Les Cahiers du GERAD ISSN: 0711–2440

Mixtures of Empirical Copulas and Weighted Coefficients of Correlation

J.-F. Plante

G–2010–41

July 2010

Mixtures of Empirical Copulas and Weighted Coefficients of Correlation

Jean-François Plante

Service de l'enseignement des méthodes quantitatives de gestion
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada, H3T 2A7

jean-francois.2.plante@hec.ca

July 2010

Les Cahiers du GERAD
G–2010–41

Abstract

The population value of coefficients of correlation based on ranks depends only on the copula underlying the true distribution. We consider data sets that share the same dependence structure, but present different margins. For instance, data may be expressed in different currencies or measured by indices that cannot be compared across populations. A mixture of empirical copulas is built, yielding weighted coefficients of correlation. The consistency of the estimates and their asymptotic distributions are derived for scalar weights. We also consider the case where data-based weights detect adaptively the similarities between the copulas underlying each population, with the idea of making a compromise between bias and variance. Simulations are used to explore the finite sample behavior of these weighted methods.

Résumé

En tant que paramètre de population, les cœfficients de corrélation basés sur les rangs dépendent seulement de la copule sous-jacente à la distribution des données. Nous considérons le paradigme où des échantillons proviennent de populations partageant une même structure de dépendance, mais affublées de lois marginales univariées différentes. Par exemple, les données pourraient être mesurées en devises différentes, ou à l'aide d'indices qui ne peuvent être comparés d'un échantillon à l'autre. Un mélange de copules empiriques est construit et des cœfficients de corrélation pondérés sont développés. La convergence et la distribution asymptotique de ces estimateurs sont déterminées sous l'hypothèse d'une pondération scalaire. Nous étudions aussi des poids utilisant les données pour s'adapter aux ressemblances entre les copules de chaque échantillon afin d'établir un compromis entre le biais et la variance. Des simulations sont utilisées pour explorer la performance de ces méthodes pondérées avec des échantillons de taille finie.

Acknowledgments: A portion of this paper constitutes part of the author's doctoral thesis supervised by Professor James V. Zidek at the University of British Columbia. I am thankful to Professor Zidek for his guidance and support in discovering the weighted likelihood and in developing these results. For partial support of this work through graduate scholarships and research grants, thanks are due to the Natural Sciences and Engineering Research Council of Canada and to the Fonds québécois de la recherche sur la nature et les technologies.

1 Introduction

To make meaningful comparisons comparable scales are needed. When the dependence structure is of prime interest, comparable scales are synonym of common marginal distributions. A natural yet arbitrary choice consists in turning all the margins into uniform distributions, yielding what we call a copula.

Let $\mathbf{X} \in \mathbb{R}^p$ be multivariate random vectors whose distribution follows the cumulative distribution function $F(\mathbf{x})$ with univariate margins $G_i(x_i)$, $i = 1, \ldots, p$. For simplicity, we assume that F and its margins are continuous. The Sklar (1959) theorem shows that there exist a unique function $C(\mathbf{u}) : [0,1]^p \to [0,1]$, called copula, such that

$$F(\mathbf{x}) = C\{G_1(x_1), \ldots, G_p(x_p)\}.$$

The function $C(\mathbf{u})$, a multivariate cumulative distribution function (CDF) with uniform univariate margins, contains all the information about the dependence structure of F.

Empirically, ranks allow to estimate a copula without modeling the marginal distributions of the data. For bivariate data (X_{i1}, X_{i2}), the rank of X_{i1} within the list of $X_{\bullet 1}$ is denoted R_{i1} and similarly for R_{i2}, forming the vectors of ranks $\mathbf{R}_i = [R_{i1}, R_{i2}]^\mathsf{T}$. The empirical copula introduced by Deheuvels (1979) allocates equal weights to rescaled vectors of ranks, e.g. $\mathbf{Y}_i = \mathbf{R}_i/n$. The asymptotic normality of the empirical copula was recently derived independently by different authors including Fermanian et al. (2004) and Tsukahara (2005).

Coefficients of correlation such as Spearman's ρ and Kendall's τ are based on ranks. Their population values depend only on the copula underlying the distribution of the data (see Table 1 for examples).

Samples of data may share the same dependence structure, but feature different marginal distributions. Consider for instance the following scenarios:

1. the dependence between skills in English and Mathematics is of interest, but grades are normalized within groups;
2. the measurements of interest come from different sources (e.g. labs) and are not calibrated to make them comparable across studies;
3. the measurements are made using different units that cannot be transformed easily (e.g. different currencies or scales);
4. only ranks are available from the data sets.

