

# Distribution and dependence-function estimation for bivariate extreme-value distributions

PETER HALL<sup>1</sup> and NADER TAJVIDI<sup>2</sup>

<sup>1</sup>*Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia*

<sup>2</sup>*Department of Mathematical Statistics, Lund Institute of Technology, Box 118, SE-22100 Lund, Sweden*

Two new methods are suggested for estimating the dependence function of a bivariate extreme-value distribution. One is based on a multiplicative modification of an earlier technique proposed by Pickands, and the other employs spline smoothing under constraints. Both produce estimators that satisfy all the conditions that define a dependence function, including convexity and the restriction that its curve lie within a certain triangular region. The first approach does not require selection of smoothing parameters; the second does, and for that purpose we suggest explicit tuning methods, one of them based on cross-validation.

*Keywords:* convex hull; cross-validation; marginal distribution; multivariate extreme-value distribution; nonparametric curve estimation; smoothing parameter; spline

## 1. Introduction

Bivariate extreme-value distributions may be estimated completely parametrically, or by a mixture of parametric and nonparametric techniques. The marginal distributions are necessarily of one of the three classical types, and so are determined by a finite number of parameters, but the link between the marginals, provided by the ‘dependence function’,  $A$ , is determined only up to the condition that it be convex, pass through the points  $P_1 = (0, 1)$  and  $P_2 = (1, 1)$ , and lie within the triangle  $\mathcal{T}$  determined by  $P_1$ ,  $P_2$  and  $P_3 = (\frac{1}{2}, \frac{1}{2})$ . See, for example, Geffroy (1958), Tiago de Oliveira (1958) and Sibuya (1960). Even smoothness (beyond that implied by convexity) is not specified. Therefore, one may either model the dependence function and fit a completely structural model to the data (see, for example, Tiago de Oliveira 1984; Tawn 1988; Smith *et al.* 1990; Hutchinson and Lai 1990, Chapter 9; Coles and Tawn 1991; Joe *et al.* 1992), or estimate the dependence function nonparametrically and the marginal distributions parametrically, and fit the resulting partly structural model.

In the present paper we propose new, nonparametric estimators of the dependence function. Existing techniques of this type include that of Pickands (1981; 1989); a modification proposed by Deheuvels (1991); kernel-type estimators suggested by Smith (1985), Smith *et al.* (1990) and Abdous *et al.* (1998); and a technique based on

exponentiation, proposed by Capéraà *et al.* (1997). These methods tend to suffer from at least one of several problems: they look unsmooth, or they do not produce estimators that satisfy the properties that a dependence function must enjoy, or they do not have good convergence rates. For example, Pickands' (1981; 1989) convex estimator of the dependence function, defined as it is in terms of the convex hull of a minimum of a maximum, can look relatively unsmooth. Deheuvels' (1991) modification, designed to overcome drawbacks of Pickands' approach, is not necessarily convex, and both it and its convex hull can protrude outside  $\mathcal{S}$ . Kernel and exponentiation estimators are not necessarily convex, and can also violate the other properties required of  $A$ .

One of our estimators is based on a simple, multiplicative modification of Pickands' (1981; 1989) approach, and produces an estimator that satisfies all the constraints that define a dependence function. Additionally, not only is it root- $n$  consistent for  $A$ , its derivative is root- $n$  consistent for the derivative of  $A$ . Our second method is founded on smoothing splines, constrained so as to satisfy the essential conditions. We suggest empirical methods for choosing the smoothing parameter.

Related work includes that of Coles and Tawn (1991; 1994), Joe *et al.* (1992), de Haan and Resnick (1993) and Einmahl *et al.* (1997) on estimating an extreme-value distribution from data on a distribution in its domain of attraction. Methods for estimating marginals of extreme-value distributions include those studied by Davis and Resnick (1984) and Dekkers and de Haan (1989). Techniques for estimating exponents of regular variation, in an extremal setting, have been studied by many authors (see, for example, Beirlant *et al.* 1996).

