Bivariate tail estimation: dependence in asymptotic independence

GERRIT DRAISMA, HOLGER DREES, ANA FERREIRA and LAURENS DE HAAN

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

\[ P(X_i > u \text{ and } Y_i > v). \]  

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

\[ X_i = a_nX_i + c_n, \quad Y_i = b_nY_i + d_n. \]
\[
\lim_{n \to \infty} F^n(a_n x + b_n, c_n y + d_n) \\
= \lim_{n \to \infty} P\left(\frac{\max\{X_1, \ldots, X_n\} - b_n}{a_n} \leq x, \frac{\max\{Y_1, \ldots, Y_n\} - d_n}{c_n} \leq y\right) \\
= G(x, y),
\]
(1.2)
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 \( F_2 \), respectively, of the limiting distribution converge to the marginals \( G_1 \) and \( G_2 \), respectively, of the limiting distribution. Taking logarithms in (1.2), one obtains
\[
\lim_{n \to \infty} nP\left\{\frac{X - b_n}{a_n} > x \text{ or } \frac{Y - d_n}{c_n} > y\right\} = -\log G(x, y),
\]
(1.3)
hence
\[
\lim_{n \to \infty} nP\left\{\frac{X - b_n}{a_n} > x \text{ and } \frac{Y - d_n}{c_n} > y\right\} = \log G(x, y) - \log G_1(x) - \log G_2(y).
\]
(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 \( \eta \), 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 \( \eta \) 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 \( \eta \) 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.
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

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

exists, for all \(x, y \geq 0\) with \(x + y > 0\), some positive functions \(q\) and \(q_1 \rightarrow 0\) as \(t \rightarrow 0\), 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) \(q(t) = t^{1/\eta}\). The function \(q_1\) is also regularly varying at zero with an index \(\tau \geq 0\). 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 \(\lim \sup q(t) / t \leq 1\). 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 \(\nu\) 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 \(\eta\) 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 \(\eta\), 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

An extended simulation study and more detailed proofs can be found in the technical report by Draisma et al. (2001).
estimate \( \eta \) 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 \( F_T \) satisfies \( 1 - F_T(t) = q(1/t) \); 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

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

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 \( \eta \) 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 \( u = T_{n,n-m}^{(n)} \).

Alternatively the Hill estimator can be used:

\[
\hat{\eta}_2 := \frac{1}{m} \sum_{i=1}^{m} \log \frac{T_{n,n-i+1}^{(n)}}{T_{n,n-m}^{(n)}}.
\]

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 \( \sqrt{m}(\hat{\eta}_1 - \eta) \), \( i = 1, 2 \), are asymptotically normal with mean 0 and variance

\[
\sigma_1^2 = (1 + \eta)^2(1 - l)(1 - 2lc_x(1, 1)c_y(1, 1)),
\]

\[
\sigma_2^2 = \eta^2(1 - l)(1 - 2lc_x(1, 1)c_y(1, 1)),
\]

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 \( m = O(n^{2\eta/((2\eta + 1) - \varepsilon)}) \) for some \( \varepsilon > 0 \). (ii) Note that instead of (2.1) the weaker condition \( \lim_{\tau \to 0} P\{1 - F_1(X) < tx \) 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 $\sqrt{m}q_1(q^{-1}(m/n)) \to c \geq 0$. 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 $\eta$ 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

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

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

with $\hat{k} := m/\hat{l}$, and $T_{n,i}^{(n,u)}$, $i = 1, \ldots, n$, 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), \quad 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

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

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 $\sigma_i^2$ 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

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

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

$$c_1(x, y) = -k_3(\rho) - k_4(x, y, \rho), \quad q_1(t) = \frac{1}{2 \log t}.$$
where
\[ k_1(\rho) = \frac{(1 - \rho^2)^{3/2}}{(1 - \rho^2)^2} (4\pi)^{-\rho/(1+\rho)}, \quad k_2(\rho) = \frac{\rho}{1 + \rho}, \]
\[ k_3(\rho) = \frac{\rho \log(4\pi) + 2}{1 + \rho} - \frac{(1 + \rho)(2 - \rho)}{1 - \rho}, \]
\[ k_4(x, y, \rho) = \log x + \log y + \frac{(\rho - 1)(\log x + \log y) + \rho \log x \log y - \rho(\log^2 x + \log^2 y)/2}{(1 - \rho^2)}. \]

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 \( \Phi \) is the standard univariate normal distribution function:
\[ f^2(t) = 2 \log t - \log(\log t) - \log(4\pi) + \frac{\log(\log t)}{2 \log t} + \frac{\log(4\pi) - 2}{2 \log t} \]
\[ + \frac{1}{2} \left( \frac{\log(\log t)}{2 \log t} \right)^2 + o\left( \left( \frac{\log(\log t)}{\log t} \right)^2 \right), \quad \text{as } t \to \infty. \]

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
\[ P\{1 - F_1(X) < 1 - F_1(x) \text{ or } 1 - F_2(Y) < 1 - F_2(y)\} \]
\[ \approx t P\{1 - F_1(X) < (1 - F_1(x))/t \text{ or } 1 - F_2(Y) < (1 - F_2(y))/t\}, \quad (3.1) \]
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 \( \{X > x \text{ and } Y > y\} \) then a different approximation holds under condition (2.1):
\[ P\{1 - F_1(X) < 1 - F_1(x) \text{ or } 1 - F_2(Y) < 1 - F_2(y)\} \]
\[ \approx t^{1/\eta} P\{1 - F_1(X) < (1 - F_1(x))/t \text{ and } 1 - F_2(Y) < (1 - F_2(y))/t\}. \quad (3.2) \]

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
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:

\[
1 - F_i(x) \approx \frac{k}{n} \left(1 + \gamma_i \frac{x - b_i(n/k)}{a_i(n/k)}\right)^{-1/\gamma_i} =: \frac{k}{n}\left(1 - F_{a_i b_i \gamma_i}(x)\right), \quad i = 1, 2,
\]

