Bernoulli **10**(2), 2004, 251–280

Bivariate tail estimation: dependence in asymptotic independence

GERRIT DRAISMA[1], HOLGER DREES[2], ANA FERREIRA[3,4] and LAURENS DE HAAN[5]

[1]*Department of Public Health, Erasmus University Medical Center, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: g.draisma@erasmusmc.nl*
[2]*Department of Mathematics, University of Hamburg, Bundesstr, 55, 20146 Hamburg, Germany. E-mail: drees@math.uni-hamburg.de*
[3]*EURANDOM, Eindhoven and* [4]*ISA, Technical University of Lisbon, Tapada da Ajuda, 1349-017 Lisboa, Portugal. E-mail: anafh@isa.utl.pt*
[5]*Econometrisch Instituut, Erasmus Universiteit, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: ldehaan@few.eur.nl*

In the classical setting of bivariate extreme value theory, the procedures for estimating the probability of an extreme event are not applicable if the componentwise maxima of the observations are asymptotically independent. To cope with this problem, Ledford and Tawn proposed a submodel in which the penultimate dependence is characterized by an additional parameter. We discuss the asymptotic properties of two estimators for this parameter in an extended model. Moreover, we develop an estimator for the probability of an extreme event that works in the case of asymptotic independence as well as in the case of asymptotic dependence, and prove its consistency.

Keywords: asymptotic normality; bivariate extreme value distribution; coefficient of tail dependence; copula; failure probability; Hill estimator; moment estimator

1. Introduction

Suppose that (X_i, Y_i), $i = 1, \ldots, n$, is a sequence of independent and identically distributed (i.i.d.) random vectors. Given two large threshold values, u and v, we are interested in estimating probabilities of the type

$$P(X_i > u \text{ and } Y_i > v). \tag{1.1}$$

For instance, if (X_i, Y_i) are the levels of two different air pollutants, the exceedance of both at some prespecified levels may represent a dangerous situation to be avoided. In financial mathematics (X_i, Y_i) may represent the losses suffered in two different investments.

Let F be the common distribution function of (X_i, Y_i) with marginal distributions F_1 and F_2. Since only large values of X_i and Y_i are involved, one would expect multivariate extreme value theory to provide the appropriate framework for systematic estimation of the above probability. To be more specific, we assume that there exist normalizing constants a_n, $c_n > 0$ and b_n, $d_n \in \mathbb{R}$ such that

1350–7265

$$\begin{aligned}
&\lim_{n\to\infty} F^n(a_n x + b_n, c_n y + d_n) \\
&\quad = \lim_{n\to\infty} P\left(\frac{\max\{X_1, \ldots, X_n\} - b_n}{a_n} \leqslant x, \frac{\max\{Y_1, \ldots, Y_n\} - d_n}{c_n} \leqslant y\right) \\
&\quad = G(x, y), \qquad (1.2)
\end{aligned}$$

in the weak sense where G is a distribution function with non-denenerate marginals (Resnick 1987, Chapter 5).

We say that the maxima of the X_i and of the Y_i are asymptotically independent if $G(x, y) = G(x, \infty)G(\infty, y)$, for all x and y. This is a rather common situation; for instance, it holds for non-degenerate bivariate normal distributions with $|\rho| < 1$. Unfortunately, in this case the limit assumption (1.2) is of little help in estimating probability (1.1). Note that under the given conditions the marginal distributions F_1 and F_2 converge to the marginals G_1 and G_2, respectively, of the limiting distribution. Taking logarithms in (1.2), one obtains

$$\lim_{n\to\infty} nP\left\{\frac{X - b_n}{a_n} > x \text{ or } \frac{Y - d_n}{c_n} > y\right\} = -\log G(x, y), \qquad (1.3)$$

hence

$$\lim_{n\to\infty} nP\left\{\frac{X - b_n}{a_n} > x \text{ and } \frac{Y - d_n}{c_n} > y\right\} = \log G(x, y) - \log G_1(x) - \log G_2(y). \qquad (1.4)$$

Therefore if the marginals of the limiting distribution are independent, that is, $G(x, y) = G_1(x)G_2(y)$, the right-hand side in (1.4) is identically zero.

In order to overcome this problem, Ledford and Tawn (1996; 1997; 1998; see also Coles *et al*. 1999) introduced a submodel in which the penultimate tail dependence is characterized by a coefficient $\eta \in (0, 1]$. More precisely, they assumed that the function $t \mapsto P(1 - F_1(X) < t \text{ and } 1 - F_2(Y) < t)$ is regularly varying at 0 with index $1/\eta$. Then $\eta = 1$ in the case of asymptotic dependence, whereas $\eta < 1$ implies asymptotic independence. Ledford and Tawn also suggested estimators for the so-called coefficient of tail dependence η, but they did not establish their asymptotic properties.

In Section 2 of the present paper we interpret an extension of Ledford's and Tawn's condition as a bivariate second-order regular variation condition, thereby generalizing an approach by Peng (1999). Then we prove the asymptotic normality of modified versions of two estimators for η proposed by Ledford and Tawn. In Section 3 we set up a procedure to estimate the probability of a failure set of type (1.1). Its consistency is established under asymptotic independence as well as under asymptotic dependence. We report the results of a simulation study in Section 4. Here we compare the performance of both the estimators for η proposed in the present paper and the estimator introduced by Peng (1999). In addition, we examine the small-sample behaviour of tests for the hypothesis $\eta = 1$ which are based on these estimators. We also study the behaviour of the estimator of a failure probability in a simple situation. In Section 5 we investigate the dependence between still water level, wave heights and wave periods at a particular point in the Dutch coastal protection system. Section 6 contains the proofs of the results of Sections 2 and 3.

An extended simulation study and more detailed proofs can be found in the technical report by Draisma *et al*. (2001).

2. Estimating the coefficient of tail dependence

Let (X, Y) be a random vector whose distribution function F has continuous marginal distribution functions F_1 and F_2. Our basic assumption is that

$$\lim_{t \downarrow 0} \left(\frac{P\{1 - F_1(X) < tx \text{ and } 1 - F_2(Y) < ty\}}{q(t)} - c(x, y) \right) \bigg/ q_1(t) =: c_1(x, y) \tag{2.1}$$

exists, for all $x, y \geqslant 0$ with $x + y > 0$, some positive functions q and $q_1 \to 0$ as $t \to 0$, and a function c_1 which is neither constant nor a multiple of c. Moreover, we assume that the convergence is uniform on $\{(x, y) \in [0, \infty)^2 | x^2 + y^2 = 1\}$.

Essentially, relation (2.1) is a second-order regular variation condition for the function Q defined by $Q(x, y) := P\{1 - F_1(X) < x \text{ and } 1 - F_2(Y) < y\}$. The function $(x, y) \mapsto Q(1 - x, 1 - y)$ is sometimes called a copula survivor function. It follows that the function q is regularly varying at zero with index $1/\eta$ for some $\eta \in (0, 1]$ – in the paper by Ledford and Tawn (1996) $q(t)$ equals $t^{1/\eta}$. The function q_1 is also regularly varying at zero with an index $\tau \geqslant 0$. Without loss of generality we may take $c(1, 1) = 1$ and $q(t) = P\{1 - F_1(X) < t \text{ and } 1 - F_2(Y) < t\}$. For these results and more information on such a second-order condition, see the Appendices in de Haan and Resnick (1993) and Draisma *et al*. (2001).

In addition, we assume that $l := \lim_{t \downarrow 0} q(t)/t$ exists. This condition is always satisfied if $\eta < 1$ or $\tau > 0$. Since $F_1(X)$ and $F_2(Y)$ are uniformly distributed, obviously $\limsup q(t)/t \leqslant 1$. Moreover, $l = 0$ if $\eta < 1$, and $l > 0$ if the marginals are asymptotically dependent.

Our assumptions imply that (2.1) holds locally uniformly on $(0, \infty)^2$. The bivariate normal distribution satisfies these conditions: see the example at the end of this section. Several other examples were given by Ledford and Tawn (1997) and Heffernan (2000).

The function c is homogeneous of order $1/\eta$, that is, $c(tx, ty) = t^{1/\eta} c(x, y)$. The measure ν defined by $\nu([0, x] \times [0, y]) = c(x, y)$ inherits this homogeneity:

$$\nu(tA) = t^{1/\eta} \nu(A) \tag{2.2}$$

for $t > 0$ and all bounded Borel sets $A \subset [0, \infty)^2$.

The parameter η has the same meaning as in Ledford and Tawn (1996; 1997), and condition (2.1) is similar to condition (2.2) in Ledford and Tawn (1997). Under the given assumptions, $l > 0$ implies asymptotic dependence and $l = 0$ implies asymptotic independence. Hence $\eta < 1$ implies asymptotic independence.

We now turn to estimators for η, given an i.i.d. sample $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$. We start with an informal introduction to the estimators of Ledford and Tawn (1996). They proposed first to standardize the marginals to the unit Fréchet distribution, using either the empirical marginal distributions (that is, using the ranks of the components) or extreme value estimators for the marginal tails, and then to estimate η as the shape parameter of the minimum of the components, for example by the maximum likelihood estimator or the Hill estimator. However, since these estimators have larger bias for Fréchet distributions than for Pareto distributions (see Drees, 1998a, 1998b), we prefer to standardize to the unit Pareto distribution using the ranks of the components.

For this purpose consider the random variable

$$T := \frac{1}{1 - F_1(X)} \wedge \frac{1}{1 - F_2(Y)}.$$

Its distribution function F_T satifies $1 - F_T(t) = q(1/t)$; in particular, $1 - F_T$ is regularly varying with index $1/\eta$. Since the marginal distribution functions F_i are unknown, we replace them with their empirical counterparts. After a small modification to prevent division by 0, this leads to

$$T_i^{(n)} := \frac{n+1}{n+1-R_i^X} \wedge \frac{n+1}{n+1-R_i^Y}, \qquad i = 1, \ldots, n,$$

with R_i^X denoting the rank of X_i among $(X_1, X_2, \ldots, X_n)$ and R_i^Y that of Y_i among $(Y_1, Y_2, \ldots, Y_n)$.

Now η can be estimated by the maximum likelihood estimator $\hat{\eta}_1$ in a generalized Pareto model, based on the largest $m+1$ order statistics of the $T_i^{(n)}$ (cf. Drees *et al*. 2004); here $m = m(n)$ denotes an intermediate sequence, that is, $m \to \infty$ and $m/n \to 0$. (Smith (1987) defined the maximum likelihood estimator in terms of excesses over a high threshold u; here we use the random threshold $u = T^{(n)}_{n,n-m}$.)

Alternatively the Hill estimator can be used:

$$\hat{\eta}_2 := \frac{1}{m} \sum_{i=1}^{m} \log \frac{T^{(n)}_{n,n-i+1}}{T^{(n)}_{n,n-m}}.$$

Note that one important advantage of the maximum likelihood estimator over the Hill estimator in the classical i.i.d. setting, namely its location invariance, is not relevant here: there is no shift after standardizing the marginals to unit Pareto (see Lemma 6.2). Since $\hat{\eta}_2$ has smaller variance, one might expect $\hat{\eta}_2$ to outperform $\hat{\eta}_1$.

Theorem 2.1 (Asymptotic normality). *Assume that* (2.1) *holds with a function c that has first-order partial derivatives* $c_x = \partial c(x, y)/\partial x$ *and* $c_y = \partial c(x, y)/\partial y$. *Suppose that m is an intermediate sequence such that* $\sqrt{m} q_1(q^{-1}(m/n)) \to 0$ *as* $n \to \infty$. *Then* $\sqrt{m}(\hat{\eta}_i - \eta)$, $i = 1, 2$, *are asymptotically normal with mean* 0 *and variance*

$$\sigma_1^2 = (1+\eta)^2 (1-l)(1 - 2l c_x(1, 1) c_y(1, 1)),$$
$$\sigma_2^2 = \eta^2 (1-l)(1 - 2l c_x(1, 1) c_y(1, 1)),$$

respectively.

Remark 2.1. (i) Since $q_1 \circ q^{-1}$ is $\eta\tau$-varying at 0, for $\tau > 0$ the condition $\sqrt{m} q_1(q^{-1}(m/n)) \to 0$ is satisfied if $m = O(n^{2\eta\tau/(2\eta\tau+1)-\varepsilon})$ for some $\varepsilon > 0$.

(ii) Note that instead of (2.1) the weaker condition $\lim_{t \to 0} P\{1 - F_1(X) < tx$ and

$1 - F_2(Y) < ty\}/q(t) - c(x, y) = O(q_1(t))$ is sufficient to prove the assertions of Theorem 2.1. However, under (2.1) similar results can be easily deduced if the intermediate sequence m is such that $\sqrt{m}q_1(q^{-1}(m/n)) \to c \geqslant 0$. In that case, usually a non-negligible bias occurs if $c > 0$ (and the present results correspond to the simpler case $c = 0$).