## 2. Distribution estimation

### 2.1. Definition of the dependence function

Let  $(X^{(1)}, X^{(2)})$  denote a random vector with bivariate extreme-value distribution function  $F$ . To express the distribution in canonical form, we note that the two marginal distributions are necessarily of extreme-value type, and so there exist finite vectors  $\theta_1, \theta_2$  of parameters, and monotone increasing transformations  $T_j = T_j(\cdot|\theta_j)$  depending only on the indicated parameters, such that  $(Y^{(1)}, Y^{(2)}) = (T_1(X^{(1)}), T_2(X^{(2)}))$  has a bivariate distribution with standard exponential marginals. Let the corresponding distribution function be  $G_0$ , and, given a vector  $x = (x^{(1)}, x^{(2)})$ , put  $G_1(x) = P(Y^{(1)} > x^{(1)}, Y^{(2)} = x^{(2)})$ . Then  $\gamma \equiv -\log G_1$  may be expressed in the form

$$\gamma(x^{(1)}, x^{(2)}) = (x^{(1)} + x^{(2)})A\left(\frac{x^{(2)}}{x^{(1)} + x^{(2)}}\right). \quad (2.1)$$

The function  $v = A(u)$ , called the *dependence function*, is necessarily convex, satisfies  $A(0) = A(1) = 1$ , and lies entirely within the triangular region  $\mathcal{S}$  defined in Section 1. We call these properties (C), and refer the reader to Pickands (1981) and Smith *et al.* (1990) for further details.

### 2.2. Estimating the dependence function

To estimate  $A$  nonparametrically, let  $Y_i = (Y_i^{(1)}, Y_i^{(2)})$ ,  $1 \leq i \leq n$ , be independent and identically distributed random variables with distribution  $G_0$ , and put  $\bar{Y}^{(\ell)} = n^{-1} \sum_i Y_i^{(\ell)}$  and  $\hat{Y}_i^{(\ell)} = Y_i^{(\ell)} / \bar{Y}^{(\ell)}$ ,  $\ell = 1, 2$ . Then

$$\hat{B}(u) \equiv n^{-1} \sum_{i=1}^n \min\{\hat{Y}_i^{(1)} / (1 - u), \hat{Y}_i^{(2)} / u\}$$

is uniformly root- $n$  consistent for  $B(u) \equiv A(u)^{-1}$  (see Section 4).

This estimator may be viewed as a modification of a proposal of Pickands (1981). It differs from an alternative modification suggested by Deheuvels (1991), in that the adjustment is multiplicative rather than additive, which has the effect of normalizing the marginal moments to the required form.

While  $\hat{B}(u)$  is not differentiable everywhere, it has infinitely many derivatives at all points except at  $\hat{Y}_i^{(2)} / (\hat{Y}_i^{(1)} + \hat{Y}_i^{(2)})$ ,  $1 \leq i \leq n$ ; and if we define  $\hat{B}'$  here by either left or right continuity, then  $\hat{B}$  equals the integral of  $\hat{B}'$ . The normalization used to compute  $\hat{Y}_i$  from  $Y_i$  forces the corresponding estimator of the dependence function to pass through the points  $(0, 1)$  and  $(1, 1)$ , and to have gradients  $-1$  and  $1$  at these respective points. It also implies that  $\hat{B}(u) \leq \min\{1/(1 - u), 1/u\}$ , or equivalently, that  $\hat{A} \equiv \hat{B}^{-1}$  lies above the lower boundary of the triangle  $\mathcal{T}$  introduced in Section 2.1. This fact, and the property  $\hat{A}(0) = \hat{A}(1) = 1$ , ensure that the greatest convex minorant,  $\tilde{A}$ , of  $\hat{A}$  satisfies conditions (C) and is differentiable at all but at most  $n$  points in  $[0, 1]$ , at each of which  $\tilde{A}$  is continuous.

Moreover, since  $\hat{A}$  is uniformly root- $n$  consistent for  $A$  on the full interval  $[0, 1]$ , then so too is  $\tilde{A}$ . If the distribution  $H$  of  $Y^{(1)} / (Y^{(1)} + Y^{(2)})$  has a bounded density (which virtually requires  $A$  to be twice differentiable in  $[0, 1]$ ) then both  $\hat{A}'$  and  $\tilde{A}'$  are uniformly root- $n$  consistent for  $A'$  on any interval  $[\varepsilon, 1 - \varepsilon]$ , for  $\varepsilon > 0$ . Also, the estimators  $\hat{A}, \tilde{A}$  have biases of order  $n^{-1}$  as estimators of  $A$ , and  $\hat{A}', \tilde{A}'$  have biases of order  $n^{-1}$  as estimators of  $A'$ . Details are given in Section 4, where it is also shown that these results continue to hold if the monotone transformations taking  $F$  to  $G_0$  are constructed empirically.

