Copulas: An Introduction
III - Inference

Johan Segers

Université catholique de Louvain (BE)
Institut de statistique, biostatistique et sciences actuarielles

Columbia University, New York City
9–11 Oct 2013

Copulas: An Introduction
III - Inference

Inference on measures of association

The empirical copula

Inference on parametric copula families

Shape-constrained inference: Extreme-value copulas

Copula models: Separating the margins and the copula

Sklar's celebrated theorem:

$$F(x_1, \ldots, x_p) = C\big(F_1(x_1), \ldots, F_p(x_p)\big)$$

Separate assumptions on C and $F_1, \ldots, F_p$:

	margins $F_1, \ldots, F_p$	
copula C	*nonparametric*	*parametric*
nonparametric	empirical copula	plug-in
shape constraints	Archimedean, extreme-value, elliptical, …	
parametric	pseudo-likelihood	likelihood

Rank-based inference

If there are no assumptions on the margins (except for continuity), copula models are invariant under component-wise increasing transformation.

Same invariance property for the estimators of copula properties?
$\Rightarrow$ rank-based inference.

- robust w.r.t. outliers
- no need to select models for the margins

Copulas: An Introduction
III - Inference

Inference on measures of association

The empirical copula

Inference on parametric copula families

Shape-constrained inference: Extreme-value copulas

Kendall's tau as a correlation

Let $(X_1, Y_1), (X_2, Y_2)$ be iid F, continuous margins. Recall

$$\begin{aligned}\tau(F) = \quad & \mathrm{P}[X_1 - X_2 \text{ and } Y_1 - Y_2 \text{ have the same sign}] \\ & - \mathrm{P}[X_1 - X_2 \text{ and } Y_1 - Y_2 \text{ have opposite signs}]\end{aligned}$$

Ex. Show that

$$\tau(F) = \mathrm{cor}\big(\mathbf{1}(X_1 \le X_2), \mathbf{1}(Y_1 \le Y_2)\big)$$

[Hint: $\mathrm{P}(X_1 \le X_2) = 1/2$ and there's lots of symmetry.]

An extension of Kendall's tau: Association between two random vectors

Let $(\boldsymbol{X}_1, \boldsymbol{Y}_1), (\boldsymbol{X}_2, \boldsymbol{Y}_2)$ be iid $(\boldsymbol{X}, \boldsymbol{Y})$ in $\mathbb{R}^{p+q}$.
As when $p = q = 1$, quantify association between $\boldsymbol{X}$ and $\boldsymbol{Y}$ via

$$\begin{aligned} \tau(\boldsymbol{X}, \boldsymbol{Y}) &= \operatorname{cor}(\mathbf{1}\{\boldsymbol{X}_1 \le \boldsymbol{X}_2\}, \mathbf{1}\{\boldsymbol{Y}_1 \le \boldsymbol{Y}_2\}) \\ &= \frac{p_{\boldsymbol{X},\boldsymbol{Y}} - p_{\boldsymbol{X}}\, p_{\boldsymbol{Y}}}{\sqrt{p_{\boldsymbol{X}}\,(1 - p_{\boldsymbol{X}})\, p_{\boldsymbol{Y}}\,(1 - p_{\boldsymbol{Y}})}} \end{aligned}$$

where

$$\begin{aligned} p_{\boldsymbol{X},\boldsymbol{Y}} &= \mathrm{P}(\boldsymbol{X}_1 \le \boldsymbol{X}_2, \boldsymbol{Y}_1 \le \boldsymbol{Y}_2), \\ p_{\boldsymbol{X}} &= \mathrm{P}(\boldsymbol{X}_1 \le \boldsymbol{X}_2), \\ p_{\boldsymbol{Y}} &= \mathrm{P}(\boldsymbol{Y}_1 \le \boldsymbol{Y}_2) \end{aligned}$$

- Depends on the law of $(\boldsymbol{X}, \boldsymbol{Y})$ only through its copula.
- Sample version: U-statistic of degree $m = 2$.

U-statistics: generalizations of sample means

Let P_X be the distribution of a random object X taking values in a space $\mathcal{X}$. Suppose we want to estimate a 'parameter' $\theta = \theta(P_X)$ of the form

$$\begin{aligned} \theta &= \mathrm{E}[g(X_1, \ldots, X_m)] \\ &= \int \cdots \int g(x_1, \ldots, x_m) \, \mathrm{d}P_X(x_1) \cdots \mathrm{d}P_X(x_m) \end{aligned}$$

where $X_1, \ldots, X_m$ are iid X and where $g : \mathcal{X}^m \to \mathbb{R}$ is given.