In order to construct confidence intervals for η or to test the hypothesis $\eta = 1$, we need consistent estimators for the unknown quantities in the asymptotic variances in Theorem 2.1.

Theorem 2.2. *Define*

$$\hat{l} := \frac{m}{n} T^{(n)}_{n,n-m},$$

$$\hat{c}_x(1, 1) := \frac{\hat{k}^{5/4}}{n}\left(T^{n,\hat{k}^{-1/4})}_{n,n-m} - T^{(n)}_{n,n-m}\right),$$

with $\hat{k} := m/\hat{l}$, *and* $T^{(n,u)}_{n,i}$, $i = 1, \ldots, n$, *the order statistics of*

$$T_i^{(n,u)} := \min\left(\frac{n+1}{n+1-R_i^X}(1+u), \frac{n+1}{n+1-R_i^Y}\right), \qquad i = 1, \ldots, n,$$

and define $\hat{c}_y(1, 1)$ *analogously to* $\hat{c}_x(1, 1)$. *If the conditions of Theorem* 2.1 *hold then*

$$\hat{l} \xrightarrow{p} l.$$

If, in addition, $\eta = 1$ *then*

$$\hat{c}_x(1, 1) \xrightarrow{p} c_x(1, 1), \qquad \hat{c}_y(1, 1) \xrightarrow{p} c_y(1, 1).$$

Moreover, let

$$\hat{\sigma}_1^2 := (1 + \hat{\eta})^2(1 - \hat{l})(1 - 2\hat{l}\hat{c}_x(1, 1)\hat{c}_y(1, 1))$$

and define $\hat{\sigma}_2^2$ *likewise. Then* $\hat{\sigma}_i^2$, $i = 1, 2$, *are consistent estimators of* σ_i^2 *for all* $\eta \in (0, 1]$.

Remark 2.2. Note that $c_y(1, 1)$ may also be estimated by $1 - \hat{c}_x(1, 1)$ if $\eta = 1$.

Example 2.1. The bivariate normal distribution with mean 0, variance 1 and correlation coefficient $\rho \notin \{1, -1\}$, satisfies (2.1) with

$$\eta = (1 + \rho)/2, \qquad c(x, y) = (xy)^{1/(1+\rho)},$$

$$q(t) = k_1(\rho)t^{2/(1+\rho)}(-\log t)^{-\rho/(1+\rho)}\left\{1 - k_2(\rho)\frac{\log(-\log t)}{2\log t}\right\},$$

$$c_1(x, y) = -k_3(\rho) - k_4(x, y, \rho), \qquad q_1(t) = \frac{1}{2\log t},$$

where

$$k_1(\rho) = \frac{(1-\rho^2)^{3/2}}{(1-\rho)^2}(4\pi)^{-\rho/(1+\rho)}, \qquad k_2(\rho) = \frac{\rho}{1+\rho},$$

$$k_3(\rho) = \frac{\rho\log(4\pi)+2}{1+\rho} - \frac{(1+\rho)(2-\rho)}{1-\rho},$$

$$k_4(x, y, \rho) = \log x + \log y + \frac{(\rho-1)(\log x + \log y) + \rho\log x \log y - \rho(\log^2 x + \log^2 y)/2}{(1-\rho^2)}.$$

This can be checked using the tail expansion of the bivariate normal distribution by Ruben (1964) as given in Ledford and Tawn (1997), combined with a sufficiently precise expansion of the function f, the inverse function of $1/(1-\Phi)$ where Φ is the standard univariate normal distribution function:

$$f^2(t) = 2\log t - \log(\log t) - \log(4\pi) + \frac{\log(\log t)}{2\log t} + \frac{\log(4\pi)-2}{2\log t}$$
$$+ \frac{1}{2}\left(\frac{\log(\log t)}{2\log t}\right)^2 + o\left(\left(\frac{\log(\log t)}{\log t}\right)^2\right), \qquad \text{as } t \to \infty.$$

3. Estimation of failure probabilities

Throughout this section we assume that the marginal distribution functions F_i of F are continuous and belong to the domain of attraction of a univariate extreme value distribution, and that condition (2.1) holds.

If we wish to estimate the probability of an extreme set of the form $\{X > x \text{ or } Y > y\}$ and we assume that F belongs to the domain of attraction of a bivariate extreme value distribution, then we can use the approximate equality

$$P\{1-F_1(X) < 1-F_1(x) \text{ or } 1-F_2(Y) < 1-F_2(y)\}$$
$$\approx tP\{1-F_1(X) < (1-F_1(x))/t \text{ or } 1-F_2(Y) < (1-F_2(y))/t\}, \qquad (3.1)$$

since for small t the right-hand side can be estimated using the empirical distribution function (de Haan and Sinha 1999). However, if the marginals are asymptotically independent and the failure set is, for example, of the form $\{X > x \textit{ and } Y > y\}$ then a different approximation holds under condition (2.1):

$$P\{1-F_1(X) < 1-F_1(x) \text{ or } 1-F_2(Y) < 1-F_2(y)\}$$
$$\approx t^{1/\eta}P\{1-F_1(X) < (1-F_1(x))/t \text{ and } 1-F_2(Y) < (1-F_2(y))/t\}. \qquad (3.2)$$

We develop an estimation procedure which works in this situation.

More generally, we aim to be able to estimate the failure probability $p_n = P\{(X, Y) \in C_n\}$ for failure regions $C_n \subset [x_n, \infty] \times [y_n, \infty]$ for some $x_n, y_n \in \mathbb{R}$ such that

$$(x, y) \in C_n \Rightarrow [x, \infty] \times [y, \infty] \subset C_n. \tag{3.3}$$

The latter property means that if an observation (x, y) causes a failure (e.g., the concentrations of two pollutants exceed maximum acceptable levels) then an event with both components larger will do so, too. Asymptotically we let both x_n and y_n converge to the right endpoint of the pertaining marginal distribution to ensure that $p_n \to 0$, that is, that we are indeed estimating the probability of an extremal event.

The basic idea is to use a generalized version of the scaling property (3.2) to inflate the transformed failure set $(1 - F_1, 1 - F_2)(C_n) := \{(1 - F_1(x), 1 - F_2(y)) | (x, y) \in C_n\}$ such that it contains sufficiently many observations and hence the empirical probability gives an accurate estimate. Since the marginal distribution functions F_i are unknown, their tails are estimated by suitable generalized Pareto distributions.

We begin by recalling from univariate extreme value theory that there exist normalizing constants $a_i(n/k) > 0$ and $b_i(n/k) \in \mathbb{R}$ such that the following generalized Pareto approximation is valid:

$$1 - F_i(x) \approx \frac{k}{n}\left(1 + \gamma_i \frac{x - b_i(n/k)}{a_i(n/k)}\right)^{-1/\gamma_i} =: \frac{k}{n}(1 - F_{a_i b_i, \gamma_i}(x)), \qquad i = 1, 2, \tag{3.4}$$

for x close to the right endpoint $F_i^{-1}(1)$. Here a_i and b_i are abbreviations for $a_i(n/k)$ and $b_i(n/k)$, respectively; and $(1 + \gamma x)^{-1/\gamma}$ is defined as ∞ if $\gamma > 0$ and $x \leqslant -1/\gamma$, and as 0 if $\gamma < 0$ and $x \geqslant -1/\gamma$. Dekkers *et al.* (1989) proposed and analysed the following estimators of the parameters a_i, b_i and γ_i. Define

$$M_j(X) := \frac{1}{k}\sum_{i=1}^{k}(\log X_{n,n-i+1} - \log X_{n,n-k})^j, \qquad j = 1, 2,$$

$$\hat{\gamma}_1 := M_1(X) + 1 - \frac{1}{2}\left(1 - \frac{(M_1(X))^2}{M_2(X)}\right)^{-1},$$

$$\hat{b}_1\left(\frac{n}{k}\right) := X_{n,n-k},$$

$$\hat{a}_1\left(\frac{n}{k}\right) := \frac{X_{n,n-k}\sqrt{3M_1(X)^2 - M_2(X)}}{\sqrt{(1 - 4\hat{\gamma}_1^-)/((1 - \hat{\gamma}_1^-)^2(1 - 2\hat{\gamma}_1^-))}} \qquad \text{with } \hat{\gamma}_1^- := \hat{\gamma}_1 \wedge 0;$$

for $\hat{\gamma}_2$, $\hat{a}_2$ and $\hat{b}_2$ replace X by Y in these formulae. The estimator $\hat{\gamma}_i$ for the extreme value index γ_i is often called a moment estimator.

Using these definitions, $nk^{-1}(1 - F_i(x))$ may be estimated by

$$1 - F_{\hat{a}_i, \hat{b}_i, \hat{\gamma}_i}(x) = \left(1 + \hat{\gamma}_i \frac{x - \hat{b}_i(n/k)}{\hat{a}_i(n/k)}\right)^{-1/\hat{\gamma}_i}.$$

Write $\mathbf{1} - \mathbf{F}(x, y)$ as a shorthand for $(1 - F_1(x), 1 - F_2(y))$; likewise $\mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\boldsymbol{\gamma}} = (1 - F_{a_1, b_1, \gamma_1}, 1 - F_{a_2, b_2, \gamma_2})$ and $\mathbf{1} - \mathbf{F}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\boldsymbol{\gamma}}} = (1 - F_{\hat{a}_1, \hat{b}_1, \hat{\gamma}_1}, 1 - F_{\hat{a}_2, \hat{b}_2, \hat{\gamma}_2})$ are functions from $\mathbb{R}^2$ to $[0, \infty]^2$. Then, in view of (3.4), the transformed failure set $nk^{-1}(\mathbf{1} - \mathbf{F}(C_n))$ can be approximated by

$$D_n := \mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}(C_n)$$

which in turn is estimated by

$$\hat{D}_n := \mathbf{1} - \mathbf{F}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}(C_n).$$

Now we may argue heuristically as follows, using a generalization of the scaling property (3.2) to inflate the transformed failure set by the factor $1/c_n$ for some $c_n \to 0$ chosen in a suitable way:

$$\begin{aligned}
p_n &= P\{1 - \mathbf{F}(X, Y) \in \mathbf{1} - \mathbf{F}(C_n)\} \\
&\approx P\left\{\frac{n}{k}(\mathbf{1} - \mathbf{F}(X, Y)) \in D_n\right\} \\
&\approx c_n^{1/\eta} P\left\{\frac{n}{k}(\mathbf{1} - \mathbf{F}(X, Y)) \in \frac{D_n}{c_n}\right\} &\quad (3.5)\\
&\approx c_n^{1/\hat{\eta}} P\{(X, Y) \in B\}_{|B=\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}(\mathbf{1}-\hat{D}_n/c_n)} \\
&\approx c_n^{1/\hat{\eta}} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{(X_i, Y_i) \in \mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}(\mathbf{1} - \hat{D}_n/c_n)\} &\quad (3.6)\\
&=: \hat{p}_n &\quad (3.7)
\end{aligned}$$

where $\hat{\eta}$ denotes one of the estimators for η examined in Section 2.

In the following, we state the exact conditions under which we will prove consistency of the estimator $\hat{p}_n$, that is, $\hat{p}_n/p_n \to 1$ in probability as $n \to \infty$. For the sake of simplicity, we will not determine the non-degenerate limit distribution of the standardized estimation error. However, employing the ideas of de Haan and Sinha (1999), one may establish asymptotic normality of $\hat{p}_n$ under more complex conditions.

To study the asymptotic behaviour of $\hat{p}_n$, we have to impose a regularity condition on the sequence of failure sets C_n, or rather on the transformed sets D_n. Note that D_n will shrink towards the origin because we are interested in extremal events. We assume that, after a suitable standardization, D_n converges in the following sense:

Condition D. *There exist a sequence $d_n \to 0$ and a measurable bounded set $A \subset [0, \infty)^2$ with $\nu(A) > 0$ such that for all $\varepsilon > 0$ one has, for sufficiently large n,*

$$A_{-\varepsilon} \subset \frac{D_n}{d_n} \subset A_{+\varepsilon}.$$

Here $A_{+\varepsilon} := \{\mathbf{x} \in [0, \infty)^2 | \inf_{\mathbf{y}\in A}\|\mathbf{x} - \mathbf{y}\| \leqslant \varepsilon\}$ and $A_{-\varepsilon} := [0, \infty)^2 \backslash (([0, \infty)^2 \backslash A)_{+\varepsilon})$ denote the outer and inner ε-neighbourhood of A with respect to the maximum norm $\|\mathbf{x} - \mathbf{y}\| = |x_1 - y_1| \vee |x_2 - y_2|$, and ν is the measure corresponding to the function c (cf. Section 2).

