Nonparametric Identification of Copula Structures

Marc G. Genton
marc.genton@kaust.edu.sa
stsda.kaust.edu.sa

King Abdullah University of Science and Technology

Based on joint work with Bo Li

May 21, 2013

KAUST

- New graduate-level university located 50 miles north of Jeddah
- On the Red Sea
- Western style campus (14 miles^2) and encourages cultural diversity
- First classes in Fall 2009
- About 900 students & 120 faculty (will grow to 2000 & 220)
- More at: www.kaust.edu.sa
- Partnership with TAMU through IAMCS
- Recruting students, postdocs and faculty in statistics
- Current President of CalTech will be next KAUST President in Fall 2013

KAUST: located in Thuwal

KAUST: campus

KAUST: main buildings

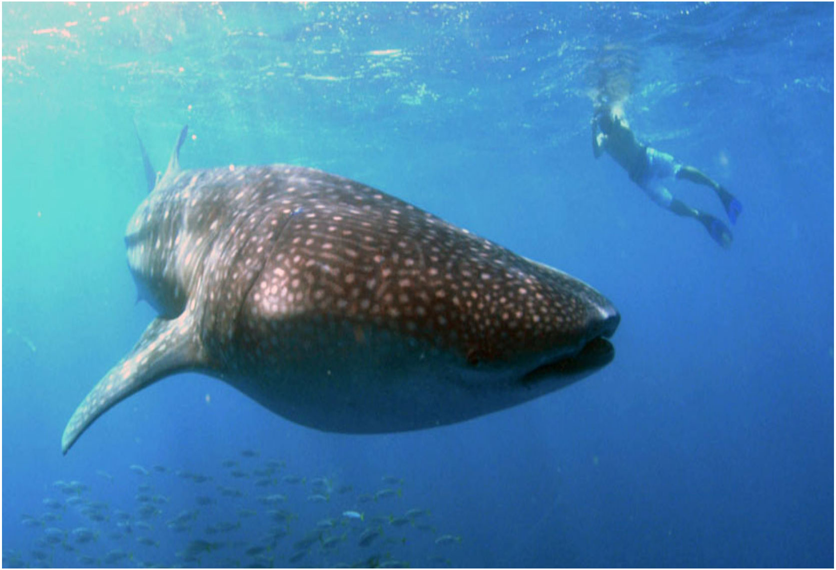
KAUST: the Beacon

KAUST: the Beacon & sunset

KAUST: on the Red Sea

KAUST: on the Red Sea

KAUST: whale sharks project



Motivation

- Many parametric copula models available in the literature
- Mikosch (2006): How does one choose a copula?
- Often choose copula models that are mathematically convenient rather than useful for the data at hand
- Goodness-of-fit tests for copulas
- We aim at testing the *structure* of copulas

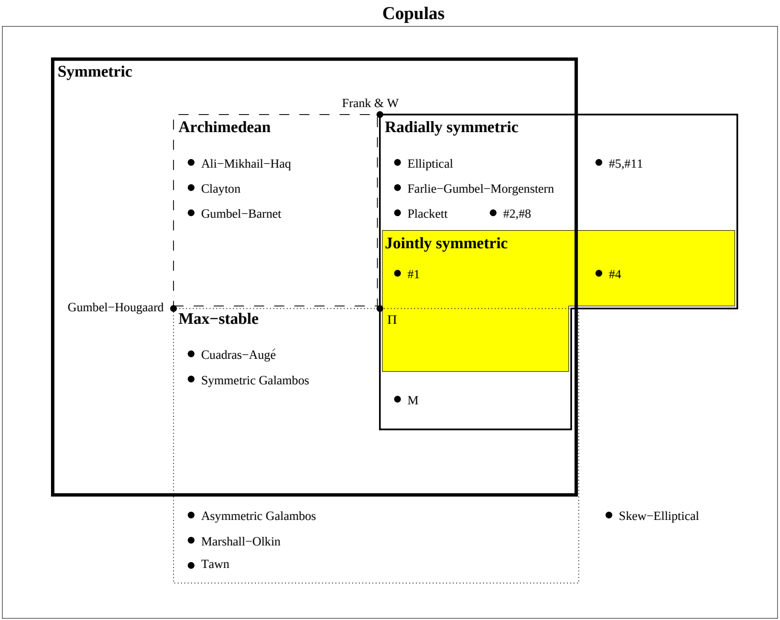
Specific Tests

- Jaworski (2010): test for associativity structure based on asymptotic distribution of pointwise copula estimator
- Bücher, Dette, and Volgushev (2011): test for extreme-value dependence
- Bücher, Dette, and Volgushev (2012): test for associativity and Archimedeanity
- Genest, Nešlehová, and Quessy (2012): test for bivariate symmetry
- We propose a unified framework for testing a variety of assumptions commonly made for the structure of copulas, including symmetry, radial symmetry, joint symmetry, associativity and Archimedeanity, and max-stability
- Our test is nonparametric and based on the asymptotic distribution of the empirical copula process

