1$, provided that $p(0)=1-p(1)=1$. It is clearly not unbiased for A, since $E(A_n)>\exp\{E(\log A_n)\}=A$ by Jensen's inequality; it may be shown to be asymptotically unbiased, however. The basic properties of A_n are summarised in the following proposition, which constitutes this paper's main theoretical contribution.

PROPOSITION 4·1. *Suppose that p is a bounded function on $[0, 1]$. The estimator A_n defined implicitly by* (2) *is an asymptotically unbiased estimator of A which is uniformly, strongly consistent.*

Naturally, one could also use Proposition 3·2 to characterise the asymptotic behaviour of the process $(A_n/A)^{n^{1/2}}$. The fact that, in large samples, $\log A_n(t)$ is approximately normal with mean $\log A(t)$ and variance $\Gamma(t)/n$ can be exploited to construct pointwise approximate confidence intervals for $\log A(t)$ or $A(t)$. This only requires consistent estimation of $\Gamma(t)$, which may be obtained by replacing H by H_n in (3). The proof that the resulting estimator is uniformly, strongly convergent is similar to that of Proposition 3·1; the details are provided by A.-L. Fougères, in an unpublished 1996 Université Paul-Sabatier doctoral dissertation.

5. COMPARISONS WITH ALTERNATIVE ESTIMATORS

A Monte Carlo experiment was carried out to compare the small-sample behaviour of A_n with two of its main nonparametric competitors, and with the maximum likelihood estimator in two parametric models. The purpose of the analysis was to evaluate the relative precision of these procedures in terms of overall and local fit, as measured by the L_1, L_2 and L_∞ distances between the true dependence function and its estimate. An auxiliary objective was to illustrate the effect of fitting true and false parametric models in different circumstances.

Specifically, estimator A_n with weight function $p(t)=1-t$ was compared with two nonparametric alternatives, namely the classical estimator of Pickands (1981) and its variant proposed by Deheuvels (1991). Two maximum likelihood estimators were also included, based on the following models.

Model 1. The Gumbel or logistic model, whose generator is $A_r(t)=\{t^r+(1-t)^r\}^{1/r}$ with $r\geqslant 1$.

Model 2. An asymmetric extension thereof discussed by Tawn (1988), in which

$$A_{\alpha,\beta,r}(t)=1-\beta+(\beta-\alpha)t+\{\alpha^r t^r+\beta^r(1-t)^r\}^{1/r}\quad (0\leqslant\alpha,\ \beta\leqslant 1,\ r\geqslant 1).$$

In the first of these models, parameter r is linked to Kendall's tau through the relation $\tau=1-1/r$, while, in the second model, a simple algebraic expression for this measure of dependence does not exist in terms of α, β and r.

For three predetermined sample sizes, $n=25$, 50 and 100, data were generated according to a factorial design. The selected factors and their levels were as follows:

(i) choice of estimator ($A = A_n$, P = Pickands, D = Deheuvels, $1 = A_{\hat{r}}$, $2 = A_{\hat{\alpha},\hat{\beta},\hat{r}}$);

(ii) presence or absence of symmetry about $\frac{1}{2}$ in the function A, as embodied by the logistic models with parameters $(\alpha, \beta) = (1, 1)$, in the case of symmetry, and $(\alpha, \beta) = (0{\cdot}78, 0{\cdot}97)$ otherwise;

(iii) degree of dependence, as measured by Kendall's tau, with $\tau = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ corresponding to $r = \frac{4}{3}$, 2, 4 and $r = 1{\cdot}42$, $2{\cdot}58$, 50 in the symmetric and asymmetric case respectively.

These factors were considered fixed, and 500 independent pseudo-random samples of size n were generated for each treatment. Data were simulated separately in the case of independence, where the symmetric and asymmetric models coincide. The dependent variable used in each analysis was the logarithm of the L_1, L_2 or L_∞ distance between A and its estimate. This transformation ensured that the standard hypotheses of the analysis of variance model were approximately met.

The analysis shows that there exists a significant third-order interaction (P-value $< 0{\cdot}0001$). Table 1 summarises the results of paired comparisons between the five estimators by level of the other two factors, in the case where $n = 100$. Here, a strict inequality means that the L_1 errors are significantly different at the 0·005 level and that they can be ordered as indicated; the symbol $\approx$ is used when a difference is not significant at the selected level. This unusual choice of threshold value is justified by the large number of observations available for each treatment.