Note that d_n and A are not determined by this condition as the former may be multiplied by a fixed factor and the latter divided by the same number. Moreover, even for given d_n the set A is determined only up to its boundary.

Condition (3.3) on C_n implies

$$(x, y) \in D_n \Rightarrow [0, x] \times [0, y] \subset D_n. \tag{3.8}$$

Example 3.1. For $C_n = [x_n, \infty] \times [y_n, \infty]$ we have $D_n = [0, 1 - F_{a_1,b_1,\gamma_1}(x_n)] \times [0, 1 - F_{a_2,b_2,\gamma_2}(y_n)]$. Hence Condition D is satisfied with $d_n = 1 - F_{a_1,b_1,\gamma_1}(x_n)$ if $(1 - F_{a_2,b_2,\gamma_2}(y_n))/(1 - F_{a_1,b_1,\gamma_1}(x_n))$ converges in $(0, \infty)$.

This example demonstrates that Condition D essentially means that the convergence of the failure set in the x- and the y-direction is balanced.

Next we need a certain rate of convergence for the marginal estimators to ensure that the transformation of the failure set does not introduce too big an error. For that purpose note that

$$R_i(t, x) := t(1 - F_i(a_i(t)x + b_i(t))) - (1 - \gamma_i x)^{-1/\gamma_i} \to 0, \qquad i = 1, 2,$$

locally iniformly for $x \in (0, \infty]$ as $t \to \infty$, since F_i belongs to the domain of attraction of an extreme value distribution (cf. (3.4)). Here we impose the following slightly stricter condition:

$$R_{x_1,x_2}(t) := \max_{i=1,2} \sup_{x_i < x < 1/((-\gamma_i)\vee 0)} |R_i(t, x)(1 + \gamma_i x)^{1/\gamma_i}| \to 0 \tag{3.9}$$

for some $-1/(\gamma_i \vee 0) < x_i < 1/((-\gamma_i) \vee 0)$, $i = 1, 2$. Observe that then (3.9) even holds for all such x_i. For example, if F_i satisfies the second-order condition

$$\frac{R_i(t, x)}{A_i(t)} \to \Psi(x)$$

for some ρ_i-varying function A_i with $\rho_i < 0$ $(i = 1, 2)$, then (3.9) holds with $R_{x_1,x_2}(t) = O(A_1(t) \vee A_2(t))$. In addition, we require that not too many order statistics are used for estimation of the marginal parameters:

$$k^{1/2} R_{x,x}\left(\frac{n}{k}\right) = O(1) \tag{3.10}$$

for some $x < 0$. Then it follows that the estimators $\hat{a}_i$, $\hat{b}_i$ and $\hat{\gamma}_i$ are $\sqrt{k}$-consistent in the following sense:

$$\left|\frac{\hat{a}_i}{a_i} - 1\right| \vee \left|\frac{\hat{b}_i - b_i}{a_i}\right| \vee |\hat{\gamma}_i - \gamma_i| = O_P(k^{-1/2}), \qquad i = 1, 2 \tag{3.11}$$

(cf. Dekkers *et al*. 1989; de Haan and Resnick 1993).

We will see that using the estimated parameters instead of the unknown true ones for the transformation of the failure sets does not cause problems provided

$$w_{\gamma_1 \wedge \gamma_2}(d_n) = o(k^{1/2}) \qquad \text{with } w_\gamma(x) := -x^\gamma \int_x^1 u^{-\gamma-1} \log u \, \mathrm{d}u. \tag{3.12}$$

Check that

$$w_\gamma(x) \sim \begin{cases} -\dfrac{1}{\gamma}\log x, & \gamma > 0 \\ \dfrac{(\log x)^2}{2}, & \gamma = 0 \\ \dfrac{x^\gamma}{\gamma^2}, & \gamma < 0, \end{cases}$$

as $x \to 0$. Though at first glance (3.12) seems rather strict a condition if one of the extreme value indices is negative, it is indeed a natural one; for without it the difference between the transformed set D_n and its estimate $\hat{D}_n$ would be at least of the same order in probability as the typical elements of D_n, namely at least of the order d_n, which of course would render impossible any further statistical inference on the failure probability.

In addition, the scaling factor c_n chosen by the statistician when applying the estimator $\hat{p}_n$ must be related to the actual scaling factor d_n as follows:

$$d_n = O(c_n), \qquad w_{\gamma_1 \wedge \gamma_2}\left(\frac{c_n}{d_n}\right) = o(k^{1/2}), \qquad \left(\frac{c_n}{d_n}\right)^{1/\eta} = o((r(n))^{1/2}), \tag{3.13}$$

with $r(n) := nq(k/n)$. In particular, (3.13) is satisifed if c_n and d_n are of the same order. Below the choice of c_n is discussed more thoroughly.

Note that the scaling property (3.2) is a consequence of approximation (2.1) and the homogeneity of the measure ν. In order to justify (3.5) in the motivation for $\hat{p}_n$ given above, we need the following condition, which applies to more general sets than just upper quadrants:

$$\sup_{B \in \overline{\mathcal{B}}_n} \left| \frac{P\{\mathbf{1} - \mathbf{F}(X, Y) \in \mathbf{1} - \mathbf{F}(B)\}}{q(kn^{-1})\nu(nk^{-1}(\mathbf{1} - \mathbf{F}(B)))} - 1 \right| \to 0, \qquad \text{as } n \to \infty, \tag{3.14}$$

where

$$\mathcal{B}_n := \left\{ \mathbf{F}^{-1}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}},\tilde{\boldsymbol{\gamma}}}\left(\mathbf{1} - \frac{\mathbf{1} - \mathbf{F}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}},\tilde{\boldsymbol{\gamma}}}(C_n)}{c_n}\right) \middle| \left\| \frac{\tilde{\mathbf{a}}}{\mathbf{a}} - 1 \right\| \vee \left\| \frac{\tilde{\mathbf{b}} - \mathbf{b}}{\mathbf{a}} \right\| \vee \|\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\| \leqslant \varepsilon_n \right\}$$

for some $\varepsilon_n \to 0$ such that $k^{1/2}\varepsilon_n \to \infty$, and

$$\overline{\mathcal{B}}_n := \mathcal{B}_n \cup \left\{ C_n, \bigcup_{B \in \mathcal{B}_m, m \geqslant n} B \right\}.$$

It will turn out (see (6.16)) that for sufficiently large n the denominator in (3.14) is strictly positive.

Notice that the convergence of the absolute value in (3.14) for sets of the form $\mathbf{1} - \mathbf{F}(B) = [0, xk/n] \times [0, yk/n]$ follows from convergence (2.1) with $t = k/n$.

Finally, to make approximation (3.6) rigorous, we need a kind of uniform law of large numbers. This is provided by the theory of Vapnik–Chervonenkis (VC) classes of sets as outlined, for example, in the monograph by Pollard (1984, Section II.4). For this we require

$$\mathcal{B} = \bigcup_{n \in \mathbb{N}} \mathcal{B}_n \text{ is a VC class.} \tag{3.15}$$

Theorem 3.1. *Suppose that Conditions* D, (3.3) *(or* (3.8)*),* (3.9), (3.10) *and* (3.12)–(3.15) *are satisfied. If* $\hat{\eta} - \eta = O_P((r(n))^{-1/2})$*,* $\log c_n = o((r(n))^{1/2})$*, and* $k(n)/n$ *is almost decreasing, which means* $\sup_{m \geqslant n} k(m)/m = O(k(n)/n)$*, then*

$$\frac{\hat{p}_n}{p_n} \to 1 \textit{ in probability.}$$

Remark 3.1. (i) In the most important case that np_n is bounded, the conditions (3.12)–(3.14) can be jointly satisifed only if $\gamma_1 \wedge \gamma_2 > -1/2$.

(ii) If the conditions of Theorem 2.1 are satisfied for $m(n) = \lfloor r(n) \rfloor$, then $\hat{\eta}_i - \eta = O_P((r(n))^{-1/2})$.

(iii) The sequence $k(n)/n$ is almost decreasing, for example, if $k(n)$ is regularly varying with index less than 1 or, more generally, has an upper Matuszewska index $\alpha \leqslant 1$ (see Bingham *et al*., 1987, Theorem 2.2.2).

The scaling factor $1/c_n$ by which the transformed failure set is inflated determines the number of large observations taken into account for the empirical probability (3.6). More precisely, according to (6.17) in the proof of Lemma 6.6, this number is of the order $r(n)/(d_n/c_n)^{1/\eta}$. Hence if d_n and c_n are of the same order and $\hat{\eta}$ is based on the largest $m(n) = \lfloor r(n) \rfloor$ order statistics of $T_i^{(n)}$, then the numbers of observations used in both steps of the estimation procedure are of the same order of magnitude, which seems quite natural.

In practice, of course, d_n and $r(n)$ are not known. However, one may conversely choose c_n such that about $r(n)$ observations lie in the inflated set $\hat{D}_n/c_n$. To be more concrete, let

$$c_n(\lambda) := \sup\left\{ c > 0 \,\middle|\, \sum_{i=1}^{n} \mathbf{1}\left\{ (X_i, Y_i) \in \mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\left(\mathbf{1} - \frac{\hat{D}_n}{c} \right) \right\} \geqslant \lambda \widehat{r(n)} \right\} \tag{3.16}$$

for some $\lambda > 0$, where

$$\widehat{r(n)} := \sum_{i=1}^{n} \mathbf{1}\{X_i > X_{n,n-k} \text{ and } Y_i > Y_{n-k,n}\}.$$

Following the lines of the proof of Theorem 3.1, one may show that the estimator $\hat{p}_n$ is consistent for p_n if one chooses $c_n = c_n(\lambda)$ and $m(n) = \widehat{r(n)}$.

Alternatively, one may copy a heuristic approach which is common in univariate extreme value statistics: one plots $\hat{p}_n$ as a function of c_n and chooses a value c_n where this graph seems sufficiently stable.

Finally, it is worth mentioning that one may use other estimators for the marginal parameters, such as the maximum likelihood estimator examined by Smith (1987) and Drees *et al*. (2004), provided these estimators converge with the same rate.

4. Simulations

We examine the small-sample behaviour of the estimators $\hat{\eta}_1$ and $\hat{\eta}_2$ for four different distributions:

(i) the bivariate Cauchy distribution ($\eta = 1$);
(ii) the bivariate extreme value distribution (BEV) with a logistic dependence function, with $\alpha = 0.75$ ($\eta = 1$) (Ledford and Tawn 1996; 1997);
(iii) the bivariate normal distribution with correlation $\rho = 0.6$ ($\eta = 0.8$);
(iv) the Morgenstern distribution with $\alpha = 0.75$ ($\eta = 0.5$) (Ledford and Tawn 1996; 1997).

From each distribution we generated 250 samples of size 1000. All calculations were carried out with the GAUSS package. For comparison we also simulated Peng's (1999) estimator of η:

$$\hat{\eta}_3 := \log 2/\log\left(\frac{S_n(m)}{S_n(\lfloor m/2\rfloor)}\right) \qquad \text{with } S_n(k) := \sum_{i=1}^{n} \mathbf{1}\{X_i > X_{n,n-k} \text{ and } Y_i > Y_{n,n-k}\}.$$

Note that the meaning of m is different here from that in the definitions of $\hat{\eta}_1$ and $\hat{\eta}_2$.

In Table 1, besides the averages, the root mean squared errors and the standard deviations of the estimates, we report the means of two different estimates of the approximate standard deviation obtained in Theorem 2.1 and Peng's Theorems 2.1 and 2.2. The first estimator, referred to as $\hat{\sigma}_i(\hat{\eta}_i)$, is defined as $\hat{\sigma}_i m^{-1/2}$ with $\hat{\sigma}_i$ defined in Theorem 2.2 for $i = 1, 2$, while for $i = 3$ we use the variance estimator proposed by Peng (1999); the second variance estimator, called $\hat{\sigma}_i(1)$, is defined similarly but the estimator for η is replaced with 1.

These estimators for the standard deviations can be used to construct two different tests for asymptotic dependence (or $\eta = 1$) with nominal size 0.05. More concretely, asymptotic dependence is accepted if $(1 - \hat{\eta}_i)/\hat{\sigma}_i(\hat{\eta}_i) \leqslant \Phi^{-1}(0.95)$ or, alternatively, if $(1 - \hat{\eta}_i)/\hat{\sigma}_i(1) \leqslant \Phi^{-1}(0.95)$, with Φ^{-1} denoting the standard normal quantile function. The proportion of simulations in which the hypothesis $\eta = 1$ is accepted is also reported in Table 1. Finally, the number of simulations in which the test statistics could not be calculated is given in the last column. For the maximum likelihood estimator this occurred when no solution of the likelihood equations could be found, for Peng's estimator $S_n(m)$ may be equal to $S_n(\lfloor m/2\rfloor)$ and the estimated variance can be negative. The values m were chosen in a range where the overall performance of the tests seems best.