As an alternative to computing the greatest convex minorant,  $\hat{A}$  may be approximated by a spline that is constrained to satisfy (C). For example, we may choose regularly spaced points  $0 = t_0 < \dots < t_m = 1$  in the interval  $[0, 1]$ , and, given a smoothing parameter  $s > 0$ , take  $\tilde{A}_s$  to be a polynomial smoothing spline of degree 3 or more which minimizes

$$\sum_{j=1}^m \{\hat{A}(t_j) - \tilde{A}_s(t_j)\}^2 + s \int_0^1 \tilde{A}_s''(t)^2 dt, \tag{2.2}$$

subject to  $\tilde{A}_s'' \geq 0$  on  $[0, 1]$ ,  $\tilde{A}_s(0) = \tilde{A}_s(1) = 1$ ,  $\tilde{A}_s'(0) \geq -1$  and  $\tilde{A}_s'(1) \leq 1$ .

The choice of  $m$  is relatively unimportant, since the series in (2.2) is intended merely as an approximation to the integral of  $\{\hat{A}(t) - \tilde{A}_s(t)\}^2$  over  $[0, 1]$ , and so  $m$  can be taken large with virtually no penalty. One approach to choosing  $s$ , which we have found to work well in

practice, is to select the value that is as small as possible consistent with no numerical difficulties arising. This is motivated by the fact that, in order to preserve root- $n$  consistency of the basic estimator  $\hat{A}$ , we wish to smooth as little as possible away from  $\hat{A}$ , subject to producing a convex estimate.

Such an approach will not lead to consistent estimation of the second derivative of  $A$ . In some instances the density estimator  $\tilde{f}_s$  associated with the extreme-value distribution estimator based on  $\tilde{A}_s$  will be required as an estimator of the true density, and for that purpose it is usually necessary for  $\tilde{A}_s''$  to accurately approximate  $A''$ . In this case we suggest computing the cross-validation criterion,

$$CV(s) = \int \tilde{f}_s(x)^2 dx - 2n^{-1} \sum_{i=1}^n \tilde{f}_{-i,s}(X_i),$$

where  $X_1, \dots, X_n$  denote the original data with distribution  $F$ , and  $\tilde{f}_{-i,s}$  represents the version of  $\tilde{f}_s$  computed from the  $(n-1)$ -sample obtained by omitting  $X_i$ .

Note that  $CV(s)$  is an almost unbiased approximation to  $\int E(\tilde{f}_s^2 - 2\tilde{f}_s f)$ , and so the value of  $s$  that results from minimizing  $CV(s)$  will asymptotically minimize  $\int E(\tilde{f}_s - f)^2$ . We have applied this method to real data in the context of estimating bivariate distributions for purposes of prediction, and found it to give good results. Usually a graph of  $y = CV(s)$  has one or at most two local minima. In the case of two minima we choose the larger of the two values of  $s$ , as suggested by experience in more conventional density estimation problems; see, for example, Hall and Marron (1991). Occasionally, again as in the classical setting,  $CV(s)$  is monotone decreasing, and there one has little choice but to experiment with different values of  $s$ .

### 3. Simulation study

We simulated data from a bivariate extreme-value distribution having a symmetric logistic model for the dependence function, i.e.

$$A(w) = \{(1-w)^r + w^r\}^{1/r}, \quad r \geq 1. \quad (3.1)$$

See Tawn (1988) for extensions of this model to asymmetric cases. We took  $r = 1, 2$  or  $3$ , and compared performances of (a) model (3.1) fitted by maximum likelihood, (b) the Pickands (1981; 1989) estimator, (c) the convex hull of Pickands' estimator, (d) the estimator proposed by Capéraà *et al.* (1997), (e) the convex hull of the latter, (f) our modification of Pickands' estimator, and (g) the convex hull of the latter. In the case of (d) and (e) we used the weight function  $p(t) = 1 - t$  employed by Capéraà *et al.* (1997) in their simulation study. Additionally, we addressed performance of constrained smoothing splines fitted to all these estimators. In this instance the spline smoothing parameter was chosen as the smallest value that did not result in numerical difficulties, as suggested in Section 2.1.

Since the focus of our attention was estimation of the dependence function,  $A$ , we took the marginal distribution to be exponential. That is, we computed  $n$  independent values of

the pair  $(Y^{(1)}, Y^{(2)})$  for which  $G_1(x) = P(Y^{(1)} > x^{(1)}, Y^{(2)} = x^{(2)})$  was given by  $G_1 \equiv \exp(-\gamma)$ , with  $\gamma$  defined in (2.1). We could equivalently have simulated from a bivariate extreme-value distribution with the same dependence function and known marginal distributions.

