AISM (2006) 58: 261–290
DOI 10.1007/s10463-005-0016-6

Michael Falk · René Michel

Testing for tail independence in extreme value models

Received: 13 July 2004 / Revised: 14 February 2005 / Published online: 4 March 2006

Abstract Let (X, Y) be a random vector which follows in its upper tail a bivariate extreme value distribution with reverse exponential margins. We show that the conditional distribution function (df) of $X + Y$, given that $X + Y > c$, converges to the df $F(t) = t^2$, $t \in [0, 1]$, as $c \uparrow 0$ if and only if X, Y are tail independent. Otherwise, the limit is $F(t) = t$. This is utilized to test for the tail independence of X, Y via various tests, including the one suggested by the Neyman–Pearson lemma. Simulations show that the Neyman–Pearson test performs best if the threshold c is close to 0, whereas otherwise it is the Kolmogorov–Smirnov test that performs best. The mathematical conditions are studied under which the Neyman–Pearson approach actually controls the type I error. Our considerations are extended to extreme value distributions in arbitrary dimensions as well as to distributions which are in a differentiable spectral neighborhood of an extreme value distribution.

Keywords Bivariate extremes · Pickands dependence function · Tail independence · Tail dependence parameter · Neyman–Pearson test · Kolmogorov–Smirnov test · Fisher's κ · Chi-square goodness-of-fit test · Differentiable spectral neighborhood · Generalized Pareto distribution

1 Introduction

Let (X, Y) be a random vector (rv) with values in $(-\infty, 0]^2$, whose distribution function (df) $H(x, y)$ coincides, for $x, y \leq 0$ close to 0, with a max-stable or extreme value df (EV) G with reverse exponential margins, i.e.,

$$G(x, 0) = G(0, x) = \exp(x), \quad x \leq 0,$$

M. Falk (✉) · R. Michel
Institute of Mathematics, University of Würzburg,
97074 Würzburg, Germany
E-mail: falk@mathematik.uni-wuerzburg.de

and

$$G^n\left(\frac{x}{n}, \frac{y}{n}\right) = G(x, y), \quad x, y \le 0, \; n \in \mathbb{N}.$$

It is well-known that G can be represented as

$$G(x, y) = \exp\left((x + y)D\left(\frac{x}{x + y}\right)\right), \quad x, y \le 0,$$

where $D : [0, 1] \to [1/2, 1]$ is the *Pickands dependence function* (see Pickands (1981), Galambos ((1987), Theorem 5.4.5), Resnick ((1987), Proposition 5.11)). It is absolutely continuous and convex, satisfies $D(0) = D(1) = 1$, and its derivative $D'(z)$ has values between -1 and 1. The cases $D(z) = 1$ and $D(z) = \max(z, 1-z)$, $z \in [0, 1]$, characterize the cases of independence and complete dependence of the margins of G. We refer to Falk and Reiss ((2005), Sect. 2), for elementary derivations of the basic properties of D.

Examples are the Marshall–Olkin df (Marshall and Olkin 1976), the bivariate Gumbel df of type B (Gumbel 1960; Johnson and Kotz 1972 p. 251), or the Hüsler–Reiss EV (Hüsler and Reiss 1989). For well-organized accounts of the current state of research of the theory of multivariate extreme value distributions we refer to Kotz and Nadarajah ((2000), Chapter 3), and to Coles ((2001), Chapter 8).

Let now $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of (X, Y). If diagnostic checks of $(X_1, Y_1), \dots, (X_n, Y_n)$ suggest X, Y to be independent in their upper tail, then modeling with dependencies leads to the over estimation of probabilities of extreme joint events. Some inference problems caused by model mis-specification are, for example, discussed in Dupuis and Tawn (2001). Testing for tail independence is, therefore, mandatory in a data analysis of extreme values.

In Sect. 2 we establish the fact that the conditional distribution of $X + Y$, given $X + Y > c$, has a limiting df $F(t) = t^2$, $t \in [0, 1]$, as $c \uparrow 0$ if and only if $D(z) = 1$, $z \in [0, 1]$—i.e., if and only if X and Y are tail independent. If D is not the constant function 1, then the limiting df is that of the uniform distribution on $[0, 1]$: $F(t) = t$, $t \in [0, 1]$.

This result will be utilized to define tests for the tail independence of X and Y which are suggested by the Neyman–Pearson lemma as well as via the goodness-of-fit tests that are based on Fisher's κ, on the Kolmogorov-Smirnov test as well as on the chi-square goodness-of-fit test, applied to the exceedances $X_i + Y_i > c$ in the sample $(X_1, Y_1), \dots, (X_n, Y_n)$. Numerous simulations which we carried out indicate that the Neyman–Pearson test (NPT) has the smallest type II error rate, closely followed by the Kolmogorov–Smirnov test and the chi-square test, whereas Fisher's κ almost fails. The NPT does not, however, control the type I error rate if the threshold c is too far away from 0. The other three tests control the type I error rate for any c. Note that a multivariate EV with arbitrary one-dimensional margins is transformed to an EV G with reverse exponential margins by simple corresponding transformations of its margins.

In accordance with the work of Geffroy (1958, 1959) and Sibuya (1960), X and Y are said to be *tail independent* or *asymptotically independent* if the *tail dependence parameter*

$$\chi := \lim_{c \uparrow 0} P(Y > c | X > c)$$

equals 0. Note that $\chi = 2(1 - D(1/2))$ and, thus, the convexity of $D(z)$ implies that $\chi = 0$ is equivalent to the condition $D(z) = 1$, $z \in [0, 1]$.

Recent attention given to the statistical properties of asymptotically independent distributions is largely a result of a series of articles by Ledford and Tawn (1996, 1997, 1998). Coles et al. (1999) give an elementary synthesis of the theory.

Peng (1999) proposed a consistent tail dependence parameter estimator and established its asymptotic normality; see also the discussion in Kotz and Nadarajah ((2000), Sect. 3.2.1). Such a result could also be used to test for the asymptotic independence of X and Y. For a review of the statistical estimation in multivariate extreme value models we refer to Kotz and Nadarajah ((2000), Sect. 3.6). For a directory of the coefficients of tail dependence we refer to Heffernan (2000).

In Sects. 3 and 4 we investigate the NPT in more detail. In Sect. 3 we determine the rate at which c has to converge to 0 such that the NPT actually controls the type I error. In Sect. 4 we compute the asymptotic power of the NPT by considering a triangular array $(X_1^{(n)}, Y_1^{(n)}), \ldots, (X_n^{(n)}, Y_n^{(n)})$ of rvs, whose tail dependence parameter χ_n converges to 0 as n increases.

In Sect. 5 we extend our considerations to distributions, which are in a certain neighborhood of an EV. In Sect. 6 we investigate generalized Pareto distributions and in Sect. 7 we consider extreme value distributions in arbitrary dimension.

2 The bivariate case

We assume in the following that the rv (X, Y) has a df $H(x, y)$, which coincides, for x, y close to 0, with the EV $G(x, y) = \exp((x + y)D(x/(x + y)))$, where D is an arbitrary Pickands dependence function.

Lemma 2.1 *We have for $c < 0$ close to 0*

$$P(X + Y \le c) = \exp(c) - c \int_0^1 \exp(cD(z)) \big(D(z) + D'(z)(1 - z)\big)\, dz.$$

Proof The following arguments have been taken from Ghoudi, Khoudraji, and Rivest (1998). The conditional df of $X + Y$, given $X = u < 0$, is for c close to 0

$$\begin{aligned} P(X + Y \le c|X = u) &= P(Y \le c - u|X = u) \\ &= \lim_{\varepsilon \downarrow 0} \frac{P(Y \le c - u, X \in [u, u + \varepsilon])}{P(X \in [u, u + \varepsilon])} \\ &= \lim_{\varepsilon \downarrow 0} \frac{G(u+\varepsilon, c-u) - G(u, c-u)}{\varepsilon} \frac{\varepsilon}{\exp(u+\varepsilon) - \exp(u)} \\ &= \frac{G(u, c - u)}{\exp(u)} \left(D\left(\frac{u}{c}\right) + D'\left(\frac{u}{c}\right) \frac{c - u}{c} \right) \\ &= \exp\left(cD\left(\frac{u}{c}\right) - u \right) \left(D\left(\frac{u}{c}\right) + D'\left(\frac{u}{c}\right) \frac{c - u}{c} \right) \end{aligned}$$

if $u > c$, and

$$P(X + Y \le c|X = u) = 1 \quad \text{if } u \le c.$$

Hence we obtain

$$\begin{aligned} &P(X+Y\le c)\\ &=\int_{-\infty}^{0} P(Y\le c-u|X=u)\exp(u)\,\mathrm{d}u\\ &=\int_{c}^{0}\exp\left(cD\left(\frac{u}{c}\right)\right)\left(D\left(\frac{u}{c}\right)+D'\left(\frac{u}{c}\right)\frac{c-u}{c}\right)\mathrm{d}u+\exp(c)\\ &=\exp(c)-c\int_0^1\exp(cD(u))\left(D(u)+D'(u)(1-u)\right)\,\mathrm{d}u. \end{aligned}$$

The following auxiliary result is actually one of the main results of the present paper.

Lemma 2.2 *We have uniformly for $t\in[0,1]$ as $c\uparrow 0$*

$$P(X+Y>ct|X+Y>c)=\begin{cases} t^2(1+O(c)), & \text{if } D(z)=1,\ z\in[0,1],\\ t(1+O(c)) & \text{elsewhere.}\end{cases}$$

Proof From Lemma 2.1 and the Taylor expansion of exponentials we obtain uniformly for $t\in[0,1]$ and c close to 0

$$\begin{aligned} &P(X+Y>ct|X+Y>c)\\ &\quad=\frac{1-\exp(ct)+ct\int_0^1\exp(ctD(u))\left(D(u)+D'(u)(1-u)\right)\,\mathrm{d}u}{1-\exp(c)+c\int_0^1\exp(cD(u))\,(D(u)+D'(u)(1-u))\,\mathrm{d}u}\\ &\quad=\frac{-ct+ct\int_0^1 D(u)+D'(u)(1-u)\,\mathrm{d}u+O\left((ct)^2\right)}{-c+c\int_0^1 D(u)+D'(u)(1-u)\,\mathrm{d}u+O\left(c^2\right)}\\ &=t(1+O(c)) \end{aligned}$$

if D is not the constant function 1. This follows from partial integration:

$$\int_0^1 D(u)+D'(u)(1-u)\,\mathrm{d}u=2\int_0^1 D(u)\,\mathrm{d}u-1\quad\in(0,1]$$

and the fact that $D(z)\in[1/2,1]$ and that $D(0)=1$.