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 \(\infty\) if \(\gamma > 0\) and \(x \leq -1/\gamma\), and as \(0\) if \(\gamma < 0\) and \(x \geq -1/\gamma\). Dekkers et al. (1989) proposed and analysed the following estimators of the parameters \(a_i\), \(b_i\) and \(\gamma_i\). Define

\[
M_j(X) := \frac{1}{k} \sum_{i=1}^{k} \left(\log X_{n,n-i+1} - \log X_{n,n-k}\right)^j, \quad j = 1, 2,
\]

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

\[
\hat{b}_1(n/k) := X_{n,n-k},
\]

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

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 \(\gamma_i\) 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 \(1 - F(x, y)\) as a shorthand for \((1 - F_1(x), 1 - F_2(y))\); likewise \(1 - F_{a,b,i} = (1 - F_{a_1, b_1, \gamma_1}, 1 - F_{a_2, b_2, \gamma_2})\) and \(1 - F_{a,b,i} = (1 - F_{a_1, b_1, \gamma_1}, 1 - F_{a_2, b_2, \gamma_2})\) are functions from \(\mathbb{R}^2\) to \([0, \infty]^2\). Then, in view of (3.4), the transformed failure set \(nk^{-1}(1 - F(C_n))\) can be approximated by
which in turn is estimated by

\[ \hat{D}_n := 1 - F_{a,b,\gamma}(C_n). \]

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:

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

where \( \hat{\eta} \) denotes one of the estimators for \( \eta \) 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 \( A_{+\varepsilon} := \{ x \in [0, \infty)^2 \mid \inf_{y \in A} \| x - y \| \leq \varepsilon \} \) and \( A_{-\varepsilon} := [0, \infty)^2 \setminus (([0, \infty)^2 \setminus A)_{+\varepsilon}) \) denote the outer and inner \( \varepsilon \)-neighbourhood of \( A \) with respect to the maximum norm \( \| x - y \| = |x_1 - y_1| \vee |x_2 - y_2| \), and \( \nu \) 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
\[
R_i(t, x) := t(1 - F_i(a_i(t)x + b_i(t))) - (1 - \gamma_i x)^{-1/\gamma_i} \to 0, \quad i = 1, 2,
\]
locally uniformly 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:
\[
R_{x_1,x_2}(t) := \max_{i=1,2} \sup_{x < x < 1/(\gamma_i x)} |R_i(t, x)(1 + \gamma_i x)^{1/\gamma_i}| \to 0 \tag{3.9}
\]
for some \(-1/\gamma_i \vee 0 < x_i < 1/((\gamma_i \vee 0) \vee 0)\), \( i = 1, 2 \). 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 \( \rho_i \)-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:
\[
k^{1/2}R_{x,x}(\frac{R}{k}) = O(1) \tag{3.10}
\]
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:
\[
\left| \frac{\hat{a}_i}{a_i} - 1 \right| \vee \left| \frac{\hat{b}_i - b_i}{a_i} \right| \vee |\hat{\gamma}_i - \gamma_i| = O_p(k^{-1/2}), \quad i = 1, 2 \tag{3.11}
\]

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
\[
w_{\gamma_1 \vee \gamma_2}(d_n) = o(k^{1/2}) \quad \text{with} \quad w_\gamma(x) := -x^\gamma \int_x^1 u^{-\gamma-1} \log u \, du. \tag{3.12}
\]
Check that

\[ w_\gamma(x) \sim \begin{cases} 
\frac{1}{\gamma} \log x, & \gamma > 0 \\
\frac{(\log x)^2}{2}, & \gamma = 0 \\
x^\gamma & \gamma < 0,
\end{cases} \]

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:

\[ d_n = O(c_n), \quad w_{\gamma_1, \gamma_2} \left( \frac{c_n}{d_n} \right) = o(k^{1/2}), \quad \left( \frac{c_n}{d_n} \right)^{1/\eta} = o((r(n))^{1/2}), \quad (3.13) \]

with \( r(n) := nq(k/n) \). In particular, (3.13) is satisfied 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 \( \nu \). 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:

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

where

\[ \mathcal{B}_n := \left\{ \mathbf{F}^{-1}_{\mathbf{a}, \mathbf{b}, \gamma} \left( 1 - \frac{1 - \mathbf{F}(C_n)}{c_n} \right) \right\} \mathbf{a} - 1 \vee \left\| \mathbf{b} - \mathbf{a} \right\| \vee \| \gamma - \gamma \| \leq \varepsilon_n \]

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

\[ \mathcal{B}_n := \mathcal{B}_n \cup \left\{ C_n, \bigcup_{B \in \mathcal{B}_{n, m}} B \right\}. \]

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 \( 1 - F(B) = [0, xk/n] \times [0, yk/n] \) 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
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 $r(n)/(d_n/c_n)^{1/9}$. 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 $T^{(n)}_i$, 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

$$c_n(\lambda) := \sup \left\{ c > 0 \left| \sum_{i=1}^{n} 1 \left\{ (X_i, Y_i) \in F_{a,b,\gamma}^{-1} \left( 1 - \frac{D_n}{c} \right) \right\} > \lambda r(n) \right\}$$

(3.16)

for some $\lambda > 0$, where

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

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 $m(n) = r(n)$.