Copula structures

- Symmetry: $C(u_1, u_2) - C(u_2, u_1) = 0$
- Radial symmetry:
  $C(u_1, u_2) - C(1 - u_1, 1 - u_2) + 1 - u_1 - u_2 = 0$
- Joint symmetry: $C(u_1, u_2) + C(u_1, 1 - u_2) - u_1 = 0$ and $C(u_1, u_2) + C(1 - u_1, u_2) - u_2 = 0$
- Archimedean: $C_\phi(u_1, \ldots, u_d) = \phi^{[-1]}\{\phi(u_1) + \cdots + \phi(u_d)\}$
- Archimedean copulas are symmetric and *associative*, i.e., for $d = 2$, $C(u_1, u_2) = C(u_2, u_1)$ and $C\{C(u_1, u_2), u_3\} = C\{u_1, C(u_2, u_3)\}$
- Archimedean copula is characterized by an associative copula with $C(u, u) < u$
- Max-stable: $C(u_1, \ldots, u_d) - C^r(u_1^{1/r}, \ldots, u_d^{1/r}) = 0$, for any $r > 0$

Copula structures



Asymptotics

- $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$ and Λ a set of user-chosen points in $[0,1]^d$ of cardinality c
- $\mathbf{X}_i = (X_{1i}, \ldots, X_{di})^{\mathrm{T}}$, $i = 1, \ldots, n$ with corresponding $\widehat{\mathbf{U}}_i = (\widehat{U}_{1i}, \ldots, \widehat{U}_{di})^{\mathrm{T}}$, where $\widehat{U}_{ki} = \frac{1}{n}\sum_{j=1}^{n} I(X_{kj} \le X_{ki})$ for $k = 1, \ldots, d$
- $D_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{k=1}^{d} I(\widehat{U}_{ki} \le u_k)$
- $\sqrt{n}\{\widehat{D}_n(\mathbf{u}) - C(\mathbf{u})\} \xrightarrow{d} U^C(\mathbf{u}) - \sum_{k=1}^{d} \frac{\partial C(\mathbf{u})}{\partial u_k} U^C(\mathbf{1}, u_k, \mathbf{1})$ where $U^C(\mathbf{u})$ is a d-dimensional pinned C-Brownian sheet, i.e. a centered Gaussian random field with $\mathrm{cov}\{U^C(\mathbf{u}_i), U^C(\mathbf{u}_j)\} = C(\mathbf{u}_i \wedge \mathbf{u}_j) - C(\mathbf{u}_i)C(\mathbf{u}_j)$
- $\sqrt{n}(\widehat{\mathbf{H}}_n - \mathbf{H}) \xrightarrow{d} N_c(\mathbf{0}, \mathbf{\Pi})$ as $n \to \infty$, where

$$
\begin{aligned}
\mathbf{\Pi}_{ij} = \quad & C(\mathbf{u}_i \wedge \mathbf{u}_j) - C(\mathbf{u}_i)C(\mathbf{u}_j) - \sum_{k=1}^{d} \frac{\partial C(\mathbf{u}_j)}{\partial u_{kj}} \left[C\{\mathbf{u}_i \wedge (\mathbf{1}, u_{kj}, \mathbf{1})\} - u_{kj} C(\mathbf{u}_i)\right] \\
& - \sum_{k=1}^{d} \frac{\partial C(\mathbf{u}_i)}{\partial u_{ki}} \left[C\{\mathbf{u}_j \wedge (\mathbf{1}, u_{ki}, \mathbf{1})\} - u_{ki} C(\mathbf{u}_j)\right] \\
& + \sum_{m=1}^{d}\sum_{n=1}^{d} \frac{\partial C(\mathbf{u}_i)}{\partial u_{mi}} \frac{\partial C(\mathbf{u}_j)}{\partial u_{nj}} \left[C\{(\mathbf{1}, u_{mi}, \mathbf{1}) \wedge (\mathbf{1}, u_{nj}, \mathbf{1})\} - u_{mi} u_{nj}\right]
\end{aligned}
$$