If $D(z)$ is the constant function 1, then we obtain uniformly for $t\in[0,1]$ and c close to 0

$$\begin{aligned} &P(X+Y>ct|X+Y>c)\\ &\quad=\frac{1-\exp(ct)+ct\exp(ct)}{1-\exp(c)+c\exp(c)}\\ &\quad=\frac{-ct-(ct)^2/2+ct(1+ct)+O\left((ct)^3\right)}{-c-c^2/2+c(1+c)+O\left(c^3\right)}\\ &\quad=t^2(1+O(c)). \end{aligned}$$

If X and Y are tail independent, then $(X+Y)/c$, conditional on $X+Y>c$, has by Lemmas 2.1 and 2.2 for c close to 0 the df

$$\begin{aligned} F_c(t) &:= P(X+Y>tc|X+Y>c) \\ &= \frac{1-(1-tc)\exp(tc)}{1-(1-c)\exp(c)} \\ &= t^2(1+O(c)), \quad 0\le t\le 1. \end{aligned} \tag{1}$$

Otherwise, the conditional df converges to the uniform df on [0,1].

Suppose now that we have n independent copies $(X_1,Y_1),\dots,(X_n,Y_n)$ of (X,Y). Fix $c<0$ and consider only those observations X_i+Y_i among the sample that satisfy $X_i+Y_i>c$. Denote these by $C_1,C_2,\dots,C_{K(n)}$ in the order of their outcome. Then $C_i/c, i=1,2,\dots$ are iid with a common df F_c, if c is large enough, and they are independent of $K(n)$, which is binomial $B(n,q)$—distributed with $q=1-(1-c)\exp(c)$. This is a consequence of Theorem 1.4.1 in Reiss (1993).

The first test which we consider is suggested by the Neyman–Pearson lemma. We have to decide, roughly, whether the df of $V_i:=C_i/c$, $i=1,2,\dots$ is equal to either the null hypothesis $F_{(0)}(t)=t^2$ or the alternative $F_{(1)}(t)=t$, $0\le t\le 1$. Assuming that these approximations of the df of $V_i:=C_i/c$ are exact and that $K(n)=m>0$, the optimal test for testing $F_{(0)}$ against $F_{(1)}$ is based on the loglikelihood ratio

$$T(V_1,\dots,V_m):=\log\left(\prod_{i=1}^m\frac{1}{2V_i}\right)=-\sum_{i=1}^m\log(V_i)-m\log(2),$$

and $F_{(0)}$ is rejected if $T(V_1,\dots,V_m)$ gets too large. Note that $-2\log(V_i)$ has a df $1-\exp(-x)$, $x\ge 0$, under $F_{(0)}$, and hence $2(T(V_1,\dots,V_m)+m\log(2))$ has the df $1-\exp(-x)\sum_{0\le j\le m-1}x^j/j!$, $x\ge 0$, under $F_{(0)}$.

The p value of the optimal test derived from the Neyman–Pearson lemma is, therefore,

$$\begin{aligned} p_{\mathrm{NP}} &= \exp\left(-2(T(V_1,\dots,V_m)+m\log(2))\right) \\ &\quad\times\sum_{0\le j\le m-1}\frac{(2(T(V_1,\dots,V_m)+m\log(2)))^j}{j!} \\ &= \exp\left(2\sum_{i=1}^m\log(V_i)\right)\sum_{0\le j\le m-1}\frac{\left(-2\sum_{i=1}^m\log(V_i)\right)^j}{j!} \\ &\approx \Phi\left(\frac{2\sum_{i=1}^m\log(V_i)+m}{m^{1/2}}\right), \end{aligned} \tag{2}$$

if m is large by the central limit theorem, where Φ denotes the df of the standard normal df.

Next we consider the goodness-of-fit tests based on C_i/c to test for the tail independence of X and Y. Conditional on $K(n)=m>0$, the rvs

$$U_i:=F_c(C_i/c)=\frac{1-(1-C_i)\exp(C_i)}{1-(1-c)\exp(c)},\quad 1\le i\le m,$$

are by Eq. 1 independent and uniformly distributed on (0, 1), if X and Y are tail independent and c is close to 0.

Consider the corresponding order statistics

$$U_{1:m} \leq \cdots \leq U_{m:m}$$

and denote, with $U_{0:m} := 0$, $U_{m+1:m} := 1$, by

$$S_j := U_{j:m} - U_{j-1:m}, \quad 1 \leq j \leq m+1$$

the $m + 1$ spacings. Consider

$$M_m := \max_{j \leq m+1} S_j$$

and

$$\kappa_m := (m+1) M_m.$$

Then, κ_m is the Fisher's κ-statistic, conditional on $K(n) = m$. It is typically used in the time series analysis for testing for white noise; see, for example, Fuller (1976). Given $K(n) = m > 0$, κ_m has the df

$$P\,(\kappa_m \leq x) = G_{m+1}\left(\frac{x}{m+1}\right),$$

where

$$G_{m+1}(x) = \sum_{j=0}^{m+1} (-1)^j \binom{m+1}{j} (\max(0, 1 - jx))^m, \qquad x > 0;$$

see, e.g. Theorem 3 in Sect. I.9 of Feller (1971).

Note that $G_{m+1}((x + \log(m+1))/(m+1)) \to_{m\to\infty} \exp(-\mathrm{e}^{-x})$, $x \in \mathbb{R}$, and that $\kappa_{K(n)} - \log(K(n) + 1) \to_{n\to\infty} \infty$ in probability if the threshold $c = c_n \uparrow 0$ satisfies $nc_n \to \infty$, $nc_n^2 \to 0$ as $n \to \infty$ and if $D(z)$ is not the constant function 1. This follows from elementary computations.

Given that $K(n) = m > 0$, the hypothesis that X and Y are tail independent is, therefore, rejected if the p value

$$p_\kappa := 1 - G_{m+1}\left(\frac{\kappa_m}{m+1}\right) = 1 - G_{m+1}\,(M_m)$$

is small, typically if $p_\kappa \leq 0.05$. A table of the critical values of Fisher's κ-test is given in Fuller (1976).

An alternative is the Kolmogorov–Smirnov test applied to $U_1, \ldots, U_m$, conditional on $K(n) = m$. Denote by $\hat{F}_m(t) := m^{-1} \sum_{i=1}^m 1_{[0,t]}(U_i)$ the empirical df of $U_1, \ldots, U_m$ and by

$$\Delta_m := m^{1/2} \sup_{t\in[0,1]} |\hat{F}_m(t) - t|$$

the Kolmogorov–Smirnov statistic. The hypothesis that X and Y are tail independent is rejected if the approximate p value

$$p_{\mathrm{KS}} := 1 - K(\Delta_m)$$

is small. By K we denote the Kolmogorov distribution. The following rule is quite common. For $m > 30$, the hypothesis of the independent tails of X, Y is rejected if $\Delta_m > c_\alpha$, where $c_{0.05} = 1.36$ and $c_{0.01} = 1.63$ are the critical values for the type I errors 0.05 and 0.01.

A further alternative is the chi-square goodness-of-fit test applied to $U_1, \dots, U_m$, conditional on $K(n) = m > 0$. Divide the interval [0, 1] into k consecutive and disjoint intervals $I_1, \dots, I_k$ and consider

$$\chi^2_{m,k} := \sum_{i=1}^{k} \frac{(m_i - mp_i)^2}{mp_i},$$

where m_i is the number of observations among $U_1, \dots, U_m$ that fall into the interval I_i and p_i is the length of I_i, $1 \le i \le k$. If m is large, i.e. if $mp_i > 5$, $1 \le i \le k$, the hypothesis that X and Y are tail independent is rejected if the approximate p value

$$p_{\chi^2} := 1 - \chi^2_{k-1}(\chi^2_{m,k})$$

is small. By χ^2_{k-1} we denote the chi-square distribution with $k - 1$ degrees of freedom.

For a general discussion of the Kolmogorov–Smirnov and the chi-square goodness-of-fit test we refer to Chapters 7 and 8 of Sheskin (2004).

The following figures exemplify numerous simulations that we carried out to evaluate the performance of each of the four tests for the tail independence defined above. Figure 1 shows quantile plots of 100 independent realizations of each of the p values p_{NP}, p_κ, p_{KS} and p_{χ^2}, based on $K(n) = m = 25$ exceedances under the hypothesis H_0 of the independence of X and Y.

The 100 p values were ordered, i.e., $p_{1:100} \le \dots \le p_{100:100}$, and the points $(i/101, p_{i:100})$, $1 \le i \le 100$, were plotted for each test. The threshold is $c = -0.5$ and the chi-squared statistic uses $k = 4$ intervals of equal length.

The three almost straight lines formed by the quantile plots pertaining to the goodness-of-fit tests in Fig. 1 indicate that each of them has the correct type I error rate. The NPT, however, does not control the type I error rate; its distribution is affected by the very small threshold $c = -0.5$. Figure 2 indicates that the value $c = -0.1$ is sufficiently close to 0 to ensure that the NPT also controls the type I error rate. The horizontal lines in each plot are drawn at the 5% level; a p value below that level usually leads to a rejection of H_0.

Next we simulate deviations from the independence and consider (X, Y) having a Marshall–Olkin df as well as a Gumbel type B df. The Marshall–Olkin df with parameter $\lambda \in [0, 1]$ is defined by

$$G_\lambda(x, y) = \exp\left(x + y - \lambda \max(x, y)\right) = \exp\left((x + y) D_\lambda\left(\frac{x}{x + y}\right)\right),$$

where

$$D_\lambda(z) = 1 - \lambda \min(z, 1 - z), \quad z \in [0, 1].$$

The Gumbel type B df with a parameter $\lambda \ge 1$ is the df

$$G_\lambda(x, y) = \exp\left(-\left((-x)^\lambda + (-y)^\lambda\right)^{1/\lambda}\right) = \exp\left((x + y) D_\lambda\left(\frac{x}{x + y}\right)\right)$$

