Comparing copulas

Denys Pommeret

ISFA, Lab. SAF

joint work with Yves Ngounou (Aix-Marseille University, lab. I2M)

ECODEP 2021

Outline

- ↪ Copulas in ecology
- ↪ Copula coefficients
- ↪ Two-sample case
- ↪ K-sample case
- ↪ Illustration

Why copulas?

Let $\mathbf{X} = (X_1, \cdots, X_p)$ be a p-dimensional continuous random variable with joint probability distribution function $F_{\mathbf{X}}$. We have

$$F_{\mathbf{X}}(x_1, \cdots, x_p) \quad = \quad C(F_1(x_1), \cdots, F_p(x_p)),$$

where F_j denote the marginal probability distribution functions of X_j, and C denotes the *copula* associated to $F_{\mathbf{X}}$.

Why copulas?

Writing

$$U_j = F_j(X_j), \qquad \text{for } j = 1, \cdots, p,$$

we have for all $u_j \in (0,1)$,

$$C(u_1, \cdots, u_p) \quad = \quad F_{\mathbf{U}}(u_1, \cdots, u_p),$$

with $\mathbf{U} = (U_1, \cdots, U_p)$, and deriving this expression p times with respect to $u_1, \cdots, u_p$, we get an expression of the *density copula*

$$c(u_1, \cdots, u_p) \quad = \quad f_{\mathbf{U}}(u_1, \cdots, u_p),$$

where $f_{\mathbf{U}}$ denotes the joint density of the vector $\mathbf{U}$.

Copulas in ecology

Very few works:

- de Valpine et al. (2014) *Ecology Letters*
- Anderson et al. (2018) *Ecology and Evolution*
- Popovic et al. (2019) *Methods in Ecology Evolution*
- Ghosh et al. (2020) *Advances in Ecological Research*

Copulas in ecology

But many applications:

- Environmental, ecological and evolutionary processes may commonly generate complex dependence structures, including asymmetric tail associations (not only correlation).
- In Ghosh et al. "We believe copula approaches are among the tools all ecologists should be considering for analysis of their data in the 21st century."

Copulas in ecology

- Liebig's law. Liebig's law is the idea that growth is controlled not by total resources but by the resource which is scarcest relative to organism needs. If, for instance, the growth of a plant depends on soil nitrogen, say N, and other factors, a plot of growth rates vs soil N may look like the next Figure: the two variables may show left-tail association: N controls plant growth, producing a clear relationship, only when it is limiting.

Copulas in ecology

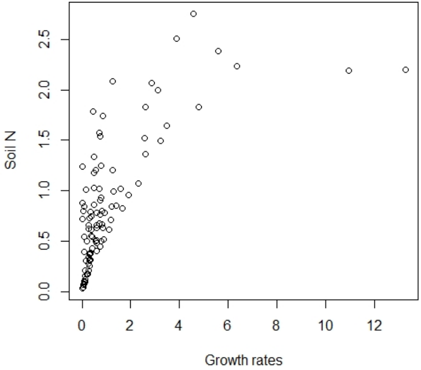

Figure – Liebig's law illustration

Copulas in ecology

- The Moran effect. If asymmetric tail associations, or other complex dependence structure, is transmitted from environmental to ecological variables, then we would expect complex dependence structure and tail associations to be a common feature of the spatial synchrony of population, community, biogeochemical and other environmentally influenced ecological variables. Synchrony attracts major interest in ecology.

Copulas in ecology

- Understanding of tail associations and complex dependence structures is useful for ecology. For instance, if populations of a pest species in different locations are all positively associated and are also more strongly related to each other in their right tails, then local outbreaks will tend to occur together, creating regional epidemics. Stronger left-tail associations in a pest species, even if overall correlation were the same, would have more benign effects.
- Causal mechanisms between variables. Conversely if two species, Sp1 and Sp2, are strong competitors, abundances of the two species can have complex relation as illustrated in the next Figure.