Table 1. *Paired comparisons of the* $\log(L_1)$ *error of three nonparametric* (A, D, P) *and two maximum likelihood* (1, 2) *estimators of the generator A of a bivariate extreme value distribution whose marginals are known.*

τ	Symmetric	Asymmetric
0·25	$1 < A < D \approx 2 < P$	$1 < A < D \approx 2 < P$
0·50	$1 < 2 < A < D < P$	$2 < A < 1 < D < P$
0·75	$1 < 2 < A < D < P$	$2 < A < D < 1 < P$

Comparisons are based on 500 pseudo-random samples of size 100 from such distributions with various degrees of dependence and presence or absence of symmetry with respect to $\frac{1}{2}$. Differences and their direction are shown only if significant at the 0·005 level.

The results shown in Table 1 are virtually identical to those obtained for other sample sizes and the two other error functions. When $n = 100$, for example, the same relations are valid for $\log(L_2)$, while for $\log(L_\infty)$ all instances of $\approx$ must be replaced by $<$ on the first line. These extensive simulations indicate that the new estimator, A, is preferable to its nonparametric competitors, D and P, in all the situations examined. Although it is always dominated by a maximum likelihood estimator, as might be expected, it is remarkable that, more often than not, A ranks second, and typically above the maximum likelihood estimator derived from the incorrect model. This shows the value of using a good nonparametric estimator as a guide for parametric model selection. Note in passing that, in situations of weak dependence, maximum likelihood estimation is sometimes unreliable. In Table 1, this is illustrated in the case where $\tau = \frac{1}{4}$ and the true model is asymmetric: despite the greater flexibility of model 2, the corresponding estimator is outperformed by the simpler, but incorrect maximum likelihood estimate based on the symmetric model, as well as by the new nonparametric estimator. Figure 1 gives an idea of the size of the L_1 error for the nonparametric estimator A, as a function of τ and sample size, for the symmetric logistic model.

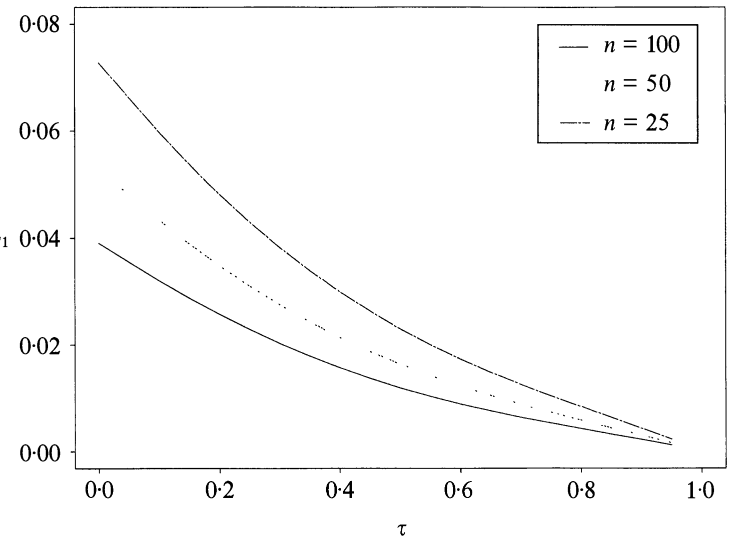

Fig. 1. Observed L_1 error of the nonparametric estimator A with weight function $p(t) = 1 - t$, based on 500 pseudo-random samples of size 100 from bivariate Gumbel extreme value distributions with different values of Kendall's tau.

6. Test for independence

It is well known that bivariate extreme value distributions are always associated in the sense of Esary, Proschan & Walkup (1967). Marshall & Olkin (1983), who established this result, pointed out that in practice the specific case of independence is often encountered. They listed several families of bivariate distributions that belong to the domain of attraction of independence; other examples were given by Tawn (1988) and Genest & Rivest (1989), among others.