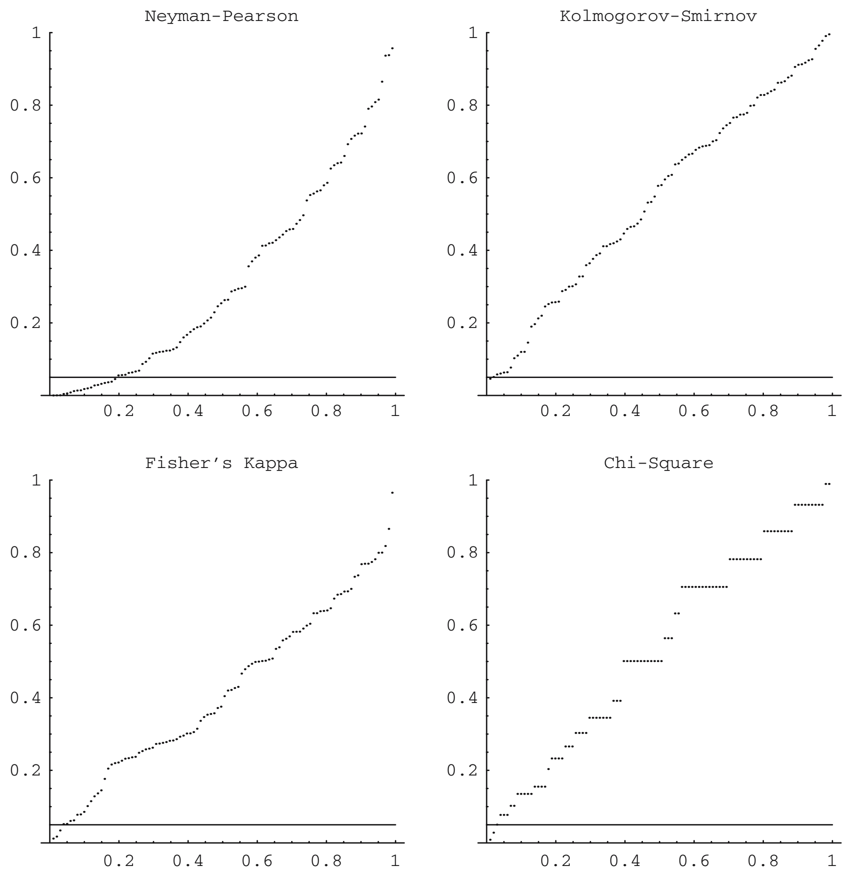

Fig. 1 Quantile plots of 100 values of p_{NP}, p_{KS}, p_κ and p_{χ^2}, with the underlying df $\exp(x+y)$, $x, y \le 0$, and 25 exceedances over the threshold $c = -0.5$

with the dependence function

$$D_\lambda(z) = \left(z^\lambda + (1-z)^\lambda\right)^{1/\lambda}, \quad z \in [0,1].$$

The plots in Figs. 3 and 4 were generated in the same way as in Fig. 1, but this time with tail dependent X, Y. Figure 3 illustrates the test for tail independence with the underlying Marshall–Olkin df with the parameter $\lambda = 0.5$, which coincides with the tail dependence parameter.

Figure 4 was generated with the underlying Gumbel type B df with the parameter $\lambda = 2$, whose tail dependence parameter is $\chi = 2 - 2^{1/\lambda} = 0.5858$. The distribution of the p values represents the performance of each of the four tests.

It turns out that the distributions of p_{NP}, p_{χ^2} and p_{KS} are now shifted to the left under dependence, i.e. their type II error rate is quite small, with the NPT for independence having the smallest error rate, followed by the Kolmogorov–Smirnov test. Since the distribution of p_κ is almost not affected, the test for the independence of X and Y based on Fisher's κ fails.

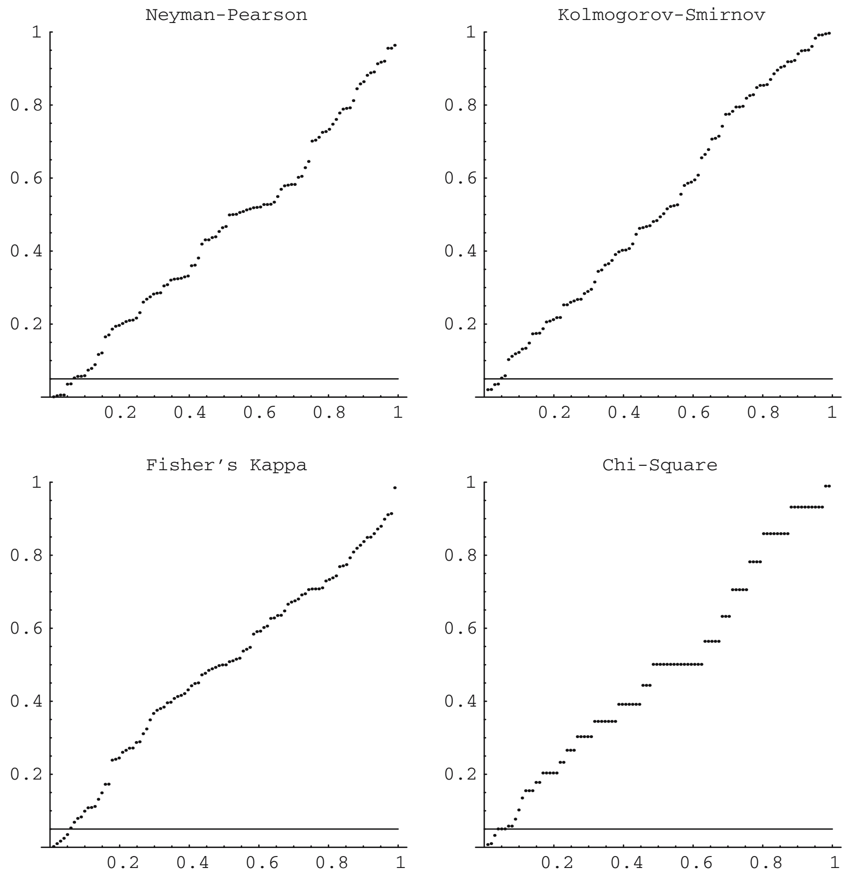

Fig. 2 Quantile plots of 100 values of p_{NP}, p_{KS}, p_κ and p_{χ^2}, with the underlying df $\exp(x+y)$, $x, y \le 0$, and 25 exceedances over the threshold $c = -0.1$

3 Type I error of the NPT

We suppose in the following again that (X, Y) is a random vector with values in $(-\infty, 0]^2$, whose df coincides, for $x, y \le 0$ close to 0, with an EV G with reverse exponential margins.

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent copies of (X, Y). Choose $c < 0$ and consider only those rvs $(X_i + Y_i)/c$ with $(X_i + Y_i)/c \le 1$. Denote these again by $V_1, V_2, \dots, V_{K(n)}$.

The limiting NPT for testing $H_0 : \ F_c(t) = t^2$ against $F_c(t) = t$, based on $V_1, \dots, V_{K(n)}$, has by (2) the approximate p value

$$p_{\mathrm{NPT}} := \Phi\left(K(n)^{-1/2} \sum_{i \le K(n)} (2\log(V_i) + 1)\right).$$

In the sequel we provide a more detailed analysis of the NPT. First we investigate the question; under which conditions on the sequence of thresholds

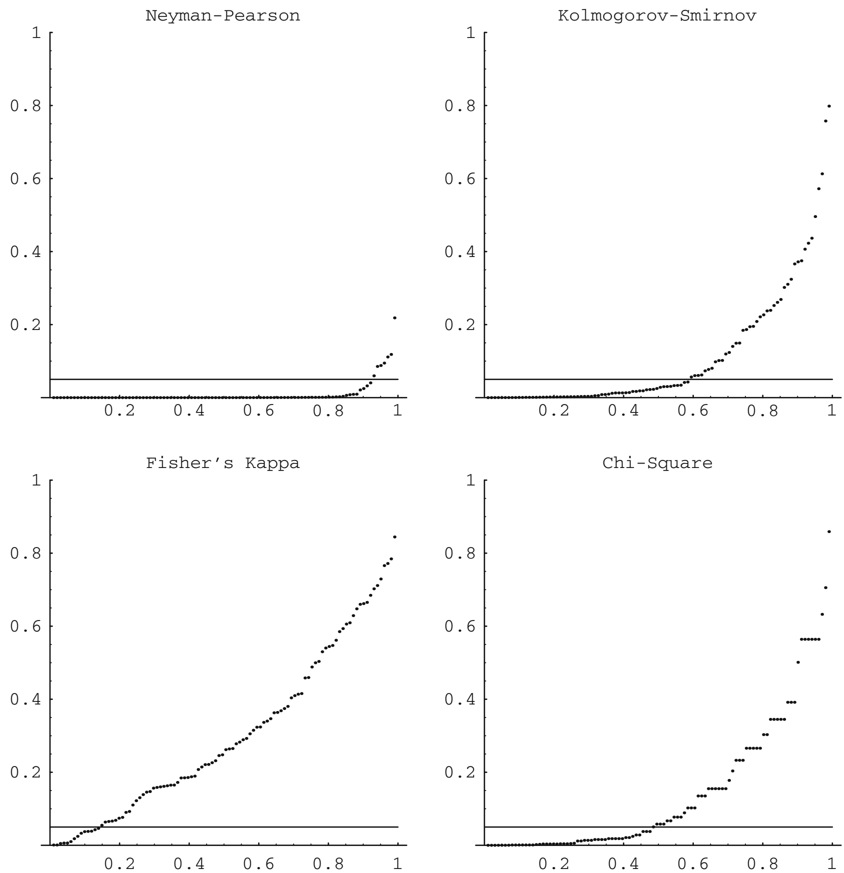

Fig. 3 Quantile plots of 100 values of p_{NP}, p_{KS}, p_κ and p_{χ^2}, with the underlying Marshall–Olkin df with $\lambda = 0.5$ and 25 exceedances over the threshold $c = -0.1$

$c = c(n) \uparrow 0$, the NPT controls the type I error, i.e., the corresponding p value is asymptotically uniformly distributed on $(0, 1)$

$$p_{\mathrm{NPT}} \to_{\mathcal{D}} U(0, 1)$$

in the case of tail independence. By $U(0, 1)$ we denote the uniform distribution on $(0, 1)$ and by $\to_{\mathcal{D}}$ the convergence in distribution as the sample size n tends to infinity. By $N(\mu, \sigma^2)$ we denote the normal distribution on the real line with mean μ and variance σ^2.

Lemma 3.1 *Suppose that* $\chi = 0$. *If* $c = c(n)$ *satisfies* $nc^2 \to_{n\to\infty} \infty$ *and* $nc^4 \to_{n\to\infty} \lambda \geq 0$, *then we obtain*

$$K(n)^{-1/2} \sum_{i \leq K(n)} (2\log(V_i) + 1) \to_{\mathcal{D}} N\left(\frac{2^{1/2}}{9}\lambda^{1/2}, 1\right).$$

The NPT controls, therefore, the type I error iff $\lambda = 0$, i.e., iff $nc^4 \to_{n\to\infty} 0$.