In Figure 1 the averages of the observed $\hat{\eta}_i$ are plotted against m, and the standard deviations of the three estimators are indicated for the Cauchy and the normal distribution.

The maximum likelihood estimator $\hat{\eta}_1$ exhibits the greatest stability with respect to the choice of m, but it is biased downward for the BEV and the normal distribution. The Hill estimator $\hat{\eta}_2$ is also biased downward for the Cauchy, the BEV and the normal distribution, and the bias increases rapidly with m. Peng's estimator is nearly unbiased for small values of m, but it shows a growing negative bias in particular for the Cauchy and the BEV distribution. The variance of the estimates is smallest for $\hat{\eta}_2$ and largest for $\hat{\eta}_3$. The estimators for the standard deviations are reasonably accurate.

Table 1 shows that the tests based on the maximum likelihood estimator $\hat{\eta}_1$ perform best.

Table 1. Mean and root mean squared errors (RMSE) of $\hat{\eta}_i$, observed standard deviation of the estimator, mean of estimates $\hat{\sigma}_i(\hat{\eta}_i)$ and $\hat{\sigma}_i(1)$, and proportion of samples in which $\eta = 1$ is accepted by a 5% test, based on $\hat{\sigma}_i(\hat{\eta}_i)$ or $\hat{\sigma}_i(1)$. The last column indicates the number of simulations where calculations failed (sample size $n = 1000$, 250 simulations)

		$\hat{\eta}_i$		Standard deviation			$\eta = 1$ accepted; test		No.
	m_n	Mean	RMSE	Obs.	$\hat{\sigma}_i(\hat{\eta}_i)$	$\hat{\sigma}_i(1)$	with $\hat{\sigma}_i(\hat{\eta}_i)$	with $\hat{\sigma}_i(1)$	failed
Maximum likelihood, $\hat{\eta}_1$									
Cauchy	80	0.98	0.18	0.18	0.17	0.17	0.89	0.92	0
	160	1.03	0.13	0.13	0.11	0.11	0.93	0.95	0
	240	1.04	0.10	0.09	0.08	0.08	0.95	0.94	4
BEV	80	0.91	0.18	0.16	0.15	0.15	0.81	0.86	0
	160	0.91	0.15	0.11	0.09	0.10	0.68	0.72	0
	240	0.90	0.13	0.09	0.07	0.07	0.55	0.58	0
Normal	80	0.72	0.18	0.17	0.13	0.15	0.36	0.38	0
	160	0.74	0.13	0.12	0.08	0.09	0.16	0.18	0
	240	0.74	0.11	0.09	0.06	0.07	0.04	0.05	0
Morgenstern	80	0.47	0.16	0.16	0.12	0.17	0.05	0.06	0
	160	0.49	0.11	0.11	0.08	0.10	0.00	0.00	0
	240	0.50	0.08	0.08	0.06	0.08	0.00	0.00	0
Hill, $\hat{\eta}_2$									
Cauchy	40	0.93	0.14	0.12	0.11	0.12	0.81	0.88	0
	80	0.89	0.14	0.08	0.08	0.09	0.57	0.63	0
	120	0.84	0.17	0.06	0.06	0.07	0.15	0.22	0
BEV	40	0.87	0.17	0.11	0.10	0.11	0.60	0.71	0
	80	0.84	0.17	0.08	0.06	0.08	0.29	0.34	0
	120	0.82	0.19	0.06	0.05	0.06	0.05	0.08	0
Normal	40	0.73	0.12	0.10	0.08	0.11	0.12	0.18	0
	80	0.74	0.09	0.07	0.05	0.07	0.01	0.01	0
	120	0.73	0.08	0.05	0.04	0.06	0.00	0.00	0
Morgenstern	40	0.51	0.07	0.07	0.07	0.13	0.00	0.00	0
	80	0.53	0.06	0.05	0.04	0.08	0.00	0.00	0
	120	0.54	0.06	0.04	0.03	0.06	0.00	0.00	0
Peng, $\hat{\eta}_3$									
Cauchy	40	1.05	0.37	0.36	0.23	0.25	0.92	1.00	6
	80	0.97	0.18	0.18	0.16	0.18	0.88	1.00	1
	120	0.88	0.17	0.12	0.11	0.14	0.67	0.97	1
BEV	40	0.96	0.23	0.23	0.20	0.23	0.90	1.00	5
	80	0.85	0.20	0.12	0.12	0.17	0.60	0.97	2
	120	0.80	0.21	0.09	0.09	0.14	0.28	0.67	0
Normal	40	0.78	0.20	0.19	0.18	0.30	0.60	1.00	2
	80	0.75	0.10	0.09	0.12	0.19	0.27	0.94	0
	120	0.74	0.09	0.07	0.09	0.14	0.05	0.27	0
Morgenstern	40	0.55	0.23	0.22	0.24	0.74	0.32	1.00	10
	80	0.54	0.11	0.11	0.12	0.37	0.03	1.00	0
	120	0.55	0.08	0.07	0.09	0.25	0.00	0.10	0

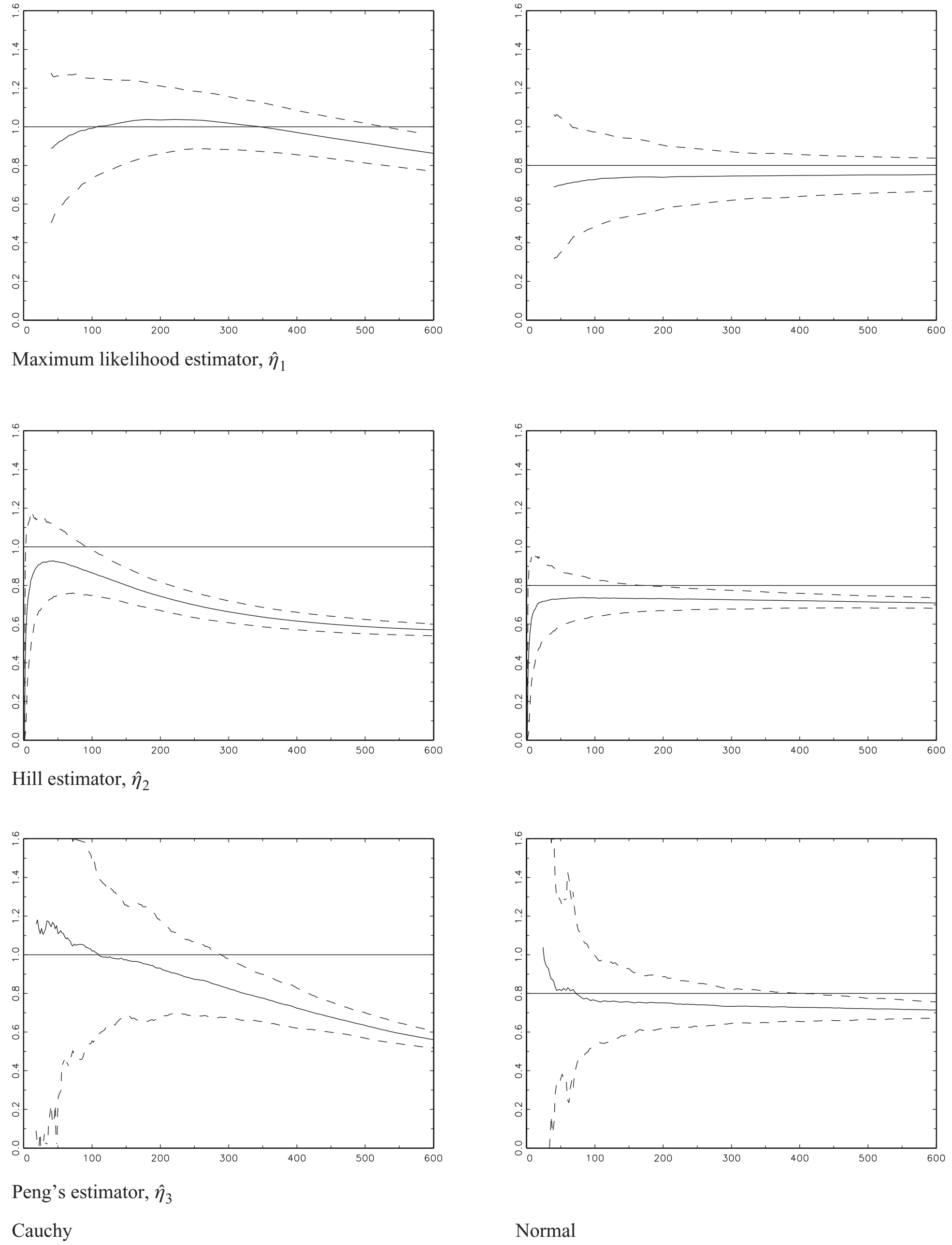

Figure 1. Estimators of η versus m for bivariate Cauchy (left) and normal distributions (right); average over 250 simulations (solid line) and ± 1.64 standard deviations (dashed lines). Horizontal line: true η.

For the Cauchy and the BEV distribution, the smaller variance and the somewhat larger bias of the Hill estimator lead to an empirical size of the test based on this estimator that is much larger than the nominal size. Conversely, the number of simulations in which the test based on Peng's estimator and $\hat{\sigma}_3(1)$ rejects the hypothesis is quite low for the normal and the Morgenstern distribution, because $\hat{\sigma}_3(1)$ is rather large.

We also studied the finite-sample behaviour of the proposed estimator of a failure probability. For this we considered failure sets of the form $[a, \infty)^2$ where a is chosen such that the failure probability p_n equals $(100n)^{-1} = 10^{-5}$ for sample size $n = 1000$. We use the maximum likelihood estimator $\hat{\eta}_1$ to estimate the coefficient of tail dependence and consider the following three estimators of p_n: $\hat{p}(\hat{\eta}) = \hat{p}_n$ as defined in (3.7) with $c_n = c_n(1)$ defined by (3.16); $\hat{p}(1) = c_n^{1-1/\hat{\eta}}\hat{p}_n$ (thus assuming $\eta = 1$); and $\hat{p} = \hat{p}(1)$ or $\hat{p}(\hat{\eta})$ depending on whether the hypothesis $\eta = 1$ is accepted or rejected by the test with standard deviation estimated by $\hat{\sigma}_1(1)$. Table 2 summarizes the main results for the failure probability estimators. The corresponding boxplots are shown in Figure 2.

For the Cauchy distribution we have asymptotic dependence, so $\hat{p}(1)$ is appropriate. As expected, $\hat{p}(\hat{\eta})$ spreads more widely than $\hat{p}(1)$ (see Figure 2).

