NONPARAMETRIC ESTIMATION OF EXTREME VALUE COPULAS IN ARBITRARY DIMENSIONS

Gordon Gudendorf and Johan Segers

Université Catholique de Louvain

January 29, 2009
Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$ i.i.d. $p$-variate random vectors with joint df $F$. Define:

$$
M_n = (M_{n1}, \ldots, M_{np})
= (\lor_{i=1}^{n} X_{i1}, \ldots, \lor_{i=1}^{n} X_{ip})
$$

If for some normalizing constants $(a_{n,1}, b_{n,1}), \ldots, (a_{n,p}, b_{n,p})$:

$$
\lim_{n \to \infty} P\left( \frac{M_{n1} - a_{n,1}}{b_{n,1}} \leq x_1, \ldots, \frac{M_{np} - a_{n,p}}{b_{n,p}} \leq x_p \right)
= \lim_{n \to \infty} F^n(a_{n,1}x_1 + b_{n,1}, \ldots, a_{n,p}x_p + b_{n,p}) = G(x_1, \ldots, x_p)
$$

then $F$ is in the domain of attraction of $G$ ($F \in D(G)$), if $G$ is a $p$-variate distribution function with nondegenerate marginals. $G$ is called a $p$-variate extreme-value distribution.
Introduction

Definition

$G$ is a **multivariate extreme value distribution** if and only if:

1. its margins $G_1, \ldots, G_p$ are univariate extreme value distributions (similar definition with $p = 1$)

2. 

$$G(x) = \exp[-l\{-\log G_1(x_1), \ldots, -\log G_p(x_p)\}] \quad (1)$$

where $l$ is the so-called stable tail dependence function.

$$l(y) = \lim_{s \downarrow 0} \frac{1}{s} \Pr[\bigcup_{j=1}^{p} \{F_j(X_j) > 1 - sy_j\}] \quad (2)$$
Introduction

Figure: Graphical interpretation of stable dependence function \( l \). Name is deduced from the L-shape of the probability domain. \( F \leftarrow: \) generalized inverse.
Definition
The restriction of \( l \) to the unit simplex

\[
\Delta_p = \{(w_1, \ldots, w_p) : \sum_{i=1}^{p} w_i = 1, \ w_j \geq 0, \ \forall j = 1, \ldots, p\}
\]

is called Pickands dependence function \( A \).

\( A \) satisfies the following properties:

- \( A(e_j) = 1 \ \forall j = 1, \ldots, p \)
- \( A \) is convex
- \( \max(w_1, \ldots, w_p) \leq A(w) \leq 1 \)

Example: Gumbel or logistic dependence function

\[
A(w) = (w_1^r + \cdots + w_p^r)^{1/r}, \quad r \geq 1
\]
Recall the definition of multivariate extreme-value distribution:

\[ G(x) = \exp[-l\{- \log G_1(x_1), \ldots, \log G_p(x_p)\}] \]

The extreme-value copula \( C \) is defined by:

\[ G(x) = C(G_1(x_1), \ldots, G_p(x_p)) \quad (3) \]

**Definition**

It follows immediately that the expression for \( C \) is given by:

\[
C(u) = \exp\{-l(- \log u_1, \ldots, - \log u_p)\} \\
= \exp\left\{\left(\sum_{j=1}^{p} \log u_j\right) \cdot A\left(\frac{\log u_1}{\sum_{j=1}^{p} \log u_j}, \ldots, \frac{\log u_{p-1}}{\sum_{j=1}^{p} \log u_j}\right)\right\}
\]

for \( 0 < u_j \leq 1 \). Such copulas are called *extreme-value copulas*. 
Introduction

An alternative definition for extreme-value copulas is given below:

Definition
A copula \( C \) is an extreme-value copula iff it is max-stable

\[
\forall t > 0 : \{ C(u_1^{1/t}, \ldots, u_p^{1/t}) \}^t = C(u_1, \ldots, u_p) \tag{4}
\]

Example: Gumbel (logistic) dependence function:
\( A(w) = (w_1^r + \cdots + w_p^r)^{1/r}, \quad r \geq 1 \) For all \( t > 0 \):

\[
C^t(u_1^{1/t}, \ldots, u_p^{1/t}) = \exp \left\{ - \left[ (-\log u_1^{1/t})^r + \cdots + (-\log u_p^{1/t})^r \right] \right\}^{t/r} \\
= \exp \left\{ - \frac{1}{t} \left[ (-\log u_1)^r + \cdots + (-\log u_p)^r \right] \right\}^{t/r} \\
= C(u_1, \ldots, u_p)
\]
Introduction

Let $X_1, \ldots, X_n$ be a $p$-variate random sample from an unknown distribution $F$ with continuous margins $F_1, \ldots, F_p$ and extreme value copula $C$.