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 $\hat{\eta}$:

$$\hat{\eta}_3 := \log 2 / \log \left( \frac{S_n(m)}{S_n([m/2])] \right)$$

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

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 $\eta$ 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 $(1 - \hat{\eta}_i)/\hat{\sigma}_i(\hat{\eta}_i) \leq \Phi^{-1}(0.95)$ or, alternatively, if $(1 - \hat{\eta}_i)/\hat{\sigma}_i(1) \leq \Phi^{-1}(0.95)$, with $\Phi^{-1}$ 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 $S_n([m/2])$ 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 $\bar{\sigma}_i(\hat{\eta}_i)$ and $\bar{\sigma}_i(1)$, and proportion of samples in which $\eta = 1$ is accepted by a 5% test, based on $\bar{\sigma}_i(\hat{\eta}_i)$ or $\bar{\sigma}_i(1)$. The last column indicates the number of simulations where calculations failed (sample size $n = 1000$, 250 simulations)

<table>
<thead>
<tr>
<th>$m_n$</th>
<th>Mean RMSE</th>
<th>Obs. $\bar{\sigma}_i(\hat{\eta}_i)$</th>
<th>$\bar{\sigma}_i(1)$</th>
<th>$\eta = 1$ accepted; test with $\bar{\sigma}_i(\hat{\eta}_i)$ with $\bar{\sigma}_i(1)$ failed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard deviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum likelihood, $\hat{\eta}_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cauchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.98 0.18 0.18 0.17 0.17 0.17 0.89 0.92 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>1.03 0.13 0.13 0.11 0.11 0.11 0.93 0.95 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>1.04 0.10 0.09 0.08 0.08 0.08 0.95 0.94 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.91 0.18 0.16 0.15 0.15 0.81 0.86 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>0.91 0.15 0.11 0.09 0.10 0.68 0.72 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>0.90 0.13 0.09 0.07 0.07 0.55 0.58 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.72 0.18 0.17 0.13 0.15 0.36 0.38 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>0.74 0.13 0.12 0.08 0.09 0.16 0.18 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>0.74 0.11 0.09 0.06 0.07 0.04 0.05 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgenstern</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.47 0.16 0.16 0.12 0.17 0.05 0.06 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>0.49 0.11 0.11 0.08 0.10 0.00 0.00 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>0.50 0.08 0.08 0.06 0.08 0.00 0.00 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hill, $\hat{\eta}_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cauchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.93 0.14 0.12 0.11 0.12 0.81 0.88 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.89 0.14 0.08 0.08 0.09 0.57 0.63 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.84 0.17 0.06 0.06 0.07 0.15 0.22 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.87 0.17 0.11 0.10 0.11 0.60 0.71 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.84 0.17 0.08 0.06 0.08 0.29 0.34 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.82 0.19 0.06 0.05 0.06 0.05 0.08 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.73 0.12 0.10 0.08 0.11 0.12 0.18 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.74 0.09 0.07 0.05 0.07 0.01 0.01 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.73 0.08 0.05 0.04 0.06 0.00 0.00 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgenstern</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.51 0.07 0.07 0.07 0.13 0.00 0.00 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.53 0.06 0.05 0.04 0.08 0.00 0.00 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.54 0.06 0.04 0.03 0.06 0.00 0.00 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peng, $\hat{\eta}_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cauchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1.05 0.37 0.36 0.23 0.25 0.92 1.00 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.97 0.18 0.18 0.16 0.18 0.88 1.00 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.88 0.17 0.12 0.11 0.14 0.67 0.97 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.96 0.23 0.23 0.20 0.23 0.90 1.00 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.85 0.20 0.12 0.12 0.17 0.60 0.97 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.80 0.21 0.09 0.09 0.14 0.28 0.67 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.78 0.20 0.19 0.18 0.30 0.60 1.00 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.75 0.10 0.09 0.12 0.19 0.27 0.94 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.74 0.09 0.07 0.09 0.14 0.05 0.27 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgenstern</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.55 0.23 0.22 0.24 0.74 0.32 1.00 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.54 0.11 0.11 0.12 0.37 0.03 1.00 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.55 0.08 0.07 0.09 0.25 0.00 0.10 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Maximum likelihood estimator, $\hat{\eta}_1$

Hill estimator, $\hat{\eta}_2$

Peng’s estimator, $\hat{\eta}_3$

Cauchy

Normal

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

Draisma et al.
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 \( \sigma_3(1) \) rejects the hypothesis is quite low for the normal and the Morgenstern distribution, because \( \sigma_3(1) \) 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 \( p_n \): \( \hat{p}(\hat{\eta}) = p_n \) as defined in (3.7) with \( c_n = c_n(1) \) defined by (3.16); \( \hat{p}(1) = c_n^{1-1/\hat{\eta}} p_n \) (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 \( \hat{\gamma}_1 \) 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)