The U-statistic estimator for θ based on a sample $X_1, \ldots, X_n$ is

$$\hat{\theta}_m = \frac{1}{n!/(n-m)!} \sum_{\substack{(i_1,\ldots,i_m)\in\{1,\ldots,n\}^m \\ \sharp\{i_1,\ldots,i_m\}=m}} g(X_{i_1}, \ldots, X_{i_m})$$

U-statistics show up everywhere

$m = 1$ Expectations and sample averages:

$$\theta = \mathrm{E}[g(X_1)]$$
$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$$

$m = 2$ Variance of a real-valued random variable X:

$$\sigma^2 = \mathrm{var}(X) = \mathrm{E}\left[\frac{1}{2}(X_1 - X_2)^2\right]$$
$$\hat{\sigma}_n^2 = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{2}(X_i - X_j)^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

the (unbiased version of) the sample variance

Hoeffding's decomposition theorem: Linear expansion of a *U*-statistic

If $\mathrm{E}[g^2(X_1, \ldots, X_m)] < \infty$, then

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{m}{\sqrt{n}} \sum_{i=1}^{n} h_1(X_i) + o_p(1)$$

where

$$h_1(x_1) = \mathrm{E}[g_{\mathrm{sym}}(x_1, X_2, \ldots, X_m)] - \theta$$

$$g_{\mathrm{sym}}(x_1, \ldots, x_m) = \frac{1}{m!} \sum_{\substack{\text{permutations} \\ (i_1, \ldots, i_m)}} g(x_{i_1}, \ldots, x_{i_m})$$

U-statistics with non-degenerate kernels are asymptotically normal

By Slutsky's lemma and the multivariate central limit theorem, Hoeffding's decomposition yields joint asymptotic normality of a vector of U-statistics:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, m^2\sigma_1^2) \qquad (n \to \infty)$$

$$\sigma_1^2 = \operatorname{var} h_1(X_1)$$

The asymptotic (co)variance(s) can be estimated consistently by

- U-statistics
- jackknife
- the sample (co)variance of

$$\hat{h}_{1,n}(X_i), \qquad i = 1, \ldots, n$$
$$\hat{h}_{1,n}(x) = U\text{-statistic}$$

The estimator of Kendall's tau is asymptotically normal

Hoeffding's decomposition yields joint asymptotic normality of

$$\sqrt{n} \begin{pmatrix} \hat{p}_{\boldsymbol{X},\boldsymbol{Y},n} - p_{\boldsymbol{X},\boldsymbol{Y}} \\ \hat{p}_{\boldsymbol{X},n} - p_{\boldsymbol{X}} \\ \hat{p}_{\boldsymbol{Y},n} - p_{\boldsymbol{Y}} \end{pmatrix}$$

with explicit 3×3 covariance matrix Σ.

From the delta method, we get asymptotic normality of

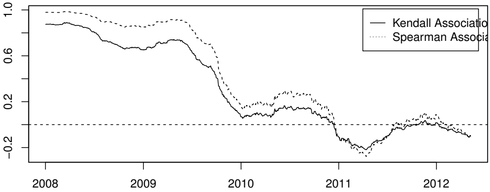
$$\sqrt{n}\big(\hat{\tau}_n(\boldsymbol{X}, \boldsymbol{Y}) - \tau(\boldsymbol{X}, \boldsymbol{Y})\big)$$

The expression for the asymptotic variance is longish, but explicit, and can be estimated consistently.

European sovereign debt crisis: North and South

- Association between north and south European bond markets changes as credit-worthiness of countries evolve
- Data: daily returns of Merrill Lynch government bond indices
  - North: France, Germany, the Netherlands
  - South: Italy, Portugal, Spain
  - From January 1, 2007 to November 15, 2012
  - Forward looking moving window of 150 days

With the crisis, association becomes negative



Inference on measures of association: Some literature

El Maache, H. and Y. Lepage (2003). Spearman's rho and Kendall's tau for multivariate data sets. In M. Moore, S. Froda, and C. Léger (Eds.), *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*, pp. 113–130. Beachwood, OH: Institute of Mathematical Statistics.

Grothe, O., J. Schnieders, and J. Segers (2013). Measuring association and dependence between random vectors. *Journal of Multivariate Analysis (to appear)*, arXiv:1107.4381.

Quessy, J.-F., M. Saïd, and A.-C. Favre (2013). Multivariate Kendall's tau for change-point detection in copulas. *The Canadian Journal of Statistics 41*, 65–82.

Schmid, F., R. Schmidt, T. Blumentritt, S. Gaisser, and M. Ruppert (2010). Copula-based measures of multivariate association. In P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik (Eds.), *Copula Theory and Its Applications*, Lecture Notes in Statistics, pp. 209–236. Berlin: Springer.

Székely, G., M. Rizzo, and N. Bakirov (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics 35*(6), 2769–2794.

Copulas: An Introduction
III - Inference

Why nonparametric inference on a copula?

Use a nonparametric estimate $\hat{C}$ of C in order to …

… perform goodness-of-fit testing and to assist in model selection

- Compare $\hat{C}$ with $C_{\hat{\theta}}$

… test for some qualitative property

- Assess (lack of) symmetry by comparing $\hat{C}(u, v)$ with $\hat{C}(v, u)$

… have a starting point for estimation of measures of association

- Apply the plug-in principle: e.g. $\widehat{\tau(C)} := \tau(\hat{C})$

…

Empirical distributions: counting points

Points $X_1, \ldots, X_n$. Empirical probability measure

$$P_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \in A)$$

Special case: empirical cdf

$$\hat{F}_n(x) = P_n((-\infty, x]) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \leq x)$$

Also multivariate: $A = (-\infty, x_1] \times \ldots \times (-\infty, x_d]$

Ex. Make a plot of the empirical cdf of a univariate sample.

- Where does it jump?
- What are the jump sizes?