Problem

Estimate $C$ or, equivalently, its Pickands dependence function $A$ in the case the margins $F_1, \ldots, F_p$ are known.

Remark:

For $p = 2$, the case with unknown margins is treated in:
Genest C. and Segers J. (2009) Rank-based inference for bivariate extreme-value copulas The Annals of Statistics

It consists in replacing $U_{ij} = F(X_{ij})$ by the rank estimator

$$\hat{U}_{ij} = \frac{{R_{ij}^x}}{n+1}.$$
Different estimators for dimension $p > 2$ can be extended to higher dimensions:

- Deheuvels (1991)
- Hall & Tajvidi (2000)
- Pickands (1981)
- ZWP (2008)

ZWP-estimator: extension of the CFG estimator to higher dimensions.
Define:

\[ Z_{ij} = \bigvee_{l \neq j} \log F_l(X_{il}) \frac{1}{w_l} - \log F_j(X_{ij}) \frac{1}{1 - w_j} + \bigvee_{l \neq j} \log F_l(X_{il}) \frac{1}{w_l} \] (5)

One can show that:

\[ P(Z_{ij} \leq z) = z + z(1 - z) \times \frac{\partial}{\partial z} \log A \left( \frac{zw_1}{1 - w_j}, \ldots, \frac{zw_{j-1}}{1 - w_j}, 1 - z, \frac{zw_{j+1}}{1 - w_j}, \ldots, \frac{zw_{p-1}}{1 - w_j} \right) \] (6)

Taking into account the boundary conditions, the preceding P.D.E. admits the following solution:

\[ \log A(w_1, \ldots, w_{p-1}) = \int_0^{1 - w_j} \frac{P(Z_{ij} \leq z) - z}{z(1 - z)} dz \] (7)

Estimator:

1. Replace \( P(Z_{ij} \leq z) \) by its empirical distribution.
2. Add weight functions \( \lambda_j(\cdot) \) to satisfy boundary conditions, i.e. \( \hat{A}(e_j) = 1 \).
ZWP-estimator

Definition
The ZWP-estimator is defined as follows:

$$\log \hat{A}_n^{ZW} (w) = \sum_{j=1}^{p} \lambda_j(w) \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{1-w_j} \frac{1(Z_{ij} \leq z) - z}{z(1-z)} dz$$

with weight functions:

- $\lambda_j(w) \geq 0 \quad \forall j = 1, \ldots, p$
- $\sum_{j=1}^{p} \lambda_j(w) = 1$

Main disadvantage: No real statistic criterion for choosing the optimal weight functions.
Naive estimator

Observe that:

$$
\xi_i(w) = -\sqrt[p]{\frac{\log(U_{ij})}{w_j}} \sim \text{Exp}(A(w)) \left( E\xi_i(w) = \frac{1}{A(w)} \right) \quad (8)
$$

Exploit the relationship between Gumbel and Exponential distributions:

$$
-\log \xi_i(w) = -\log \left( \frac{\xi_i(w)}{1/A(w)} \right) + \log A(w)
$$

\underline{Standard Gumbel}

From where we deduce that:

$$
\log A(w) = E(- \log \xi_i(w) - \gamma)
$$
Naive estimator

Definition

Naive estimator inspired from Segers(2007):

\[
\log \left( \hat{A}_n(w) \right) = -\frac{1}{n} \sum_{i=1}^{n} \log (\xi_i(w)) - \gamma
\]  

(9)

with \( \gamma = 0.577 \ldots \) (= mean of standard Gumbel)

Naive in the sense that this estimator is unbiased but does not necessarily satisfy boundary conditions, i.e.