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \hat{\gamma}_1 )</th>
<th>( \hat{\gamma}_2 )</th>
<th>( \eta )</th>
<th>( \hat{p}(\hat{\eta}) )</th>
<th>( \hat{p}(1) )</th>
<th>( \hat{p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cauchy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( \times 10^{-5} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.89</td>
<td>0.95</td>
<td>0.91</td>
<td>0.1310</td>
<td>0.2996</td>
<td>0.2914</td>
</tr>
<tr>
<td>80</td>
<td>0.98</td>
<td>1.02</td>
<td>0.99</td>
<td>0.3738</td>
<td>0.5056</td>
<td>0.4810</td>
</tr>
<tr>
<td>160</td>
<td>0.91</td>
<td>0.96</td>
<td>1.03</td>
<td>1.0815</td>
<td>0.7973</td>
<td>0.7973</td>
</tr>
<tr>
<td>240</td>
<td>1.00</td>
<td>1.04</td>
<td>1.04</td>
<td>1.6677</td>
<td>1.1440</td>
<td>1.1285</td>
</tr>
<tr>
<td>Morgenstern</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>( \times 10^{-5} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.96</td>
<td>0.99</td>
<td>0.44</td>
<td>0.3277</td>
<td>11.7065</td>
<td>0.3277</td>
</tr>
<tr>
<td>80</td>
<td>1.01</td>
<td>1.03</td>
<td>0.46</td>
<td>0.3754</td>
<td>27.0215</td>
<td>0.3754</td>
</tr>
<tr>
<td>160</td>
<td>0.94</td>
<td>0.97</td>
<td>0.48</td>
<td>0.6287</td>
<td>57.5168</td>
<td>0.6287</td>
</tr>
<tr>
<td>240</td>
<td>1.00</td>
<td>1.02</td>
<td>0.50</td>
<td>0.7640</td>
<td>83.7799</td>
<td>0.7640</td>
</tr>
<tr>
<td>Normal</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>( \times 10^{-5} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>-0.13</td>
<td>-0.15</td>
<td>0.68</td>
<td>0.0000</td>
<td>0.0600</td>
<td>0.0000</td>
</tr>
<tr>
<td>80</td>
<td>-0.17</td>
<td>-0.20</td>
<td>0.71</td>
<td>0.0000</td>
<td>0.0024</td>
<td>0.0000</td>
</tr>
<tr>
<td>160</td>
<td>-0.13</td>
<td>-0.13</td>
<td>0.73</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>240</td>
<td>-0.17</td>
<td>-0.20</td>
<td>0.75</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Exponential/Normal</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>( \times 10^{-5} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>-0.00</td>
<td>0.02</td>
<td>0.68</td>
<td>0.0054</td>
<td>0.8557</td>
<td>0.0212</td>
</tr>
<tr>
<td>80</td>
<td>0.04</td>
<td>0.06</td>
<td>0.71</td>
<td>0.0723</td>
<td>2.2344</td>
<td>0.0975</td>
</tr>
<tr>
<td>160</td>
<td>0.01</td>
<td>0.04</td>
<td>0.73</td>
<td>0.2384</td>
<td>3.7091</td>
<td>0.2467</td>
</tr>
<tr>
<td>240</td>
<td>0.04</td>
<td>0.06</td>
<td>0.75</td>
<td>0.3623</td>
<td>6.1357</td>
<td>0.3623</td>
</tr>
</tbody>
</table>
the estimates for $\gamma_1$ and $\gamma_2$ 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 $\gamma_i$ 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.

**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)
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 $H_mO$, 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](image1.png)

**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 \( \eta \) 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:

\[
S_1(x, y) := \sum_{i=1}^{n} 1\{U_i \leq x \text{ and } V_i \leq y\}, \quad S_2(x, y) := \sum_{i=1}^{n} 1\{U_i \leq x \text{ or } V_i \leq y\}. \tag{6.1}
\]

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

\[
\mathbb{E}\{W_1(x_1, y_1)W_1(x_2, y_2)\} = c(x_1 \wedge x_2, y_1 \wedge y_2),
\]

\[
\mathbb{E}\{W_2(x_1, y_1)W_2(x_2, y_2)\} = x_1 \wedge x_2 + y_1 \wedge y_2 - lc(x_1, y_1) - lc(x_2, y_2) + lc(x_1 \vee x_2, y_1 \vee y_2),
\]

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

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

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

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 \( l > 0 \),

\[
W(x, y) = \frac{1}{\sqrt{l}}(W_2(x, 0) + W_2(0, y) - W_2(x, y)) - \sqrt{l}c(x, y)W_2(x, 0) - \sqrt{l}c(x, y)W_2(0, y),
\]

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

\[
\sqrt{m} \left( S_1(kn^{-1}x, kn^{-1}y) - c(x, y) \right) \xrightarrow{D} W_1(x, y). \tag{6.2}
\]
Similarly, one obtains
\[
\sqrt{k} \left( \frac{S_2(kn^{-1}x, kn^{-1}y)}{k} - (x + y - lc(x, y)) \right) \Rightarrow W_2(x, y).
\] (6.3)

This implies
\[
\sqrt{k} \left( \frac{1}{k} \sum_{i=1}^{n} 1 \left\{ U_i \leq \frac{k}{n} \right\} - x \right) \Rightarrow W_2(x, 0).
\]

Note that the generalized inverse of \(x \mapsto k^{-1} \sum_{i=1}^{n} 1 \{ U_i \leq k/n \}\) equals \(x \mapsto (n/k)U_{n,[kx]}\).

Vervaat’s (1972) lemma yields
\[
\sqrt{k} \left( \frac{n}{k} U_{n,[kx]} - x \right) \Rightarrow -W_2(x, 0)
\]
\[
\sqrt{k} \left( \frac{n}{k} V_{n,[ky]} - y \right) \Rightarrow -W_2(0, y).
\] (6.4)

For \(l = 0\), we have \(m = o(k)\) and hence
\[
\sqrt{m} \left( \frac{n}{k} U_{n,[kx]} - x \right) \overset{p}{\rightarrow} 0
\]
\[
\sqrt{m} \left( \frac{n}{k} V_{n,[ky]} - y \right) \overset{p}{\rightarrow} 0.
\]

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

Denote by \(Q_n\) the tail empirical quantile function pertaining to \(T_i^{(n)}\), \(1 \leq i \leq n\), that is,
\[
Q_n(t) := T_i^{(n)}_{n, n-\lfloor nt \rfloor}, \quad 0 < t < n/m.
\]