In work not presented here we investigated the ‘naive’ estimator, defined to equal  $-\log \hat{G}_1(1 - u, u)$  where  $\hat{G}_1$  is the conventional empirical estimator of  $G_1$ , the latter defined in Section 2.1. For the naive estimator, however, mean square error was inferior to that of the next best approach by a factor of more than 1.7, reducing to at least 1.2 after spline smoothing, for each  $r$ . Other approaches considered in the literature, for example kernel methods, do not always enjoy root- $n$  consistency, and their performance depends very much on choice of bandwidth.

Table 1 gives mean integrated square errors (MISE), for example  $\int(\tilde{A} - A)^2$ , based on 200 independent simulations for methods (a)–(g) in the cases  $n = 25, 50$  and  $100$ . Interestingly, our results suggest that taking the convex hull of an estimator does not generally improve performance, apparently since non-convexity results from high-variance fluctuations at occasional point estimates that are a substantial distance from the true dependence function.

**Table 1.** Mean integrated square errors in simulation study. Monte Carlo approximations to mean integrated square errors, multiplied by  $10^5$ , of different estimators of the dependence function when the true function is bivariate symmetric logistic, defined in (3.1). Results are for  $r = 1, 2, 3$  and  $n = 25, 50, 100$ . The different methods are (a) model (3.1), (b) Pickands’ estimator, (c) convex hull of Pickands, (d) estimator in Capéraà *et al.*, (e) convex hull of Capéraà *et al.*, (f) modification of Pickand’s estimator, and (g) convex hull of the modified Pickand’s estimator.

	$n = 25$			$n = 50$			$n = 100$		
	$r = 1$	$r = 2$	$r = 3$	$r = 1$	$r = 2$	$r = 3$	$r = 1$	$r = 2$	$r = 3$
Method									
(a)	197	64	14	110	34	8	42	14	4
(b)	5614	2829	3331	2034	1547	1261	1172	712	567
(c)	7229	2611	2775	2588	1388	1049	1430	671	477
(d)	889	102	35	568	49	20	307	29	10
(e)	1188	95	41	666	57	25	373	32	12
(f)	1351	138	33	614	77	18	366	37	11
(g)	1861	139	46	815	70	24	453	38	15
Smoothed spline of									
(b)	784	919	1020	396	728	487	215	334	220
(c)	525	769	1055	282	637	490	135	327	230
(d)	303	82	21	177	37	12	97	24	8
(e)	286	66	22	167	39	14	97	26	11
(f)	447	104	21	240	62	12	130	31	9
(g)	401	73	24	232	49	16	107	30	15

Of the nonparametric estimators listed in the top panel of the table, not involving spline smoothing, only the convex hulls of the Pickands and the modified Pickands estimators are guaranteed to satisfy the essential properties (C) of a dependence function, and of these approaches the latter has greatest accuracy in terms of MISE. Importantly, if a dependence function estimate does not satisfy (C) then, if one wishes to estimate upper quantiles of the bivariate distribution function, which is arguably the application of greatest interest, it is not possible to use the dependence function estimate without modification.

The results in the bottom panel of Table 1 show that constrained spline smoothing consistently improves performance at the same time as guaranteeing conditions (C). The overall winner is constrained spline smoothing of the convex hull of the Capéraà *et al.* (1997) estimate, although usually only by a small margin over constrained spline smoothing of the convex hull of the simpler, modified Pickands approach.

An advantage of parametric methods is that they permit modelling of covariates and other structural features, thereby giving increased information, as well as increased accuracy, when the model is correct. Moreover, in the context of function estimation they usually admit relatively simple estimates of variability. By way of contrast, nonparametric methods often do not. Using the methods and results discussed in Section 4, one can develop asymptotic approximations to variability, but they are very complex, particularly when (as is almost always the case in practice) parameters of the marginal distributions are estimated from the same data. Nonparametric bootstrap methods offer a potential way out of this difficulty, but, for example in the case where  $\tilde{A}_s$  is constructed so as to provide a good estimator of  $\tilde{f}_s$ , they do not capture bias well. From a statistical viewpoint the most attractive approach is arguably to use the nonparametric estimator to suggest an approximate parametric model, to fit that to data, and to assess variability of the nonparametric estimator by simulating from the fitted model.