\[
\hat{A}_n(e_j) \neq 1 \quad \forall j = 1, \ldots, p
\]
Lemma (Gudendorf & Segers, 2009)

Let $w \in \Delta_p$ and the constraints on the weight functions be still valid, then:

$$\log \hat{A}_n^{ZWp}(w) = \log \hat{A}_n(w) - \sum_{j=1}^{p} \lambda_j(w) \log \hat{A}_n(e_j)$$  \hspace{1cm} (10)
Remember:

$$\log \left( \hat{A}_n(w) \right) = -\frac{1}{n} \sum_{i=1}^{n} \log (\xi_i(w)) - \gamma$$ \hspace{1cm} (11)

The preceding estimator:

$$\log \left( \hat{A}_n^{ZWP}(w) \right) = \log \left( \hat{A}_n(w) \right) - \sum_{j=1}^{p} \lambda_j(w) \log \hat{A}_n(e_j)$$

⇒ Formula resembles the intercept estimator in linear regression model. Or expressed alternatively:

$$\log \left( \hat{A}_n(w) \right) = \log \left( \hat{A}_n^{ZWP}(w) \right) + \sum_{j=1}^{p} \lambda_j(w) \log \hat{A}_n(e_j)$$
Independent of the values for the weight functions $\lambda_j(\cdot)$, we have that $\log \hat{A}_n^{ZWP}$ is an unbiased estimator:

$$E \log \hat{A}_n^{ZWP}(w) = \log A(w)$$

because

$$E \log \hat{A}_n(e_j) = E(- \log \xi_i(e_j) - \gamma) = \log A(e_j) + \gamma - \gamma = \log 1 + 0 = 0$$
Consider the following regression model:

\[- \log \{ \xi_i(w) \} - \gamma = \lambda_0 + \lambda_1 (- \log \xi_i(e_1) - \gamma) \]

\[+ \cdots + \lambda_p (- \log \xi_i(e_p) - \gamma) + \epsilon_i \] (12)

where we deduce that the intercept can be considered as an estimate for \( \log A(w) \).

**Definition (Gudendorf & Segers, 2009)**

We define the OLS-estimator for the stable tail dependence function \( A(w) \) as follows:

\[ \hat{\lambda}_0 = \log \hat{A}_n^{OLS}(w) = \log \hat{A}_n(w) - \sum_{j=1}^{p} \hat{\lambda}_j(w) \log \hat{A}_n(e_j) \] (13)
Properties:

1. Constraints on weight functions no longer necessary and also not satisfied by OLS.

2. **Big advantage:** Gauss-Markov theorem guarantees efficiency in the class of linear unbiased estimators.

3. The OLS estimator $\log \hat{A}_{n}^{OLS}(w)$ satisfies the boundary conditions.
Simulations

Logistic (Gumbel) dependence function:

\[ A(w) = (w_1^5 + w_2^5 + w_3^5)^{1/5}, \quad (14) \]

Our simulations:

- 1000 data sets (Monte-Carlo)
- Sample size \( n = 25, 50, 75, 100 \)
- Estimated for values \( w_1 = w_2 \)
Figure: Black: OLS, Red: ZWP, Green: Pickands, Blue: Deheuvels, Blue(light): Hall & Tajvidi
Figure: Black: OLS, Red: ZWP, Green: Pickands, Blue: Deheuvels, Blue(light): Hall & Tajvidi
Deheuvels, P. (1991): *On the limiting behaviour of Pickands estimator for bivariate extreme-value distributions*. Statistics & Probability Letters 12(5), pp. 429-439.

Hall P. and Tajvidi N. (2000): *Distribution and dependence-function estimation for bivariate extreme-value distributions*, Bernoulli 6(5), pp. 835-844.

Pickands J.: *Multivariate extreme value distributions*, Proceedings of the 43rd Session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981), Volume 49, pp. 859-878, 894-902.

Segers J. (2007), *Nonparametric inference for bivariate extreme-value copulas*, In M. Ahsanullah and S. Kirmani (Eds.), Extreme Value Distributions, Chapter 9, pp. 181-203. Nova Science Publishers, Inc.
Zhang D., Wells M.T, Peng L. (2008), *Nonparametric estimation of the dependence function for multivariate extreme value distribution.*, Journal of Multivariate Analysis 99(4), 577-588.