The following lemma is central to the proof of the asymptotic normality of estimators for \(\eta\) 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 \(\tilde{W}\) 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 \(t_0, \varepsilon > 0\),
\[
\sup_{0 < t \leq t_0} \left| m^{1/2} \left( \frac{k}{n} Q_n(t) - t^{-\eta} \right) - \eta t^{-(\eta + 1)} \tilde{W}(t) \right| = o_P(1).
\]
Proof. First, check that
\[
\sum_{i=1}^{n} 1\{T_i^{(n)} > x\} = \sum_{i=1}^{n} 1\{R_i^{X} > (n+1)(1-1/x)\text{ and } R_i^{Y} > (n+1)(1-1/x)\}
\]
\[
= \sum_{i=1}^{n} 1\{U_i < U_{n,(n+1)/x} \text{ and } V_i < V_{n,(n+1)/x}\} \text{ a.s.}
\]
with the convention \( U_{n,n+1} = V_{n,n+1} = 1 \). Hence
\[
F_n(x) := \frac{1}{n} \sum_{i=1}^{n} 1\{k/(n+1)T_i^{(n)} > x\} = \frac{1}{n} S_1(U_{n,[k/x]} - V_{n,[k/x]}),
\]
where \( f(x-) \) denotes the left-hand limit of \( f \) at \( x \). From Lemma 6.1 one readily obtains that
\[
m^{1/2}\left(\frac{F_n(x)}{q(k/n)} - x^{-1/\eta}\right)_{0 < x < \beta} \to (W(1/x, 1/x))_{0 < x < \infty}
\]
\[
\Rightarrow m^{1/2}\left(\frac{F_n(x^{-\eta})}{q(k/n)} - x\right)_{0 < x < \infty} \to (W(x^\eta, x^\eta))_{0 < x < \infty} := \ol{W}
\]
\[
\Rightarrow m^{1/2}(\frac{F_n^{-1}(q(k/n)t)^{-1/\eta}}{t})_{0 < t < \infty} \to -\ol{W}
\]
weakly in \( D(0, \infty) \), where in the last step Vervaat’s (1972) lemma is used. For this, note that \( \ol{W} \) 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 \( \delta \)-method yields, for suitable versions,
\[
F_n^{-1}(q(k/n)t) = t^{-\eta}(1 + m^{-1/2}\eta t^{-1}\ol{W}(t) + o(m^{-1/2})) \text{ a.s.}
\]
uniformly on compact intervals bounded away from 0.

Next, note that \( F_n^{-1}(q(k/n)t) = k/nQ_n(t) = O(1/m) \) uniformly and \( \sup_{0 < t < \beta} t^{-1/2+\varepsilon} |\ol{W}(t)| = o_P(1) \) as \( \beta \downarrow 0 \) by the law of the iterated logarithm and the aforementioned representation of \( \ol{W} \). Thus it suffices to prove that, for all \( \delta > 0 \),
\[
\lim_{\beta \downarrow 0} \limsup_{n \to \infty} P\left\{ \sup_{0 < t < \beta} m^{1/2}t^{\eta+1/2+\varepsilon}\left|\frac{k}{n+1}Q_n(t) - t^{-\eta}\right| > \delta \right\} = 0. \tag{6.5}
\]

Here we will only consider
\[
P\left\{ \sup_{0 < t < \beta} m^{1/2}t^{\eta+1/2+\varepsilon}\left(\frac{k}{n+1}Q_n(t) - t^{-\eta}\right) > \delta \right\}
\]
\[
\leq P\left\{ \exists i \leq m\delta + 1 : \frac{k}{n+1}T_{n,n-i+1}^{(n)} > x_{i,n} \text{ and } x_{i,n} < k \right\} \tag{6.6}
\]
with
\[ x_{i,n} := \left( \frac{i}{m} \right)^{-\eta} + \delta m^{-1/2} \left( \frac{i}{m} \right)^{-\left(\eta + 1/2 + \epsilon\right)}. \]

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
\[
P \left\{ \exists 1 \leq i \leq m9 + 1 : \sum_{j=1}^{n} 1 \{ T_j > (1/U_{n,[k/x_{i,n}]}) \wedge (1/V_{n,[k/x_{i,n}]}) \} \geq i \text{ and } x_{i,n} < k \right\}.
\]

We now distinguish two different ranges of \( i \)-values.

According to Shorack and Wellner (1986, Theorem 10.3.1), for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that eventually, with probability greater than \( 1 - \epsilon \),
\[
(1/U_{n,[k/x_{i,n}]}) \wedge (1/V_{n,[k/x_{i,n}]}) \geq \frac{n}{k} x_{i,n} \delta \geq \delta L k^{-1} m^{\eta} n^{1-\eta} (n/i)^\eta,
\]
for all \( i \leq i_n := \lfloor (\delta m^i/L)^{1/(1+\epsilon)} \rfloor \), with \( x_{i,n} < k \).

Since \( q^{-1} \) is \( \eta \)-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 \( k/n = o((m/n)^{\eta+\epsilon}) \) and \( F_T^{-1}(1-t) = o(t^{\eta+\epsilon}) \) as \( t \downarrow 0 \) for all \( t > 0 \). Hence the right-hand side of (6.7) is of larger order than \( F_T^{-1}(1-2i/\delta Ln) \).

If \( \eta = 1 \), in view of (2.1) and Lemma 2.1 of Drees (1998a), we have
\[
\sup_{x < 1} x^{i-1} \left| \frac{q'(x)}{q(t)} - x \right| = o(q_1(t)).
\]

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

Hence, for all \( \eta \), it follows that
\[
\limsup_{n \to \infty} P \left\{ \exists 1 \leq i \leq i_n : \sum_{j=1}^{n} 1 \{ T_j > (1/U_{n,[k/x_{i,n}]}) \wedge (1/V_{n,[k/x_{i,n}]}) \} \geq i \text{ and } x_{i,n} < k \right\} 
\leq \limsup_{n \to \infty} P \left\{ \exists 1 \leq i \leq i_n : T_{n,n-i+1} > \frac{n}{k} x_{i,n} \delta \right\} + \bar{\epsilon}
\leq \limsup_{n \to \infty} P \left\{ \max_{1 \leq i \leq m+1} \frac{T_{n,n-i+1}}{F_T^{-1}(1-2i/\delta Ln)} > 1 \right\} + \bar{\epsilon}
< 2\bar{\epsilon}
\]
for sufficiently large \( L \), where for the last step we again use Theorem 10.3.1 of Shorack and Wellner (1986).