Copulas in ecology

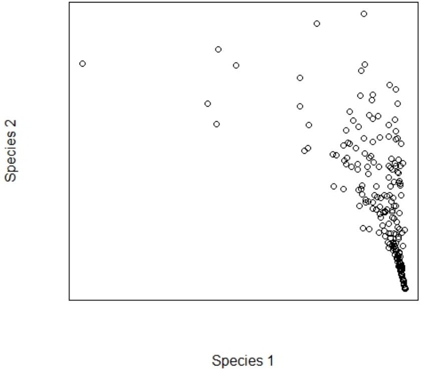

Figure – Sp1 is the dominant competitor: when Sp1 is abundant, Sp2 is necessarily rare because it is suppressed

Example: phytoplankton

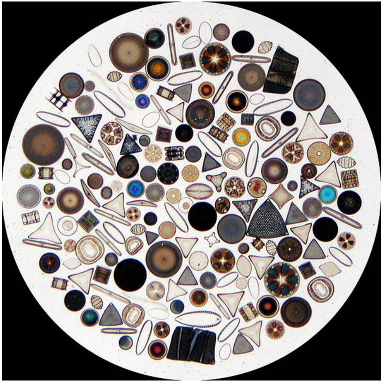

Figure – Phytoplankton can produce 80% of the oxygen

Example: the phytoplankton paradox

Phytoplankton do not satisfy the classical laws of ecology and it is extensively studied to try to understand climatic and environmental evolution.

- For instance Automated Flow Cytometry (FCM) with hourly sampling strategies generates significant phytoplankton datasets.
- To this will be added the new altimetric satellite SWOT (Surface Water Ocean Topography) launched in 2023.

Context

Data $\nearrow$ & Dimension $\nearrow$

$\hookrightarrow$ How to analysis simultaneously such data (from different captors, with different scales, with different shapes)?
$\hookrightarrow$ How can we compare the comportment of various phenomena, possibly paired?
$\hookrightarrow$ How can we classify groups with similar dependance (but not necessary with the same distribution)?
$\hookrightarrow$ How can we compare various copulas simultaneously?

$\hookrightarrow$ **We propose a K-sample test of comparison**

A very short review

- One-sample case: many testing methods have been proposed within the frame of parametric families of copulas (see Can et al. 2020, Bernoulli).
- Two-sample case: the important reference is the work of Remillard and Scaillet (2009, JMVA). It is adapted to the case of paired populations and a package is available
- K-sample case ($K > 2$): there is a theoretical work of Bouzebda et al. (2011, Math. Meth. Stat.) who tried to extend Remillard and Scaillet in the independent case. But the limit distribution of their statistic seems intractable for $K > 2$. More recently Derumigny et al. (2021, Arxiv) tackled the testing problem for conditional copulas.

Hypotheses testing

Assume that we observe K iid samples, possibly paired, with associated copulas denoted by $C_1, \cdots, C_K$. We consider the problem of testing the equality

$$H_0 : \qquad C_1 = \cdots = C_K$$

against H_1 : there exists $1 \leq k \neq k' \leq K$ such that $C_k \neq C_{k'}$.

Copulas coefficients

We have the following L^2 Legendre decomposition

$$f_{\mathbf{U}}(u_1, \cdots, u_p) \quad = \quad \sum_{j_1, \cdots, j_p \in \mathbb{N}} \rho_{j_1, \cdots, j_p} L_{j_1}(u_1) \cdots L_{j_p}(u_p),$$

where

$$\rho_{j_1, \cdots, j_p} \quad = \quad \mathbb{E}(L_{j_1}(U_1) \cdots L_{j_d}(U_p)),$$

as soon as

$$\int_0^1 \cdots \int_0^1 f_{\mathbf{U}}(u_1, \cdots, u_p)^2 du_1 \cdots du_p \quad < \quad \infty.$$

Copulas coefficients

Write $\mathbf{j} = (j_1, \cdots, j_p)$ and $\mathbf{0} = (0, \cdots, 0)$. From the previous equalities we get, for all $u_1, \cdots, u_p \in (0,1)$:

$$\begin{aligned} c(u_1, \cdots, u_p) &= 1 + \sum_{\mathbf{j} \in \mathbb{N}_*^p} \rho_{\mathbf{j}} L_{j_1}(u_1) \cdots L_{j_p}(u_p), \\ C(u_1, \cdots, u_p) &= u_1 \ u_2 \cdots u_p + \sum_{\mathbf{j} \in \mathbb{N}_*^p} \rho_{\mathbf{j}} I_{j_1}(u_1) \cdots I_{j_p}(u_p), \end{aligned}$$

where

$$I_j(u) = \int_0^u L_j(x) dx.$$

Clearly the sequence $(\rho_{\mathbf{j}})_{\mathbf{j} \in \mathbb{N}_*^p}$ characterizes the copula and we call it the *copula coefficients*. Then the comparison of copulas consists in the comparison of these coefficients.