We develop weighted methods that can handle such data. Suppose that we have samples from m sources of data where ranks are computed only within each sample, yielding m empirical copulas or m coefficients of correlations. We define weighted mixtures of these objects. We first propose scalar weights for the cases where all data are assumed to share the same copula. In the case where copulas may differ, data-based weights allow to use all the data, making a compromise between bias and precision.

Background definitions and notation are presented in Section 2. Weighted empirical copulas are introduced in Section 3. Weighted versions of the coefficients of correlations based on ranks are presented in Section 4. Finally, Section 5 presents simulation results to illustrate the use of these weighted methods and explore their performance on finite samples.

2 Background and Notation

Suppose that p-dimensional data are available from m different populations believed to have similar dependence structures (i.e. similar copulas). For any fixed $k \in \mathbb{N}$, we observe n_{ik} data points from Population $i \in \{1, \ldots, m\}$. The index k is used to monitor increasing sample sizes when studying asymptotic results. Explicitly,

$$\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_{ik}} \overset{iid}{\sim} F_i$$

are observed, hence a total of $N_k = \sum_{i=1}^m n_{ik}$ data, where $\mathbf{X}_{ij} = [X_{ij1}, \ldots, X_{ijp}]^\mathsf{T}$ is a vector in p dimensions and F_i are continuous. The sample sizes n_{ik} are assumed to be non-decreasing with k. By Sklar's (1959)

Theorem, there exists a unique copula underlying the distribution F_i; we denote it by $C_i(\mathbf{u})$, where $\mathbf{u} = [u_1, \ldots, u_p]^\mathsf{T}$ is a vector in $[0,1]^p$. That unique copula is a cumulative distribution function with uniform margins such that $F_i(\mathbf{x}) = C_i\{G_{i1}(x_1), \ldots, G_{ip}(x_p)\}$ where $G_{i1}, \ldots, G_{ip}$ are the marginal distributions of F_i.

Let $\mathbf{R}_{ij}^k = [R_{ij1}^k, \ldots, R_{ijp}^k]^\mathsf{T}$ be the ranks associated with the vectors $\mathbf{X}_{ij}$, $j = 1, \ldots, n_{ik}$. For fixed i and ℓ, the list of values $X_{i1\ell}, \ldots, X_{in_{ik}\ell}$ is sorted and $R_{ij\ell}^k$ is the rank of $X_{ij\ell}$ in that list. Since F_i are continuous, ties cannot occur with probability 1.

Empirical Copula

The empirical copula, uses ranks to estimate C_i:

$$\hat{C}_{ik}(\mathbf{u}) = \frac{1}{n_{ik}} \sum_{j=1}^{n_{ik}} \prod_{\ell=1}^{p} \mathbb{1}\left(\frac{R_{ij\ell}^k}{n_{ik}} \le u_\ell\right)$$

for $\mathbf{u} = [u_1, \ldots, u_p]^\mathsf{T}$. The indicator variable $\mathbb{1}(\bullet)$ is equal to one if all the elements of its argument are true and equal to 0 otherwise. The empirical copula puts a weight of $1/n_{ik}$ on the points of the grid

$$\left\{\frac{1}{n_{ik}}, \frac{2}{n_{ik}}, \ldots, 1\right\} \times \cdots \times \left\{\frac{1}{n_{ik}}, \frac{2}{n_{ik}}, \ldots, 1\right\}$$

corresponding to an observed combination of ranks. There is exactly one such point in every $(p-1)$-dimensional slice of the grid (rows and columns in 2 dimensions). Consequently, the univariate margins of the empirical copula $\hat{C}_{ik}$ are uniformly distributed on the points $\{1/n_{ik}, 2/n_{ik}, \ldots, 1\}$.

Deheuvels (1979) shows that

$$\sqrt{\frac{n_{ik}}{\log\log n_{ik}}} \sup_{\mathbf{u}\in[0,1]^p} |\hat{C}_{ik}(\mathbf{u}) - C_i(\mathbf{u})| \to 0 \tag{1}$$

almost surely as $k \to \infty$ (since $n_{ik} \to \infty$ then). Fermanian et al. (2004) show that $\sqrt{n_{ik}}\{\hat{C}_{ik}(\mathbf{u}) - C_i(\mathbf{u})\}$ converges weakly to a Brownian sheet whose variance depends on C_i and its partial first-order derivatives. Although they hold for an arbitrary number of dimensions, the results of Fermanian et al. (2004) are presented for bivariate copulas only. Tsukahara (2005) credits Fermanian et al. (2004) for the discovery and expresses the same results in p dimensions.