Let \( y_{i,n} := \frac{n}{k} x_{i,n} - \delta nk^{-1/2} x_{i,n}^{3/2+\epsilon} \).
for some $\ell \in (0, \varepsilon)$ and $\tilde{\delta} > 0$. Using

$$
\lim \limsup_{n \to \infty} \mathbb{P} \left\{ \sup_{0 < r < \delta} \frac{k^{1/2} r^{3/2 + \ell}}{nU_{n,k} |k|} | t^{-1} > \tilde{\delta} \right\} = 0
$$

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

$$
\lim \limsup_{n \to \infty} \mathbb{P} \left\{ \exists i_n < i \leq m\delta + 1 : \sum_{j=1}^{n} \mathbb{1} \{ T_j > (1/U_{n,k_j\delta}) \wedge (1/V_{n,k_j\delta}) \} \geq i \right\} 
\leq \lim \limsup_{n \to \infty} \mathbb{P} \left\{ \exists i_n < i \leq m\delta + 1 : m^{1/2} \left( \frac{i}{m} \right)^{\eta + 1/2 + \ell} \left( \frac{k}{n} \right) T_{n,n-i+1} - \left( \frac{i}{m} \right)^{-\eta} > \delta/2 \right\} = 0. \tag{6.9}
$$

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 asymptotically negligible.

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

**Proof of Theorem 2.1** (asymptotic normality of $\bar{\eta}_1$ and $\bar{\eta}_2$). 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 $\bar{\eta}_1$ and $\bar{\eta}_2$ follows from Lemma 6.2 exactly as in Drees (1998a, Example 4.1) and Drees (1998b, Example 3.1) using the $\delta$-method. The asymptotic variance is given by

$$
\int_0^1 \int_0^1 \text{cov}(\bar{W}(s), \bar{W}(t))(st)^{-(\eta+1)} \nu_\eta(ds)\nu_\eta(dt)
$$

with $\nu_\eta(dt) := (\eta + 1)^2 (\eta^2 - (2\eta + 1)t^{2\eta})/\eta dt + (\eta + 1)\varepsilon_1(dt)$ for the maximum likelihood estimator $\bar{\eta}_1$ and $\nu_\eta(dt) := \eta(t^\eta dt - \varepsilon_1(dt))$ in case of the Hill estimator. (Here $\varepsilon_1$ denotes the Dirac measure at 1.) Now using the homogeneity of order 1 of the covariance function which implies $\int_0^1 \text{cov}(\bar{W}(s), \bar{W}(t))(st)^{-1} ds = \int_0^1 \text{cov}(\bar{W}(u), \bar{W}(1))u^{-1} du$, one obtains $(\eta + 1)^2 \text{var}(\bar{W}(1))$ and $\eta^2 \text{var}(\bar{W}(1))$, 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,

$$
\hat{l} = \frac{m}{k} \frac{k}{n} Q_n(1) \to l
$$

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

$$
\frac{k}{n} \frac{t^{(s,u)}}{n Q_n(1)} = \left( \frac{t}{c(1 + u, 1)} \right)^{-\eta} + O_p(m^{-1/2}).
$$
Hence, if $\eta = 1$, then

$$\hat{c}_s(1, 1) = \frac{k}{\bar{k}}(\bar{k}^{1/4}(c(1 + 1/k^{1/4}, 1) - c(1, 1)) + O_P(k^{1/4}m^{-1/2})) P_{n, c}(1, 1).$$

The consistency of $\hat{c}_s(1, 1)$ 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

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

and thus

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

Together with the analogous result for $\hat{c}_y(1, 1)$ and the consistency of $\hat{l}$ and $\hat{\eta}$, 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:

$$p_n = P\{1 - F(X, Y) \in 1 - F(C_n)\}$$

(3.14) $q\left(\frac{k}{n}\right) v\left(\frac{n}{\bar{k}}(1 - F(C_n))\right)$

Lemma 6.4 $q\left(\frac{k}{n}\right) v(D_n)$

(2.2) $C_n^{1/\eta} q\left(\frac{k}{n}\right) v\left(\frac{D_n}{c_n}\right)$

Cor. 6.3 $C_n^{1/\eta} q\left(\frac{k}{n}\right) v\left(1 - F_{ab,\gamma}\left(F^{-1}_{ab,\gamma}\left(1 - \hat{D}_n\right)\right)\right)$

Lemma 6.5 $C_n^{1/\eta} q\left(\frac{k}{n}\right) v\left(\frac{n}{\bar{k}}\left(1 - F\left(F^{-1}_{ab,\gamma}\left(1 - \hat{D}_n\right)\right)\right)\right)$

(3.14) $C_n^{1/\eta} P\{1 - F(X, Y) \in 1 - F(B)\}_{b = F^{-1}_{ab,\gamma}(1 - D_n/c_n)}$

Lemma 6.6 $C_n^{1/\eta} \sum_{i=1}^n 1\left\{(X_i, Y_i) \in F^{-1}_{ab,\gamma}\left(1 - \hat{D}_n/c_n\right)\right\}$

$\sim \hat{p}_n.$

(6.10)
Lemma 6.3. Let \( a = a(n), \, \tilde{a} > 0, \, b, \, \tilde{b}, \, \gamma, \, \tilde{\gamma} \in \mathbb{R} \) denote sequences such that

\[
\left| \frac{\tilde{a}}{a} - 1 \right| \lor \left| \frac{\tilde{b} - b}{a} \right| \lor |\tilde{\gamma} - \gamma| = O(\varepsilon_n)
\]

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 \( w_\gamma \) defined in (3.12). Then

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

uniformly for \( 0 \leq x \leq \lambda_n \).