Ranks and the empirical cdf

Univariate sample $X_1, \ldots, X_n$. Evaluate $\hat{F}_n$ at the data:

$$\begin{aligned}\hat{F}_n(X_i) &= \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}(X_k \leq X_i) \\ &= \frac{1}{n} R_{i,n}\end{aligned}$$

with $R_{i,n}$ the rank of X_i among $X_1, \ldots, X_n$

Ex. What is the empirical cdf of the points $\hat{F}_n(X_1), \ldots, \hat{F}_n(X_n)$?

The empirical copula: Empirical distributions inside and outside

Assume continuous margins $F_1, \ldots, F_d \Rightarrow$ no ties. Recall

$$C(\boldsymbol{u}) = \mathrm{P}[F_1(X_1) \le u_1, \ldots, F_d(X_d) \le u_d]$$

Sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from $F = C(F_1, \ldots, F_d)$.

1. Replace population by empirical probability measure:

$$\tilde{C}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{F_1(X_{i1}) \le u_1, \ldots, F_d(X_{id}) \le u_d\}$$

2. Replace unknown margins by empirical cdfs: empirical copula

$$\hat{C}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\hat{F}_{n,1}(X_{i1}) \le u_1, \ldots, \hat{F}_{n,d}(X_{id}) \le u_d\}$$

Properties of the empirical copula

- Multivariate cdf supported on n points in the grid $\{1/n, 2/n, \ldots, 1\}^d$
- Margins: discrete uniform on $\{1/n, 2/n, \ldots, 1\}$
- Not a copula! But close...
- Based on ranks
- Invariant under component-wise increasing transformations of the data

Ex. Make a picture of the support of the empirical copula of a bivariate sample of size $n = 4$.

Alternative versions of the empirical copula

The versions below differ from $\hat{C}_n$ no more than $O(1/n)$:

- Avoid boundary problems by dividing by $n+1$ rather than by n in the definition of the marginal empirical cdfs
- Obtain a genuine copula by eiter
  - smoothing out point masses to obtain uniform $(0,1)$ margins, i.e. subtract from each component an independent Uniform$(0,1/n)$ random variable: checkerboard copula
  - or convoluting $\hat{C}_n$ with a kernel with standard deviation $O(1/n)$ and transforming back to Uniform$(0,1)$ margins
- Simplify asymptotic analysis by using generalized inverses:

$$\boldsymbol{u} \mapsto \hat{F}_n\big(\hat{F}_{n,1}^{\leftarrow}(u_1), \ldots, \hat{F}_{n,d}^{\leftarrow}(u_d)\big)$$

Not knowing the margins makes a difference

If margins were known, we could estimate $C(\boldsymbol{u})$ by

$$\tilde{C}_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{F_1(X_{i1}) \le u_1, \ldots, F_d(X_{id}) \le u_d\}$$

By the central limit theorem,

$$\sqrt{n}\big(\tilde{C}_n(\boldsymbol{u}) - C(\boldsymbol{u})\big) \xrightarrow{d} N\big(0, C(\boldsymbol{u})\,(1 - C(\boldsymbol{u}))\big)$$

Is this still true if we replace $\tilde{C}_n$ by $\hat{C}_n$, i.e. F_j by $\hat{F}_{n,j}$? — *No!*

Ex. Compare the variances of $\tilde{C}_n$ and $\hat{C}_n$ via a simulation study.

- Which of the two has the smaller variance?
- Find an intuitive explanation.

The empirical copula process: A key that opens many doors

- View the collection of random variables

$$\mathbb{C}_n(\boldsymbol{u}) = \sqrt{n}\big(\hat{C}_n(\boldsymbol{u}) - C(\boldsymbol{u})\big), \qquad \boldsymbol{u} \in [0,1]^d$$

  as a stochastic process indexed by $[0,1]^d$: the empirical copula process.
- Knowledge of the limit behavior of $\mathbb{C}_n$ helps to find the limit distribution of any statistic based upon $\hat{C}_n$.
  - Spearman's rho $\rho_S(C) = 12 \int_{[0,1]^2} C(u,v)\,\mathrm{d}(u,v) - 3$, then

$$\sqrt{n}\big(\rho_S(\hat{C}_n) - \rho_S(C)\big) = 12 \int_{[0,1]^2} \mathbb{C}_n(u,v)\,\mathrm{d}(u,v)$$

  - The delta method allows to deal with non-linear functionals in C which are Hadamard differentiable.

Weak convergence of the empirical copula process follows from the functional delta method

Consider the *copula* mapping sending a cdf F to its copula C:

$$\Phi : F \mapsto F(F_1^{\leftarrow}, \ldots, F_d^{\leftarrow}) = C$$

Then

$$\sqrt{n}\big(\hat{C}_n - C\big) = \sqrt{n}\big(\Phi(\hat{F}_n) - \Phi(F)\big)$$

If Φ is (Hadamard-)differentiable at F with derivative $\dot{\Phi}_F$, we find

$$\ldots = \dot{\Phi}_F\big(\underbrace{\sqrt{n}(\hat{F}_n - F)}_{=\alpha_n}\big)$$

with α_n the ordinary empirical process.