Remark 2.1 *Let $\mathcal{U}_i(\mathbf{u})$ be a p-dimensional centered Gaussian random field with covariance function $C_i(\mathbf{u} \wedge \mathbf{v}) - C_i(\mathbf{u})C_i(\mathbf{v})$, where $\wedge$ is the component-wise minimum. Such a random field is called a p-dimensional pinned C_i-Brownian sheet.*

Theorem 2.1 (Tsukahara (2005)) *Assume that $C_i(\mathbf{u})$ is differentiable with continuous partial derivatives $\partial C_i(\mathbf{u})/\partial u_\ell$ for $\ell = 1, \ldots, p$ and let $[\mathbf{1}, u_\ell, \mathbf{1}]^\mathsf{T}$ represent a vector of ones, except for the ℓ^{th} element who is equal to the ℓ^{th} element of $\mathbf{u}$. Then as $k \to \infty$, the random variable $\sqrt{n_{ik}}\{\hat{C}_{ik}(\mathbf{u}) - C_i(\mathbf{u})\}$ converges weakly to the random field $\mathcal{U}_i(\mathbf{u}) - \sum_{\ell=1}^{p}\{(\partial/\partial u_\ell)\, C_i(\mathbf{u})\}\mathcal{U}_i([\mathbf{1}, u_\ell, \mathbf{1}]^\mathsf{T})$.*

Coefficients of Correlation Based on Ranks

For bivariate data, coefficients of correlation based on ranks measure concordance of the data. Table 1 contains some examples of estimates and their population values. A simplified notation is used in the tables where for a fixed population i and a fixed k, we write $n = n_{ik}$, $\hat{C}(u_1, u_2) \equiv \hat{C}_{ik}(\mathbf{u})$, $(R_j, S_j) = \mathbf{R}_{ij}^k$ and $C(u, v) = C_i([u, v]^\mathsf{T})$. That notation is the most commonly seen in the literature.

More details on Spearman's ρ, Kendall's τ, Gini's γ and Blomqvist's β can be found in Nelsen (1999). Blest's coefficients were first introduced by Blest (2000), then further developed by Genest & Plante (2003). Pinto da Costa & Soares (2005) studied the same coefficients of correlation and rediscovered independently some of the results published by Genest & Plante (2003).

The coefficients in Table 1 are asymptotically Normal (with the possible exception of Gini's γ whose asymptotic distribution as a coefficient of correlation seems unknown). This property can be used to build confidence intervals or to test for independence.

3 Mixtures of Empirical Copulas

Let $\boldsymbol{\lambda}_k = [\lambda_{1k}, \ldots, \lambda_{mk}]^\mathsf{T}$ be nonnegative weights such that $\sum_{i=1}^m \lambda_{ik} = 1$ for all $k \in \mathbb{N}$ and let

$$\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) = \sum_{i=1}^m \lambda_{ik} \hat{C}_{ik}(\mathbf{u})$$

be a mixture of the empirical copulas based on the m available samples.

3.1 Scalar Weights

We first consider the paradigm where the m populations are known (or assumed) to share a common dependence structure, i.e. $C_i(\mathbf{u}) \equiv C(\mathbf{u})$, but their marginal distributions are not comparable.

The choice $\lambda_{ik} = n_{ik}/N_k$ is of special interest as it will allocate an equal weight to each datum. However, we consider arbitrary scalar weights, as long as each datum's contribution tends to 0 as $k \to \infty$, as per Assumption 3.1.

Assumption 3.1 *To ensure that all sample sizes increase at a similar rate, we assume that* $\limsup_k \dfrac{N_k}{n_{ik}} < \infty$ *for* $i = 1, \ldots, m$. *This also implies that* $A_k = \displaystyle\sum_{i=1}^m \frac{\lambda_{ik}^2 N_k}{n_{ik}}$ *is finite for all* k.

The estimate $\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u})$ is consistent, but its rate of convergence is controlled by the smallest of the m samples.

Theorem 3.1 $\sqrt{n_{i'_k k}/\log\log n_{i'_k k}}\, \sup_{\mathbf{u}\in[0,1]^p} |\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) - C(\mathbf{u})| \to 0$ *almost surely as* $k \to \infty$ *where* $i'_k = \arg\min_{i=1,\ldots,m} n_{ik}$.

Since each of the $\hat{C}_{ik}(\mathbf{u})$ are defined on independent samples, the asymptotic distribution of $\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u})$ can easily be derived from Theorem 2.1.

