INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2012/12

NONPARAMETRIC INFERENCE FOR MAX-STABLE DEPENDENCE

SEGERS, J.

Nonparametric inference for max-stable dependence

Johan Segers*

Discussion of "Statistical Modelling of Spatial Extremes",
by A. C. Davison, S. Padoan and M. Ribatet,
to appear in *Statistical Science*

April 2, 2012

The choice for parametric techniques in the discussion article is motivated by the claim that for multivariate extreme-value distributions, "owing to the curse of dimensionality, nonparametric estimation has essentially been confined to the bivariate case" (Section 2.3). Thanks to recent developments, this is no longer true if data take the form of multivariate maxima, as is the case in the article. A wide range of nonparametric, rank-based estimators and tests are nowadays available for extreme-value copulas. Since max-stable processes have extreme-value copulas, these methods are applicable for inference on max-stable processes too. The aim of this note is to make the link between extreme-value copulas and max-stable processes explicit and to review the existing nonparametric inference methods.

1 Extreme-value copulas

Let the random variables $Y_1, \ldots, Y_D$ represent the maxima in a given year of a spatial process (e.g. rainfall) that is observed at a finite number of sites, $x_1, \ldots, x_D$, in a region $\mathcal{X}$ in space $\mathbb{R}^p$ (typically, $p = 2$). Let $F_1, \ldots, F_D$ be the marginal cumulative distribution functions, assumed to be continuous. In the article, these are assumed to be univariate generalized extreme-value distributions, an assumption that will not be needed here.

The random variables $U_d = F_d(Y_d)$ are uniformly distributed on the interval $(0, 1)$ and the joint cumulative distribution function C of the vector $U_1, \ldots, U_D$ is the copula of the random vector $Y_1, \ldots, Y_D$:

$$C(u_1, \ldots, u_D) = \Pr(U_1 \le u_1, \ldots, U_D \le u_D), \quad (1)$$

for $0 \le u_d \le 1$. The requirement that the random vector $Y_1, \ldots, Y_D$ is max-stable entails

$$C^m(u_1^{1/m}, \ldots, u_D^{1/m}) = C(u_1, \ldots, u_D) \quad (2)$$

for all $m > 0$. In [18], it was shown that (2) holds if, and only if,

$$C(u_1, \ldots, u_D) = \exp\{-r\, A(v_1, \ldots, v_D)\}. \quad (3)$$

where $r = -\sum_{d=1}^{D} \log u_d$ and $v_d = -r^{-1} \log u_d$. The domain of the Pickands dependence function A is the unit simplex, $\mathcal{S}_D = \{v \in [0,1]^D : \sum_d v_d = 1\}$. A necessary and sufficient condition for a function A on $\mathcal{S}_D$ to be a Pickands dependence function is that

$$A(v_1, \ldots, v_D) = \int_{\mathcal{S}_D} \max(v_1 s_1, \ldots, v_D s_D)\, \mathrm{d}M(s_1, \ldots, s_D), \quad (4)$$

for a Borel measure M on $\mathcal{S}_D$ verifying the constraints $\int_{\mathcal{S}_D} s_d \, \mathrm{d}M(s_1, \ldots, s_D) = 1$ for all $d \in \{1, \ldots, D\}$. In particular, A is convex and $\max(v_1, \ldots, v_D) \le A(v_1, \ldots, v_D) \le v_1 + \cdots + v_D$. In dimension $D = 2$, these two properties completely characterize Pickands dependence functions (but not if $D \ge 3$).

*Université catholique de Louvain, Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: johan.segers@uclouvain.be. Funding was provided by IAP research network grant P6/03 of the Belgian government (Belgian Science Policy) and by "Projet d'actions de recherche concertées" number 07/12-002 of the Communauté française de Belgique, granted by the Académie universitaire de Louvain.

2 Max-stable models

The representation in (3)–(4) is valid for general max-stable copulas and therefore also holds for the finite-dimensional distributions of the max-stable processes considered in Section 6 in the article. The purpose of this section is to make this relation explicit.