Copulas coefficients

By the previous expansions, testing the null hypothesis H_0 remains to test the equality of all copulas coefficients, that is

$$H_0 : \rho_{\mathbf{j}}^{(1)} = \quad \cdots \quad = \rho_{\mathbf{j}}^{(K)}, \quad \forall \mathbf{j} \in \mathbb{N}_*^p,$$

where $\rho^{(k)}$ stands for the copula coefficients associated to C_k. We propose a test statistic based on the estimation of these quantities.

Estimation procedure

We consider K random vectors, namely

$$\mathbf{X}^{(1)} = (X_1^{(1)}, \cdots, X_p^{(1)}), \cdots, \mathbf{X}^{(K)} = (X_1^{(K)}, \cdots, X_p^{(K)}),$$

with joint cdf $\mathbf{F}^{(1)}, \cdots, \mathbf{F}^{(K)}$, and with associated copulas $C_1, \cdots, C_K$, respectively. Assume that we observe K iid samples from $\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(K)}$, possibly paired, denoted by

$$(X_{i,1}^{(1)}, \cdots, X_{i,p}^{(1)})_{i=1,\cdots,n_1}, \cdots, (X_{i,1}^{(K)}, \cdots, X_{i,p}^{(K)})_{i=1,\cdots,n_K}.$$

We assume that

$$\text{for all } 1 \leq k < \ell \leq K, \qquad n_k/(n_k + n_\ell) \to a_{k\ell} > 0.$$

We denote by $F_j^{(k)}$ the marginal cdf of the jth component of $\mathbf{X}^{(k)}$ and we write

$$U_{i,j}^{(k)} = F_j^{(k)}(X_{i,j}^{(k)}).$$

Estimation procedure

For testing H_0 we first estimate the copula coefficients by

$$\widehat{\rho}^{(k)}_{j_1\cdots j_p} \quad = \quad \frac{1}{n_k}\sum_{i=1}^{n_k} L_{j_1}(\widehat{U}^{(k)}_{i,1})\cdots L_{j_p}(\widehat{U}^{(k)}_{i,p})),$$

where

$$\widehat{U}^{(k)}_{i,j} = \widehat{F}^{(k)}_j(X^{(k)}_{i,j}),$$

and $\widehat{F}$ denotes the empirical distribution functions associated to F.

Estimation procedure

Our test procedure is based on the sequences of differences

$$r_{\mathbf{j}}^{(\ell,m)} := \widehat{\rho}_{\mathbf{j}}^{(\ell)} - \widehat{\rho}_{\mathbf{j}}^{(m)}, \qquad \text{for } 1 \le \ell \le m \le K, \text{ and } \mathbf{j} \in \mathbb{N}_*^p.$$

Estimation procedure

In order to select automatically the number of copula coefficients we introduce the notion of *total order* defined for any vector $\mathbf{j} = (j_1, \cdots, j_p)$ by

$$|\mathbf{j}| = j_1 + \cdots + j_p,$$

and for any integer $d > 1$ we write

$$\begin{aligned} \mathcal{S}(d) &= \{\mathbf{j} \in \mathbb{N}^p; |\mathbf{j}| = d \\ &\qquad \text{and there exists } k \neq k' \text{ such that } j_k > 0 \text{ and } j_{k'} > 0\}. \end{aligned}$$

Estimation procedure

The set $\mathcal{S}(d)$ contains all non null integers $\mathbf{j} = (j_1, \cdots, j_p)$ with total order d and such that $j_k < d$ for $k = 1, \cdots, p$. We denote by $c(d) = \binom{d}{d+p-1} - p$ the cardinal of $\mathcal{S}(d)$ and we define the order $ord(\mathbf{j}, d)$ of $\mathbf{j} \in \mathcal{S}(d)$ as follows:

$$\begin{aligned}
\mathbf{j} = (d-1, 1, 0, \cdots, 0) \quad &\Rightarrow \quad ord(\mathbf{j}, d) = 1 \\
\mathbf{j} = (d-1, 0, 1, \cdots, 0) \quad &\Rightarrow \quad ord(\mathbf{j}, d) = 2 \\
&\dots \\
\mathbf{j} = (0, \cdots, 0, 2, d-2) \quad &\Rightarrow \quad ord(\mathbf{j}, d) = c(d) - 1 \\
\mathbf{j} = (0, \cdots, 0, 1, d-1) \quad &\Rightarrow \quad ord(\mathbf{j}, d) = c(d).
\end{aligned}$$