The empirical copula process converges weakly to a Gaussian process with continuous trajectories

In an appropriate function space, jointly in $\boldsymbol{u} \in [0, 1]^d$,

$$\sqrt{n}\big(\hat{C}_n(\boldsymbol{u}) - C(\boldsymbol{u})\big) \xrightarrow{d} \alpha(\boldsymbol{u}) - \underbrace{\sum_{j=1}^{d} \alpha(1, \ldots, 1, u_j, 1, \ldots, 1)\, \frac{\partial C(\boldsymbol{u})}{\partial u_j}}_{\text{price for not knowing the margins}} \quad (\lim \hat{C}_n)$$

where $\{\alpha(\boldsymbol{u}) : \boldsymbol{u} \in [0, 1]^d\}$ is a collection of zero-mean Gaussian random variables with

$$\operatorname{cov}\big(\alpha(\boldsymbol{u}), \alpha(\boldsymbol{v})\big) = \operatorname{cov}\Big(\mathbf{1}(\boldsymbol{U} \le \boldsymbol{u}), \mathbf{1}(\boldsymbol{U} \le \boldsymbol{v})\Big) \quad (\operatorname{cov}\alpha)$$

Assumption: the partial derivatives $\partial C(\boldsymbol{u}) / \partial u_j$ exist and are continuous on the domain $\{\boldsymbol{u} \in [0, 1]^d : 0 < u_j < 1\}$.

The weak limit of the empirical copula process: Understanding its covariance function

Ex. Calculate the covariance on the rhs of $(\mathrm{cov}\alpha)$.

Ex. Calculate the variance of the random variable on the rhs on $(\lim \hat{C}_n)$.

Ex. The distribution on the rhs of $(\lim \hat{C}_n)$ is zero-mean normal with variance stemming from the previous exercise. Compare the asymptotic distribution of $\mathbb{C}_n$ with the finite-sample distribution computed from Monte Carlo simulations.

Ex. Use the central limit theorem to show that for $\tilde{C}_n$ (known margins) rather than $\hat{C}_n$, the limit in $(\lim \hat{C}_n)$ would just be $\alpha(\boldsymbol{u})$.

Empirical copulas: Some literature I

Bücher, A. and S. Volgushev (2013). Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis 119*, 61–70.

Fermanian, J.-D., D. Radulović, and M. H. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli 10*, 847–860.

Genest, C. and J. Segers (2010). On the covariance of the asymptotic empirical copula process. *Journal of Multivariate Analysis 101*, 1837–1845.

Rémillard, B. and O. Scaillet (2009). Testing for equality between two copulas. *Journal of Multivariate Analysis 100*(3), 377–386.

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics 4*(5), 912–923.

Segers, J. (2012). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli 18*(3), 764–782.

Stute, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *The Annals of Probability 12*(2), 361–379.

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

Empirical copulas: Some literature II

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

van der Vaart, A. and J. A. Wellner (2007). Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems*, Volume 55 of *IMS Lecture Notes–Monograph Series*, pp. 234–252.

Copulas: An Introduction
III - Inference

Inference on measures of association

The empirical copula

Inference on parametric copula families

Shape-constrained inference: Extreme-value copulas

Why parametric copula families?

- Avoid curse of dimension: more accurate inference.
- Hopefully interpretable parameters in ‘natural’ model.
- Allow for covariates.
- Can help solving identifiability issues:
  - Discrete data
  - Censoring

Estimation strategies: there's choice

Moment-type estimators. For instance, if the map $\theta \mapsto \tau(C_\theta)$ is one-to-one, define

$$\hat{\theta}_n = \theta(\hat{\tau}_n)$$

Minimum-distance estimators. For instance w.r.t. an L^2 norm,

$$\hat{\theta}_n = \arg\min_\theta \int (\hat{C}_n - C_\theta)^2$$

Likelihood based procedures.

Depending on the assumptions on the margins, we obtain a parametric or a semiparametric model

Consider a parametric copula family $\{c_\theta : \theta \in \Theta\}$, with $\Theta \subset \mathbb{R}^k$.

The final model for f depends on the assumptions on the margins:

- If we just assume the margins to be absolutely continuous, the model is semiparametric: joint pdf

$$f(x_1, \ldots, x_d) = \underbrace{c_\theta\big(F_1(x_1), \ldots, F_d(x_d)\big)}_{\text{parametric}} \underbrace{f_1(x_1) \ldots f_d(x_d)}_{\text{nonparametric}}$$

- If we assume parametric models $\{f_j(\,\cdot\,; \theta_j) : \eta_j \in H_j\}$ for the margins, the model is fully parametric: joint pdf

$$\begin{aligned} &f(x_1, \ldots, x_d; \theta, \eta) \\ &\quad = \underbrace{c_\theta\big(F_1(x_1; \eta_1), \ldots, F_d(x_d; \eta_d)\big)}_{\text{parametric}} \underbrace{f_1(x_1; \eta_1) \ldots f_d(x_d; \eta_d)}_{\text{nonparametric}} \end{aligned}$$

The maximum likelihood estimator: maximizing the loglikelihood

$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} f(\,\cdot\,; \theta, \eta)$.
Maximum likelihood estimator: Joint optimisation over θ and η

$$(\hat{\theta}_n, \hat{\eta}_n) = \arg\max_{\theta,\eta} \sum_{i=1}^{n} \Big\{ \log c_\theta\big(F_1(X_{i1}; \eta_1), \ldots, F_d(X_{id}; \eta_d)\big) + \sum_{j=1}^{d} \log f_j(X_{ij}; \eta_j) \Big\}$$