For the normal distribution the main problem is to estimate the marginals. In particular,

Table 2. Median of $\hat{\gamma}_i$ and of estimated failure probabilities; 'Exponential/Normal' indicates a bivariate normal distribution with marginals standardized to exponential distribution (true failure probability $p_n = 10^{-5}$, sample size $n = 1000$, 250 simulations)

k	$\hat{\gamma}_1$	$\hat{\gamma}_2$	η	$\hat{p}(\hat{\eta})$	$\hat{p}(1)$	$\hat{p}$
Cauchy	1	1	1		$\times 10^{-5}$	
40	0.89	0.95	0.91	0.1310	0.2996	0.2914
80	0.98	1.02	0.99	0.3738	0.5056	0.4810
160	0.91	0.96	1.03	1.0815	0.7973	0.7973
240	1.00	1.04	1.04	1.6677	1.1440	1.1285
Morgenstern	1	1	0.5		$\times 10^{-5}$	
40	0.96	0.99	0.44	0.3277	11.7065	0.3277
80	1.01	1.03	0.46	0.3754	27.0215	0.3754
160	0.94	0.97	0.48	0.6287	57.5168	0.6287
240	1.00	1.02	0.50	0.7640	83.7799	0.7640
Normal	0	0	0.8		$\times 10^{-5}$	
40	−0.13	−0.15	0.68	0.0000	0.0600	0.0000
80	−0.17	−0.20	0.71	0.0000	0.0024	0.0000
160	−0.13	−0.13	0.73	0.0000	0.0000	0.0000
240	−0.17	−0.20	0.75	0.0000	0.0000	0.0000
Exponential/Normal	0	0	0.8		$\times 10^{-5}$	
40	−0.00	0.02	0.68	0.0054	0.8557	0.0212
80	0.04	0.06	0.71	0.0723	2.2344	0.0975
160	0.01	0.04	0.73	0.2384	3.7091	0.2467
240	0.04	0.06	0.75	0.3623	6.1357	0.3623

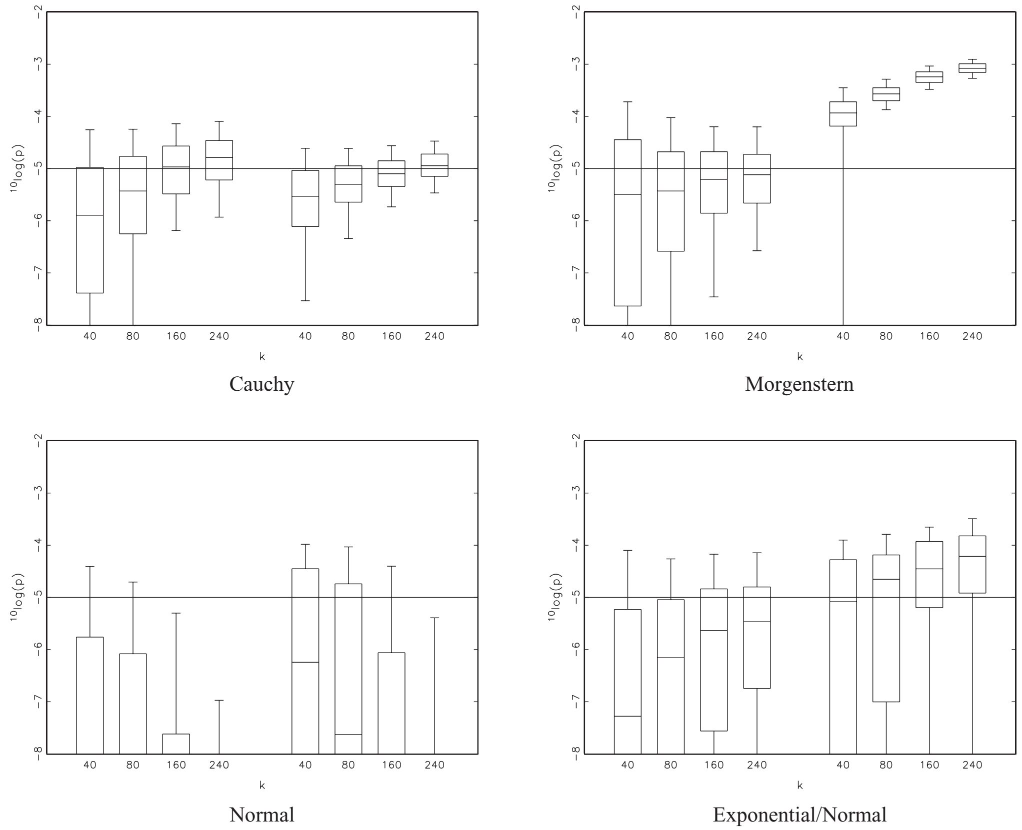

Figure 2. Each panel shows boxplots indicating the 5th, 25th, 50th, 75th and 95th percentiles of $\hat{p}(\hat{\eta})$ (left) and $\hat{p}(1)$ (right) for different values of *k*; the horizontal line shows the true failure probability $p_n = 10^{-5}$ (sample size $n = 1000$, 250 simulations)

the estimates for γ_1 and γ_2 are often negative. This implies a finite right endpoint of the marginal distributions and in quite many simulations the failure area lies outside the support of the distribution, leading to an estimated failure probability equal to 0.

When the marginals are first transformed to the exponential distribution, the estimators of the marginal parameters are much more accurate with $\hat{\gamma}_i$, $i = 1, 2$, close to 0, and the estimators for p_n perform much better. Nevertheless, in several simulations $\hat{p}_n = 0$ when one or both estimates of γ_i are negative. Here the estimator $\hat{p}(1)$, which assumes $\eta = 1$, overestimates the probability, while $\hat{p}(\hat{\eta})$ underestimates it.

The Morgenstern distribution has asymptotically independent marginals. The estimator $\hat{p}(\hat{\eta})$ is slightly biased downward, whereas $\hat{p}(1)$ has a strong positive bias. Estimating the marginals does not cause problems here as the Morgenstern distribution has extreme value (Fréchet) marginals.

5. An application: dependence of sea state parameters

In the course of the Neptune project, we studied the joint distribution of three sea state variables and its consequences for the sea wall at Petten. The data set, supplied by the Dutch National Institute for Marine and Coastal management, consists of date, time and sea characteristics recorded from 1979 to 1991, at three-hourly intervals at the Eierland station, 20 km off the Dutch coast. To obtain (nearly) independent observations of wave height HmO, wave period Tpb and still water level SWL, the maximum values of each of these state variables in distinct storm events are considered (see de Valk, 1994, for details). De Haan and de Ronde (1998) estimated the failure probability of the 'Pettemer zeewering' assuming asymptotic dependence between the variables. Figure 3 shows a scatterplot of HmO and SWL and illustrates the estimation of the corresponding coefficient of tail dependence. While the test based on Peng's estimator and the estimator $\hat{\sigma}_3(1)$ of the standard deviation accepts the hypothesis of asymptotic dependence at the 5% level, the

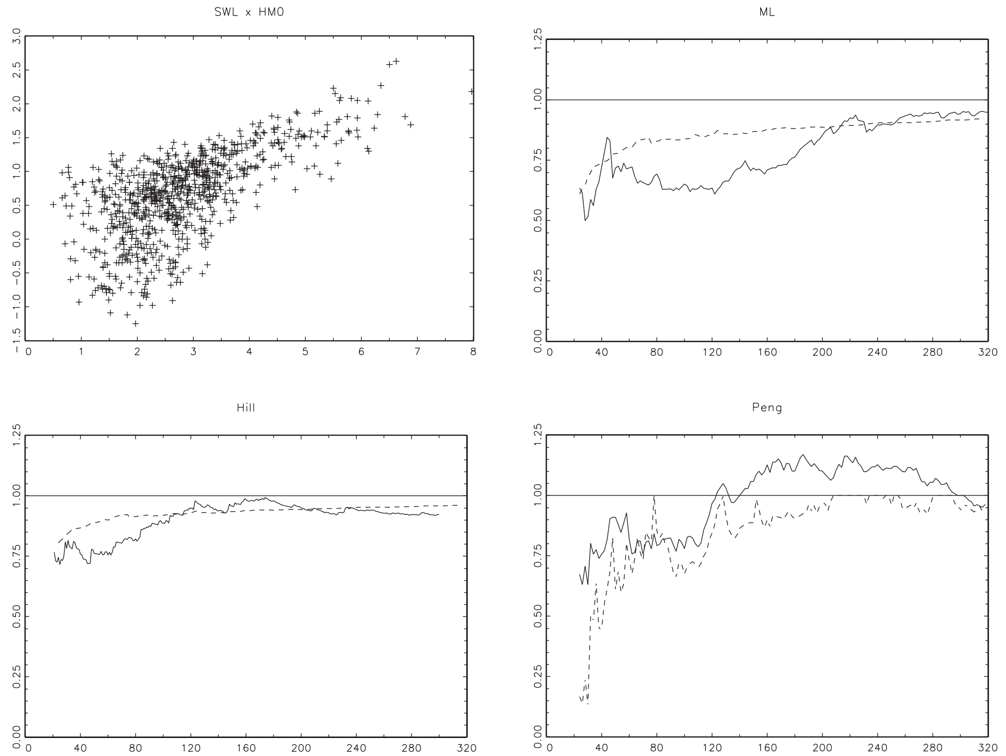

Figure 3. From top left to bottom right: scatterplot of still water level SWL versus wave height HmO, the maximum likelihood estimator $\hat{\eta}_1$, Hill's estimator $\hat{\eta}_2$ and Peng's estimator $\hat{\eta}_3$; $\hat{\eta}_i$ versus m (solid line) and upper boundary of critical region of 5% test for $\eta = 1$ (dotted line). Horizontal line: $\eta = 1$.

maximum likelihood estimator suggests this hypothesis should be rejected, because for m between 60 and 160, where the curve of estimates for η is most stable, the estimates lie in the critical region. Also the test based on the Hill estimator rejects the hypothesis for small values of m. In view of the results from the simulation study reported in Section 4, it seems plausible to assume asymptotic independence between the wave heights and the still water level.

6. Proofs for Sections 2 and 3

Define uniformly distributed random variables $U_i := 1 - F_1(X_i)$ and $V_i := 1 - F_2(Y_i)$ and denote the pertaining order statistics by $U_{n,i}$ and $V_{n,i}$, with the convention $U_{n,0} = V_{n,0} = 0$.

We will use the following notation:

$$S_1(x, y) := \sum_{i=1}^{n} \mathbf{1}\{U_i \leqslant x \text{ and } V_i \leqslant y\}, \qquad S_2(x, y) := \sum_{i=1}^{n} \mathbf{1}\{U_i \leqslant x \text{ or } V_i \leqslant y\}. \tag{6.1}$$

Let $W_1(x, y)$ and $W_2(x, y)$ be Gaussian processes with mean zero and covariance structure given by

$$\mathrm{E}\{W_1(x_1, y_1)W_1(x_2, y_2)\} = c(x_1 \wedge x_2, y_1 \wedge y_2),$$

$$\mathrm{E}\{W_2(x_1, y_1)W_2(x_2, y_2)\} = x_1 \wedge x_2 + y_1 \wedge y_2 - lc(x_1, y_1) - lc(x_2, y_2) + lc(x_1 \vee x_2, y_1 \vee y_2),$$

respectively. Moreover, let $k = \lceil nq^{-1}(m/n) \rceil$, so that $m/k \to l$.

Lemma 6.1. *Under the conditions of Theorem* 2.1,

$$\sqrt{m}\left(\frac{S_1(U_{n,\lfloor kx \rfloor}, V_{n,\lfloor ky \rfloor})}{m} - c(x, y)\right) \xrightarrow{D} W(x, y).$$

Here, and below, $\xrightarrow{D}$ *denotes convergence in distribution in* $D([0, \infty)^2)$*, and* $W(x, y)$ *is a Gaussian process with mean zero and covariance structure depending on* l*: if* $l = 0$,

$$W(x, y) = W_1(x, y);$$

if $l > 0$,

$$\begin{aligned} W(x, y) = {} & \frac{1}{\sqrt{l}}(W_2(x, 0) + W_2(0, y) - W_2(x, y)) \\ & - \sqrt{l}c_x(x, y)W_2(x, 0) - \sqrt{l}c_y(x, y)W_2(0, y), \end{aligned}$$

where the term in the first line of the right-hand side has the same distribution as $W_1(x, y)$.

Proof. From Peng (1999), Huang (1992) and Einmahl (1997, Theorem 3.1), it follows that

$$\sqrt{m}\left(\frac{S_1(kn^{-1}x, kn^{-1}y)}{m} - c(x, y)\right) \xrightarrow{D} W_1(x, y). \tag{6.2}$$

Similarly, one obtains

$$\sqrt{k}\left(\frac{S_2(kn^{-1}x,\ kn^{-1}y)}{k} - (x + y - lc(x,\ y))\right) \xrightarrow{D} W_2(x,\ y). \tag{6.3}$$

This implies

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n} \mathbf{1}\left\{U_i \leqslant \frac{k}{n}x\right\} - x\right) \xrightarrow{D} W_2(x,\ 0).$$

Note that the generalized inverse of $x \mapsto k^{-1}\sum_{i=1}^{n}\mathbf{1}\{U_i \leqslant k/nx\}$ equals $x \mapsto (n/k)U_{n,\lfloor kx\rfloor}$. Vervaat's (1972) lemma yields

$$\sqrt{k}\left(\frac{n}{k}U_{n,\lfloor kx\rfloor} - x\right) \xrightarrow{D} -W_2(x,\ 0)$$

$$\sqrt{k}\left(\frac{n}{k}V_{n,\lfloor ky\rfloor} - y\right) \xrightarrow{D} -W_2(0,\ y). \tag{6.4}$$

For $l = 0$, we have $m = o(k)$ and hence

$$\sqrt{m}\left(\frac{n}{k}U_{n,\lfloor kx\rfloor} - x\right) \xrightarrow{p} 0$$

$$\sqrt{m}\left(\frac{n}{k}V_{n,\lfloor ky\rfloor} - y\right) \xrightarrow{p} 0.$$

Therefore, the assertion follows from (6.2) and the differentiability of c.

In the case $m/k \to l$ with $l > 0$, one may derive the result in a similar fashion using

$$S_1(U_{n,\lfloor kx\rfloor},\ V_{n,\lfloor ky\rfloor}) = \lfloor kx\rfloor + \lfloor ky\rfloor - S_2(U_{n,\lfloor kx\rfloor},\ V_{n,\lfloor ky\rfloor})$$

(cf. Peng 1999). □

Denote by Q_n the tail empirical quantile function pertaining to $T_i^{(n)}$, $1 \leqslant i \leqslant n$, that is,

$$Q_n(t) := T^{(n)}_{n,n-\lfloor mt\rfloor}, \qquad 0 < t < n/m.$$

The following lemma is central to the proof of the asymptotic normality of estimators for η based on largest order statistics of $T_i^{(n)}$.