## 4. Theoretical properties

### 4.1. Root- $n$ consistent estimation of dependence function

We adopt notation from Section 2. In the present section we prove that  $\hat{A}$  and its greatest convex minorant,  $\tilde{A}$ , are uniformly root- $n$  consistent for  $A$ :

$$\sup_{0 \leq u \leq 1} |\hat{A}(u) - A(u)| + \sup_{0 \leq u \leq 1} |\tilde{A}(u) - A(u)| = O_p(n^{-1/2}). \quad (4.1)$$

Furthermore, we show that if the distribution  $H$  of  $Y^{(1)}/(Y^{(1)} + Y^{(2)})$  has a bounded density then, for each  $\varepsilon \in (0, \frac{1}{2}]$ ,

$$\sup_{\varepsilon \leq u \leq 1-\varepsilon} |\hat{A}'(u) - A'(u)| + \sup_{\varepsilon \leq u \leq 1-\varepsilon} |\tilde{A}'(u) - A'(u)| = O_p(n^{-1/2}). \quad (4.2)$$

(Both  $\hat{A}$  and  $\tilde{A}$  are differentiable except at the points  $\hat{W}_i \equiv \hat{Y}_i^{(2)}/(\hat{Y}_i^{(1)} + \hat{Y}_i^{(2)})$ , or at a subset of these points in the case of  $\tilde{A}$ . At  $\hat{W}_i$  we choose to define  $\hat{A}'$  and  $\tilde{A}'$  by left continuity.)

To derive (4.1) and (4.2), put  $V_i(u) \equiv \min\{Y_i^{(1)}/(1-u), Y_i^{(2)}/u\}$  and observe that  $V_i$  is differentiable everywhere except at  $u = W_i \equiv Y_i^{(2)}/(Y_i^{(1)} + Y_i^{(2)})$ , where we suggest defining  $V_i'$  by left continuity:

$$V_i'(u) = \begin{cases} Y_i^{(1)}/(1-u)^2 & \text{if } u \leq W_i, \\ -Y_i^{(2)}/u^2 & \text{if } u > W_i. \end{cases}$$

Let  $\check{B} = n^{-1}\sum_i V_i$  and  $\check{B}' = n^{-1}\sum_i V_i'$ . Then,  $\check{B}, \check{B}'$  are unbiased for  $B, B'$  respectively,

$$\sup_{0 \leq u \leq 1} |\check{B}(u) - B(u)| + \sup_{0 \leq u \leq 1} |\check{B}'(u) - B'(u)| = O_p(n^{-1/2}), \tag{4.3}$$

and both  $n^{1/2}(\check{B}' - B')$  and  $n^{1/2}(\check{B} - B)$  converge weakly, in the space  $D[0, 1]$  of left-continuous functions on  $[0, 1]$  with right-hand limits, to Gaussian processes. Compare Pickands (1989) and Deheuvels (1991).

Put  $M_1 = \{\max(\bar{Y}^{(1)}, \bar{Y}^{(2)})\}^{-1}$  and  $M_2 = \{\min(\bar{Y}^{(1)}, \bar{Y}^{(2)})\}^{-1}$ , and note that, since the variables  $Y_i^{(j)}$  ( $1 \leq i \leq n$ ) are independent and exponentially distributed with unit mean,  $M_j = 1 + O_p(n^{-1/2})$ . Also,  $M_1\check{B} \leq \check{B} \leq M_2\check{B}$ . Hence, (4.3) implies that  $\sup|\check{B} - B| = O_p(n^{-1/2})$ . This gives the first part of (4.1), and the second part then follows since  $A$  is itself convex.

It is not true, however, that  $\hat{A}'$  and  $\tilde{A}'$  are necessarily root- $n$  consistent for  $A'$ . Indeed,  $\hat{A}'(1) = -\hat{A}'(0) = 1$ , regardless of whether  $A'$  satisfies these constraints. Nevertheless, (4.2) holds if  $A$  is twice differentiable, which may be proved as follows. Define  $\delta = \max_j |(\bar{Y}^{(j)})^{-1} - 1| = O_p(n^{-1/2})$ ,  $\hat{W}_i = \hat{Y}_i^{(2)}/(\hat{Y}_i^{(1)} + \hat{Y}_i^{(2)})$ ,

$$\hat{V}_i'(u) = \begin{cases} \hat{Y}_i^{(1)}/(1-u)^2 & \text{if } u \leq \hat{W}_i \\ -\hat{Y}_i^{(2)}/u^2 & \text{if } u > \hat{W}_i, \end{cases}$$