Consider the simple max-stable process

$$Z(x) = \max_{j \geq 1} [S_j \max\{0, W_j(x)\}], \qquad x \in \mathbb{R}^p, \quad (5)$$

where $\{S_j\}_{j=1}^{\infty}$ are the points of a Poisson process on $\mathbb{R}_+$ with rate $s^{-2}\mathrm{d}s$ and where $W_1, W_2, \ldots$ are iid replicates of a stationary stochastic process W on $\mathbb{R}^p$, independent of the previous Poisson process, and such that $\mathrm{E}[W^+(x)] = 1$, where we write $W^+(x) = \max\{0, W(x)\}$. Particular cases of this model include the so-called Smith model [24], the Schlather model [22] and the Brown–Resnick model [12].

The stationary, marginal distribution of $Z(x)$ in (5) is unit-Fréchet and the joint distribution function of the vector $Z(x_1), \ldots, Z(x_d)$ is given by

$$\begin{aligned} &\Pr[Z(x_1) \leq z_1, \ldots, Z(x_D) \leq z_D] \\ &= \exp[-\mu(\{(s, w) : \max_d (w^+(x_d)/z_d) > 1/s\})], \end{aligned}$$

for $z_d > 0$, where μ is the intensity measure of the Poisson point process $\{(S_j, W_j)\}_{j=1}^{\infty}$. A simple calculation shows that

$$\begin{aligned} &\Pr[Z(x_1) \leq z_1, \ldots, Z(x_D) \leq z_D] \\ &\qquad = \exp\bigl(-\mathrm{E}[\max_d \{W^+(x_d)/z_d\}]\bigr). \end{aligned}$$

As a consequence, the copula of $Z(x_1), \ldots, Z(x_D)$ is given by the extreme-value copula with Pickands dependence function

$$A(v_1, \ldots, v_D) = \mathrm{E}[\max_d \{v_d \, W^+(x_d)\}] \qquad (6)$$

for $(v_1, \ldots, v_D) \in \mathcal{S}_D$. As illustrated by the computations in [5], the integral arising in (6) can rarely be calculated analytically, unless $D = 2$.

To recover the spectral measure M in (4) from the distribution of the stochastic process W, let $R = \sum_d W^+(x_d)$. On the event $R > 0$, consider the random vector $(W^+(x_1), \ldots, W^+(x_D))/R$. Then

$$\begin{aligned} &\mathrm{d}M(s_1, \ldots, s_D) = \\ &\quad \Pr(R > 0) \; \mathrm{E}[R \mid \forall d : W^+(x_d)/R = s_d; R > 0] \\ &\qquad\qquad \Pr[\forall d : W^+(x_d)/R \in \mathrm{d}s_d \mid R > 0]. \end{aligned}$$

3 Estimation

Nonparametric estimators of the Pickands dependence function are surprisingly easy to construct and calculate. The starting point is the simple fact that for $u \in [0, 1]$ and for $(v_1, \ldots, v_D) \in \mathcal{S}_D$, the extreme-value copula C with Pickands dependence function A satisfies

$$C(u^{v_1}, \ldots, u^{v_D}) = u^{A(v_1, \ldots, v_D)}, \qquad (7)$$

as can be verified from (3). Using (7), the function A can be recovered from the copula C in various ways, for instance, through integrals of the form

$$\begin{aligned} &\int_0^1 f\bigl(C(u^{v_1}, \ldots, u^{v_D})\bigr) \, g(u) \, \mathrm{d}u \\ &= \int_0^1 f(u^\alpha) \, g(u) \, \mathrm{d}u, \qquad \alpha = A(v_1, \ldots, v_D) \quad (8) \end{aligned}$$

for well-chosen functions f and g. Plugging estimators for C and solving for α then yields estimators for A.