Lemma 6.2. *Under the conditions of Theorem* 2.2 *there exist suitable versions of* Q_n, *a suitable process* $\overline{W}$ *equal in distribution to a standard Brownian motion if* $l = 0$ *and to* $x \mapsto W(x,\ x)$ *if* $l > 0$ *such that, for all* $t_0,\ \varepsilon > 0$,

$$\sup_{0<t\leqslant t_0} t^{\eta+1/2+\varepsilon}\left| m^{1/2}\left(\frac{k}{n}Q_n(t) - t^{-\eta}\right) - \eta t^{-(\eta+1)}\overline{W}(t)\right| = o_P(1).$$

Proof. First, check that

$$\sum_{i=1}^{n} \mathbf{1}\{T_i^{(n)} > x\} = \sum_{i=1}^{n} \mathbf{1}\{R_i^X > (n+1)(1-1/x) \text{ and } R_i^Y > (n+1)(1-1/x)\}$$

$$= \sum_{i=1}^{n} \mathbf{1}\{U_i < U_{n,\lceil (n+1)/x \rceil} \text{ and } V_i < V_{n,\lceil (n+1)/x \rceil}\} \text{ a.s.}$$

with the convention $U_{n,n+1} = V_{n,n+1} = 1$. Hence

$$\overline{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\frac{k}{n+1} T_i^{(n)} > x\right\} = \frac{1}{n} S_1\left(U_{n,\lceil k/x \rceil}-,\ V_{n,\lceil k/x \rceil}-\right),$$

where $f(x-)$ denotes the left-hand limit of f at x. From Lemma 6.1 one readily obtains that

$$m^{1/2}\left(\frac{\overline{F}_n(x)}{q(k/n)} - x^{-1/\eta}\right)_{0<x<\partial} \to (W(1/x,\ 1/x))_{0<x<\infty}$$

$$\Rightarrow m^{1/2}\left(\frac{\overline{F}_n(x^{-\eta})}{q(k/n)} - x\right)_{0<x<\infty} \to (W(x^{\eta},\ x^{\eta}))_{0<x<\infty} =: \overline{W}$$

$$\Rightarrow m^{1/2}((\overline{F}_n^{-1}(q(k/n)t))^{-1/\eta} - t)_{0<t<\infty} \to -\overline{W}$$

weakly in $D(0, \infty)$, where in the last step Vervaat's (1972) lemma is used. For this, note that $\overline{W}$ has almost surely continuous sample paths, because by the definition of W it is a Brownian motion for $l = 0$ and can be represented as a sum of Brownian motions if $l > 0$. Thus the δ-method yields, for suitable versions,

$$\overline{F}_n^{-1}(q(k/n)t) = t^{-\eta}(1 + m^{-1/2}\eta t^{-1}\overline{W}(t) + o(m^{-1/2})) \text{ a.s.}$$

uniformly on compact intervals bounded away from 0.

Next, note that $\overline{F}_n^{-1}(q(k/n)t) = k/nQ_n(t) = O(1/m)$ uniformly and $\sup_{0<t<\vartheta} t^{-1/2+\varepsilon}$ $|\overline{W}(t)| = o_P(1)$ as $\vartheta \downarrow 0$ by the law of the iterated logarithm and the aforementioned representation of $\overline{W}$. Thus it suffices to prove that, for all $\delta > 0$,

$$\lim_{\vartheta \downarrow 0} \limsup_{n\to\infty} P\left\{\sup_{0<t\leqslant\vartheta} m^{1/2} t^{\eta+1/2+\varepsilon}\left|\frac{k}{n+1} Q_n(t) - t^{-\eta}\right| > \delta\right\} = 0. \tag{6.5}$$

Here we will only consider

$$P\left\{\sup_{0<t\leqslant\vartheta} m^{1/2} t^{\eta+1/2+\varepsilon}\left(\frac{k}{n+1} Q_n(t) - t^{-\eta}\right) > \delta\right\}$$

$$\leqslant P\left\{\exists 1 \leqslant i \leqslant m\vartheta + 1 : \frac{k}{n+1} T_{n,n-i+1}^{(n)} > x_{i,n} \text{ and } x_{i,n} < k\right\} \tag{6.6}$$

with

$$x_{i,n} := \left(\frac{i}{m}\right)^{-\eta} + \delta m^{-1/2}\left(\frac{i}{m}\right)^{-(\eta+1/2+\varepsilon)}.$$

The other inequality can be treated in a similar way.

Let $T_i := (1/U_i) \wedge (1/V_i)$. Then the right-hand side of (6.6) can be bounded by

$$P\left\{\exists 1 \leqslant i \leqslant m\vartheta + 1 : \sum_{j=1}^{n} \mathbf{1}\{T_j > (1/U_{n,\lceil k/x_{i,n}\rceil}) \wedge (1/V_{n,\lceil k/x_{i,n}\rceil})\} \geqslant i \text{ and } x_{i,n} < k\right\}.$$

We now distinguish two different ranges of i-values.

According to Shorack and Wellner (1986, Theorem 10.3.1), for all $\bar{\varepsilon} > 0$ there exists $\bar{\delta} > 0$ such that eventually, with probability greater than $1 - \bar{\varepsilon}$,

$$(1/U_{n,\lceil k/x_{i,n}\rceil}) \wedge (1/V_{n,\lceil k/x_{i,n}\rceil}) \geqslant \frac{n}{k} x_{i,n}\bar{\delta} \geqslant \bar{\delta} L k^{-1} m^{\eta} n^{1-\eta}(n/i)^{\eta}, \tag{6.7}$$

for all $i \leqslant i_n := \lfloor (\delta m^{\varepsilon}/L)^{1/(1/2+\varepsilon)} \rfloor$, with $x_{i,n} < k$.

Since q^{-1} is η-varying at 0 and the quantile function F_T^{-1} of T_i is $(-\eta)$-varying at 1, in the case $\eta < 1$, we have $k/n = o((m/n)^{\eta+\iota})$ and $F_T^{-1}(1-t) = o(t^{-(\eta+\iota)})$ as $t \downarrow 0$ for all $\iota > 0$. Hence the right-hand side of (6.7) is of larger order than $F_T^{-1}(1 - 2i/\bar{\delta}Ln))$.

If $\eta = 1$, in view of (2.1) and Lemma 2.1 of Drees (1998a), we have

$$\sup_{x \leqslant 1} x^{\iota - 1}\left|\frac{q(tx)}{q(t)} - x\right| = o(q_1(t)).$$

Apply this bound with $t = k/n$ and $x = i/(\bar{\delta}Lm)$ to obtain $1 - F_T(x_{i,n}\bar{\delta}n/k) \leqslant 2i/(\bar{\delta}Ln)$, since $x_{i,n} \geqslant Lm/i$ and $(i/m)^{1-\iota}q_1(k/m) = o(m^{1/2}q_1(k/n)i/m) = o(i/m)$ uniformly for $1 \leqslant i \leqslant i_n$.

Hence, for all η, it follows that

$$\begin{aligned}
&\limsup_{n\to\infty} P\left\{\exists 1 \leqslant i \leqslant i_n : \sum_{j=1}^{n} \mathbf{1}\{T_j > (1/U_{n,\lceil k/x_{i,n}\rceil}) \wedge (1/V_{n,\lceil k/x_{i,n}\rceil})\} \geqslant i \text{ and } x_{i,n} < k\right\} \\
&\quad \leqslant \limsup_{n\to\infty} P\left\{\exists 1 \leqslant i \leqslant i_n : T_{n,n-i+1} > \frac{n}{k} x_{i,n}\bar{\delta}\right\} + \bar{\varepsilon} \\
&\quad \leqslant \limsup_{n\to\infty} P\left\{\max_{1\leqslant i \leqslant m+1} \frac{T_{n,n-i+1}}{F_T^{-1}(1 - 2i/(\bar{\delta}Ln))} > 1\right\} + \bar{\varepsilon} \\
&\quad < 2\bar{\varepsilon}
\end{aligned} \tag{6.8}$$

for sufficiently large L, where for the last step we again use Theorem 10.3.1 of Shorack and Wellner (1986).

Let

$$y_{i,n} := \frac{n}{k} x_{i,n} - \tilde{\delta} n k^{-3/2} x_{i,n}^{3/2+\iota}$$

for some $\iota \in (0, \varepsilon)$ and $\tilde{\delta} > 0$. Using

$$\lim_{\vartheta \downarrow 0} \limsup_{n \to \infty} P\left\{ \sup_{0 < t \leqslant \vartheta} k^{1/2} t^{3/2+\iota} \left| \frac{k}{nU_{n,\lceil kt \rceil}} - t^{-1} \right| > \tilde{\delta} \right\} = 0$$

(Drees 1998a, Theorem 2.1) instead of (6.7), one can conclude by similar arguments to those above that

$$\begin{aligned} &\lim_{\vartheta \downarrow 0} \limsup_{n \to \infty} P\left\{ \exists i_n < i \leqslant m\vartheta + 1 : \sum_{j=1}^{n} \mathbf{1}\{T_j > (1/U_{n,\lceil k/x_{i,n} \rceil}) \wedge (1/V_{n,\lceil k/x_{i,n} \rceil})\} \geqslant i \right\} \\ &\quad \leqslant \lim_{\vartheta \downarrow 0} \limsup_{n \to \infty} P\left\{ \exists i_n < i \leqslant m\vartheta + 1 : m^{1/2} \left(\frac{i}{m} \right)^{\eta + 1/2 + \varepsilon} \left(\frac{k}{n} T_{n,n-i+1} - \left(\frac{i}{m} \right)^{-\eta} \right) > \delta/2 \right\} \\ &\quad = 0. \end{aligned} \tag{6.9}$$

In the last step, we again use Theorem 2.1 of Drees (1998a), where (2.1) implies Condition 1 of that paper and $m^{1/2} q_1(k/n) \to 0$ ensures that the bias is aymptotically negligible.

Combining (6.8) and (6.9), one arrives at (6.5). □

Proof of Theorem 2.1 *(asymptotic normality of $\hat{\eta}_1$ and $\hat{\eta}_2$)*. Note that this approximation is analogous to the approximation of the tail empirical quantile function established in Drees (1998a) in the classical situation of i.i.d. random variables. Hence, the asymptotic normality of $\hat{\eta}_1$ and $\hat{\eta}_2$ follows from Lemma 6.2 exactly as in Drees (1998a, Example 4.1) and Drees (1998b, Example 3.1) using the δ-method. The asymptotic variance is given by

$$\int_0^1 \int_0^1 \mathrm{cov}(\overline{W}(s), \overline{W}(t))(st)^{-(\eta+1)} \nu_\eta(\mathrm{d}s) \nu_\eta(\mathrm{d}t)$$

with $\nu_\eta(\mathrm{d}t) := (\eta + 1)^2 (t^\eta - (2\eta + 1)t^{2\eta})/\eta \, \mathrm{d}t + (\eta + 1)\varepsilon_1(\mathrm{d}t)$ for the maximum likelihood estimator $\hat{\eta}_1$ and $\nu_\eta(\mathrm{d}t) := \eta(t^\eta \, \mathrm{d}t - \varepsilon_1(\mathrm{d}t))$ in case of the Hill estimator. (Here ε_1 denotes the Dirac measure at 1.) Now using the homogeneity of order 1 of the covariance function which implies $\int_0^t \mathrm{cov}(\overline{W}(s), \overline{W}(t))(st)^{-1} \, \mathrm{d}s = \int_0^1 \mathrm{cov}(\overline{W}(u), \overline{W}(1)) u^{-1} \, \mathrm{d}u$, one obtains $(\eta + 1)^2 \, \mathrm{var}(\overline{W}(1))$ and $\eta^2 \, \mathrm{var}(\overline{W}(1))$, respectively, as asymptotic variance and thus the assertion, using $c_x(1, 1) + c_y(1, 1) = 1/\eta$.

Proof of Theorem 2.2. By Lemma 6.2,

$$\hat{l} = \frac{m}{k} \cdot \frac{k}{n} Q_n(1) \xrightarrow{P} l$$

and $k/\hat{k} = k/nQ_n(1) \to 1$ in probability.