Proposition 3·2 can be exploited to construct a test of independence, based on Tawn's measure of association, $2\{1 - A(\frac{1}{2})\}$, for, in the case where $A \equiv 1$ and $p(t) = 1 - t$, formula (3) reduces to

$$\gamma(t) = 2t(1-t)\{\eta(t) + \eta(1-t) - \log(t)\log(1-t)\},$$

with $\eta(t) = (1-t)\sum_{k=1}^{\infty} t^{k-1}/k^2$. This suggests rejecting the null hypothesis at level α whenever

$$T_n = -\{n/\gamma(\tfrac{1}{2})\}^{\frac{1}{2}} \log A_n(\tfrac{1}{2}) \approx -(n/0{\cdot}342)^{\frac{1}{2}} \log A_n(\tfrac{1}{2})$$

exceeds the quantile of order $1 - \alpha$ of the standard normal distribution. Based on Table 2, which reports the results of 100 000 replications of the normalised score T_n, it would appear that the asymptotic normal approximation is quite good, even in small samples. This contrasts with the behaviour of the normalised score statistic for independence in the logistic model, as considered by Tawn (1988).

Table 2. *Significance level (in percent) for the test of independence based on statistic T_n, as estimated from* 100 000 *replications of pseudo-random samples of size* $n = 25, 50, 100$.

	Nominal levels		
n	10%	5%	2·5%
25	9·18	4·01	1·62
50	9·53	4·29	1·88
100	9·60	4·55	2·09

7. Discussion

The developments presented here raise at least three additional issues that would deserve further attention. First, it would be of interest to devise an appropriate estimation procedure for the optimal weight function involved in the definition of estimator A_n. Secondly, the assumption of known marginals could be relaxed. Thirdly, multivariate extensions might be envisaged.

Of these three problems, the question of how to handle the case of unknown marginals is perhaps the most pressing. Since the parametric form of the extreme value marginals is known, a rough-and-ready solution along the lines of Gong & Samaniego (1981) would be to estimate these margins via maximum likelihood and to act as if they were fixed thereafter. In view of the numerical difficulties associated with the estimation of parameters in such models (Prescott & Walden, 1980), it is not clear that this solution would be particularly efficient or reliable in small samples. An alternative would be to use Bayesian methods, as suggested by Coles & Powell (1996). Another option inspired by recent work of Genest, Ghoudi & Rivest (1995) would consist of estimating the marginals by their empirical distribution functions F_n and G_n, and then computing an estimator of A based on the pseudo-observations $U_i = F_n(X_i)$ and $V_i = G_n(Y_i)$. Since the latter would then be functions of the ranks only, this approach would have the merit of being fully nonparametric, but the induced dependence between the U_i's and V_i's would make the mathematical analysis harder.

Acknowledgement

Research funds in partial support of this work were granted by the Natural Sciences and Engineering Research Council of Canada and by the Fonds pour la formation de chercheurs et l'aide à la recherche du Gouvernement du Québec.

Appendix 1

Proof of Proposition 3·1

The unbiasedness of $\log A_n^0(t)$ as an estimator of $\log A(t)$ is an immediate consequence of the fact that $E(H_n) = H$, upon interchanging the order of integration in the expression

$$E\{\log A_n^0(t)\} = E\left\{\int_0^t \frac{H_n(z) - z}{z(1-z)}\, dz\right\}.$$

This application of Fubini's theorem is justified by the fact that $|D(t)| \leqslant 1$ for all $0 \leqslant t \leqslant 1$, so that

$$\int_0^t \int_0^1 \frac{|1_{[0,z]}(x) - z|}{z(1-z)}\, dH(x)\, dz = \int_0^t \{2 + (1-2z)D(z)\}\, dz < \infty.$$

Next, it must be shown that

$$S = \sup_{t \in [0,1]} \left| \int_0^t \frac{H_n(z) - H(z)}{z(1-z)} \, dz \right| \to 0$$

almost surely as $n \to \infty$. If $\mathscr{H}$ denotes the support of H and if $0 < \alpha < 1$, then

$$S \leqslant \sup_{t \in \mathscr{H}} \frac{|H_n(t) - H(t)|}{[H(t)\{1 - H(t)\}]^\alpha} \int_0^1 \left[\frac{H(z)\{1 - H(z)\}}{z(1-z)} \right]^\alpha \{z(1-z)\}^{\alpha - 1} \, dz.$$

Observe that the latter integral is finite, since

$$\frac{H(z)\{1 - H(z)\}}{z(1-z)} = \{1 + (1-z)D(z)\}\{1 - zD(z)\} \leqslant 4, \tag{A1·1}$$

from the boundedness of D. The result then follows from an application of Lai's theorem, as stated on p. 410 of Shorack & Wellner (1986), with $\psi(t) = \{t(1-t)\}^{-\alpha}$.