$$v_i(u) = \begin{cases} Y_i^{(1)}/(1-u)^2 & \text{if } u \leq W_i, \\ -Y_i^{(2)}/u^2 & \text{if } u > W_i \end{cases}$$

and  $b(u) = n^{-1}\sum_i v_i(u)$ . Since  $|\hat{V}_i'(u) - v_i(u)| \leq \delta(Y_i^{(1)} + Y_i^{(2)})/\varepsilon^2$ , uniformly in  $u \in \mathcal{T}_\varepsilon \equiv [\varepsilon, 1 - \varepsilon]$ , and also  $\hat{B}' = n^{-1}\sum_i \hat{V}_i'$ , then

$$\sup_{u \in \mathcal{T}_\varepsilon} |\hat{B}'(u) - b(u)| = O_p(n^{-1/2}) \tag{4.4}$$

uniformly in  $u \in \mathcal{T}_\varepsilon$ .

Put  $\Delta = (\bar{Y}^{(2)} - \bar{Y}^{(1)})/\bar{Y}^{(1)} = O_p(n^{-1/2})$ , in which notation

$$W_i - \hat{W}_i = \frac{W_i(1 - W_i)\Delta}{1 + (1 - W_i)\Delta}.$$

It follows that if  $n \geq 4C^2$  and  $\delta \leq Cn^{-1/2}$  then  $|\hat{W}_i - W_i| \leq Cn^{-1/2}$ , whence, for  $\delta \leq Cn^{-1/2}$ ,

$$\begin{aligned} n\varepsilon^2|\check{B}'(u) - b(u)| &\leq \sum_{i=1}^n (Y_i^{(1)} + Y_i^{(2)})I\{(W_i - u)(\hat{W}_i - u) \leq 0\} \\ &\leq \sum_{i=1}^n (Y_i^{(1)} + Y_i^{(2)})I(|W_i - u| \leq Cn^{-1/2}) \equiv S(u), \end{aligned}$$

say. Since  $H$  has a bounded density then  $S(u) = O_p(n^{1/2})$  uniformly in  $u \in \mathcal{T}_\varepsilon$ . Hence, by (4.4),  $\sup_{u \in \mathcal{T}_\varepsilon} |\check{B}' - \hat{B}'| = O_p(n^{-1/2})$ . From this result and the second part of (4.3) we deduce that  $\sup_{u \in \mathcal{T}_\varepsilon} |\check{B}' - B'| = O_p(n^{-1/2})$ . And from this fact and the first part of (4.1) we see that  $\sup_{u \in \mathcal{T}_\varepsilon} |\check{B}'\check{B}^{-2} - B'B^{-2}| = O_p(n^{-1/2})$ , which is equivalent to the first part of (4.2).

Next we derive the second part of (4.2). Let  $u_0 \in \mathcal{T}_\varepsilon$ . From the definitions of  $\tilde{A}$  and  $\hat{A}$  we see that either (a)  $\tilde{A} \equiv \hat{A}$  in an interval  $[u_0 - \varepsilon_0, u_0 + \varepsilon_0]$  for some (stochastic)  $\varepsilon_0 > 0$ ; or (b) there exists a largest interval  $\tilde{\mathcal{I}} = \tilde{\mathcal{I}}(u_0) \equiv [u_1, u_2]$  (again, stochastic) such that  $\tilde{A} \leq \hat{A}$  on  $\tilde{\mathcal{I}}$ ,  $\tilde{A} = \hat{A}$  at the ends of  $\tilde{\mathcal{I}}$ , and  $\tilde{A}$  is linear on  $\tilde{\mathcal{I}}$ . In case (a) we have  $\tilde{A}'(u_0) = \hat{A}'(u_0)$ , and so

$$|\tilde{A}'(u_0) - A'(u_0)| \leq \sup_{u \in \mathcal{T}_\varepsilon} |\hat{A}'(u) - A'(u)|. \tag{4.5}$$