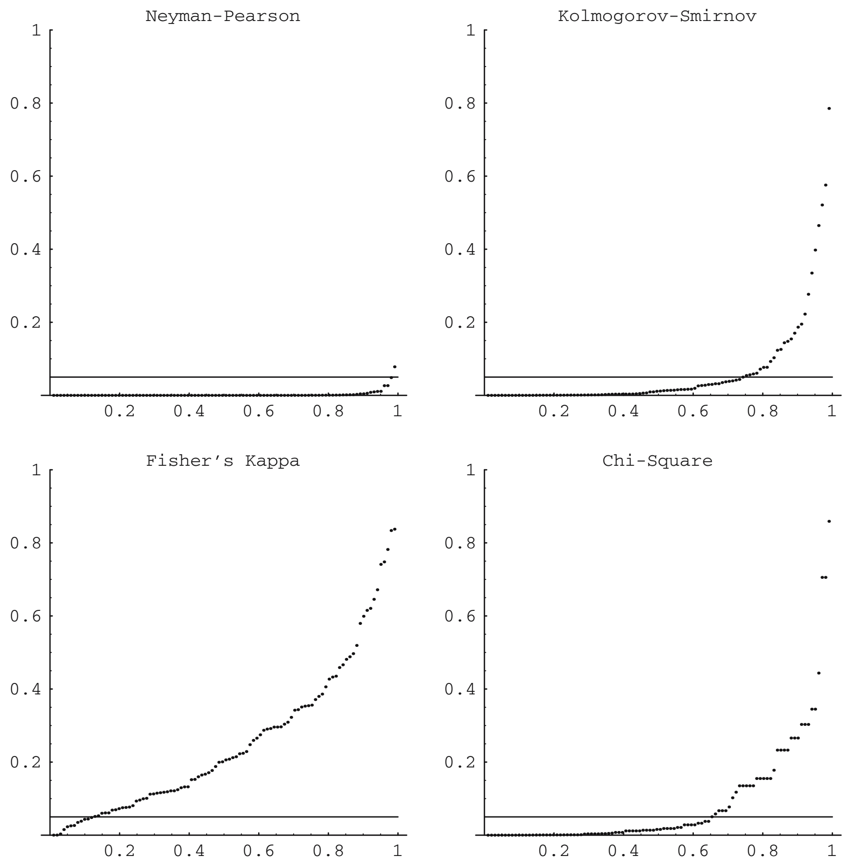

Fig. 4 Quantile plots of 100 values of p_{NP}, p_{KS}, p_κ and p_{χ^2}, with the underlying Gumbel df with $\lambda = 2$ and 25 exceedances over the threshold $c = -0.1$

Corollary 3.1 *We have under the conditions of the preceding lemma*

$$p_{\mathrm{NPT}} \to_{\mathcal{D}} U(0, 1) \quad \textit{iff } \lambda = 0.$$

The fact that p_{NPT} converges to 0 in probability, if the underlying continuously differentiable dependence function D is different from the constant function 1, will be established in a more general setup in Sect. 5; see the discussion after Corollary 5.2.

Proof The assertion in Lemma 3.1 will be a consequence of a series of facts, which we list in the following.

From Lemma 2.1 we obtain in the case $\chi = 0$

$$P(X + Y \geq ct) = 1 - \exp(ct)(1 - ct), \quad 0 \leq t \leq 1,$$

if c is close enough to 0. The density of V is, consequently, in the case $\chi = 0$

$$\begin{aligned} f_c(t) &:= \frac{\partial}{\partial t}\frac{P(X+Y \geq ct)}{P(X+Y \geq c)} \\ &= \frac{c^2 t \exp(ct)}{1-\exp(c)(1-c)} \\ &= 2t\left(1 + c\left(t - \frac{2}{3}\right) + O\left(c^2\right)\right), \end{aligned} \tag{3}$$

uniformly for $t \in [0, 1]$.

We have, moreover,

$$p_c := P(X+Y \geq c) = 1 - \exp(c)(1-c) = \frac{c^2}{2}(1 + O(c)). \tag{4}$$

Further, we have by (3) for $k \in \mathbb{N}$

$$\begin{aligned} E((2\log(V))^k) &= 2^k \int_0^1 \log^k(t) f_c(t)\,\mathrm{d}t \\ &= 2^k \int_0^1 \log^k(t) 2t \left(1 + c\left(t - \frac{2}{3}\right) + O\left(c^2\right)\right)\,\mathrm{d}t \\ &= 2^{k+1}\int_0^1 \log^k(t) t\,\mathrm{d}t + c2^{k+1}\int_0^1 \log^k(t) t\left(t - \frac{2}{3}\right)\mathrm{d}t + O\left(c^2\right) \\ &= (-1)^k k!\left(1 + c\left(\left(\frac{2}{3}\right)^{k+1} - \frac{2}{3}\right)\right) + O\left(c^2\right). \end{aligned} \tag{5}$$

In particular we obtain that

$$\begin{aligned} E(2\log(V)) &= -1 + \frac{2}{9}c + O\left(c^2\right), \\ \mathrm{Var}(2\log(V)) &= 1 - \frac{8}{27}c + O\left(c^2\right), \end{aligned}$$

and the third moment of $2\log(V)$ is uniformly bounded in c.

From the fact that $K(n)/(np_c) \to_{n\to\infty} 1$ in probability and, that thus, $K(n)/(nc^2/2) \to_{n\to\infty} 1$ in probability, the independence of $K(n)$ and $V_1, V_2, \ldots$ and the Berry–Esseen theorem one concludes

$$
\begin{aligned}
&K(n)^{-1/2} \sum_{i \le K(n)} (2\log(V_i) + 1) \\
&\quad = K(n)^{-1/2} \sum_{i \le K(n)} (2\log(V_i) - E(2\log(V))) + K(n)^{1/2}(E(2\log(V)) + 1) \\
&\quad = K(n)^{-1/2} \sum_{i \le K(n)} (2\log(V_i) - E(2\log(V))) + K(n)^{1/2}\left(\frac{2}{9}c + O\left(c^2\right)\right) \\
&\quad \to_{\mathcal{D}} N\left(\lambda^{1/2}\frac{2^{1/2}}{9}, 1\right),
\end{aligned}
$$

which completes the proof of Lemma 3.1.

4 Asymptotic power of the NPT

In this section we investigate the asymptotic power of the NPT. To this end, we consider a triangular array $(X_i^{(n)}, Y_i^{(n)})$, $i \le n$, $n \in \mathbb{N}$, of iid rvs in each row, such that $(X_i^{(n)}, Y_i^{(n)})$ follows an EV with a dependence function D_n defined below, which converges to 1 as n increases. We then compute the limiting distribution, as n increases, of the p value p_{NPT}, evaluated for these rvs in the n-th row.

The convex combination of two dependence functions is again a dependence function. Choose, therefore, an arbitrary dependence function D different from the constant 1 and consider

$$
D_n(z) := 1 - \vartheta_n + \vartheta_n D(z) = 1 - \vartheta_n(1 - D(z))
$$

with

$$
\vartheta_n := \vartheta n^{-1/2},
$$

where $\vartheta \ge 0$ is an arbitrary number. Hence, D_n is a convex combination of D and the constant 1, if n is large enough such that $\vartheta_n \le 1$.

For the corresponding tail dependence parameter we obtain

$$
\chi_n = 2(1 - D_n(1/2)) = \vartheta_n 2(1 - D(1/2)) = \vartheta_n \chi,
$$

where $\chi = 2(1 - D(1/2))$ is the tail dependence parameter corresponding to D.

In what follows we will compute the asymptotic power of the NPT. Precisely, we will compute the asymptotic distribution of the approximate p value

$$
p_{\mathrm{NPT}} = \Phi\left(K(n)^{-1/2} \sum_{i \le K(n)} \left(2\log(V_i^{(n)}) + 1\right)\right).
$$

Theorem 4.1 *If $c = c(n)$ satisfies $nc^2 \to_{n\to\infty} \infty$ and $nc^4 \to_{n\to\infty} 0$, then we obtain*

$$
p_{\mathrm{NPT}} \to_{\mathcal{D}} \Phi\left(\xi - \vartheta 2^{1/2} \int_0^1 1 - D(z)\,\mathrm{d}z\right),
$$

where ξ is a standard normal distributed rv.

Proof $V^{(n)}$ has by Lemma 2.1 the density

$$\begin{aligned}
g_{\vartheta_n}(t) &= \frac{\partial}{\partial t} P\left(V^{(n)} \le t\right) \\
&= \frac{1}{p_c(n)} \frac{\partial}{\partial t} P\left(X^{(n)} + Y^{(n)} \ge tc\right) \\
&= \frac{1}{p_c(n)} \left(-c\exp(ct) + c\int_0^1 \exp\left(ctD_n(z)\right)\left(1 + tcD_n(z)\right) u_n(z)\,\mathrm{d}z\right) \\
&=: \frac{1}{p_c(n)} \tilde{g}_{\vartheta_n}(t), \quad 0 \le t \le 1,
\end{aligned}$$

where

$$u_n(z) := D_n(z) + D_n'(z)(1-z) = D_n(z) + \vartheta_n D'(z)(1-z), \quad 0 \le z \le 1,$$

and

$$\begin{aligned}
p_c(n) &= P\left(X_i^{(n)} + Y_i^{(n)} \ge c\right) \\
&= 1 - \exp(c) + c\int_0^1 \exp\left(cD_n(z)\right)\left(D_n(z) + D_n'(z)(1-z)\right)\,\mathrm{d}z.
\end{aligned}$$

Now consider

$$\tilde{f}_c(t) := p_c f_c(t) = c^2 t\exp(ct), \quad 0 \le t \le 1,$$

where p_c, f_c are defined in (3), (4). Then we have uniformly for $t \in [0, 1]$

$$\frac{\tilde{g}_{\vartheta_n}(t)}{\tilde{f}_c(t)} - 1 = \frac{\vartheta_n}{tc}\left(-2\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right). \tag{6}$$

This can be observed as follows. Note that $\int_0^1 u_n(z)\,\mathrm{d}z = 2\int_0^1 D_n(z)\,\mathrm{d}z - 1$, $D_n(z) - 1 = -\vartheta_n(1 - D(z))$ and that $D_n'(z) = \vartheta_n D'(z)$:

$$\begin{aligned}
&\frac{\tilde{g}_{\vartheta_n}(t)}{\tilde{f}_c(t)} - 1 \\
&= \frac{-c\exp(ct) + c\int_0^1 \exp\left(ctD_n(z)\right)\left(1 + ctD_n(z)\right) u_n(z)\,\mathrm{d}z}{c^2 t\exp(ct)} - 1 \\
&= \frac{1}{ct}\left(-1 + \int_0^1 \exp\left(ct\left(D_n(z) - 1\right)\right)\left(1 + ctD_n(z)\right) u_n(z)\,\mathrm{d}z - ct\right)
\end{aligned}$$

$$= \frac{1}{ct}\left(-1+\int_0^1 (1+ct\,(D_n(z)-1))\,(1+ct D_n(z))\,u_n(z)\,\mathrm{d}z \right.$$

$$\left. - ct + O\left(c^2\vartheta_n^2\right)\right)$$

$$= \frac{1}{ct}\left(-1+\int_0^1 u_n(z)+ct\,(D_n(z)-1)\,u_n(z)+ct D_n(z)u_n(z)\,\mathrm{d}z \right.$$

$$\left. - ct + O\left(c^2\vartheta_n\right)\right)$$

$$= \frac{1}{ct}\left(2\int_0^1 D_n(z)-1\,\mathrm{d}z + O\,(c\vartheta_n)\right)$$

$$= \frac{\vartheta_n}{ct}\left(-2\int_0^1 1-D(z)\,\mathrm{d}z + O(c)\right),$$

which is (6).