Asymptotic normality: under regularity conditions,

$$\sqrt{n}(\hat{\theta}_n - \theta, \hat{\eta}_n - \theta) \xrightarrow{d} N\big(0, I^{-1}(\theta, \eta)\big)$$

with $I(\theta, \eta)$ the Fisher information matrix

Avoiding high-dimensional optimization, treat margins and copula separately

1. Perform inference for each margin separately:

$$\hat{\eta}_{n,j} = \arg\max_{\eta_j} \sum_{i=1}^{n} \log f_j(x_j; \eta_j), \qquad j \in \{1, \ldots, d\}$$

2. Pretend margin parameters are known and estimate θ:

$$\hat{\theta}_n = \arg\max_{\theta} \sum_{i=1}^{n} \log c_\theta\big(F_1(X_{i1}; \hat{\eta}_{n,1}), \ldots, F_d(X_{id}; \hat{\eta}_{n,d})\big)$$

- Easier to compute than full maximum likelihood estimator.
- Asymptotically normal too.
- A (little) less efficient than the maximum likelihood estimator.

Even simpler computationally: composite likelihoods

Sometimes, even the copula density is hard to compute

- High-dimensional extreme-value copulas (spatial extremes)

If the parameter (vector) θ is determined by the pairwise distributions, replace $\log c_\theta(\boldsymbol{u})$ by a weighted sum of bivariate log densities:

$$\boldsymbol{u} \mapsto \sum_{1 \le j_1 < j_2 \le d} w_{j_1 j_2} \, \log c_{j_1 j_2}(u_{j_1}, u_{j_d}; \theta)$$

Resulting estimators are still asymptotically normal,
the asymptotic (co)variances depending on the weights $w_{j_1 j_2} \ge 0$

Similar idea for pairwise copula constructions:
perform inference pair by pair

Pseudo-likelihood estimator: Estimate margins by empirical cdfs

Semiparametric model for a d-variate density f:

- No assumptions on the marginal pdfs $f_1, \ldots, f_d$
- Copula density c belongs to a parametric family $\{c_\theta : \theta \in \Theta\}$

Log-likelihood for θ given iid sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d \sim f$:

$$\theta \mapsto \sum_{i=1}^{n} \left\{ \log c\big(F_1(X_{i1}), \ldots, F_d(X_{id}); \theta\big) + \sum_{j=1}^{d} f_j(X_{ij}) \right\}$$

Maximum pseudo-likelihood estimator:

$$\hat{\theta}_n = \arg\max_{\theta} \sum_{i=1}^{n} \log c\big(\hat{F}_{n,1}(X_{i1}), \ldots, \hat{F}_{n,d}(X_{id}); \theta\big)$$

with $\hat{F}_{n,j}$ the j-th marginal empirical cdf

Properties of the maximum pseudo-likelihood estimator

+ Based on ranks
+ Asymptotically normal
– In general *not* semiparametrically efficient
  - Exception: Gaussian copula models with certain correlation structures
  - Efficiency loss, if any, is most of the time rather small

? Semiparametrically *efficient* procedure?

Borrowing strength from parametric procedures: Sieve estimator

1. For each margins, choose a sequence of nested parametric models that are dense in the family of all distributions
   - E.g. normal mixtures with m components
2. For a finite sample, fix a parametric model for each margin and estimate parameters (margins and copula) by maximum likelihood
3. Asymptotically, let the marginal models change with the sample size
   - E.g. $m = m_n \to \infty$

\+ Asymptotically normal

\+ Semiparametrically efficient

– Not rank-based

– Requires potentially influential choice of marginal parametric models

? Semiparametrically efficient *rank-based* procedure?

Parametric inference: Some literature

Chen, X., Y. Fan, and V. Tsyrennikov (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association 101*, 1228–1240.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82*, 543–552.

Hobæk Haff, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli 19*, 462–491.

Kojadinovic, I. and J. Yan (2010). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics 47*, 52–63.

Lawless, J. and Y. Yilmaz (2011). Comparison of semiparametric maximum likelihood estimation and two-stage semiparametric estimation in copula models. *Computational Statistics and Data Analysis 55*, 2446–2455.

Copulas: An Introduction
III - Inference

Inference on measures of association

The empirical copula

Inference on parametric copula families

Shape-constrained inference: Extreme-value copulas

Shape-constrained inference problems show up naturally for copulas

Certain copula families are described in terms of lower-dimensional functions subject to shape constraints:

- Archimedean copulas:

$$C(u, v) = \psi\big(\psi^{-1}(u) + \psi^{-1}(v)\big)$$

  with $\psi : [0, 1] \to [0, \infty]$ decreasing, convex, and $\psi(0) = 1$ and $\psi(\infty) = 0$

- Extreme-value copulas:

$$C(u, v) = (uv)^{A(t)}, \qquad t = \frac{\log(v)}{\log(uv)}$$

  with $\max(t, 1 - t) \leq A(t) \leq 1$ and A is convex

- Elliptical copulas, Archimax copulas, …

$\Rightarrow$ nonparametric, shape-constrained inference

Multivariate extreme-value copulas

A copula C is an extreme value copula if

$$C(\boldsymbol{u}) = \exp\{-\ell(-\log u_1, \ldots, -\log u_d)\}, \quad 0 < u_j \leq 1,$$

with stable tail dependence function

$$\ell(\boldsymbol{y}) = \int_{\Delta_{d-1}} \max(y_1 v_1, \ldots, y_d v_d) \, H(\mathrm{d}\boldsymbol{v}),$$

and spectral measure H satisfying

$$\int_{\Delta_{d-1}} v_j \, H(\mathrm{d}\boldsymbol{v}) = 1, \quad j \in \{1, \ldots, d\}.$$