Estimation procedure

For instance, in the bivariate case, that is $p = 2$, we have

- if $d = 2$ there is only one possibility: $\mathbf{j} = (j_1, j_2) = (1, 1)$ with $ord(\mathbf{j}, 2) = 1$. The cases $(2, 0)$ or $(0, 2)$ are excluded.
- if $d = 3$ there are two possibilities: $\mathbf{j} = (2, 1)$ with $ord(\mathbf{j}, 3) = 1$ and $\mathbf{j} = (1, 2)$ with $ord(\mathbf{j}, 3) = 2$. The cases $\mathbf{j} = (0, 3)$ and $\mathbf{j} = (3, 0)$ are excluded.

Two-sample test

We first propose to test $H_0 : C_1 = C_2$.
We restrict our attention to the paired case and we write $n_1 = n_2 = n$.
For $1 \leq k \leq c(2)$ we define

$$T_{2,k}^{(1,2)} = n \sum_{\mathbf{j} \in \mathcal{S}(2); ord(\mathbf{j},2) \leq k} (r_{\mathbf{j}}^{(1,2)})^2,$$

and for $d > 2$ and $1 \leq k \leq c(d)$,

$$T_{d,k}^{(1,2)} \quad = \quad T_{d-1,c(d-1)}^{(1,2} + n \sum_{\mathbf{j} \in \mathcal{S}(d); ord(\mathbf{j},d) \leq k} (r_{\mathbf{j}}^{(1,2)})^2.$$

Two-sample test

To simplify notation we write such a sequence of statistics as

$$V_1^{(1,2)} = T_{2,1}^{(1,2)};\ V_2^{(1,2)} = T_{2,2}^{(1,2)}; \cdots\ V_k^{(1,2)} = \cdots$$

How to choose k?

Two-sample test: data driven procedure

We set

$$D(n) := \min \Big\{ \underset{1\le k\le d(n)}{\text{argmax}} \, (V_k^{(1,2)} - k\log(n)) \Big\},$$

where $d(n) \to +\infty$ as $n \to +\infty$.

Our test statistic is $V_{D(n)}^{(1,2)}$.

We assume that

(A) $d(n)^{(5p-3)} = o(\log(n))$

Two-sample test: results

Theorem

Let assumption **(A)** *holds. Then, under H_0, $D(n)$ converges in Probability towards 1 as $n \to +\infty$.*

⇒ The asymptotic distribution of $V_{D(n)}^{(1,2)}$ is that of $V_1^{(1,2)} = T_{2,1}^{(1,2)} = n(r_{\mathbf{j}}^{(1,2)})^2$, with $\mathbf{j} = (1, 1, 0, \cdots, 0)$.

Two-sample test: results

Theorem

Let assumption **(A)** *holds and assume that* $\mathbf{j} = (1, 1, 0, \cdots, 0)$*. Then, under* H_0*,* $\sqrt{n} r_{\mathbf{j}}^{(1,2)}$ *converges in law towards a central normal distribution with variance*

$$\begin{aligned}
\sigma^2(1,2) \quad = \quad & \mathbb{V}\Big(L_1(U_1^{(1)}) L_1(U_2^{(1)}) - L_1(U_1^{(2)}) L_1(U_2^{(2)}) \\
& +2\sqrt{3} \int\int \left(\mathbb{I}(X_1^{(1)} \leq x) - F_1^{(1)}(x)\right) L_1(F_2^{(1)}(y)) dF^{(1)}(x,y) \\
& -2\sqrt{3} \int\int \left(\mathbb{I}(X_1^{(2)} \leq x) - F_1^{(2)}(x)\right) L_1(F_2^{(2)}(y)) dF^{(2)}(x,y) \\
& +2\sqrt{3} \int\int \left(\mathbb{I}(X_2^{(1)} \leq y) - F_2^{(1)}(y)\right) L_1(F_1^{(1)}(x)) dF^{(1)}(x,y) \\
& -2\sqrt{3} \int\int \left(\mathbb{I}(X_2^{(2)} \leq y) - F_2^{(2)}(y)\right) L_1(F_1^{(2)}(x)) dF^{(2)}(x,y) \Big)
\end{aligned}$$