We have, moreover,

$$\frac{p_c(n)}{p_c}-1 = \frac{\vartheta_n}{c}\left(-4\int_0^1 1-D(z)\,\mathrm{d}z + O(c)\right). \tag{7}$$

This follows by repeating the arguments in the derivation of (6):

$$\frac{p_c(n)}{p_c}-1$$

$$= \frac{1-\exp(c)+c\int_0^1 \exp\left(cD_n(z)\right)u_n(z)\,\mathrm{d}z}{1-\exp(c)(1-c)}-1$$

$$= \frac{c\int_0^1 \exp\left(c\,(D_n(z)-1)\right)u_n(z)\,\mathrm{d}z - c}{\exp(-c)-(1-c)}$$

$$= \frac{c\int_0^1 u_n(z)-1\,\mathrm{d}z + c^2\int_0^1 (D_n(z)-1)\,u_n(z)\,\mathrm{d}z + O\left(c^3\vartheta_n^2\right)}{\frac{c^2}{2}(1+O(c))}$$

$$= \frac{2c\int_0^1 D_n(z)-1\,\mathrm{d}z + c^2\int_0^1 (D_n(z)-1)\,u_n(z)\,\mathrm{d}z + O\left(c^3\vartheta_n^2\right)}{\frac{c^2}{2}(1+O(c))}$$

$$= \frac{\vartheta_n}{c}\,\frac{-4\int_0^1 1-D(z)\,\mathrm{d}z + O(c)}{1+O(c)}$$

which is (7).

From (3), (5) and (7) we obtain

$$\begin{aligned}
&E\left(2^k \log^k\left(V^{(n)}\right)\right)\\
&= \int_0^1 2^k \log^k(t) g_{\vartheta_n}(t)\,\mathrm{d}t\\
&= \frac{p_c}{p_c(n)} \int_0^1 2^k \log^k(t) \left(1 + \left(\frac{\tilde{g}_{\vartheta_n}(t)}{\tilde{f}_c(t)} - 1\right)\right) f_c(t)\,\mathrm{d}t\\
&= \frac{1}{1+\left(\frac{p_c(n)}{p_c} - 1\right)} \int_0^2 2^k \log^k(t) \left(1 - \frac{\vartheta_n}{ct}\left(2\int_0^1 1 - D(z)\,\mathrm{d}z\right.\right.\\
&\qquad \left.\left. + O(c)\right)\right) f_c(t)\,\mathrm{d}t\\
&= \frac{c}{c - \vartheta_n\left(4\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)} \int_0^1 2^k \log^k(t) f_c(t)\,\mathrm{d}t\\
&\quad - \frac{\vartheta_n\left(2\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)}{c - \vartheta_n\left(4\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)} \int_0^1 2^{k+1} \log^k(t)(1 + O(c))\,\mathrm{d}t\\
&= \frac{c}{c - \vartheta_n\left(4\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)} (-1)^k k! \left(1 + c\left(\left(\frac{2}{3}\right)^{k+1} - \frac{2}{3}\right) + O\left(c^2\right)\right)\\
&\quad - \frac{\vartheta_n\left(2\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)}{c - \vartheta_n\left(4\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)} (-1)^k k!(1 + O(c)).
\end{aligned}$$

Hence, we have in particular

$$E\left(2\log\left(V^{(n)}\right)\right) + 1 = \frac{-\vartheta_n\left(2\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)}{c - \vartheta_n\left(4\int_0^1 1 - D(z)\,\mathrm{d}z + O(c)\right)} + O(c)$$

and

$$\mathrm{Var}\left(2\log\left(V^{(n)}\right)\right) \to_{n\to\infty} 1;$$

note that $\vartheta_n/c \to_{n\to\infty} 0$.

From the fact that $K(n)/(np_c(n)) \to_{n\to\infty} 1$ in probability, where $p_c(n)/\left(c^2/2\right) \to_{n\to\infty} 1$ by (7), the independence of $K(n)$ and $V_1^{(n)}, V_2^{(n)}, \ldots$ and the

Berry–Esseen theorem we now obtain

$$
\begin{aligned}
&K(n)^{-1/2} \sum_{i \le K(n)} \left(2 \log\left(V_i^{(n)}\right) + 1\right) \\
&= K(n)^{-1/2} \sum_{i \le K(n)} \left(2 \log\left(V_i^{(n)}\right) - E\left(2 \log\left(V^{(n)}\right)\right)\right) \\
&\quad + K(n)^{1/2} \left(E\left(2 \log\left(V^{(n)}\right)\right) + 1\right) \\
&\to_{\mathcal{D}} N\left(-\vartheta 2^{1/2} \int_0^1 1 - D(z)\, dz, 1\right).
\end{aligned}
$$

This completes the proof of Theorem 4.1.

5 Differentiable spectral neighborhood of an EV

In this section we extend the results of Sects. 2 and 3 to a rv (X, Y), whose df H is in a certain neighborhood of an EV G with the dependence function D. A typical example is Mardia (1970) df

$$M(x, y) = \frac{1}{\exp(-x) + \exp(-y) - 1}, \quad x,\ y \le 0.$$

The df M has reverse exponential margins, but it is not max–stable. Precisely, it is tail equivalent with $G(x, y) = \exp(x + y)$, i.e.,

$$\lim_{x+y \uparrow 0} \frac{1 - M(x, y)}{1 - \exp(x + y)} = 1$$

and its margins are, therefore, tail independent:

$$\lim_{c \uparrow 0} P(Y > c | X > c) = 0.$$

The neighborhood of an EV $G(x, y) = \exp\left((x + y) D(x/(x + y))\right)$ will be defined in terms of the *spectral decomposition* of an arbitrary df $H(x, y), x,\ y \le 0$: Consider for $z \in [0, 1]$ and $c \le 0$

$$H_z(c) := H(c(z, 1 - z)) = P\left(\max\left(\frac{X}{z}, \frac{Y}{1 - z}\right) \le c\right),$$

where the rv (X, Y) has the df H. With z kept fixed, $H_z(\cdot)$ is a univariate df on $(-\infty, 0]$, and H_1, H_0 are the marginal dfs of X and Y. The df H is obviously uniquely determined by the set of univariate dfs $\{H_z(\cdot) :\ z \in [0, 1]\}$. This is the spectral decomposition of H, which has turned out to be quite a helpful tool in the multivariate extreme value theory; see Falk and Reiss (2005).

We require in the following that the partial derivatives

$$h_z(c) := \frac{\partial}{\partial c} H_z(c) \quad \text{and} \quad g_z(c) := \frac{\partial}{\partial z} H_z(c) \tag{8}$$

of $H_z(c)$ exist for c close to 0 and any $z \in [0, 1]$ and that they are continuous.

We require, moreover, that $h_z(c)$ satisfies the expansion

$$h_z(c) = a(z) + cA(z) + O(c^2) \quad \text{as } c \uparrow 0 \tag{9}$$

uniformly for $z \in [0, 1]$, where $a : [0, 1] \to [0, \infty)$ satisfies $a(0) = a(1) = 1$ and $A : [0, 1] \to \mathbb{R}$ is an integrable function.

From Theorem 3.1 in Falk and Reiss (2005) we obtain that $a(z)$ is actually a Pickands dependence function, i.e., $a(z) = D(z)$, $z \in [0, 1]$, and, thus, we could equivalently replace the function $a(z)$ in condition (9) by $D(z)$. From Theorem 3.1 in Falk and Reiss (2005) we obtain, moreover, that $H(x, y)$ is in the bivariate domain of attraction of the EV G with the dependence function $D = a$, which means, that

$$\lim_{n\to\infty} H^n\left(\left(\frac{x}{n}, \frac{y}{n}\right)\right) = \exp\left((x + y)D\left(\frac{x}{x + y}\right)\right), \quad x, y \le 0.$$

Finally, we have

$$\lim_{c\uparrow 0} P(Y > c|X > c) = 2(1 - D(1/2)),$$

and, thus, we have the tail independence of X and Y if and only if $D(z) = 1$, $z \in [0, 1]$.

Note that the condition (9) immediately implies that H_z satisfies the von Mises condition

$$\lim_{c\uparrow 0} \frac{-ch_z(c)}{1 - H_z(c)} = 1,$$

which, in turn, implies that H_z is in the (univariate) domain of attraction of $\exp(x)$, $x \le 0$, for any $z \in [0, 1]$; see, e.g. Resnick ((1987), Proposition 1.16).

A df H is now said to be in the *differentiable spectral neighborhood* of the EV G with a dependence function D if its spectral decomposition satisfies conditions (8) and (9) with $a(z) = D(z)$.

Note that G itself has the spectral decomposition

$$G_z(c) = \exp(cD(z)), \quad c \le 0,\ z \in [0, 1],$$

and hence it satisfies conditions (8) and (9), if $D'(z)$ is continuous, with

$$h_z(c) = \exp(cD(z))D(z) = D(z) + cD(z)^2 + O(c^2)$$

and

$$g_z(c) = \exp(cD(z))cD'(z).$$

Mardia's df satisfies, for instance, conditions (8) and (9) with $a(z) = 1$ and $A(z) = 2 - z^2 - (1 - z)^2$.

The following result extends Lemma 2.1 to a df H, which satisfies condition (8). Recall that $H_1(c) = H(c(1, 0))$, $c \le 0$, is the first marginal distribution of H.

Lemma 5.1 *Suppose that the df* $\mathrm{H}(x, y)$, $x, y \le 0$, *of* (X, Y) *satisfies the condition (8). Then we have for c close to 0*

$$P(X + Y \le c) = H_1(c) - \int_0^1 ch_z(c) + g_z(c)(1 - z)\,\mathrm{d}z.$$

Proof Repeating the arguments in the proof of Lemma 2.1, we obtain for $0 > u > c$

$$\begin{aligned} &P(X + Y \le c | X = u) \\ &= \frac{1}{h_1(u)} \lim_{\varepsilon \downarrow 0} \frac{H(u+\varepsilon, c-u) - H(u, c-u)}{\varepsilon} \\ &= \frac{1}{h_1(u)} \lim_{\varepsilon \downarrow 0} \frac{H_{\frac{u+\varepsilon}{c+\varepsilon}}(c+\varepsilon) - H_{\frac{u}{c}}(c)}{\varepsilon} \\ &= \frac{1}{h_1(u)} \left(h_{\frac{u}{c}}(c) + g_{\frac{u}{c}}(c) \frac{c-u}{c^2} \right) \end{aligned}$$

by making use of Taylor's formula and the continuity of the partial derivatives of $H_z(c)$.