A natural estimator for C is the empirical copula. Let $(Y_{i1}, \ldots, Y_{iD})$, for $i \in \{1, \ldots, n\}$, be an independent random sample from a distribution with continuous margins and copula C. The empirical copula is defined as

$$\begin{aligned} &C_n(u_1, \ldots, u_D) \\ &= \frac{1}{n} \sum_{i=1}^n I\{F_{n1}(Y_{i1}) \leq u_1, \ldots, F_{nD}(Y_{iD}) \leq u_D\}, \end{aligned} \qquad (9)$$

where F_{nd} is the (marginal) empirical distribution function of $Y_{1d}, \ldots, Y_{nd}$. Being based on multivariate ranks, the empirical copula is invariant under monotone transformations of the data. The empirical copula goes back to the seminal paper by Rüschendorf [21] and has been studied and applied intensively, such as recently in [23, 25, 26].

Inserting the empirical copula into (8) and solving for $A(v_1, \ldots, v_D)$ produces simple and (almost) explicit estimators. Particular instances are the Pickands estimator [6, 9, 11, 19] and the Capéraà–Fougères–Genest estimator [4, 9, 11]. The bivariate versions of these estimators are special cases of the weighted estimator in [17]. Minimum-distance estimators are another instance of this technique

[3]. Standard errors can be obtained via resampling [2, 20] or via empirical likelihood [17].

A drawback of the nonparametric estimators of A is that they typically do not produce valid Pickands dependence functions—remember the representation in (4) that such functions must satisfy. A way to overcome this issue is by projecting a possibly invalid pilot estimator A_n onto the family of Pickands dependence functions, yielding

$$A_n^{\mathrm{proj}} = \operatorname*{arg\,min}_{A \in \mathcal{A}_D} \int_{\mathcal{S}_D} (A_n - A)^2, \qquad (10)$$

where $\mathcal{A}_D$ denotes the family of all Pickands dependence functions in dimension D and where the integral is with respect to some measure on $\mathcal{S}_D$. In general, the infinite-dimensional least-squares problem in (10) does not admit an explicit solution. Approximate solutions can be obtained by performing the minimization over the (finite-dimensional) class of Pickands dependence functions with discrete spectral measures supported on a given, finite grid [7, 11].

Finally, note that a nonparametric estimator A_n can be transformed into a parametric one via minimum-distance or projection techniques: in (10), replace $\mathcal{A}_D$ by the parametric model of interest. If the model happens to be specified via the point process representation (5), then this technique requires the calculation of the Pickands dependence function A via (6).

4 Testing

Nonparametric methods are particularly suitable for hypothesis testing. Of special interest are the hypothesis of max-stability in general and the goodness of fit of a parametric model in particular. In most cases, critical values are computed via resampling methods.

Even if the data at hand are vectors of component-wise maxima, it is a good idea to test whether it is safe to assume that the underlying distribution is max-stable, in particular, when the end-goal is to perform prediction and/or extrapolation. For bivariate extreme-value copulas, the first two moments of the random variable $W = C(U_1, U_2) = F(Y_1, Y_2)$ happen to satisfy a particular linear relation. The sample moments of the random variables $W_{n1}, \ldots, W_{nn}$ defined by

$$W_{ni} = \frac{1}{n} \sum_{t=1}^{n} I(Y_{t1} \le Y_{i1}, Y_{t2} \le Y_{i2})$$

can therefore be converted to a test statistic for the null hypothesis of max-stability [1, 10].

Another approach for testing max-stability is by comparing the empirical copula C_n in (9) with the extreme-value copula that has a given estimator A_n as its Pickands dependence function. For the bivariate case, Cramér–von Mises tests based on the Pickands and Capéraà–Fougères–Genest estimators are described in [3, 15].

Finally, the adequacy of the hypothesis of max-stability can be tested by directly exploiting the copula max-stability relation (2) through a comparison of $C_n(u_1, \ldots, u_D)$ with $C_n^m(u_1^{1/m}, \ldots, u_D^{1/m})$ for various values of $m > 0$. Cramér-von Mises type test statistics turn out to be particularly effective [13].