Unit simplex:

$$\Delta_{d-1} = \{(w_1, \ldots, w_d) \in [0, 1]^d : w_1 + \cdots + w_d = 1\}.$$

The stable tail dependence function is determined by the Pickands dependence function

$$C(\boldsymbol{u}) = \exp\{-\ell(-\log u_1, \ldots, -\log u_d)\}$$

1. Homogeneity: $\ell(c\,y_1, \ldots, c\,y_d) = c\,\ell(y_1, \ldots, y_d)$ for $c > 0$.
2. Bounds: $\max(y_1, \ldots, y_d) \le \ell(y_1, \ldots, y_d) \le y_1 + \cdots + y_d$.

It follows that ℓ is determined by its Pickands dependence function A

$$\ell(y_1, \ldots, y_d) = (y_1 + \cdots + y_d)\, A(w_1, \ldots, w_{d-1}),$$

$$\text{where} \qquad w_j = \frac{y_j}{y_1 + \cdots + y_d} \in \Delta_{d-1}.$$

The Pickands dependence function A: An integral transform of a measure

The restriction of ℓ to the unit simplex Δ_{d-1} is given by $A : \Delta_{d-1} \to [1/d, 1]$ and is known as the Pickands dependence function

$$A(\boldsymbol{w}) = \int_{\Delta_{d-1}} \max(w_1\, v_1, \ldots, w_d\, v_d)\, H(\mathrm{d}\boldsymbol{v})$$

with spectral measure H as defined before. Necessarily

1. A is convex;
2. $\max(w_1, \ldots, w_d) \le A(\boldsymbol{w}) \le 1$;
3. and thus $A(\boldsymbol{e}_j) = 1$, for $\boldsymbol{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$.

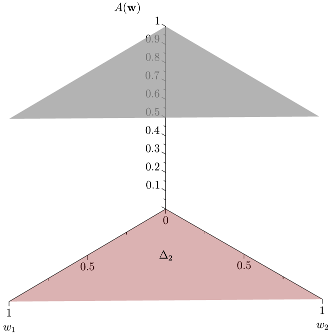
Except if $d = 2$, these properties do *not* characterize the class of Pickands dependence functions.

Plotting the Pickands dependence function A: Independence copula

Independence: $A(\boldsymbol{w}) = 1$

$$C(\boldsymbol{u}) = \exp\left\{ \left(\sum_{j=1}^{3} \log u_j \right) \underbrace{A(\ldots)}_{=1} \right\}$$

$$= u_1 \cdots u_3.$$



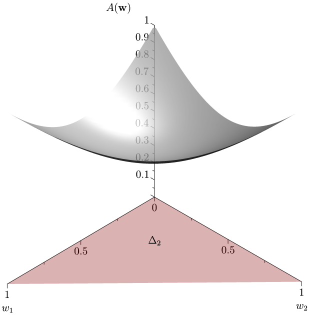
Plotting the Pickands dependence function A: Gumbel copula

Gumbel aka logistic

$$A(\boldsymbol{w}) = (w_1^\theta + w_2^\theta + w_3^\theta)^{1/\theta},$$

with $\theta \geq 1$.

$$C(\boldsymbol{u}) = \exp\Big\{ - \big((-\log u_1)^\theta + \cdots + (-\log u_3)^\theta\big)^{1/\theta} \Big\}$$

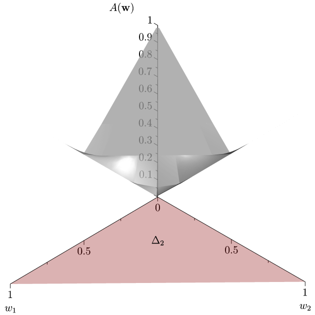


Plotting the Pickands dependence function A: Fréchet–Hoeffding upper bound

FH upper bound

$$A(\boldsymbol{w}) = \max(w_1, w_2, w_3).$$

$$\begin{aligned} C(\boldsymbol{u}) &= \exp\{-\max(-\log u_1, -\log u_2, -\log u_3)\} \\ &= \min(u_1, \ldots, u_3) \end{aligned}$$



The Pickands dependence function as the rate of an exponential distribution

Suppose for the moment the margins $F_1, \ldots, F_d$ are *known*. Put

$$\begin{aligned} \boldsymbol{U_i} &= (U_{i,1}, \ldots, U_{i,d}) \\ &= (F_1(X_{i,1}), \ldots, F_d(X_{i,d})) \end{aligned}$$

For $\boldsymbol{w} \in \Delta_{d-1}$, define

$$\xi_i(\boldsymbol{w}) = \min\left(-\frac{\log U_{i,1}}{w_1}, \ldots, -\frac{\log U_{i,d}}{w_d} \right).$$

The distribution of $\xi_i(\boldsymbol{w})$ is exponential with mean $1/A(\boldsymbol{w})$:

$$\begin{aligned} \mathrm{P}[\xi_i(\boldsymbol{w}) > x] &= \mathrm{P}[U_{i,1} < e^{-w_1 x}, \ldots, U_{i,d} < e^{-w_d x}] \\ &= C(e^{-w_1 x}, \ldots, e^{-w_d x}) = e^{-x A(\boldsymbol{w})}. \end{aligned}$$

The exponential representation suggests nonparametric estimators for A

Let $\hat{\xi}_{i,n}(\boldsymbol{w})$ be as $\xi_i(\boldsymbol{w})$, with F_j replaced by $\hat{F}_{j,n}$: rank-based.