To find the asymptotic distribution of the process $n^{\frac{1}{2}}(\log A_n^0 - \log A)$, it is necessary to call on a theorem of Mason and van Zwet (Shorack & Wellner, 1986, p. 501). According to this result, there exists a probability space on which sit jointly a sequence (Z_n^*) having the same distribution as the sequence (Z_n) and a sequence (B_n) of Brownian bridge processes such that, assuming $Z_n^* = Z_n$ without loss of generality, one has

$$n^{\frac{1}{2}}\{H_n(z) - H(z)\} = B_n\{H(z)\} 1_{[1/n, 1-1/n]}\{H(z)\} + [H(z)\{1 - H(z)\}]^{\frac{1}{2} - \nu} O_p(n^{-\nu})$$

for all $0 \leqslant \nu < \frac{1}{2}$. One may thus write

$$\begin{aligned} n^{\frac{1}{2}}\{\log A_n^0(t) - \log A(t)\} &= \int_0^t n^{\frac{1}{2}} \frac{H_n(z) - H(z)}{z(1-z)} \, dz \\ &= \int_0^t \frac{B_n\{H(z)\}}{z(1-z)} 1_{[1/n, 1-1/n]}\{H(z)\} \, dz + O_p(n^{-\nu}) \int_0^t \frac{[H(z)\{1 - H(z)\}]^{\frac{1}{2} - \nu}}{z(1-z)} \, dz. \end{aligned}$$

Observe that, because of (A1·1), the second term in this summand is $O_p(n^{-\nu})$, so that the processes

$$t \mapsto n^{\frac{1}{2}}\{\log A_n^0(t) - \log A(t)\}, \quad t \mapsto \int_0^t \frac{B_n\{H(z)\}}{z(1-z)} 1_{[1/n, 1-1/n]}\{H(z)\} \, dz$$

have the same asymptotic behaviour. Now as it turns out, their limiting distribution is precisely that of the process

$$t \mapsto \int_0^t \frac{B\{H(z)\}}{z(1-z)} \, dz, \tag{A1·2}$$

defined in terms of an arbitrary Brownian bridge B. Note that the latter integral exists for all $0 \leqslant t \leqslant 1$, because one can write $B\{H(z)\} = W\{H(z)\} - H(z)W(1)$ in terms of a Brownian motion $W(t)$ and $W\{H(z)\} = O[\{-2H(z) \log H(z)\}^{\frac{1}{2}}]$ in a neighbourhood of the origin; see for example Csörgő & Révész (1981, p. 26).

To complete the proof, it remains to show that (A1·2) is a centred Gaussian process with the desired covariance function. Again, this requires an application of Fubini's theorem. In the case of the expectation, this is justified by the fact that

$$\int_0^t \frac{E|B\{H(z)\}|}{z(1-z)} \, dz \leqslant E|\zeta| \int_0^t \frac{[H(z)\{1 - H(z)\}]^{\frac{1}{2}}}{z(1-z)} \, dz < \infty,$$

where ζ is a $\mathscr{N}(0, 1)$ random variable. A similar argument may also be invoked to handle the covariance.

Appendix 2

Proof of Proposition 4·1

It is immediate from Proposition 3·2 that A_n is a uniformly, strongly consistent estimator of A. To show that it is asymptotically unbiased, fix $0 \leqslant t \leqslant 1$ and let

$$T_t = \exp[p(t) \log A_1^0(t) + \{1 - p(t)\} \log A_1^1(t)]$$

be a positive random variable based on a single observation from distribution H. Since $Z_1, \ldots, Z_n$ form a random sample from H, it is clear that

$$\lim_{n \to \infty} E\{A_n(t)\} = \lim_{n \to \infty} \| T_t \|_{1/n},$$

where $\| T_t \|_\alpha = E(T_t^\alpha)$ for arbitrary $0 < \alpha < \infty$.

Suppose for a moment that $\| T_t \|_\alpha < \infty$ for one such α. It then follows from Exercise 5d of Rudin (1974, p. 74) that

$$\lim_{n \to \infty} E\{A_n(t)\} = \exp\{E(\log |T_t|)\} = \exp[E\{\log A_1(t)\}] = A(t)$$

in view of Proposition 3·2.