Theorem 3.2 *The random variable* $\sqrt{N_k/A_k}\left\{\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) - C(\mathbf{u})\right\}$ *converges weakly to the random field* $\mathcal{U}(\mathbf{u}) - \sum_{\ell=1}^p \{(\partial/\partial u_\ell) C(\mathbf{u})\} \mathcal{U}([\mathbf{1}, u_\ell, \mathbf{1}]^\mathsf{T})$ *as* $k \to \infty$ *where* $\mathcal{U}(\mathbf{u})$ *is a random field with covariance structure* $C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u})C(\mathbf{v})$.

Remark 3.1 *The choice of weights has an effect on the asymptotic distribution of the copula. Simple calculus may be used to show that* $\lambda_{ik} = n_{ik}/N_k$ *minimizes* A_k, *hence yielding the least variable estimate of* $\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u})$. *This choice corresponds to allocating an equal weight to each datum and yields* $A_k = 1$.

Remark 3.2 *Let* $\hat{\beta}_{\boldsymbol{\lambda}_k} = \sum_{i=1}^m \lambda_{ik} \hat{\beta}_{ik} = 4\hat{C}_{\boldsymbol{\lambda}k}(1/2, 1/2) - 1$ *be the weighted Blomqvist coefficient. As* $k \to \infty$, $\sqrt{N_k/A_k}(\hat{\beta}_{\boldsymbol{\lambda}_k} - \beta)$ *converges weakly to a centered Normal variable with variance* $-\beta^2 + \{4(d_1 - 1)(d_2 - 1) - 2\}\beta + \left\{(2d_1 - 1)^2 + (2d_2 - 1)^2 + 1\right\}$ *where* $d_i = (\partial/\partial u_i) C(u_1, u_2)|_{u_1 = u_2 = \frac{1}{2}}$. *Under the assumption of independence,* $\beta = 0$ *and the variance is* 1.

Note that our asymptotic paradigm involves a fixed number of populations whose sample sizes increase to ∞. The convergence would not hold for an infinite number of small populations. For instance, a mixture based on infinitely many samples of size 10 will still have $\hat{C}(1/20, 1/20) = 0$.

Table 1: Empirical estimates of different coefficients of correlation and their population values.

Usual Name	Empirical Estimate	Population Value
Spearman	$\hat{\rho}_n = -3\,\dfrac{n+1}{n-1} + \dfrac{12}{n(n+1)(n-1)} \displaystyle\sum_{i=1}^{n} R_i S_i$	$\rho = 12 \displaystyle\int uv \,\mathrm{d}C(u,v) - 3$
Kendall	$\hat{\tau}_n = \dbinom{n}{2}^{-1} \displaystyle\sum_{1 \le i < j \le n} \operatorname{sign}(R_i - R_j) \operatorname{sign}(S_i - S_j)$	$\tau = 4 \displaystyle\int C(u,v) \,\mathrm{d}C(u,v) - 1$
Gini	$\hat{\gamma}_n = \dfrac{1}{\lfloor n^2/2 \rfloor} \displaystyle\sum_{i=1}^{n} \lvert R_i + S_i - n - 1 \rvert - \lvert R_i - S_i \rvert$	$\gamma = \displaystyle\int \lvert u + v - 1 \rvert - \lvert u - v \rvert \,\mathrm{d}C(u,v)$
Blomqvist	$\hat{\beta}_n = 4\hat{C}\left(\dfrac{1}{2}, \dfrac{1}{2}\right) - 1$	$\beta = 4\, C\left(\dfrac{1}{2}, \dfrac{1}{2}\right) - 1$
Blest	$\hat{\nu}_n = \dfrac{2n+1}{n-1} - \dfrac{12}{n(n+1)^2(n-1)} \displaystyle\sum_{i=1}^{n} (n + 1 - R_i)^2 S_i$	$\nu = 2 - 12 \displaystyle\int (1-u)^2 v \,\mathrm{d}C(u,v)$
	$\widehat{\overline{\nu}}_n = -\dfrac{2n+1}{n-1} + \dfrac{12}{n(n+1)^2(n-1)} \displaystyle\sum_{i=1}^{m} R_i^2 S_i$	$\overline{\nu} = -2 + 12 \displaystyle\int u^2 v \,\mathrm{d}C(u,v)$
Symmetrized Blest	$\hat{\xi}_n = -\dfrac{4n+5}{n-1} + \dfrac{6}{n(n+1)(n-1)} \displaystyle\sum_{i=1}^{n} R_i S_i \left(4 - \dfrac{R_i + S_i}{n+1}\right)$	$\xi = -4 + 6 \displaystyle\int uv(4 - u - v) \,\mathrm{d}C(u,v)$
	$\widehat{\overline{\xi}}_n = -\dfrac{2n+1}{n-1} + \dfrac{6}{n(n+1)^2(n-1)} \displaystyle\sum_{i=1}^{n} R_i S_i \,(R_i + S_i)$	$\overline{\xi} = -2 + 6 \displaystyle\int uv(u+v) \,\mathrm{d}C(u,v)$