- Pickands (1981)

$$\frac{1}{\hat{A}^P(\boldsymbol{w})} = \frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_{i,n}(\boldsymbol{w}).$$

- Capéraà, Fougères and Genest (1997)

$$\log \hat{A}^{CFG}(\boldsymbol{w}) = -\frac{1}{n}\sum_{i=1}^{n} \log \hat{\xi}_{i,n}(\boldsymbol{w}) - \gamma$$

with the Euler–Mascheroni constant $\gamma = 0.5772\ldots$.

Connection with the empirical copula

If margins are unknown, estimate them by the empirical distribution functions and proceed as before. For every $\boldsymbol{w} \in \Delta_{d-1}$, we have:

$$n^{1/2}\left(\frac{1}{\hat{A}_n^{\mathrm{P}}(\boldsymbol{w})} - \frac{1}{A(\boldsymbol{w})}\right) = \int_0^1 \mathbb{C}_n(u^{w_1}, \ldots, u^{w_p})\, \frac{\mathrm{d}u}{u},$$

$$n^{1/2}(\log \hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - \log A(\boldsymbol{w})) = \int_0^1 \mathbb{C}_n(u^{w_1}, \ldots, u^{w_p}) \frac{\mathrm{d}u}{u \log u}.$$

where

$$\hat{C}_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\left(\hat{U}_{i,1} \le u_1, \ldots, \hat{U}_{i,d} \le u_d\right) \qquad \text{empirical copula}$$

$$\mathbb{C}_n = \sqrt{n}\left(\hat{C}_n - C\right), \qquad \text{empirical copula process}$$

The estimator satisfies a functional central limit theorem

Under reasonable smoothness conditions on A,

$$\mathbb{C}_n \xrightarrow{d} \mathbb{C} \qquad \text{in } \ell^\infty([0,1]^p)$$

As a consequence, in the space $\left(\mathcal{C}(\Delta_{d-1}), \|\cdot\|_\infty\right)$

$$\sqrt{n}\left(\hat{A}_n^{\mathrm{P}}(\boldsymbol{w}) - A(\boldsymbol{w})\right) \xrightarrow{d} -A^2(\boldsymbol{w}) \int_0^1 \mathbb{C}(u^{w_1}, \ldots, u^{w_d}) \, \frac{\mathrm{d}u}{u}$$

$$\sqrt{n}\left(\hat{A}_n^{\mathrm{CFG}}(\boldsymbol{w}) - A(\boldsymbol{w})\right) \xrightarrow{d} A(\boldsymbol{w}) \int_0^1 \mathbb{C}(u^{w_1}, \ldots, u^{w_d}) \frac{\mathrm{d}u}{u \, \log u}$$

How to ensure that the obtained estimator is a valid Pickands dependence function?

Nonparametric estimators do not necessarily provide valid estimates for A.

- A should be convex;
- Bounds: $\max(w_1, \ldots, w_d) \leq A(\boldsymbol{w}) \leq 1$;
- If $d \geq 3$, the previous conditions do not even characterize the set of Pickands dependence functions.

Integral representation of Pickands functions

The class of Pickands dependence functions $\mathcal{A}$:
The collection of all functions A on Δ_{d-1} such that

$$A(\boldsymbol{w}) = \int_{\Delta_{d-1}} \max(w_1 v_1, \ldots, w_d v_d) \, H(\mathrm{d}\boldsymbol{v}), \qquad \boldsymbol{w} \in \Delta_{d-1}$$

for some Borel measure H defined on Δ_{d-1} satisfying

$$\int_{\Delta_{d-1}} v_j \, \mathrm{d}H(\boldsymbol{v}) = 1, \qquad j \in \{1, \ldots, d\}.$$

$\mathcal{A}$ is a closed convex subset of the Hilbert space of $L^2(\Delta_{d-1})$

Enforce the shape constraints on a pilot estimate by projecting it onto the appropriate set of functions

Initial (nonparametric) estimator $\hat{A}_n$ for A.

Projection $\hat{A}^{\mathrm{Pr}}$ of $\hat{A}_n$ on $\mathcal{A}$ w.r.t. the norm $\|\cdot\|_2$:

$$\hat{A}^{\mathrm{Pr}} = \Pi(\hat{A}_n|\mathcal{A}) = \arg\min_{A \in \mathcal{A}} \|\hat{A}_n - A\|_2.$$

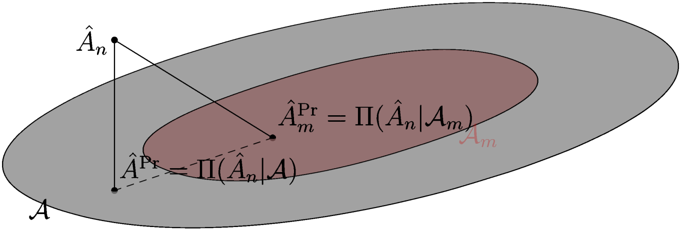
But $\mathcal{A}$ is an infinite-dimensional set
$\Rightarrow$ How to implement the projection?