Since $P(X + Y \le c | X = u) = 1$ if $u \le c$, we obtain by integration and substitution, for c close to 0

$$\begin{aligned} &P(X + Y \le c) \\ &= H_1(c) + \int_c^0 h_{\frac{u}{c}}(c) + g_{\frac{u}{c}}(c) \frac{c-u}{c^2}\,\mathrm{d}u \\ &= H_1(c) - \int_0^1 ch_z(c) + g_z(c)(1 - z)\,\mathrm{d}z. \end{aligned}$$

Corollary 5.1 1. *We obtain for* (X, Y) *with the df* $H(x, y)$, $x, y \le 0$, *in a differentiable spectral neighborhood of* $\exp(x + y)$

$$F_c(t) = P(X + Y > ct | X + Y > c) = t^2\,(1 + O(c))$$

as $c \uparrow 0$, *uniformly for* $t \in [0, 1]$, *provided that* $3 \int A(z)\,dz > A(0) + A(1)$. *This condition is actually rather weak, since we have in general the inequality*

$$A(z) \ge A(1)z^2 + A(0)(1 - z)^2, \quad z \in [0, 1],$$

and, hence, $3 \int_0^1 A(z)\,dz \ge A(0) + A(1)$ *anyway.*

2. *If the df* H *of* (X, Y) *satisfies conditions (8) and (9) with* $a(z) = D(z)$, *which is not the constant function 1, then we obtain*

$$F_c(t) = P(X + Y > ct | X + Y > c) = t\,(1 + O(c))$$

as $c \uparrow 0$, *uniformly for* $t \in [0, 1]$.

Proof From Lemma 5.1 we obtain for c close to 0

$$P(X+Y>ct|X+Y>c) \tag{10}$$
$$= \frac{\int_{ct}^{0} h_1(x)\,\mathrm{d}x + \int_0^1 cth_z(ct) + g_z(ct)(1-z)\,\mathrm{d}z}{\int_{c}^{0} h_1(x)\,\mathrm{d}x + \int_0^1 ch_z(c) + g_z(c)(1-z)\,\mathrm{d}z}$$
$$= \frac{I}{II},$$

where

$$I := \int_{ct}^{0} 1 + xA(1) + O\left(x^2\right)\,\mathrm{d}x + \int_0^1 ct\left(D(z) + ctA(z) + O\left((ct)^2\right)\right)\,\mathrm{d}z$$
$$+ \int_0^1 g_z(ct)(1-z)\,\mathrm{d}z,$$
$$II := \int_{c}^{0} 1 + xA(1) + O\left(x^2\right)\,\mathrm{d}x + \int_0^1 c\left(D(z) + cA(z) + O\left(c^2\right)\right)\,\mathrm{d}z$$
$$+ \int_0^1 g_z(c)(1-z)\,\mathrm{d}z.$$

Using partial integration we obtain

$$\int_0^1 g_z(c)(1-z)\,\mathrm{d}z$$
$$= H_z(c)(1-z)\Big|_0^1 + \int_0^1 H_z(c)\,\mathrm{d}z$$
$$= -H_0(c) + \int_0^1 H_z(c)\,\mathrm{d}z$$
$$= 1 - H_0(c) - \int_0^1 1 - H_z(c)\,\mathrm{d}z$$
$$= \int_c^0 h_0(y)\,\mathrm{d}y - \int_0^1 \int_c^0 h_z(u)\,\mathrm{d}u\,\mathrm{d}z$$
$$= \int_c^0 1 + yA(0) + O\left(y^2\right)\,\mathrm{d}y - \int_0^1 \int_c^0 D(z) + A(z)u + O\left(u^2\right)\,\mathrm{d}u\,\mathrm{d}z$$
$$= c\left(\int_0^1 D(z)\,\mathrm{d}z - 1\right) + \frac{c^2}{2}\left(\int_0^1 A(z)\,\mathrm{d}z - A(0)\right) + O\left(c^3\right)$$

and

$$\int_0^1 g_z(ct)(1-z)\,\mathrm{d}z$$

$$= ct\left(\int_0^1 D(z)\,\mathrm{d}z - 1\right) + \frac{(ct)^2}{2}\left(\int_0^1 A(z)\,\mathrm{d}z - A(0)\right) + O\left((ct)^3\right).$$

Substituting the above two expansions in Eq. (10), we obtain

$$P(X+Y>ct|X+Y>c)$$
$$= \frac{2ct\left(\int_0^1 D(z)\,\mathrm{d}z - 1\right) + \frac{(ct)^2}{2}\left(3\int_0^1 A(z)\,\mathrm{d}z - A(1) - A(0)\right) + O((ct)^3)}{2c\left(\int_0^1 D(z)\,\mathrm{d}z - 1\right) + \frac{c^2}{2}\left(3\int_0^1 A(z)\,\mathrm{d}z - A(1) - A(0)\right) + O(c^3)}.$$

Finally, one obtains from elementary computations for arbitrary $t \in (0,\infty)$, if $D(z) = 1$, $z \in [0,1]$

$$0 \le \lim_{c\uparrow 0} \frac{P(Y>c|X>tc)}{|c|} = \frac{A\left(\frac{t}{t+1}\right)(t+1)^2 - A(0) - A(1)t^2}{2t}$$

and, hence,

$$A\left(\frac{t}{t+1}\right) \ge \frac{A(0)}{(t+1)^2} + A(1)\left(\frac{t}{t+1}\right)^2.$$

Considering $z = t/(t+1)$, we obtain

$$A(z) \ge A(0)(1-z)^2 + A(1)z^2, \quad z \in [0,1],$$

which completes the proof. □

To ensure that df F_c has a density with respect to the Lebesgue measure, we refine condition (9) on h_c as follows. We require the expansion

$$h_z(c) = a(z) + cA(z) + r(z,c), \tag{11}$$

uniformly for $z \in [0,1]$ and c close to 0, where $a : [0,1] \to [0,\infty]$ satisfies $a(0) = a(1) = 1$, $A : [0,1] \to \mathbb{R}$ is an integrable function and $r(z,c) = O(c^2)$ uniformly for $z \in [0,1]$. We require, moreover, that $B_z(c) := (\partial/\partial c)r(z,c)$ exists for $z \in [0,1]$ and that $B_z(c) = O(c)$ uniformly for $z \in [0,1]$.

Lemma 5.2 *Suppose that conditions* (8) *and* (11) *are satisfied. Then*

$$f_c(t) = \frac{\partial}{\partial t}F_c(t) = \frac{\partial}{\partial t}P(X+Y \ge ct|X+Y \ge c)$$

exists for $t \in (0,1)$ *and c close to 0. We have, moreover,*

$$f_c(t) = \begin{cases} 1 + O(c), & \text{if } a(z) \text{ in (11) is not the constant function } 1 \\ 2t(1+O(c)), & \text{if } a(z) = 1 \text{ and } 3\int_0^1 A(z)\,dz > A(0) + A(1). \end{cases}$$

Proof By Lemma 5.1 we have for $t \in (0,1)$ and c close to 0

$$P(X+Y \geq ct) = 1 - H_1(ct) + \int_0^1 cth_z(ct) + g_z(ct)(1-z)\,\mathrm{d}z.$$

From the proof of Corollary 5.1 we obtain

$$\int_0^1 g_z(ct)(1-z)\,\mathrm{d}z = 1 - H_0(ct) - \int_0^1 1 - H_z(ct)\,\mathrm{d}z.$$

As a consequence we obtain for $t \in (0,1)$ and c, ε close to 0

$$\begin{aligned} &F_c(t+\varepsilon) - F_c(t) \\ &= \frac{P(X+Y \geq c(t+\varepsilon)) - P(X+Y \geq ct)}{P(X+Y \geq c)} \\ &=: \frac{I}{II}, \end{aligned}$$

where

$$II = 2c\int_0^1 a(z) - 1\,\mathrm{d}z + \frac{c^2}{2}\int_0^1 3A(z) - A(0) - A(1)\,\mathrm{d}z + O\left(c^3\right) \tag{12}$$

and

$$\begin{aligned} I &= H_1(ct) - H_1(c(t+\varepsilon)) + c(t+\varepsilon)\int_0^1 h_z(c(t+\varepsilon))\,\mathrm{d}z - ct\int_0^1 h_z(ct)\,\mathrm{d}z \\ &+ H_0(ct) - H_0(c(t+\varepsilon)) + \int_0^1 H_z(c(t+\varepsilon)) - H_z(ct)\,\mathrm{d}z \\ &= H_1(ct) - H_1(c(t+\varepsilon)) + c\varepsilon\int_0^1 h_z(c(t+\varepsilon))\,\mathrm{d}z + \varepsilon c^2 t\int_0^1 A(z)\,\mathrm{d}z \\ &+ ct\int_0^1 r(z, c(t+\varepsilon)) - r(z, ct)\,\mathrm{d}z \\ &+ H_0(ct) - H_0(c(t+\varepsilon)) + \int_0^1 H_z(c(t+\varepsilon)) - H_z(ct)\,\mathrm{d}z. \end{aligned}$$

Hence, we obtain

$$\lim_{\varepsilon\to 0}\frac{I}{\varepsilon} = -h_1(ct)c + c\int_0^1 h_z(ct)\,\mathrm{d}z + c^2 t\int_0^1 A(z)\,\mathrm{d}z + c^2 t\int_0^1 B_z(ct)\,\mathrm{d}z$$
$$-h_0(ct)c + c\int_0^1 h_z(ct)\,\mathrm{d}z$$
$$= -2c - A(1)c^2 t + 2c\int_0^1 a(z) + ctA(z)\,\mathrm{d}z + c^2 t\int_0^1 A(z)\,\mathrm{d}z$$
$$-A(0)c^2 t + O\left(c^3 t^2\right)$$
$$= 2c\int_0^1 a(z) - 1\,\mathrm{d}z + c^2 t\int_0^1 3A(z) - A(0) - A(1)\,\mathrm{d}z + O\left(c^3 t^2\right).$$

Altogether we get

$$\lim_{\varepsilon\to 0}\frac{F_c(t+\varepsilon) - F_c(t)}{\varepsilon}$$
$$= \frac{2c\int_0^1 a(z) - 1\,\mathrm{d}z + c^2 t\int_0^1 3A(z) - A(0) - A(1)\,\mathrm{d}z + O\left(c^3 t^2\right)}{2c\int_0^1 a(z) - 1\,\mathrm{d}z + \frac{c^2}{2}\int_0^1 3A(z) - A(0) - A(1)\,\mathrm{d}z + O\left(c^3\right)},$$

which implies the assertion in Lemma 5.2. □

Suppose now that the df H of (X, Y) satisfies conditions (8) and (11), and denote by V a rv with df F_c. Then we obtain for c close to 0 and $k \in \mathbb{N}$

$$E\left((2\log(V))^k\right) = 2^k\int_0^1 \log^k(t) f_c(t)\,\mathrm{d}t$$
$$= \begin{cases} 2^k(-1)^k k! + O(c), & \text{if } f_c(t) = 1 + O(c) \\ (-1)^k k! + O(c), & \text{if } f_c(t) = 2t(1 + O(c)). \end{cases}$$

The following result shows that the NPT controls the type I error if the underlying df is in a differentiable spectral neighborhood of $G(x, y) = \exp((x + y)D(x/(x+y)))$ with $D = 1$. If $D \neq 1$, then the p value converges to 0 in probability as the sample size n increases. Note that $P(X + Y \geq c)$ is by (12) of order c^2, if $a(z) = 1$ in condition (11), and of order c, if $a(z)$ is not the constant function 1. This explains the different conditions on the sequence $c = c(n)$ of thresholds in the two parts of the following result.