Two-sample test: results

In order to normalize the test, write

$$\widehat{\sigma}^2(1,2) = \frac{1}{n}\sum_{i=1}^{n}\Big(M_{i,1} - M_{i,2} - \overline{M}_1 + \overline{M}_2\Big)^2,$$

with

$$\overline{M}_s = \frac{1}{n}\sum_{i=1}^{n} M_{i,s}, \quad \text{for} \quad s = 1,2$$

where

$$\begin{aligned} M_{i,s} = L_1(\widehat{U}_{i,1}^{(s)})L_1(\widehat{U}_{i,2}^{(s)}) \quad &+ \quad \frac{2\sqrt{3}}{n}\sum_{k=1}^{n}\Big(\mathbb{I}(X_{i,1}^{(s)} \le X_{k,1}^{(s)}) - \widehat{U}_{k,1}^{(s)}\Big)L_1(\widehat{U}_{k,2}^{(s)}) \\ &+ \frac{2\sqrt{3}}{n}\sum_{k=1}^{n}\Big(\mathbb{I}(X_{i,2}^{(s)} \le X_{k,2}^{(s)}) - \widehat{U}_{k,2}^{(s)}\Big)L_1(\widehat{U}_{k,1}^{(s)}) \end{aligned}$$

Two-sample test: results

Proposition

Under H_0,

$$\widehat{\sigma}^2(1,2)) \xrightarrow{\mathbb{P}} \sigma^2(1,2).$$

We then deduce the limit distribution under the null.

Corollary

Let assumption **(A)** *holds. Then under H_0, $V_{D(n)}^{(1,2)}/\widehat{\sigma}^2(1,2)$ converges in law towards a chi-squared distribution χ^2_1 as $n \to +\infty$.*

K-sample test

Write $\mathbf{n} = (n_1, \cdots, n_K)$. We restrict our attention to the paired case here, fixing then $n_1 = n_2 = \cdots = n_K := n$. Write

$$\mathcal{V}(K) \quad = \quad \{(\ell, m) \in \mathbb{N}^2; 1 \leq \ell < m \leq K\}.$$

Clearly $\mathcal{V}(K)$ contains $v(K) = K(K-1)/2$ elements which represent the pairs of populations that we want to compare and that can be ordered as follows: we write $(\ell, m) <_{\mathcal{V}} (\ell', m')$ if $\ell < \ell'$, or $\ell = \ell'$ and $m < m'$, and we denote by $ord_{\mathcal{V}}(\ell, m)$ the rank of (ℓ, m) in $\mathcal{V}(K)$.

K-sample test

This can be seen as a natural order of the elements of the upper triangle of a $(K-1)\times(K-1)$ matrix as represented below:

$$\begin{array}{ccccc} (1,2) & (1,3) & \cdots & \cdots & (1,K) \\ & (2,3) & \cdots & \cdots & (2,K) \\ & & \ddots & & \\ & & & & (K-1,K) \end{array}$$

We see at once that $ord_{\mathcal{V}}(1,2)=1, ord_{\mathcal{V}}(1,3)=2$ and more generally, for $\ell, m \in \mathcal{V}(K)$ we have

$$ord_{\mathcal{V}}(\ell,m)=K(l-1)-\frac{l(l+1)}{2}+m.$$

K-sample test

Using the previous two-sample statistics we construct an embedded series of statistics as

$$\begin{aligned}
V_1 &= V_{D(n)}^{(1,2)} \\
V_2 &= V_{D(n)}^{(1,2)} + V_{D(n)}^{(1,3)} \\
\dots & \\
V_{v(K)} &= V_{D(n)}^{(1,2)} + \cdots + V_{D(n)}^{(K-1,K)}.
\end{aligned}$$