The goodness of fit of a parametric model can be tested by comparing the fitted parametric estimator for A with a nonparametric one [8]. For max-stable models arising through the point process representation in (5), the function A has to be computed through the relation (6). Shape constraints such as exchangeability can be tested similarly [16].

5 Conclusion

Nonparametric methods yield an attractive alternative inference method for max-stable dependence. Estimators and test statistics of the Pickands dependence function are (almost) explicit, even in the general, multivariate case. Moreover, as the procedures are based upon the ranks of the data only, the step of modeling the margins can be skipped (which is not to be confused with the false statement that the uncertainty on the margins has been eliminated altogether). Many of the methods described in this contribution are implemented in the R package copula [14].

References

[1] Ben Ghorbal, M., C. Genest, and J. Nešlehová (2009). On the test of Ghoudi, Khoudraji, and Rivest

for extreme-value dependence. *The Canadian Journal of Statistics 37*(4), 534–552.

[2] Bücher, A. and H. Dette (2010). A note on bootstrap approximations for the empirical copula process. *Statistics & Probability Letters 80*(23-24), 1925–1932.

[3] Bücher, A., H. Dette, and S. Volgushev (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics 39*(4), 1963–2006.

[4] Capéraà, P., A.-L. Fougères, and C. Genest (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika 84*, 567–577.

[5] de Haan, L. and T. T. Pereira (2006). Spatial extremes: Models for the stationary case. *The Annals of Statistics 34*, 146–168.

[6] Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters 12*(5), 429–439.

[7] Fils-Villetard, A., A. Guillou, and J. Segers (2008). Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics 36*(3), 369–382.

[8] Genest, C., I. Kojadinovic, J. Nešlehová, and J. Yan (2011). A goodness-of-fit test for extreme-value copulas. *Bernoulli 17*, 253–275.

[9] Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *Annals of Statistics 37*(5B), 2990–3022.

[10] Ghoudi, K., A. Khoudraji, and L.-P. Rivest (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *The Canadian Journal of Statistics 26*(1), 187–197.

[11] Gudendorf, G. and J. Segers (2011). Nonparametric estimation of multivariate extreme-value copulas. Technical Report 2011/18, Université catholique de Louvain, ISBA. http://arxiv.org/abs/1107.2410.

[12] Kabluchko, Z., M. Schlather, and L. de Haan (2009). Stationary max-stable fields associated to negative definite functions. *Annals of Probability 37*, 2042–2065.

[13] Kojadinovic, I., J. Segers, and J. Yan (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *The Canadian Journal of Statistics 39*(4), 703–720.

[14] Kojadinovic, I. and J. Yan (2010a). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software 34*(9), 1–20. http://www.jstatsoft.org/v34/i09/.

[15] Kojadinovic, I. and J. Yan (2010b). Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis 101*(9), 2234–2249.

[16] Kojadinovic, I. and J. Yan (2012). A nonparametric test of exchangeability for extreme-value and left-tail decreasing bivariate copulas. *The Scandinavian Journal of Statistics*, to appear.

[17] Peng, L., L. Qian, and J. Yang (2011). Weighted estimation of dependence function for an extreme-value distribution. *Bernoulli*, to appear.

[18] Pickands, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, Volume 49, pp. 859–878, 894–902. With a discussion.

[19] Pickands, III, J. (1989). Multivariate negative exponential and extreme value distributions. In *Extreme value theory (Oberwolfach, 1987)*, Volume 51 of *Lecture Notes in Statistics*, pp. 262–274. New York: Springer.

[20] Rémillard, B. and O. Scaillet (2009). Testing for equality between two copulas. *J. Multivar. Anal. 100*(3), 377–386.

[21] Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics 4*, 912–923.

[22] Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes 5*(1), 33–44.

[23] Segers, J. (2011). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, to appear.

[24] Smith, R. L. (1990). Max-stable processes and spatial extremes. Unpublished.

[25] Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics 33*(3), 357–375.

[26] van der Vaart, A. and J. Wellner (2007). Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems*, pp. 234–252. Institute of Mathematical Statistics.