To show the existence of an appropriate α, first note that

$$\begin{aligned} E(T_t^\alpha) &= \int_0^1 \exp\left[\alpha(t) \int_0^t \frac{1_{[0,z]}(x) - z}{z(1-z)} dz - \alpha\{1 - p(t)\} \int_t^1 \frac{1_{[0,z]}(x) - z}{z(1-z)} dz\right] dH(x) \\ &= t^\alpha \int_0^t \left(\frac{1-x}{x}\right)^{\alpha p(t)} dH(x) + (1-t)^\alpha \int_t^1 \left(\frac{x}{1-x}\right)^{\alpha - \alpha p(t)} dH(x) \end{aligned} \qquad (A2·1)$$

and that, in the case where $H(x) \equiv x$, both summands are clearly finite whenever

$$\max\{\alpha p(t), \alpha - \alpha p(t)\} < 1.$$

For any such α, it turns out that $E(T_t^\alpha)$ is finite for any other H, as the following argument shows.

From the properties of A, there must exist $0 < a \leqslant b < 1$ such that $H(x) < x$ for $x < a$ and $H(x) > x$ for $x > b$. Introduce two auxiliary distribution functions

$$H_a(x) = \frac{H(x)}{a} 1_{[0,a)}(x) + 1_{[a,1]}(x), \quad H_b(x) = 1_{[0,b)}(x) - 1 + \frac{H(x) - b}{1 - b} 1_{[b,1]}(x),$$

and observe that H_a is stochastically larger than the uniform distribution on $[0, a]$, while H_b is stochastically smaller than the uniform distribution on $[b, 1]$. As the integrands in (A2·1) are monotone, it follows that

$$a \int_0^a \left(\frac{1-x}{x}\right)^{\alpha p(t)} dH_a(x) \leqslant \int_0^a \left(\frac{1-x}{x}\right)^{\alpha p(t)} dx,$$

$$(1-b) \int_b^1 \left(\frac{x}{1-x}\right)^{\alpha - \alpha p(t)} dH_b(x) \leqslant \int_b^1 \left(\frac{x}{1-x}\right)^{\alpha - \alpha p(t)} dx,$$

as an application of Theorem 1.2.2 of Stoyan (1983). The boundedness of $E(T_t^\alpha)$ is now an immediate consequence of this observation.

References

Coles, S. G. & Powell, E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *Int. Statist. Rev.* **64**, 119–36.

Coles, S. G. & Tawn, J. A. (1991). Modelling multivariate extreme events. *J. R. Statist. Soc.* B **53**, 377–92.

Coles, S. G. & Tawn, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design. *Appl. Statist.* **43**, 1–48.

Csörgő, M. & Révész, P. (1981). *Strong Approximation in Probability and Statistics*. New York: Academic Press.

Deheuvels, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statist. Prob. Lett.* **12**, 429–39.

Esary, J. D., Proschan, F. & Walkup, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38**, 1466–74.

Genest, C., Ghoudi, K. & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–52.

Genest, C. & Rivest, L.-P. (1989). A characterization of Gumbel's family of extreme value distributions. *Statist. Prob. Lett.* **8**, 207–11.

Gong, G. & Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Ann. Statist.* **9**, 861–9.

Haan, L. de (1985). Extremes in higher dimensions: The model and some statistics. In *Proc. 45th Session Int. Statist. Inst.*, Amsterdam, paper 26.3. Amsterdam: International Statistical Institute.

Hutchinson, T. P. & Lai, C. D. (1990). *Continuous Bivariate Distributions, Emphasising Applications*. Adelaide: Rumsby Scientific Publishing.

Joe, H., Smith, R. L. & Weissman, I. (1992). Bivariate threshold models for extremes. *J. R. Statist. Soc.* B **54**, 171–83.

Marshall, A. W. & Olkin, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Prob.* **11**, 168–77.

Pickands, J. (1981). Multivariate extreme value distributions. *Bull. Int. Statist. Inst.* 859–78.

Prescott, P. & Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Biometrika* **67**, 723–4.

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

Rudin, W. (1974). *Real and Complex Analysis*, 2nd ed. New York: McGraw-Hill.

Shorack, G. R. & Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Smith, R. L., Tawn, J. A. & Yuen, H. K. (1990). Statistics of multivariate extremes. *Int. Statist. Rev.* **58**, 47–58.

Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*. New York: Wiley.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75**, 397–415.

Tiago de Oliveira, J. (1989). Intrinsic estimation of the dependence structure for bivariate extremes. *Statist. Prob. Lett.* **8**, 213–8.

[*Received December* 1995. *Revised December* 1996]