K-sample test

How to select the convenient statistic?
To choose automatically the appropriate number k we introduce a second penalization procedure, mimicking the Schwarz criteria procedure

$$s(\mathbf{n}) = \min\left\{ \underset{1\leq k\leq v(K)}{\mathsf{argmax}} \big(V_k - k\log(n)\big)\right\}.$$

K-sample test: results

Theorem

Assume that **(A)** *holds. Then under H_0, $s(\mathbf{n})$ converges in probability towards 1 as $n \to +\infty$.*

Corollary

Assume that **(A)** *holds. Then under H_0, $V_{s(\mathbf{n})}/\widehat{\sigma}^2(1,2))$ converges in law towards a χ^2_1 distribution.*

Then our final data driven test statistic is given by

$$V \quad = \quad V_{s(\mathbf{n})}/\widehat{\sigma}^2(1,2)).$$

Alternatives

We consider the following series of alternative hypothesis:

$$H_1(1) : C_1 \neq C_2,$$

and for $k > 1$:

$$H_1(k) : C_i = C_j \text{ for } ord_{\mathcal{V}}(i,j) < k \quad \text{and} \quad C_i \neq C_j \text{ for } ord_{\mathcal{V}}(i,j) = k,$$

with $1 < k \leq v(K)$. The hypothesis $H_1(k)$ means that the ith and jth populations such that $ord_{\mathcal{V}}(i,j) = k$ are the first (in the sense of the order in $\mathcal{V}(K)$) with different unknown components.

Theorem

Assume that **(A)** *holds. Then under $H_1(k)$, $s(\mathbf{n})$ converges in probability towards k, as $\mathbf{n} \to +\infty$, and V tends to $+\infty$, that is, $\mathbb{P}(V < \epsilon) \to 0$ for all $\epsilon > 0$.*

Sketch of the proof

- We want to show that $\mathbb{P}(D(n) > 1) \to 0$ as n tends to infinity.
- We obtain

$$\mathbb{P}_0\Big(D(n) > 1\Big) \leq \mathbb{P}_0\Big(n \sum_{\mathbf{j} \in \mathcal{H}^*(d(n))} (r_{\mathbf{j}}^{(1,2)})^2 \geq \log(n)\Big),$$

- We decompose $r_{\mathbf{j}}^2$ as follows:

$$(r_{\mathbf{j}}^{(1,2)})^2 \quad = \quad ((\widehat{\rho}_{\mathbf{j}}^{(1)} - \rho_{\mathbf{j}}^{(1)}) - (\widehat{\rho}_{\mathbf{j}}^{(2)} - \rho_{\mathbf{j}}^{(2)}))^2$$

Sketch of the proof

We decompose

$$\widehat{\rho}_{\mathbf{j}}^{(1)} - \rho_{\mathbf{j}}^{(1)} = (\widehat{\rho}_{\mathbf{j}}^{(1)} - \tilde{\rho}_{\mathbf{j}}^{(1)}) + (\tilde{\rho}_{\mathbf{j}}^{(1)} - \rho_{\mathbf{j}}^{(1)})$$

and we use Legendre polynomials properties and the Glivenko-Cantelli Theorem to conclude.

Sketch of the proof

To show the asymptotic normality we adapt a proof of Bhuchongkul (1964, Ann. Math. Stat.).

Sketch of the proof

We generalize the result to the K-sample easily since the number of samples is fixed.

$\hookrightarrow$ we could extend the result with $K = K(n) \to \infty$.

The proof of the convergence under alternatives mimics the first main result.

$\hookrightarrow$ finally, we could consider contiguous alternatives.

Illustration: Iris data

Fifty observations of four measures $(X_1, \cdots, X_4)$: SepalLength, SepalWidth, PetalLength, and PetalWidth, for each of three Species = three populations: Setosa, Versicolor, and Virginica.