Test Statistics

- $\mathbf{Af}(\mathbf{G}) = \mathbf{0}$ for a contrast matrix $\mathbf{A}$ and $\mathbf{f} = (f_1, \ldots, f_s)^{\mathrm{T}}$
- $\mathrm{TS1} = n\{\mathbf{Af}(\widehat{\mathbf{G}}_n)\}^{\mathrm{T}}(\mathbf{AB}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{BA}^{\mathrm{T}})^{-1}\{\mathbf{Af}(\widehat{\mathbf{G}}_n)\}$
  $\mathbf{B}$ is defined as $\mathbf{B}_{ij} = \partial f_j / \partial \mathbf{G}_i$, $i = 1, \ldots, c$, $j = 1, \ldots, s$
- $\mathrm{TS2} = n\{\mathbf{Af}(\widehat{\mathbf{H}}_n)\}^{\mathrm{T}}(\mathbf{AB}^{\mathrm{T}}\boldsymbol{\Pi}\mathbf{BA}^{\mathrm{T}})^{-1}\{\mathbf{Af}(\widehat{\mathbf{H}}_n)\}$
- Due to the asymptotic normality of $\widehat{\mathbf{G}}_n$ and $\widehat{\mathbf{H}}_n$, $\mathrm{TS1} \xrightarrow{d} \chi^2_q$ and $\mathrm{TS2} \xrightarrow{d} \chi^2_q$ asymptotically, where q is the row rank of $\mathbf{A}$

Simulations

- Evaluate our testing procedures for various structures of copulas
  - symmetry, radial symmetry, joint symmetry, associativity, and max-stability
- Compare our method to other tests that have been developed for assessing particular structures
  - Symmetry test – Genest et al. (2012),
  - Associativity test – Bücher et al. (2012)
  - Max-stability test – Bücher et al. (2011)

Settings: 1,000 replicates, with unknown marginals, nominal level of all tests is 5%

Symmetry test

How to convert a symmetric copula to asymmetric?

- Following Genest et al. (2012) (GNQ), an asymmetric version of a copula, $C(u_1, u_2)$, can be defined at all $(u_1, u_2) \in [0, 1]^2$ by Khoudraji's device (Khoudraji, 1985): $K_\delta(u_1, u_2) = u_1^\delta C(u_1^{1-\delta}, u_2)$, for $\delta \in (0, 1)$
- Khoudraji's device provides little asymmetry for Kendall's $\tau \leq 1/2$, and the maximum asymmetry occurs around $\delta = 1/2$

Symmetry test

Table: Sizes and powers of the test of symmetry in the same setting as Genest et al. (2012)

			Clayton			Gaussian			Gumbel		
			n			n			n		
	δ	τ	250	500	1000	250	500	1000	250	500	1000
S		.25	0.154	0.098	0.078	0.164	0.096	0.063	0.154	0.112	0.075
I	0	0.5	0.076	0.057	0.037	0.069	0.047	0.039	0.068	0.053	0.056
Z		.75	0.013	0.003	0.008	0.009	0.005	0.001	0.002	0.004	0.005
E	0	0*	0.160	0.104	0.075	—	—	—	—	—	—
		0.5	0.251	0.299	0.607	0.174	0.222	0.349	0.200	0.262	0.477
	$\frac{1}{4}$	0.7	0.813	0.990	1.000	0.495	0.865	1.000	0.610	0.929	0.999
P		0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
O		0.5	0.344	0.381	0.616	0.323	0.434	0.739	0.460	0.689	0.958
W	$\frac{1}{2}$	0.7	0.835	0.997	1.000	0.892	0.997	1.000	0.971	1.000	1.000
E		0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
R		0.5	0.235	0.189	0.230	0.322	0.378	0.634	0.474	0.670	0.940
	$\frac{3}{4}$	0.7	0.423	0.572	0.882	0.726	0.945	0.999	0.839	0.997	1.000
		0.9	0.938	0.999	1.000	0.995	1.000	1.000	0.990	1.000	1.000

*This indicates the sizes for the independent copula Π ($\delta = \tau = 0$)

Symmetry test

Remarks:

- Under small and moderate τ, the sizes converge to the nominal value as the sample size increases
- With large τ, the sizes are somewhat below the nominal level
- All the powers increase as the sample size increases
- Compared to GNQ, some of our powers at small sample sizes are not as good as their powers, but our powers with large sample sizes are superior to theirs