Corollary 5.2 *Suppose that the df H of (X, Y) satisfies conditions (8) and (11).*

1. *If $c = c(n)$ satisfies $nc^2 \to_{n\to\infty} \infty$ and $nc^4 \to_{n\to\infty} 0$, then we obtain*

$$p_{\mathrm{NPT}} = \Phi\left(K(n)^{-1/2}\sum_{i\leq K(n)}(2\log(V_i) + 1)\right) \to_{\mathcal{D}} U(0, 1),$$

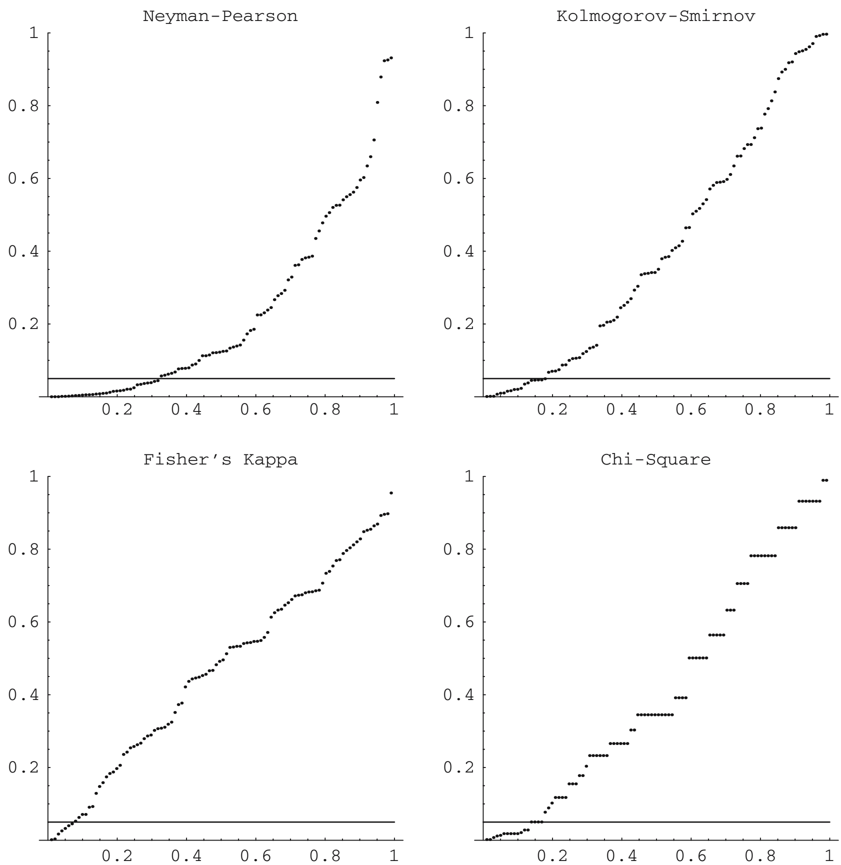

Fig. 5 Quantile plots of 100 values of p_{NPT}, p_{KS}, p_κ and p_{χ^2}, with the underlying Mardia df and 25 exceedances over the threshold $c = -0.5$

if $a(z) = 1$ *in condition (11) and* $3\int_0^1 A(z)\,dz > A(0) + A(1)$.

2. *If* $a(z)$ *is not the constant function 1, then we have*

$$p_{\mathrm{NPT}} \to_{n\to\infty} 0$$

in probability, if $c = c(n)$ *satisfies* $nc \to_{n\to\infty} \infty$ *and* $nc^3 \to_{n\to\infty} 0$.

Note that an EV G with the dependence function D satisfies condition (11) with $a(z) = D(z)$, $A(z) = D^2(z)$ and

$$r(z, c) = D(z)\left(\exp(cD(z)) - 1 - cD(z)\right).$$

The preceding result implies, therefore, in particular that the p value p_{NPT} converges to 0 as $n \to \infty$, if the underlying df G is an EV with a continuously differentiable dependence function D, which is different from the constant function 1. This result supplements Corollary 3.1.

Figures 5 and 6 display quantile plots of 100 values of p_{NPT}, p_{χ^2}, p_{KS} and p_κ, which are based on 25 exceedances C_i/c, $1 \le i \le 25$, over the threshold $c = -0.5$

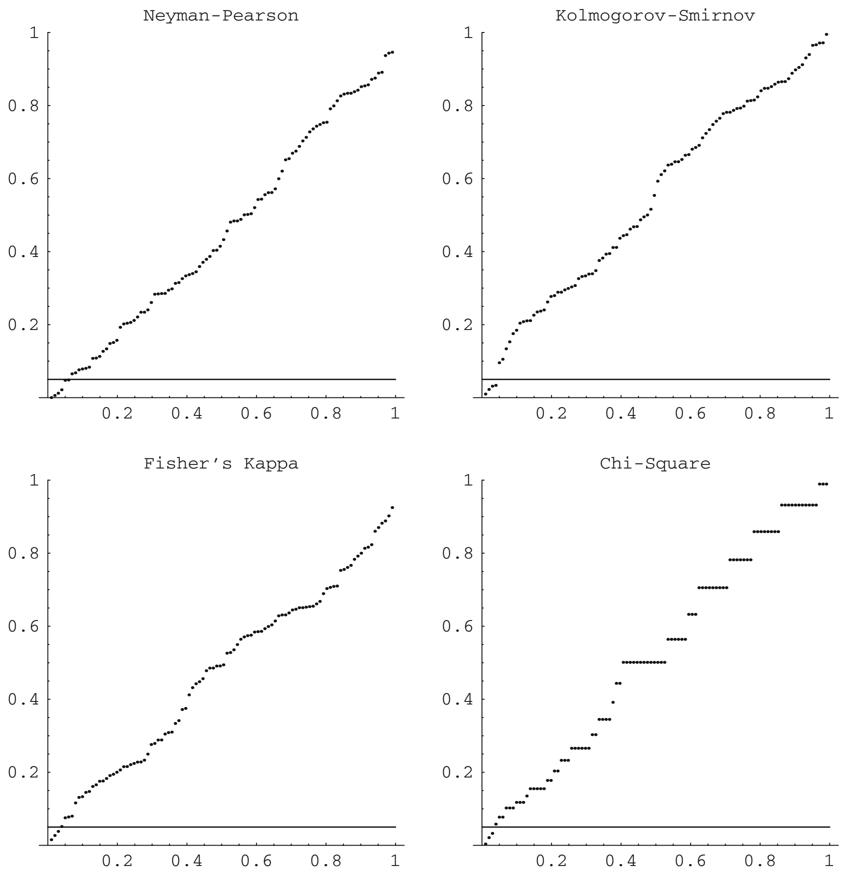

Fig. 6 Quantile plots of 100 values of p_{NPT}, p_{KS}, p_κ and p_{χ^2}, with the underlying Mardia df and 25 exceedances over the threshold $c = -0.1$

in Fig. 5 and over the threshold $c = -0.1$ in Fig. 6, with the underlying Mardia distribution. Recall that Mardia's distribution has independent tails. The simulations show that the threshold $c = -0.5$ seriously affects the type I error rate of the NPT as well as of the Kolmogorov–Smirnov and the chi-square test. This is due to the fact that the threshold $c = -0.5$ is not close enough to 0 to give approximately the independence of the tails. For the threshold $c = -0.1$, the simulations indicate that all four tests for the tail independence of the margins have the correct type I error rate. These example illustrate the practical significance of the choice of the threshold c.

6 Generalized Pareto distributions

Along with a bivariate EV G comes a *generalized Pareto df* (GP)

$$W(x, y) := 1 + \log(G(x, y)) = 1 + (x + y)D\left(\frac{x}{x + y}\right),$$

defined for all $x, y \le 0$ with $\log(G(x, y)) \ge -1$. Like in the univariate case, the class of GPs plays an important role for instance in multivariate peaks-over-threshold models (POT) or for the rate of convergence of multivariate maxima. The EV G is in particular in the *spectral δ–neighborhood* of W with $\delta = 1$:

$$1 - G_z(c) = (1 - W_z(c))\,(1 + O(c)),$$

uniformly for $z \in [0, 1]$ as $c \uparrow 0$, where $H_z(c) = H(c(z, 1 - z)),\ c \le 0,\ z \in [0, 1]$, denotes again the spectral decomposition of a df H defined on $(-\infty, 0]^2$. Note that each univariate margin of W is the uniform distribution on $(-1, 0)$.

Let the random variables U, V have df F_U and F_V and let

$$\chi(q) := P\left(V > F_V^{-1}(q) | U > F_U^{-1}(q)\right)$$

be the tail dependence parameter at the level $q \in (0, 1)$; see Reiss and Thomas ((2001), Eq. 2.57). We have

$$\chi(q) = \chi + O(1 - q)$$

if (U, V) follows a bivariate EV, and

$$\chi(q) = \chi = 2(1 - D(1/2)), \quad q \ge 1/2,$$

for bivariate GPs; see (9.24) and (10.8) in Reiss and Thomas (2001).

Tail independence $\chi = 0$ is, therefore, characterized for a GP $W(x, y) = 1 + (x + y)D(x/(x + y))$ by $\chi(q) = 0$ for large values of q or, equivalently, by $D(z) = 1$, $z \in [0, 1]$, and, hence, by $W(x, y) = 1 + x + y$. Note, however, that $W(x, y) = 1 + x + y$ is the df of $(U, -1 - U)$, where U is uniformly distributed on $(-1, 0)$, i.e., we have the tail independence $\chi = 0$ for a GP if and only if we have complete dependence $V = -1 - U$, which sounds a bit weird.

Assume now that the df $H(x, y)$ of (U, V) coincides, for x, y close to 0, with the general GP $W(x, y) = 1 + (x + y)D(x/(x + y))$. Repeating the arguments in the proof of Lemma 2.1 one obtains the following result.

Lemma 6.1 *We have for $c < 0$ close to 0*

$$P(U + V \le c) = 1 + 2c\left(1 - \int_0^1 D(u)\,\mathrm{d}u\right).$$

The preceding result shows that we have $P(U + V \ge c) = 0$ for c close to 0 if and only if $D(z) = 1$, i.e., if and only if U and V are tail independent. If they are not tail independent, then the conditional distribution of $(U + V)/c$, given that $U + V > c$, is for c close to 0 the uniform distribution on $[0, 1]$:

$$P(U + V > tc | U + V > c) = t, \quad t \in [0, 1].$$

Testing for the tail independence of U, V in the case of an upper GP tail is, therefore, equivalent to testing for $P(U + V \ge c) = 0$ for some $c < 0$.