Now assume we are in case (b). Suppose first that either  $A'(u_0) \leq 0$  and  $A'(0) \neq -1$ , or  $A'(u_0) > 0$  and  $A'(1) \neq 1$ ; together, we call this subcase (i). Then it may be proved that, for some  $0 < \varepsilon_1 < \varepsilon$ , the event  $\mathcal{E}$  that  $u_1, u_2 \in \mathcal{T}_{\varepsilon_1} \equiv [\varepsilon_1, 1 - \varepsilon_1]$ , for all  $u_0 \in \mathcal{T}_\varepsilon$ , holds with probability tending to 1 as  $n \rightarrow \infty$ . We shall assume below that  $\mathcal{E}$  holds. The fact that  $\tilde{A}$  is the largest convex minorant of  $\hat{A}$  implies that  $\tilde{A}'(u_1+) \geq \tilde{A}'(u_0) \geq \tilde{A}'(u_2-)$ . (To appreciate why, note that if either of these inequalities fails then  $\tilde{A}$  will protrude below  $\hat{A}$  at the corresponding end of the line segment that comprises  $\tilde{A}$  over  $[u_1, u_2]$ .) Also, by convexity of  $A$ ,  $A'(u_1) \leq A'(u_0) \leq A'(u_2)$ . Hence,

$$\tilde{A}'(u_0) - A'(u_0) \begin{cases} \leq \hat{A}'(u_1+) - A'(u_1) \\ \geq \hat{A}'(u_2-) - A'(u_2), \end{cases}$$

which, since  $A'$  is continuous, implies (4.5) with  $\mathcal{T}_\varepsilon$  there replaced by  $\mathcal{T}_{\varepsilon_1}$ .

Finally, consider the previously omitted subcases of case (b), where either (ii)  $A'(u_0) \leq 0$  and  $A'(0) = -1$  or (iii)  $A'(u_0) > 0$  and  $A'(1) = 1$ . In subcase (ii), minor modifications of the argument leading to the first part of (4.2) show that, for all  $\varepsilon_1 > 0$ ,

$$\sup_{0 \leq u \leq 1 - \varepsilon_1} |\hat{A}'(u) - A'(u)| = O_p(n^{-1/2}).$$

The argument leading to subcase (i), but this time allowing  $u_1$  to vary anywhere in  $[0, 1 - \varepsilon_1]$  for some  $\varepsilon_1 > 0$ , may be employed to prove that

$$|\tilde{A}'(u_0) - A'(u_0)| \leq \sup_{0 \leq u \leq 1 - \varepsilon_1} |\hat{A}'(u) - A'(u)| = O(n^{-1/2}). \tag{4.6}$$

The second part of (4.2) follows from (4.5), its counterpart with  $\mathcal{T}_\varepsilon$  replaced by  $\mathcal{T}_{\varepsilon_1}$ , (4.6), and the analogue of (4.6) in subcase (iii).

### 4.2. Biases of $\hat{A}$ , $\hat{A}'$ , $\tilde{A}$ , $\tilde{A}'$

We claim that if  $A$  has three bounded derivatives then the biases are  $O(n^{-1})$ . (In the cases of  $\hat{A}'$  and  $\tilde{A}'$  these bounds to bias apply only interior to  $[0, 1]$ .) To derive the bounds in the cases of  $\hat{A}$  and  $\tilde{A}$ , note that by Taylor expansion,  $E(\hat{A} - A) = O(n^{-1})$  and  $E(\tilde{A} - A) = O(n^{-1})$  if the same bounds apply to  $E(\hat{B} - B)$  and  $E(\tilde{B} - B)$ . (A subsidiary argument, taking the Taylor expansion to higher-order terms and using relatively crude bounds there, may be employed to justify this step.) The  $O(n^{-1})$  bounds for the biases of  $\hat{B}$  and  $\tilde{B}$  follow on noting that, after direct calculation,

$$E\{\hat{V}_1(u) - V_1(u)\} = O(n^{-1}), \quad E\{\hat{V}'_1(u) - V'_1(u)\} = O(n^{-1}).$$

To obtain the latter results, first take expectations conditional on  $Y_2, \dots, Y_n$ , and use Taylor expansion to express the left-hand sides as smooth functions of  $\Delta^{(j)} \equiv n^{-1}(Y_2^{(j)} + \dots + Y_n^{(j)}) - 1$ ,  $j = 1, 2$ ; and then take expectations in the distributions of  $\Delta^{(j)}$ .

Arguments in the cases of  $\tilde{A}$  and  $\tilde{A}'$  are more complex, although note that we may write  $E(\tilde{A}) = E(\hat{A}_{\text{conv}}) + O(n^{-1})$  and  $E(\tilde{A}') = E(\hat{A}'_{\text{conv}}) + O(n^{-1})$ , where  $\hat{A}_{\text{conv}}$  equals the greatest convex minorant of  $B^{-1}\{1 - B^{-1}(\hat{B} - B)\}$ .