Solution

Dense sequence of finite-dimensional subclasses $\mathcal{A}_m \subset \mathcal{A}$

$$\hat{A}^{\mathrm{Pr}}_m = \Pi(\hat{A}_n|\mathcal{A}_m) = \arg\min_{A \in \mathcal{A}_m} \|\hat{A}_n - A\|_2, \qquad m \in \mathbb{N}$$

Enforce the shape constraints on a pilot estimate by projecting it onto the appropriate set of functions

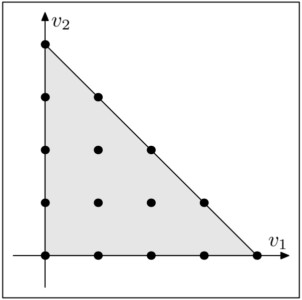


Finite-dimensional approximations: Discrete spectral measures supported on a grid

$$\mathcal{V}_{d,m} = \{\boldsymbol{v} \in \Delta_{d-1} : \boldsymbol{v} = (k_1/m, \ldots, k_d/m), \quad k_j \in \{0, \ldots, m\}, \ \ k_1 + \cdots + k_d = m\}$$

$$A_m(\boldsymbol{w}) \in \mathcal{A}_m \quad \text{if} \quad H = \sum_{\boldsymbol{v} \in \mathcal{V}_{d,m}} H(\{\boldsymbol{v}\})\, \delta_{\boldsymbol{v}}.$$

$$A_m(\boldsymbol{w}) = \sum_{\boldsymbol{v} \in \mathcal{V}_{d,m}} H(\{\boldsymbol{v}\}) \max\{w_1 v_1, \ldots, w_p v_p\}.$$



Computing the projection on the subclass: Solving a quadratic program

Find $\boldsymbol{h} = (\boldsymbol{h}_{\boldsymbol{v}})_{\boldsymbol{v} \in \mathcal{V}_{d,m}}$ minimizing the least-squares criterion

$$\arg\min_{\boldsymbol{h}} \int_{\Delta_{d-1}} \left(\hat{A}_n(\boldsymbol{w}) - \hat{A}_m^{\mathrm{Pr}}(\boldsymbol{w}) \right)^2 \mathrm{d}w_1 \ldots \mathrm{d}w_{d-1},$$

where

$$\hat{A}^{\mathrm{Pr}}(\boldsymbol{w}) = \sum_{\boldsymbol{v} \in \mathcal{V}_{d,m}} h_{\boldsymbol{v}} \max(w_1 v_1, \ldots, w_d v_d), \quad \boldsymbol{w} \in \Delta_{d-1},$$

satisfying the linear constraints

$$h_{\boldsymbol{v}} \geq 0, \quad \forall \boldsymbol{v} \in \mathcal{V}_{d,m}$$

$$\hat{A}^{\mathrm{Pr}}(\boldsymbol{e_j}) = \sum_{\boldsymbol{v} \in \mathcal{V}_{d,m}} h_{\boldsymbol{v}} \max(e_{j,1} v_1, \ldots, e_{j,d} v_d) = 1, \quad \forall j \in \{1, \ldots, d\}$$

Asymptotic distribution of the projected estimator: Project on the tangent cone

If

$$\sqrt{n}(\hat{A}_n - A) \xrightarrow{d} \mathbb{A}, \quad \text{in} \quad L^2(\Delta_{d-1}),$$

then, provided $m = m_n \to \infty$ such that $\sqrt{n}/m_n \to 0$,

$$\sqrt{n}(\hat{A}_m^{\mathrm{Pr}} - A) \xrightarrow{d} \Pi\big(\mathbb{A} \mid \mathrm{T}_{\mathcal{A}}(A)\big) \quad \text{in} \quad L^2(\Delta_{d-1}), \qquad n \to \infty,$$

with $\mathrm{T}_{\mathcal{A}}$ the tangent cone of $\mathcal{A}$ at A:

$$\mathrm{T}_{\mathcal{A}} = \overline{\{\lambda(\tilde{A} - A) : \lambda \geq 0, \tilde{A} \in \mathcal{A}\}}$$

Open problem: Limit distribution in $\big(\mathcal{C}(\Delta_{d-1}), \|\cdot\|_\infty\big)$?

Shape-constrained inference on copulas: Some literature

Berghaus, B., A. Bücher, and H. Dette (2013). Minimum distance estimators of the Pickands dependence function and related tests of multivariate extreme-value dependence. *Journal de la Société Française de Statistique 154*, 116–137.

Gijbels, I., M. Omelka, and D. Sznajder (2010). Positive quadrant dependence tests for copulas. *Canadian Journal of Statistics 38*(4), 555–581.

Gudendorf, G. and J. Segers (2012). Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference 142*, 373–385.

Guillotte, S. and F. Perron (2008). A Bayesian estimator for the dependence function of a bivariate extreme-value distribution. *The Canadian Journal of Statistics 36*(3), 383–396.

Lambert, P. (2007). Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics & Data Analysis 51*, 6307–6320.