In the same way as in Lemma 6.2, one can prove that

$$\frac{k}{n} T^{(n,u)}_{n,n-\lfloor mt \rfloor} = \left(\frac{t}{c(1 + u, 1)} \right)^{-\eta} + O_P(m^{-1/2}).$$

Hence, if $\eta = 1$, then

$$\hat{c}_x(1, 1) = \frac{\hat{k}}{k}(\hat{k}^{1/4}(c(1 + k^{-1/4}, 1) - c(1, 1)) + O_P(k^{1/4}m^{-1/2})) \xrightarrow{P} c_x(1, 1).$$

The consistency of $\hat{c}_y(1, 1)$ can be proved in a similar way, so that the consistency of $\hat{\sigma}_i^2$ follows readily in that case.

Likewise, if $\eta < 1$, we have

$$\hat{c}_x(1, 1) = (\eta c_x(1, 1) + O_P(k^{1/4}m^{-1/2}))(1 + o_P(1)),$$

and thus

$$\hat{l}^{1/2}\hat{c}_x(1, 1) = o_P(1) + O_P(m^{1/2}k^{-3/4}) = o_P(1).$$

Together with the analogous result for $\hat{c}_y(1, 1)$ and the consistency of $\hat{l}$ and $\hat{\eta}_i$, this implies $\hat{\sigma}_i^2 \to \sigma_i^2$ in probability. □

The proof of Theorem 3.1 will be given in several steps. The following sequence of equalities and asymptotic (in probability) equivalences provides an overview of the line of reasoning:

$$\begin{aligned}
p_n &= P\{\mathbf{1} - \mathbf{F}(X, Y) \in \mathbf{1} - \mathbf{F}(C_n)\} \\
&\overset{(3.14)}{\sim} q\left(\frac{k}{n}\right)\nu\left(\frac{n}{k}(\mathbf{1} - \mathbf{F}(C_n))\right) \\
&\overset{\text{Lemma 6.4}}{\sim} q\left(\frac{k}{n}\right)\nu(D_n) \\
&\overset{(2.2)}{=} c_n^{1/\eta} q\left(\frac{k}{n}\right)\nu\left(\frac{D_n}{c_n}\right) \\
&\overset{\text{Cor. 6.3}}{\sim} c_n^{1/\eta} q\left(\frac{k}{n}\right)\nu\left(\mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}\left(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\left(\mathbf{1} - \frac{\hat{D}_n}{c_n}\right)\right)\right) \\
&\overset{\text{Lemma 6.5}}{\sim} c_n^{1/\eta} q\left(\frac{k}{n}\right)\nu\left(\frac{n}{k}\left(\mathbf{1} - \mathbf{F}\left(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\left(\mathbf{1} - \frac{\hat{D}_n}{c_n}\right)\right)\right)\right) \\
&\overset{(3.14)}{\sim} c_n^{1/\eta} P\{\mathbf{1} - \mathbf{F}(X, Y) \in \mathbf{1} - \mathbf{F}(B)\}_{|B = \mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}(\mathbf{1} - \hat{D}_n/c_n)} \\
&\overset{\text{Lemma 6.6}}{\sim} c_n^{1/\eta} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{(X_i, Y_i) \in \mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\left(\mathbf{1} - \frac{\hat{D}_n}{c_n}\right)\right\} \\
&\sim \hat{p}_n.
\end{aligned} \tag{6.10}$$

Lemma 6.3. *Let $a = a(n)$, $\tilde{a} > 0$, b, $\tilde{b}$, γ, $\tilde{\gamma} \in \mathbb{R}$ denotes sequences such that*

$$\left|\frac{\tilde{a}}{a} - 1\right| \vee \left|\frac{\tilde{b} - b}{a}\right| \vee |\tilde{\gamma} - \gamma| = O(\varepsilon_n)$$

for some $\varepsilon_n \downarrow 0$. Suppose that the sequence $\lambda_n > 0$ is bounded and satisfies $\varepsilon_n \log \lambda_n \to 0$ and $\varepsilon_n w_\gamma(\lambda_n) \to 0$, with w_γ defined in (3.12). Then

$$1 - F_{\tilde{a},\tilde{b},\tilde{\gamma}}(F^{-1}_{a,b,\gamma}(1 - x)) = x + o(\lambda_n) \tag{6.11}$$

uniformly for $0 \leqslant x \leqslant \lambda_n$.

Proof. First, note that

$$T(x) := 1 - F_{\tilde{a},\tilde{b},\tilde{\gamma}}(F^{-1}_{a,b,\gamma}(1 - x)) = \left[1 + \tilde{\gamma}\frac{a}{\tilde{a}}\left(\frac{x^{-\gamma} - 1}{\gamma} + \frac{b - \tilde{b}}{a}\right)\right]^{-1/\tilde{\gamma}},$$

where, as usual, $(x^{-\gamma} - 1)/\gamma := -\log x$ if $\gamma = 0$. We now distinguish three cases.

$\gamma > 0$. Then

$$\begin{aligned} T(x) &= (1 + (1 + O(\varepsilon_n))(x^{-\gamma} - 1 + O(\varepsilon_n)))^{-(1+O(\varepsilon_n))/\gamma} \\ &= (x^{-\gamma}(1 + O(\varepsilon_n)) + O(\varepsilon_n))^{-(1+O(\varepsilon_n))/\gamma} \\ &= x \exp(O(\varepsilon_n)\log x)(1 + o(1)) \end{aligned}$$

uniformly for $0 \leqslant x \leqslant \lambda_n$. For $\lambda_n\varepsilon_n \leqslant x \leqslant \lambda_n$,

$$|\log x|\varepsilon_n \leqslant (|\log \lambda_n| + |\log \varepsilon_n|)\varepsilon_n \to 0,$$

and thus $T(x) = x(1 + o(1)) = x + o(\lambda_n)$ uniformly. Otherwise, that is, for $0 \leqslant x < \lambda_n\varepsilon_n$,

$$T(x) \leqslant T(\lambda_n\varepsilon_n) = \lambda_n\varepsilon_n(1 + o(1)) = o(\lambda_n) = x + o(\lambda_n)$$

by the monotonicity of T.

$\gamma < 0$. Choose $\delta_n \to 0$ such that $\varepsilon_n(\lambda_n\delta_n)^\gamma \to 0$ and hence also $\varepsilon_n \log \delta_n \to 0$ (e.g., $\delta_n = (\varepsilon_n\lambda_n^\gamma)^{-1/(2\gamma)}$). Then, uniformly for $\lambda_n\delta_n \leqslant x \leqslant \lambda_n$,

$$T(x) = x^{1+O(\varepsilon_n)}(1 + O(\varepsilon_n) + O(\varepsilon_n(\lambda_n\delta_n)^\gamma))^{-(1+O(\varepsilon_n))/\gamma} = x(1 + o(1)),$$

and again (6.11) follows from the monotonicity of T.

$\gamma = 0$. Note that $\tilde{\gamma}|\log x| \to 0$ uniformly for $\lambda_n\varepsilon_n \leqslant x \leqslant \lambda_n$. Hence a Taylor expansion of log yields

$$T(x) = \exp\left(-\frac{1}{\tilde{\gamma}}\log(1+\tilde{\gamma}(1+O(\varepsilon_n))(-\log x + O(\varepsilon_n)))\right)$$
$$= \exp\left(-\frac{1}{\tilde{\gamma}}[\tilde{\gamma}(1+O(\varepsilon_n))(-\log x + O(\varepsilon_n)) + O(\tilde{\gamma}^2(\log x + O(\varepsilon_n))^2]\right)$$
$$= x\exp(O(\varepsilon_n)\log x + O(\varepsilon_n) + O(\varepsilon_n\log^2 x))$$
$$= x(1+o(1)),$$

and thus the assertion follows by the aforementioned arguments. □

Remark 6.1. For fixed sequences a, b and γ, assertion (6.11) even holds uniformly for

$$(\tilde{a}, \tilde{b}, \tilde{\gamma}) \in M(\varepsilon_n) := \left\{(\bar{a}, \bar{b}, \bar{\gamma}) \in (0, \infty)\times\mathbb{R}^2 \,\middle|\, \left|\frac{\bar{a}}{a} - 1\right| \vee \left|\frac{\bar{b}-b}{a}\right| \vee |\bar{\gamma}-\gamma| \leqslant \varepsilon_n\right\}. \quad (6.12)$$

Corollary 6.1. *If Condition* D, (3.8) *and* (3.11)–(3.12) *are satisfied then, for all* $\delta > 0$,

$$P\left\{A_{-\delta} \subset \frac{\hat{D}_n}{d_n} \subset A_{+\delta}\right\} \to 1.$$

Proof. Since the set A is bounded, there exists $L > 0$ such that $D_n \subset [0, d_n L]^2$ for all sufficiently large n. Because of (3.12), one can find a sequence $\varepsilon_n \to 0$ such that $k^{-1/2} = o(\varepsilon_n)$ and the conditions of Lemma 6.3 hold for $\lambda_n = d_n L$. Then $P\{(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\boldsymbol{\gamma}}) \in (M(\varepsilon_n))^2\} \to 1$ with $M(\varepsilon_n)$ defined in (6.12), and Lemma 6.3 yields

$$\sup_{(x,y)\in D_n} \|\mathbf{1} - \mathbf{F}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{F}^{-1}_{\mathbf{a},\mathbf{b},\boldsymbol{\gamma}}(\mathbf{1} - (x, y))) - (x, y)\| \leqslant \frac{\delta}{2} d_n \quad (6.13)$$

with probability tending to 1. Thus, in view of $\hat{D}_n = \mathbf{1} - \mathbf{F}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{F}^{-1}_{\mathbf{a},\mathbf{b},\boldsymbol{\gamma}}(\mathbf{1} - D_n))$ and Condition D,

$$P\left\{\frac{\hat{D}_n}{d_n} \subset \left(\frac{D_n}{d_n}\right)_{+\delta/2} \subset A_{+\delta}\right\} \to 1.$$

On the other hand, by the definition of the inner neighbourhood of a set, $(x, y) \in (D_n/d_n)_{-\delta/2}$ implies $(x + \delta/2, y + \delta/2) \in D_n/d_n$. Since, in view of (6.13),

$$d_n(x, y) \leqslant \mathbf{1} - \mathbf{F}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\boldsymbol{\gamma}}}\left(\mathbf{F}^{-1}_{\mathbf{a},\mathbf{b},\boldsymbol{\gamma}}\left(\mathbf{1} - d_n\left(x + \frac{\delta}{2}, y + \frac{\delta}{2}\right)\right)\right)$$

componentwise, (3.8) shows that $d_n(x, y) \in \hat{D}_n$. Hence, again by Condition D,

$$P\left\{A_{-\delta} \subset \left(\frac{D_n}{d_n}\right)_{-\delta/2} \subset \frac{\hat{D}_n}{d_n}\right\} \to 1.$$

□

Corollary 6.2. *If the conditions of Corollary* 6.1 *hold, and* (3.13) *also holds, then, for all* $\delta > 0$,

$$P\left\{A_{-\delta} \subset \frac{c_n}{d_n}\left(\mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}\left(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\left(\mathbf{1} - \frac{\hat{D}_n}{c_n}\right)\right)\right) \subset A_{+\delta}\right\} \to 1.$$

Proof. According to Corollary 6.1, there exists $L > 0$ such that $P\{\hat{D}_n/c_n \subset [0, \lambda_n]^2\} \to 1$ for $\lambda_n := Ld_n/c_n$. It follows from (3.11) and (3.13) that $\lambda_n^{\hat{\gamma}_i} = \lambda_n^{\gamma_i}(1 + o_P(1))$, $i = 1, 2$. Hence, one may apply Lemma 6.3 with $(a, b, \gamma) = (\hat{a}_i, \hat{b}_i, \hat{\gamma}_i)$ and $(\tilde{a}, \tilde{b}, \tilde{\gamma}) = (a_i, b_i, \gamma_i)$ to obtain

$$\sup_{(x,y)\in\hat{D}_n/c_n} \|\mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}(\mathbf{1} - (x, y))) - (x, y)\| \leqslant \frac{\delta}{2}\frac{d_n}{c_n}$$

with probability tending to 1, for all $\delta > 0$. Now one may conclude the proof following the lines of the previous proof. □

Corollary 6.3. *Under the conditions of Corollary* 6.2*,*

$$\nu\left(\mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}\left(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\left(\mathbf{1} - \frac{\hat{D}_n}{c_n}\right)\right)\right) = \nu\left(\frac{D_n}{c_n}\right)(1 + o_P(1)).$$