### 4.3. Case where marginal distributions are estimated

In practice, the transformation  $T_j = T_j(\cdot|\theta_j)$ , used to compute  $Y_i^{(j)} = T_j(X_i^{(j)})$  from the data, is computed empirically by substituting a root- $n$  consistent estimator  $\hat{\theta}_j$  for  $\theta_j$ . After this substitution the random variables  $Y_i$  are not, strictly speaking, exactly independent, although their correlation is  $O(n^{-1})$ . However, this does not affect convergence rates, as may be seen by a longer argument where  $T_j(\cdot|\hat{\theta}_j)$  is Taylor-expanded in  $\hat{\theta}_j$ , around  $\theta_j$ . It does influence the asymptotic distribution, however, although we do not address that aspect here.

## Acknowledgements

We are grateful to Dr Liang Peng and a reviewer for helpful comments. Nader Tajvidi's research was supported by a grant from the Swedish Foundation for International Cooperation in Research and Higher Education.

## References

- Abdous, B., Ghoudi, K. and Khoudraji, A. (1998) Non-parametric estimation of the limit dependence function of multivariate extremes. Manuscript.
- Beirlant, J., Vynckier, P. and Teugels, J. (1996) Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, **91**, 1659–1667.
- Capéreaà, P., Fougères, A.-L. and Genest, C. (1997) Estimation of bivariate extreme value copulas. *Biometrika*, **84**, 567–577.

- Coles, S.G. and Tawn, J.A. (1991) Modelling multivariate extreme events. *J. Roy. Statist. Soc. Ser. B*, **53**, 377–392.
- Coles, S.G. and Tawn, J.A. (1994) Statistical methods for multivariate extremes: an application to structural design. *Appl. Statist.*, **43**, 1–48.
- Davis, R. and Resnick, S.I. (1984) Tail estimates motivated by extreme value theory. *Ann. Statist.*, **12**, 1467–1487.
- de Haan, L. and Resnick, S.I. (1993) Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models*, **9**, 275–309.
- Dekkers, A.L.M. and de Haan, L. (1989) On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.*, **17**, 1795–1832.
- Deheuvels, P. (1991) On the limiting behaviour of the Pickands estimator for bivariate extreme-value distributions. *Statist. Probab. Lett.*, **12**, 429–439.
- Einmahl, J., de Haan, L. and Sinha, A.K. (1997) Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* To appear.
- Geffroy, J. (1958) Contribution à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris*, **7**, 37–185.
- Hall, P. and Marron, J.S. (1991) Local minima in cross-validation functions. *J. Roy. Statist. Soc. Ser. B*, **53**, 245–252.
- Hutchinson, T.P. and Lai, C.D. (1990) *Continuous Bivariate Distributions, Emphasising Applications*. Adelaide: Rumbsey Scientific Publishing.
- Joe, H., Smith, R.H. and Weissman, I. (1992) Bivariate threshold methods for extremes. *J. Roy. Statist. Soc. Ser. B*, **54**, 171–183.
- Pickands, J. (1981) Multivariate extreme value distributions. In *Bulletin of the International Statistical Institute: Proceedings of the 43rd Session (Buenos Aires)*, pp. 859–878. Voorburg, Netherlands: ISI.
- Pickands, J. (1989) Multivariate negative exponential and extreme value distributions. In J. Hüsler and R.-D. Reiss (eds), *Extreme Value Theory*, Lecture Notes in Statist. 51, pp. 859–878. Berlin: Springer-Verlag.
- Sibuya, M. (1960) Bivariate extreme statistics. *Ann. Inst. Statist. Math.*, **11**, 195–210.
- Smith, R.L. (1985) Statistics of extreme values. In *Bulletin of the International Statistical Institute: Proceedings of the 45th Session (Amsterdam)*, Vol. 4, Paper 26.1. Voorburg, Netherlands: ISI.
- Smith, R.L., Tawn, J.A. and Yuen, H.K. (1990) Statistics of multivariate extremes. *Internat. Statist. Rev.*, **58**, 47–58.
- Tawn, J.A. (1988) Bivariate extreme value theory: models and estimation. *Biometrika*, **75**, 397–415.
- Tiago de Oliveira, J. (1958) Extremal distributions. *Rev. Fac. Cienc. Lisboa Ser. 2 A Mat.*, **VII**, 215–227.
- Tiago de Oliveira, J. (1984) Bivariate models for extremes. In J. Tiago de Oliveira (ed.), *Statistical Extremes and Applications*, pp. 131–153. Dordrecht: Reidel.

Received December 1998 and revised April 1999