7 Extension to a general dimension

In this section we extend the results of Sect. 2 to an EV in arbitrary dimension d. Let, therefore, in the sequel $X = (X_1, \dots, X_d)$ be a rv with values in $(-\infty, 0]^d$, whose df $H(x_1, \dots, x_d)$ coincides for $(x_1, \dots, x_d)$ close to 0 with an EV $G(x_1, \dots, x_d)$ with reverse exponential margins, i.e.,

$$G(0, \dots, 0, x_k, 0, \dots, 0) = \exp(x_k), \quad x_k \le 0, \ 1 \le k \le d,$$

and

$$G^n\left(\frac{x_1}{n}, \dots, \frac{x_d}{n}\right) = G(x_1, \dots, x_d).$$

Again, G can be represented as

$$G(x_1, \dots, x_d) = \exp\left(\left(\sum_{k \le d} x_k\right) D\left(\frac{x_1}{\sum_{k \le d} x_k}, \dots, \frac{x_{d-1}}{\sum_{k \le d} x_k}\right)\right),$$

where $D : \{(z_1, \dots, z_{d-1}) \in [0, 1]^{d-1} : \sum_{k \le d-1} z_k \le 1\} \to [1/d, 1]$ is the Pickands dependence function in d dimensions. It is continuous, convex, and the cases $D(z_1, \dots, z_{d-1}) = 1$ and $D(z_1, \dots, z_{d-1}) = \max(z_1, \dots, z_{d-1}, 1 - \sum_{k \le d-1} z_k)$ characterize the cases of independence and complete dependence of the margins; refer to Falk and Reiss (2005) for details.

Note that for $c < 0$ close to 0

$$P\left(\sum_{k \le d} X_k \le c\right) = \exp(c) \sum_{0 \le j \le d-1} \frac{(-c)^j}{j!}$$

if $X_1, \dots, X_d$ are independent in their upper tails, i.e., if $D(z_1, \dots, z_d)$ is the constant function 1. The following lemma is, therefore, an immediate consequence.

Lemma 7.1 *If $X_1, \dots, X_d$ are independent in their upper tails, then we have uniformly for $t \in [0, 1]$ as $c \uparrow 0$*

$$\begin{aligned} F_{c,d}(t) &:= P\left(\sum_{k \le d} X_k > ct \,\middle|\, \sum_{k \le d} X_k > c\right) \\ &= \frac{1 - \exp(ct) \sum_{0 \le j \le d-1} (-ct)^j / j!}{1 - \exp(c) \sum_{0 \le j \le d-1} (-c)^j / j!} \\ &= t^d (1 + O(c)). \end{aligned}$$

Suppose next that $X_1, \dots, X_d$ are not tail independent, i.e., suppose that the dependence function $D(z_1, \dots, z_{d-1})$ is not a constant. If D has continuous partial derivatives of order d, then there exists $c_0 < 0$ such that $\sum_{k \le d} X_k$ has a density

$$f(c) = \text{const} + O(c), \quad c_0 \le c < 0,$$

on $(c_0, 0)$, see Falk and Reiss (2005).

If we assume that const $\neq 0$, then we obtain

$$P\left(\sum_{k\le d} X_k > ct \,\Big|\, \sum_{k\le d} X_k > c\right) = t(1 + O(c))$$

uniformly for $t \in [0, 1]$ as $c \uparrow 0$.

As in Sect. 2 we can, therefore, test for the tail independence of $X_1, \dots, X_d$ by testing for the uniform distribution of

$$U_i := F_{c,d}(C_i/c) = \frac{1 - \exp(C_i)\sum_{0\le j\le d-1}(-C_i)^j/j!}{1 - \exp(c)\sum_{0\le j\le d-1}(-c)^j/j!}, \quad 1 \le i \le K(n),$$

where $C_i = \sum_{k\le d} X_k^{(i)}$, $i \le K(n)$, are those observations among a sample $(X_1^{(j)}, \dots, X_d^{(j)})$, $1 \le j \le n$, of independent copies of $(X_1, \dots, X_d)$, where $\sum_{k\le d} X_k^{(i)} > c$. The conclusions in Sect. 2 on the application of Fisher's κ, the Kolmogorov–Smirnov test and the chi-square goodness-of-fit test now carry over.

Conditional on $K(n) = m$, the optimal test suggested by the Neyman–Pearson lemma for testing $F_{(0)}(t) = t^d$ against $F_{(1)}(t) = t, 0 \le t \le 1$, based on $V_i = C_i/c$, $1 \le i \le m$, uses the loglikelihood ratio

$$\begin{aligned} T(V_1, \dots, V_m) &= \log\left(\prod_{i=1}^m \frac{1}{dV_i^{d-1}}\right) \\ &= -(d-1)\sum_{i=1}^m \log(V_i) - m\log(d). \end{aligned}$$

Note that $-d\log(V_i)$ has the df $1 - \exp(-x)$, $x \ge 0$, under $F_{(0)}$ and, hence, $(d/(d-1))(T(V_1, \dots, V_m) + m\log(d))$ has the df $1 - \exp(-x)\sum_{0\le j\le m-1} x^j/j!$ under $F_{(0)}$.

The p value of the optimal test derived from the Neyman–Pearson lemma is now

$$\begin{aligned} p_{NP} &= \exp\left(-\frac{d}{d-1}(T(V_1, \dots, V_m) + m\log(d)\right) \\ &\quad \times \sum_{0\le j\le m-1} \frac{\left(\frac{d}{d-1}(T(V_1, \dots, V_m) + m\log(d))\right)^j}{j!} \\ &= \exp\left(d\sum_{i=1}^m \log(V_i)\right) \sum_{0\le j\le m-1} \frac{\left(-d\sum_{i=1}^m \log(V_i)\right)^j}{j!} \\ &\approx \Phi\left(\frac{d\sum_{i=1}^m \log(V_i) + m}{m^{1/2}}\right) \end{aligned}$$

if m is large.

Note that by the preceding results the points $(C_i/c)^d$, $1 \le i \le K(n)$, should tend to be uniformly distributed on $[0, 1]$ in the case of independence, whereas otherwise they should tend to accumulate near 0. This could be utilized for a quick visual check for tail independence.

By Proposition 5.27 in Resnick (1987) we have the joint tail independence of $X_1, \dots, X_d$, i.e., $D(z) = 1$, $0 \le z \le 1$, if and only if we have the pairwise tail independence $\lim_{c \uparrow 0} P(X_j > c | X_i > c) = 0$, $i \ne j$. The above tests in the multivariate setup can be viewed, therefore, as being simultaneous tests for pairwise tail independence.

Consider next an EV $G_{\alpha_1,\dots,\alpha_d}$, where the i-th marginal G_i is an arbitrary standard extreme value df

$$G_i(x) = \exp(\psi_{\alpha_i}(x)), \quad 1 \le i \le d,$$

where

$$\psi_\alpha(x) := \begin{cases} -(-x)^\alpha, & x < 0, \quad \text{if } \alpha > 0, \\ -x^\alpha, & x > 0, \quad \text{if } \alpha < 0, \\ -\exp(-x), & x \in \mathbb{R}, \quad \text{if } \alpha = 0, \end{cases}$$

defining, thus, the family of (reverse) Weibull, Fréchet and the Gumbel distribution $\exp(\psi_\alpha(x))$. Up to a location or scale shift, $G_{\alpha_1,\dots,\alpha_d}$ is the family of possible d–dimensional EV.

Note that

$$G_{\alpha_1,\dots,\alpha_d}\left(\psi_{\alpha_1}^{-1}(x_1), \dots, \psi_{\alpha_d}^{-1}(x_d)\right) = G_{1,\dots,1}(x_1, \dots, x_d), \quad x_i < 0, \ 1 \le i \le d,$$

where $G_{1,\dots,1} = G$ has reverse exponential margins.

If the df of the rv $(X_1, \dots, X_d)$ coincides in its upper tail with $G_{\alpha_1,\dots,\alpha_d}$, then the df of $(\psi_{\alpha_1}(X_1), \dots, \psi_{\alpha_d}(X_d))$ coincides ultimately with G. We can test, therefore, for the tail independence of $(X_1, \dots, X_d)$ by applying the preceding results to $\sum_{i \le d} \psi_{\alpha_i}(X_i)$ in place of $\sum_{i \le d} X_i$.

Acknowledgements The authors would like to thank the referees for their constructive remarks, from which the paper has benefited a lot.

References

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, New York: Springer.

Coles, S.G., Heffernan, J.E., Tawn, J.A. (1999). Dependence measures for extreme value analyses. *Extremes*, *2*, 339–365.

Dupuis, D.J., Tawn, J.A. (2001). Effects of mis–specification in bivariate extreme value problems. *Extremes*, *4*, 315–330.

Falk, M., Reiss, R.-D. (2005). On Pickands coordinates in arbitrary dimensions. *Journal of Multivariate Analysis 92*, 426–453.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Vol II*, 2nd ed. New York: John Wiley.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* New York: John Wiley.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Malabar: Krieger.

Geffroy, J. (1958). Contribution à la théorie des valeurs extrêmes. *Publications de l'Institut de Statistique de l'Université de Paris 7*, 37–121.

Geffroy, J. (1959). Contribution à la théorie des valeurs extrêmes, II. *Publications de l'Institut de Statistique de l'Université de Paris 8*, 3–65.
Ghoudi, K., Khoudraji, A., Rivest, L.P. (1998). Statistical properties of couples of bivariate extreme–value copulas. *Canadian Journal of Statistics 26*, 187–197.
Gumbel, E.J. (1960). Bivariate exponential distribution. *Journal of the American Statistical Association 55*, 698–707.
Heffernan, J.E. (2000). A directory of coefficients of tail dependence. *Extremes 3*, 279–290.
Hüsler, J., Reiss, R.–D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters 7*, 283–286.
Johnson, N.L., Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.
Kotz, S., Nadarajah, S. (2000). *Extreme Value Distributions*, London: Imperial College Press.
Ledford, A., Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*, 169–187.
Ledford, A., Tawn, J.A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society B, 59*, 475–499.
Ledford, A., Tawn, J.A. (1998). Concomitant tail behaviour for extremes. *Advances in Applied Probability 30*, 197–215.
Mardia, K.V. (1970). *Families of Bivariate Distributions.* London: Griffin.
Marshall, A.W. and Olkin, I. (1976). A multivariate exponential distribution. *Journal of the American Statistical Association 62*, 30–44.
Peng, L. (1999). Estimation of the coefficients of tail dependence in bivariate extremes. *Statistics and Probability Letters 43*, 399–409.
Pickands III, J. (1981). Multivariate extreme value distributions. In: *Proceedings of the 43th Session of the International Statistical Institute (Buenos Aires)*, pp 859–878.
Reiss, R.–D. (1993). *A Course on Point Processes,* Springer Series in Statistics. New York: Springer.
Reiss, R.–D., Thomas, M. (2001). *Statistical Analysis of Extreme Values,* 2nd ed. Basel: Birkhäuser.
Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*, applied Probability, Vol. 4. New York: Springer.
Sheskin, D.J. (2004). *Parametric and Nonparametric Statistical Procedures*, 3rd. ed, Boca Raton: Chapmann & Hall.
Sibuya, M. (1960). Bivariate extreme statistics, *Annals of the Institute of Statistical Mathematics, 11*, 195–210.