$\hookrightarrow$ dimension $p = 4$ and $K = 3$

Illustration: Iris data

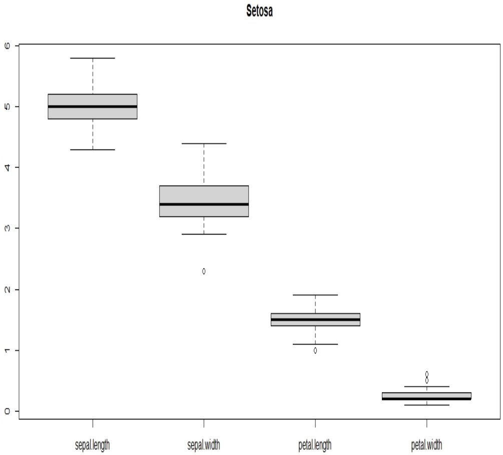

Figure – Setosa population

Illustration: Iris data

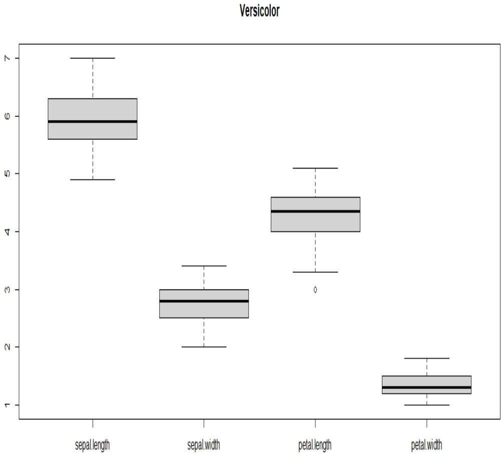

Figure – Versicolor population

Illustration: Iris data

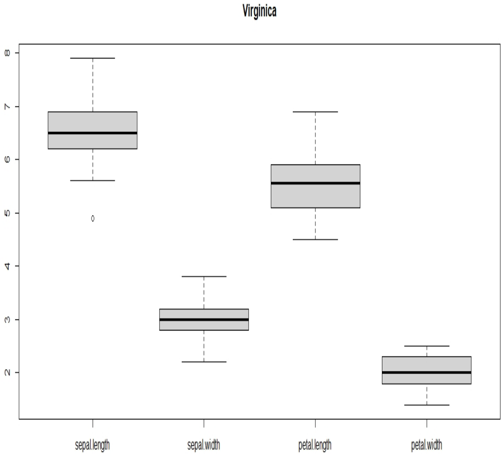

Figure – Virginica population

Illustration: Iris data

Dhar et al. (2014, Bernoulli) shown that multivariate normal distributions seem to fit the data well for all three Iris species. Looking at their mean parameters the 4-dimensional joint distributions seem different but that does not tell us about their dependence structures.

$\hookrightarrow$ We test the equality of the dependence between the four variables SepalLength, SepalWidth, PetalLength, and PetalWidth in the three-sample case:

$$H_0 : C_1 = C_2 = C_3$$

Illustration: Iris data

Results for the global test:

- $pvalue \approx 0 \ (10^{-12})$
- $StatisticsRank : D(n) = 2$
- $Statisticvalue : V = 49.9$

$\hookrightarrow$ we clearly reject the equality of the three dependence structures.

Illustration: Iris data

In case of reject we can process to an "2×2 ANOVA" type procedure as follows:

- The pvalues:

	Setosa	Versicolor	Virginica
Setosa	1	10^{-9}	0.0016
Versicolor	10^{-9}	1	0.70
Virginica	0.0016	0.70	1

Illustration: Iris data

- The statistics values:

	Setosa	Versicolor	Vriginica
Setosa	1	34.85	10
Versicolor	34.85	1	0.18
Virginica	10	0.18	1

Clustering application

We can use the K-sample procedure to make clustering. Assume we observe K populations.

- We first compare the two closest population (say A and B).
- If $C_A = C_B$ we have a first group: (A, B)
- We then search the closest new population from A and B (say C)
- If $C_A = C_B = C_C$ we have a new group: (A, B, C)
- Until the test is rejected. In this case the group is formed. The procedure continues with the other populations.

If we apply this clustering algorithm on the iris dataset we obtain two groups: $\{Versicolor, Virginica\}$ and $\{Setosa\}$.

Conclusion

- The first simulation study shows a very good power and a very easy of use of the method (as good or better for the two sample case, no competitor for the K-sample case...).
- A package is in progress.
- Copulas coefficients are related to Kendall's tau and Spearman's Rho. It opens a way to multivariate extensions, estimations and testing procedure.
- The method can be easily adapted to obtain a test for independence. This is a work in progress.

Thank you for your attention !