Proof. Denote the boundary of the set A by ∂A. Condition (3.8) implies a slightly weaker version for A, namely $(x, y) \in A \Rightarrow [0, x) \times [0, y) \subset A$. Hence, $\lambda \cdot \partial A \subset A$ for all $\lambda \in (0, 1)$, and these sets are pairwise disjoint. Since ν is homogeneous in the sense of (2.2) and $\nu(A) < \infty$ by the boundedness of A, it follows that $\nu(\partial A) = 0$. Moreover, $A_{+\delta} \backslash A_{-\delta} \downarrow \partial A$ as $\delta \downarrow 0$, so that $\nu(A_{+\delta} \backslash A_{-\delta}) \to 0$. Thus Corollary 6.2 and Condition D yield

$$\nu\left(\frac{c_n}{d_n}\left(\mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}\left(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\left(\mathbf{1} - \frac{\hat{D}_n}{c_n}\right)\right)\right)\right) \to \nu(A)$$

and $\nu(D_n/d_n) \to \nu(A)$. Now the assertion is an obvious consequence of the homogeneity (2.2). □

Lemma 6.4. *If Condition* D*,* (3.8) *and* (3.9) *hold, then*

$$\nu(D_n) = \nu\left(\frac{n}{k}(\mathbf{1} - \mathbf{F}(C_n))\right)(1 + o(1)).$$

Proof. There exists $L > 0$ such that $D_n \subset [0, d_n L]^2$ for all sufficiently large n. Choose arbitrary $-1/(\gamma_i \vee 0) < x_i < 1/((-\gamma_i) \vee 0)$, $i = 1, 2$. Then, by (3.9), for all $(x, y) \in D_n$,

$$\frac{n}{k}(\mathbf{1} - \mathbf{F}(\mathbf{F}^{-1}_{\mathbf{a},\mathbf{b},\gamma}(\mathbf{1} - (x, y)))) = (x(1 + \delta_x), y(1 + \delta_y)), \tag{6.14}$$

with $|\delta_x| \vee |\delta_y| \leqslant R_{x_1,x_2}(n/k)$, for sufficiently large n. According to (3.8), the left-hand side of (6.14) is an element of $D_n(1 + R_{x_1,x_2}(n/k))$. Thus, by the definition of D_n,

$$\frac{n}{k}(\mathbf{1}-\mathbf{F}(C_n)) \subset D_n\Big(1+R_{x_1,x_2}\Big(\frac{n}{k}\Big)\Big).$$

Likewise, (6.14) together with (3.8) implies

$$D_n\Big(1-R_{x_1,x_2}\Big(\frac{n}{k}\Big)\Big) \subset \frac{n}{k}(\mathbf{1}-\mathbf{F}(C_n))$$

eventually. Now the assertion is obvious from the homogeneity property (2.2). □

Lemma 6.5. *Under Condition* D, (3.8)*,* (3.9) *and* (3.11)–(3.13) *one has*

$$\nu\bigg(\bigg(\mathbf{1}-\mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}\bigg(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\bigg(\mathbf{1}-\frac{\hat{D}_n}{c_n}\bigg)\bigg)\bigg)\bigg) = \nu\bigg(\frac{n}{k}\bigg(\mathbf{1}-\mathbf{F}\bigg(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}\bigg(\mathbf{1}-\frac{\hat{D}_n}{c_n}\bigg)\bigg)\bigg)\bigg)(1+o_P(1)).$$

Proof. The proof is very much the same as that for Lemma 6.4, with D_n replaced by $\mathbf{1}-\mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}(\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\gamma}}(\mathbf{1}-\hat{D}_n/c_n))$. Note that, by the boundedness of d_n/c_n and the assertion of Corollary 6.2, this set is eventually bounded. Hence (3.9) is applicable for sufficiently small x_1 and x_2. □

Lemma 6.6. *If the conditions of Theorem* 3.1 *are satisfied, then*

$$\sup_{B\in\mathcal{B}_n}\left|\frac{n^{-1}\sum_{i=1}^{n}\mathbf{1}\{\mathbf{1}-\mathbf{F}(X_i,\,Y_i)\in\mathbf{1}-\mathbf{F}(B)\}}{P\{\mathbf{1}-\mathbf{F}(X,\,Y)\in\mathbf{1}-\mathbf{F}(B)\}}-1\right| \to 0 \textit{ in probability}.$$

Proof. We will apply Theorem 5.1 of Alexander (1987). To check the conditions of this uniform law of large numbers, first note that every set $B\in\mathcal{B}_n$ can be represented as

$$B=\mathbf{F}^{-1}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}},\tilde{\gamma}}\bigg(\mathbf{1}-\frac{\mathbf{1}-\mathbf{F}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}},\tilde{\gamma}}(C_n)}{c_n}\bigg) \tag{6.15}$$

with $(\tilde{\mathbf{a}},\tilde{\mathbf{b}},\tilde{\gamma})\in(M(\varepsilon_n))^2$ (cf. (6.12)). Therefore the arguments of the proofs for Lemma 6.5 and Corollary 6.3 show that

$$\begin{aligned}\nu\Big(\frac{n}{k}(\mathbf{1}-\mathbf{F}(B))\Big) &= \nu(\mathbf{1}-\mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}(B))(1+o(1)) = \nu\Big(\frac{D_n}{c_n}\Big)(1+o(1))\\ &= \Big(\frac{d_n}{c_n}\Big)^{1/\eta}\nu(A)(1+o(1))\end{aligned} \tag{6.16}$$

uniformly for $B\in\mathcal{B}_n$ (cf. Remark 6.1). Now (3.14) leads to

$$P\{\mathbf{1}-\mathbf{F}(X,\,Y)\in\mathbf{1}-\mathbf{F}(B)\} = q\Big(\frac{k}{n}\Big)\Big(\frac{d_n}{c_n}\Big)^{1/\eta}\nu(A)(1+o(1)) \tag{6.17}$$

uniformly. In particular, there exists n_0 such that $P\{\mathbf{1}-\mathbf{F}(X,\,Y)\in\mathbf{1}-\mathbf{F}(B)\}<1/2$ for all $n\geqslant n_0$ and all $B\in\mathcal{B}_n$.

Next, note that

$$\overline{B}_t := \bigcup_{\substack{B\in\mathcal{B}_n, n\geqslant n_0,\\ P\{\mathbf{1}-\mathbf{F}(X,Y)\in\mathbf{1}-\mathbf{F}(B)\}(1-P\{\mathbf{1}-\mathbf{F}(X,Y)\in\mathbf{1}-\mathbf{F}(B)\})\leqslant t}} B \subset \bigcup_{\substack{B\in\mathcal{B}_n, n\geqslant n_0,\\ P\{\mathbf{1}-\mathbf{F}(X,Y)\in\mathbf{1}-\mathbf{F}(B)\}\leqslant 2t}} B. \tag{6.18}$$

In view of (6.15), one may prove as in Corollary 6.2 that, for all $\delta > 0$, eventually $\mathbf{1} - \mathbf{F}_{\mathbf{a},\mathbf{b},\gamma}(B) \subset A_{+\delta} d_n/c_n$ for all $B \in \mathcal{B}_n$. Hence, it follows as in the proof of Lemma 6.4 that

$$\frac{n}{k}(\mathbf{1} - \mathbf{F}(B)) \subset \frac{d_n}{c_n} A_{+\delta}(1 + o(1)) \tag{6.19}$$

uniformly for $B \in \mathcal{B}_n$.

Let $n(t) := \min\{n \geqslant n_0 | q(k/n)(d_n/c_n)^{1/\eta}\nu(A) \leqslant 3t\}$, which tends to ∞ as t tends to 0. Combining (6.17)–(6.19), we arrive at

$$\mathbf{1} - \mathbf{F}(\overline{B}_t) \subset \bigcup_{n\geqslant n(t)} \frac{k(n)d_n}{nc_n} A_{+\delta}(1 + o(1)) \subset 2 \sup_{n\geqslant n(t)} \frac{k(n)d_n}{nc_n} A_{+\delta}$$

for sufficiently small t. By (3.14), the regularity condition on $k(n)$ and the definition of $n(t)$, it follows that

$$\begin{aligned} P\{\mathbf{1} - \mathbf{F}(X,\ Y) \in \mathbf{1} - \mathbf{F}(\overline{B}_t)\} &= O\left(q\left(\frac{k(n(t))}{n(t)}\right)\left(\frac{n(t)}{k(n(t))} \sup_{n\geqslant n(t)} \frac{k(n)d_n}{nc_n}\right)^{1/\eta}\right) \\ &= O\left(q\left(\frac{k(n(t))}{n(t)}\right)\left(\frac{d_n}{c_n}\right)^{1/\eta}\right) \\ &= O(t). \end{aligned}$$

Since $\mathcal{B}_n$ is a VC class, Theorem 5.1 of Alexander (1987) yields

$$\sup\left\{ \left| \frac{n^{-1}\sum_{i=1}^{n} \mathbf{1}\{\mathbf{1} - \mathbf{F}(X_i,\ Y_i) \in \mathbf{1} - \mathbf{F}(B)\}}{P\{\mathbf{1} - \mathbf{F}(X,\ Y) \in \mathbf{1} - \mathbf{F}(B)\}} - 1 \right| \middle| B \in \mathcal{B}_n,\ P\{\mathbf{1} - \mathbf{F}(X,\ Y) \in \mathbf{1} - \mathbf{F}(B)\} \geqslant \varepsilon_n \right\} \to 0,$$

provided $n\varepsilon_n \to \infty$. Because of (6.17) and the last assumption of (3.13), the choice $\varepsilon_n = q(k/n)(d_n/c_n)^{1/\eta}\nu(A)/2$ leads to the assertion. □

Proof of Theorem 3.1. Now the consistency of $\hat{p}_n$ can be proven as shown in (6.10). To this end, note that, because of (3.11), $\mathbf{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1} - \hat{D}_n/c_n)$ belongs to $\mathcal{B}_n$ with probability tending to 1 and that $\log c_n = o((r(n))^{1/2})$ implies $c_n^{1/\hat{\eta}} = c_n^{1/\eta}(1 + o_P(1))$ since $\hat{\eta}$ was assumed $\sqrt{r(n)}$-consistent for η. □

Acknowledgements

The Neptune project was financed by the European Union (grant MAS2-CT94-0081). Holger Drees was partly supported by a Heisenberg fellowship, granted by the Deutsche Forschungsgemeinschaft. The research of Ana Ferreira and Laurens de Haan was partly supported by FCT/POCTI/FEDER. We are grateful for the comments and suggestions made by an associate editor and a referee that helped to improve the presentation of our results.

References

Alexander, K.S. (1987) Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probab. Theory Related Fields*, **75**, 379–423.

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*. Cambridge: Cambridge University Press.

Coles, S., Heffernan, J. and Tawn, J. (1999) Dependence measures for extreme value analysis. *Extremes*, **2**, 339–366.

de Haan, L. and Resnick, S.I. (1993) Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models*, **9**, 275–309.

de Haan, L. and de Ronde, J. (1998) Sea and wind: multivariate extremes at work. *Extremes*, **1**, 7–45.

de Haan, L. and Sinha, A.K. (1999) Estimating the probability of a rare event. *Ann. Statist.*, **27**, 732–759.

de Valk, C.F. (1994) Selection of storm and estimation of exceedance frequencies of significant wave height for five North Sea locations. Technical report H1931, Delft Hydraulics.

Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989) A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, **17**, 1833–1855.

Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2001) Dependence in independence. Technical report, EURANDOM, Eindhoven.

Drees, H. (1998a) On smooth statistical tail functionals. *Scand. J. Statist.*, **25**, 187–210.

Drees, H. (1998b) A general class of estimators of the extreme value index. *J. Statist. Plann. Inference*, **66**, 95–112.

Drees, H., Ferreira, A. and de Haan, L. (2004) On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* To appear.

Einmahl, J.H.J. (1997) Poisson and Gaussian approximation of weighed local empirical processes. *Stochastic Process. Appl.*, **70**, 31–58.

Heffernan, J.E. (2000) A directory of coefficients of tail dependence. *Extremes*, **3**, 279–290.

Huang, X. (1992) *Statistics of Bivariate Extreme Values*. Doctoral thesis, Erasmus University Rotterdam. TIR series 22, Tinbergen Institute, Rotterdam.

Ledford, A. and Tawn, J.A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika*, **83**, 169–187.

Ledford, A. and Tawn, J.A. (1997) Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B*, **59**, 475–499.

Ledford, A. and Tawn, J.A. (1998) Concomitant tail behaviour for extremes. *Adv. Appl. Probab.*, **30**, 197–215.

Peng, L. (1999) Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Probab. Lett.*, **43**, 399–409.

Pollard, D. (1984) *Convergence of Stochastic Processes*. New York: Springer-Verlag.

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

Ruben, H. (1964) An asymptotic expansion for the multivariate normal distribution and Mill's ratio. *J. Res. Nat. Bureau Standards*, **68**, 3–11.

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

Smith, R.L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174–1207.

Vervaat, W. (1972) Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **23**, 249–253.

Received May 2001 and revised April 2003