Radial and joint symmetry test

Table: Sizes and powers of the test of **radial** and **joint** symmetry

	τ		Radial n=250	Radial n=500	Radial $n = 10^3$		Joint n=250	Joint n=500	Joint $n = 10^3$
Π	0	SIZE	0.057	0.067	0.058	Size	0.077	0.083	0.063
Frank	1/4	SIZE	0.066	0.072	0.065	POWER	0.995	1.000	1.000
	1/2	SIZE	0.080	0.073	0.056	POWER	1.000	1.000	1.000
	3/4	SIZE	0.027	0.038	0.049	POWER	1.000	1.000	1.000
Gaussian	1/4	SIZE	0.088	0.064	0.050	POWER	0.989	1.000	1.000
	1/2	SIZE	0.091	0.058	0.056	POWER	1.000	1.000	1.000
	3/4	SIZE	0.026	0.045	0.031	POWER	1.000	1.000	1.000
Clayton	1/4	POWER	0.372	0.624	0.930	POWER	0.993	1.000	1.000
	1/2	POWER	0.801	0.983	1.000	POWER	1.000	1.000	1.000
	3/4	POWER	0.861	0.999	1.000	POWER	1.000	1.000	1.000
Gumbel	1/4	POWER	0.118	0.164	0.307	POWER	0.882	0.996	1.000
	1/2	POWER	0.216	0.358	0.681	POWER	1.000	1.000	1.000
	3/4	POWER	0.178	0.378	0.754	POWER	1.000	1.000	1.000

(1) Sizes are close to 5% for all different τ even at small sample sizes, and powers increase as the sample sizes increase (2) Powers corresponding to joint symmetry are much higher than those for radial symmetry, because joint symmetry is more stringent than radial symmetry

Associativity test

Table: Sizes and powers of the test of associativity in the same setting as Bücher et al. (2012)(hereafter BDV12). Nominal levels are 5% and 10%

		τ	n =200 5%	n =200 10%	n =500 5%	n =500 10%	n =1000 5%	n =1000 10%
SIZE	Clayton	1/3	0.073	0.131	0.096	0.159	0.089	0.154
	Clayton	2/3	0.019	0.042	0.038	0.066	0.032	0.076
	Gumbel	1/3	0.056	0.104	0.093	0.170	0.101	0.181
	Gumbel	2/3	0.015	0.032	0.012	0.038	0.029	0.055
	Ordinal_A	1/3	0.012	0.023	0.022	0.034	0.027	0.059
	Ordinal_A	2/3	0.009	0.018	0.022	0.039	0.019	0.037
	Ordinal_B	1/3	0.013	0.025	0.014	0.025	0.017	0.042
	Ordinal_B	2/3	0.013	0.023	0.011	0.029	0.016	0.033
POWER	t(df=1)	1/3	0.570	0.701	0.976	0.988	1.000	1.000
	t(df=1)	2/3	0.166	0.260	0.658	0.773	0.986	0.994
	Aneglog	1/4	0.130	0.205	0.121	0.198	0.135	0.221
	Aneglog	1/2	0.435	0.555	0.718	0.797	0.901	0.942

Associativity test

Remarks:

- The overall pattern of our results is very similar to BDV12
- In all cases but two, the sizes tend to be smaller than the nominal sizes, which is not detrimental and is similar to that in BDV12
- Since our test still relies on the asymptotic distribution of the test statistic, it is more powerful with large samples
- The test for the other component of Archimedeanity, $C(u, u) < u$ versus $C(u, u) = u$, completely follows the testing procedure in BDV12

Max-stability test

Table: Sizes and powers of the test of max-stability in the same setting as Bücher et al. (2011). Nominal levels are 5% and 10%

		τ	n =200 5%	n =200 10%	n =500 5%	n =500 10%	n =1000 5%	n =1000 10%
SIZE	Π	0	0.099	0.166	0.069	0.121	0.045	0.100
	Gumbel	0.25	0.109	0.172	0.067	0.124	0.053	0.093
		0.5	0.092	0.152	0.048	0.107	0.046	0.094
		0.75	0.023	0.056	0.023	0.038	0.038	0.058
POWER	Clayton	0.25	0.603	0.709	0.954	0.976	1.000	1.000
		0.5	0.947	0.969	1.000	1.000	1.000	1.000
		0.75	0.926	0.960	1.000	1.000	1.000	1.000
	Frank	0.25	0.191	0.289	0.308	0.435	0.708	0.810
		0.5	0.260	0.391	0.625	0.743	0.975	0.989
		0.75	0.228	0.344	0.657	0.786	0.987	0.995
	Gaussian	0.25	0.176	0.268	0.219	0.317	0.402	0.528
		0.5	0.165	0.243	0.303	0.435	0.685	0.795
		0.75	0.048	0.107	0.177	0.253	0.438	0.557