Proof. First, note that

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

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

\( \gamma > 0 \). Then

\[
T(x) = (1 + (1 + O(\varepsilon_n))(x^{-\gamma} - 1 + O(\varepsilon_n)))^{-(1 + O(\varepsilon_n))/\gamma}
\]

\[
= (x^{-\gamma}(1 + O(\varepsilon_n)) + O(\varepsilon_n))^{-(1 + O(\varepsilon_n))/\gamma}
\]

\[
= x \exp(O(\varepsilon_n)\log x)(1 + o(1))
\]

uniformly for \( 0 \leq x \leq \lambda_n \). For \( \lambda_n \varepsilon_n \leq x \leq \lambda_n \),

\[
|\log x|\varepsilon_n \leq (|\log \lambda_n| + |\log \varepsilon_n|)\varepsilon_n \to 0,
\]

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

\[
T(x) \leq T(\lambda_n \varepsilon_n) = \lambda_n \varepsilon_n(1 + o(1)) = o(\lambda_n) = x + o(\lambda_n)
\]

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 \( \lambda_n \delta_n \leq x \leq \lambda_n \),

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

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

\( \gamma = 0 \). Note that \( \tilde{\gamma}|\log x| \to 0 \) uniformly for \( \lambda_n \varepsilon_n \leq x \leq \lambda_n \). Hence a Taylor expansion of \( \log \) yields...
Bivariate tail estimation

\[
T(x) = \exp \left( -\frac{1}{\gamma} \log(1 + \gamma(1 + O(\varepsilon_n))(-\log x + O(\varepsilon_n))) \right)
\]

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

\[
= x \exp(O(\varepsilon_n)\log x + O(\varepsilon_n) + O(\varepsilon_n \log^2 x))
\]

\[
x = x(1 + o(1)),
\]

and thus the assertion follows by the aforementioned arguments. \(\square\)

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

\[
(a, b, \gamma) \in M(\varepsilon_n) := \left\{ (\bar{a}, \bar{b}, \bar{\gamma}) \in (0, \infty) \times \mathbb{R}^2 \left| \begin{array}{l} \bar{a} - 1 \quad \bar{b} - b \quad |\bar{\gamma} - \gamma| \leq \varepsilon_n \end{array} \right. \right\}.
\] (6.12)

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

\[
P\left\{ A_{-\delta} \subset \frac{D_n}{d_n} \subset A_{+\delta} \right\} \to 1.
\]

**Proof.** Since the set \(A\) is bounded, there exists \(L > 0\) such that \(D_n \subset [0, d_nL]^2\) 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 \(\lambda_n = d_nL\). Then \(P\{(\hat{a}, \hat{b}, \hat{\gamma}) \in (M(\varepsilon_n))^2 \} \to 1\) with \(M(\varepsilon_n)\) defined in (6.12), and Lemma 6.3 yields

\[
\sup_{(x, y) \in D_n} \|1 - F_{a, b, \gamma}(F_{a, b, \gamma}^{-1}(1 - (x, y)) - (x, y))\| \leq \frac{\delta}{2} d_n
\] (6.13)

with probability tending to 1. Thus, in view of \(\hat{D}_n = 1 - F_{a, b, \gamma}(F_{a, b, \gamma}^{-1}(1 - D_n))\) 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),

\[
d_n(x, y) \leq 1 - F_{a, b, \gamma}(F_{a, b, \gamma}^{-1} \left( 1 - d_n \left( x + \frac{\delta}{2}, y + \frac{\delta}{2} \right) \right))
\]

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

\[
P\left\{ A_{-\delta} \subset \left( \frac{D_n}{d_n} \right)_{-\delta/2} \subset \frac{\hat{D}_n}{d_n} \right\} \to 1.
\] \(\square\)
Corollary 6.2. If the conditions of Corollary 6.1 hold, and (3.13) also holds, then, for all \( \delta > 0 \),

\[
P\left\{ A_{-\delta} \subset \frac{c_n}{d_n} \left( 1 - \mathbf{F}_{a,b,\gamma} \left( \mathbf{F}_{a,b,\gamma}^{-1} \left( 1 - \frac{D_n}{c_n} \right) \right) \right) \subset A_{+\delta} \right\} \to 1.
\]

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

\[
\sup_{(x,y) \in D_n/c_n} \| 1 - \mathbf{F}_{a,b,\gamma} \left( \mathbf{F}_{a,b,\gamma}^{-1} \left( 1 - (x, y) \right) \right) - (x, y) \| \leq \frac{\delta d_n}{2c_n}
\]

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

Corollary 6.3. Under the conditions of Corollary 6.2,

\[
v\left( 1 - \mathbf{F}_{a,b,\gamma} \left( \mathbf{F}_{a,b,\gamma}^{-1} \left( 1 - \frac{D_n}{c_n} \right) \right) \right) = v\left( \frac{D_n}{c_n} \right) (1 + o_P(1)).
\]

**Proof.** Denote the boundary of the set \( A \) by \( \partial A \). 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 \( v \) is homogeneous in the sense of (2.2) and \( v(A) < \infty \) by the boundedness of \( A \), it follows that \( v(\partial A) = 0 \). Moreover, \( A_{+\delta} \setminus A_{-\delta} \setminus \partial A \) as \( \delta \downarrow 0 \), so that \( v(A_{+\delta} \setminus A_{-\delta}) \to 0 \). Thus Corollary 6.2 and Condition D yield

\[
v\left( \frac{c_n}{d_n} \left( 1 - \mathbf{F}_{a,b,\gamma} \left( \mathbf{F}_{a,b,\gamma}^{-1} \left( 1 - \frac{D_n}{c_n} \right) \right) \right) \right) \to v(A)
\]

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

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

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

**Proof.** There exists \( L > 0 \) such that \( D_n \subset [0, d_nL]^2 \) 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 \((x, y) \in D_n,\)

\[
\frac{n}{k} \left( 1 - \mathbf{F}(\mathbf{F}_{a,b,\gamma}^{-1}(1 - (x, y))) \right) = (x(1 + \delta_x), y(1 + \delta_y)),
\]

with \(|\delta_x| \vee |\delta_y| \leq R_{\alpha_{x,2}}(n/k)\), for sufficiently large \( n \). According to (3.8), the left-hand side of (6.14) is an element of \( D_n(1 + R_{\alpha_{x,2}}(n/k)) \). Thus, by the definition of \( D_n,\)
Likewise, (6.14) together with (3.8) implies
\[ D_n \left( 1 - R_{n_1,n_2} \left( \frac{M}{C} \right) \right) \subset \frac{M}{C} (1 - F(C_n)) \]
eventually. Now the assertion is obvious from the homogeneity property (2.2). \( \square \)

**Lemma 6.5.** Under Condition D, (3.8), (3.9) and (3.11)–(3.13) one has
\[ v \left( \left( 1 - F_{a, b, \gamma} \left( 1 - \frac{D_n}{c_n} \right) \right) \right) = v \left( \left( 1 - F \left( \frac{F^{-1}_{a, b, \gamma} \left( 1 - \frac{D_n}{c_n} \right) }{c_n} \right) \right) \right) (1 + o(1)). \]

**Proof.** The proof is very much the same as that for Lemma 6.4, with \( D_n \) replaced by \( 1 - F_{a, b, \gamma}(1 - D_n/c_n) \). 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 \( x_2 \). \( \square \)

**Lemma 6.6.** If the conditions of Theorem 3.1 are satisfied, then
\[ \sup_{B \in B_n} \left| \frac{n^{-1} \sum_{i=1}^{n} 1 \{ 1 - F(X_i, Y_i) \in 1 - F(B) \}}{P(1 - F(X, Y) \in 1 - F(B))} - 1 \right| \to 0 \text{ in probability.} \]

**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 \( B \in B_n \) can be represented as
\[ B = F_{a, b, \gamma}^{-1} \left( \frac{1 - F_{a, b, \gamma}(C_n)}{c_n} \right) \]
with \( (a, b, \gamma) \in (M(\xi_n))^2 \) (cf. (6.12)). Therefore the arguments of the proofs for Lemma 6.5 and Corollary 6.3 show that
\[ v \left( \frac{M}{C} (1 - F(B)) \right) = v(1 - F_{a, b, \gamma}(B))(1 + o(1)) = v \left( \frac{D_n}{c_n} \right) (1 + o(1)) \]
\[ = \left( \frac{d_n}{c_n} \right)^{1/\eta} v(A)(1 + o(1)) \]
(6.16)
uniformly for \( B \in B_n \) (cf. Remark 6.1). Now (3.14) leads to
\[ P(1 - F(X, Y) \in 1 - F(B)) = q \left( \frac{k}{n} \right) \left( \frac{d_n}{c_n} \right)^{1/\eta} v(A)(1 + o(1)) \]
(6.17)
uniformly. In particular, there exists \( n_0 \) such that \( P(1 - F(X, Y) \in 1 - F(B)) < 1/2 \) for all \( n \geq n_0 \) and all \( B \in B_n \).
Next, note that
\[ B_t := \bigcup_{B \in B_n, n \geq n_0, \frac{P(1 - F(X, Y) \leq 1 - F(B))}{1 - P(1 - F(X, Y) \leq 1 - F(B))} \leq t} B. \]
(6.18)

In view of (6.15), one may prove as in Corollary 6.2 that, for all \( \delta > 0 \), eventually \( 1 - F(B) \subset A+\delta d_n/c_n \) for all \( B \in B_n \). Hence, it follows as in the proof of Lemma 6.4 that
\[ \frac{n}{k} (1 - F(B)) \leq \frac{d_n}{c_n} A+\delta (1 + o(1)) \]
uniformly for \( B \in B_n \).

Let \( n(t) := \min\{n \geq n_0 | q(k/n)(d_n/c_n)^{1/\eta} \nu(A) \leq 3t\} \), which tends to \( \infty \) as \( t \) tends to 0. Combining (6.17)–(6.19), we arrive at
\[ 1 - F(B_t) \subset \bigcup_{n \geq n(t)} \frac{k(n)d_n}{nc_n} A+\delta (1 + o(1)) \leq 2 \sup_{n \geq n(t)} \frac{k(n)d_n}{nc_n} A+\delta \]
for sufficiently small \( t \). By (3.14), the regularity condition on \( k(n) \) and the definition of \( n(t) \), it follows that
\[
P\{1 - F(X, Y) \in 1 - F(B_t)\} = O\left( q\left( \frac{k(n(t))}{n(t)} \right) \left( \frac{n(t)}{k(n(t))} \sup_{n \geq n(t)} \frac{k(n)d_n}{nc_n} \right)^{1/\eta} \right) = O\left( q\left( \frac{k(n(t))}{n(t)} \right) \left( \frac{d_n}{c_n} \right)^{1/\eta} \right) = O(t).
\]
Since \( B_n \) is a VC class, Theorem 5.1 of Alexander (1987) yields
\[
\sup\left\{ n^{-1} \sum_{i=1}^{n} 1\{1 - F(X_i, Y_i) \in 1 - F(B)\} - 1 \left| B \in B_n, P\{1 - F(X, Y) \in 1 - F(B)\} \geq \varepsilon_n \right\}
\]
\[
\rightarrow 0,
\]
provided \( n\varepsilon_n \rightarrow \infty \). 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. \( \square \)
**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), \( F^{1/(1-\alpha)}(1 - D_n/c_n) \) belongs to \( \mathcal{B}_n \) with probability tending to 1 and that \( \log c_n = o((r(n))^{1/2}) \) implies \( c_n^{1/\eta} = c_n^{1/\eta}(1 + o_P(1)) \) since \( \eta \) was assumed \( \sqrt{r(n)} \)-consistent for \( \eta \).

\[ \square \]

**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**


Received May 2001 and revised April 2003