Remarks:

- Although the sizes of our test are somewhat off for $n = 200$, they appear to converge to their corresponding nominal levels when n becomes larger
- The powers increase dramatically as the sample size increases, particularly if the powers begin with low values with small sample sizes
- The powers of our test with small sample sizes are not impressive compared to Bücher et al. (2011), but they become satisfactory with large samples due to the asymptotic characteristics of the test statistics

The size of Λ

- The choice of testing points in the set Λ can introduce uncertainty to the testing results
- In general, a larger number of testing points leads to greater power, but excessively many testing points can destroy the size of the test
- Data adaptive testing points procedure
  - A primitive design of testing points: grid expanded by an equally spaced sequence in [0,1]
  - Choose only the testing points that anchor in a region of data abundance, while abandon the ones in areas of data scarcity
  - Particularly important when the data concentrate in a certain area

The size of Λ: an example of DATP

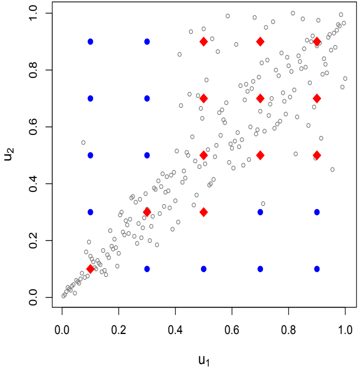


Figure: Illustration of data adaptive testing points (DATP) in the max-stability test: solid points and diamonds are primitive testing points, diamonds are testing points after applying DATP (Clayton, $n = 200$)

The powers are 0.807 (0.926) at level 5% and 0.864 (0.960) at level 10% with Primitive testing points (DATP)

Applications: Nutritional Habits Survey Data

- Collected by the U.S. Department of Agriculture in 1985 as part of a survey on nutritional habits of $n = 737$ women with ages ranging from 25 to 50 years
- Five variables of daily intakes were measured: calcium (in mg), iron (in mg), protein (in g), vitamin A (in μg), and vitamin C (in mg)
- Genest et al. (2012) used a Cramér-von Mises statistic to test for bivariate symmetry of the pairwise copulas

Applications: Nutritional Habits Survey Data

Table: P-values of test of symmetric copula structure on pairs of variables

Variable	Calcium	Iron	Protein	Vitamin A	Vitamin C
Calcium		0.061	0.013	0.002	0.591
Iron			0.766	0.025	0.086
Protein				0.254	0.123
Vitamin A					0.813

- The general pattern of our p-values is similar to the one found by Genest et al. (2012)
- However, our test does not reject a symmetric copula structure at a 5% level for the pairs (Calcium, Iron), (protein, vitamin A) and (iron, vitamin C), although the p-value for the test on (Calcium, Iron) is really on the boundary
- Our conclusion on these three pairs also remains the same when different sets of testing points are used

Applications: S&P 500 and DAX Return Data

- 396 observations of two major stock indices during 2009 and 2010: the US American S&P 500 and the German DAX
- The goal is to identify the underlying copula structure in order to propose a suitable parametric copula model for the dependence of those two indices

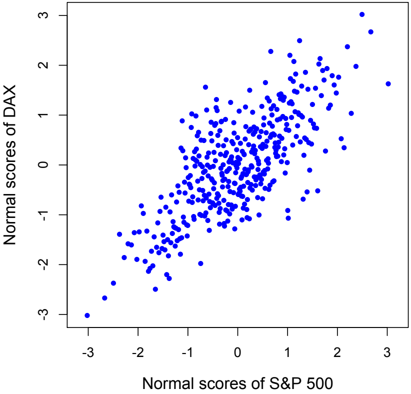


Figure: Scatter plot of normal scores for S&P 500 and DAX returns

Applications: S&P 500 and DAX Return Data

- P-values of our tests of: symmetry (0.172), radial symmetry (0.082), joint symmetry (0.000), max-stability (0.009), and associativity (0.113)
- The results suggest that, at the 5% level, symmetry, radial symmetry, and associativity of the copula structures cannot be rejected. The Archimedeanity test based on Bücher et al. (2012) also fails to reject this property even at the 10% level
- Only Frank Copula has those three structures
- Our method provides guidance for selecting a parametric copula model. e.g., the Student's t copula, is symmetric and radially symmetric, but not Archimedean, and thus is screened out

Discussion and open problems

- Unified framework to test structure of copulas
- Implement bootstrap procedure for those tests
- Asymptotics under dependence
- Applications to time series and spatial statistics
- Other interesting structures
- Extension to vine copulas