3.2 Adaptive Weights

Assuming that the dependence structure underlying the data is identical in all the samples may not always be appropriate. This does not mean that the samples do not contain relevant information.

In this section, we adopt a paradigm similar to Wang (2001) and Wang & Zidek (2005) in the context of the weighted likelihood. We suppose that one of the populations (Population 1) is identified as having the target dependence structure. The other samples are likely to feature a similar structure, but their copula may not be identical. In this context, we build adaptive weights that use the data to determine the weight allocated to each sample, trying to compromise between the bias and the reduced variance brought by additional data.

Let $C_i(\mathbf{u})$ be the copula underlying the distribution of X_{ij}. We do not assume that the C_i are identical, but we do not suppose that they are different either. Let also M_k be a discrete probability measure allocating a weight of $1/n_{1k}^p$ to each point of the grid $\mathcal{G}_k = \{1/n_{1k}, 2/n_{1k}, \ldots, 1\} \times \cdots \times \{1/n_{1k}, 2/n_{1k}, \ldots, 1\}$. The results of this section will hold for all adaptive weights $\boldsymbol{\mu}_k = [\mu_{1k}, \ldots, \mu_{mk}]^\mathsf{T}$ that respect Assumption 3.2. Adaptive weights are denoted $\boldsymbol{\mu}_k$ to avoid confusion with the scalar weights in the previous section. The adaptively weighted empirical copula is thus written as $C_{\boldsymbol{\mu}_k}$.

Assumption 3.2 $\int \{\hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u})\}^2 \,\mathrm{d}M_k(\mathbf{u}) \to 0$ *almost surely as* $k \to \infty$.

Theorem 3.3 *We have uniform convergence of the adaptively weighted empirical copula:*

$$\sup_{\mathbf{u}\in[0,1]^p} \left|\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) - C_1(\mathbf{u})\right| \to 0$$

almost surely as $k \to \infty$.

Remark 3.3 *Let* $\hat{\beta}_{\boldsymbol{\mu}_k} = \sum_{i=1}^m \mu_{ik}\hat{\beta}_{ik}$ *be the adaptively weighted Blomqvist coefficient. Then,* $\hat{\beta}_{\boldsymbol{\mu}_k} \to \beta_1$ *almost surely as* $k \to \infty$.

Lemma 3.1 *If* $g(\mathbf{u})$ *is a bounded function, then*

$$\int g(\mathbf{u}) \,\mathrm{d}\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) = \mathrm{E}\{g(\mathbf{U}_k)\} \to \mathrm{E}\{g(\mathbf{U})\} = \int g(\mathbf{u}) \,\mathrm{d}C_1(\mathbf{u})$$

almost surely as $k \to \infty$.

Plante (2007, 2008) clarifies the ties between the weighted likelihood and mixtures of empirical distributions and suggests a data-based nonparametric criterion to determine the weight that should be allocated to each population. We extend the so-called Minimum Averaged Mean Square Error (MAMSE) weights to copulas.

Let us define

$$P_k(\boldsymbol{\lambda}) = \int \left[\left|\hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u})\right|^2 + \sum_{i=1}^m \lambda_i^2 \widehat{\mathrm{var}}\left\{\hat{C}_{ik}(\mathbf{u})\right\} \right] \mathrm{d}M_k(\mathbf{u}). \tag{2}$$

The variance term in $P_k(\boldsymbol{\lambda})$ plays the role of a penalty that fosters using data from all the populations rather than limiting the inference to the population of interest. Since the asymptotic variance of the empirical copula (see Theorem 2.1) depends on the true copula $C_i(\mathbf{u})$ and its derivatives, we consider a very rough estimate thereof given by

$$\widehat{\mathrm{var}}\{\hat{C}_{ik}(\mathbf{u})\} \approx \widetilde{\mathrm{var}}\{\hat{C}_{ik}(\mathbf{u})\} = \frac{1}{n_{ik}}\hat{C}_{ik}(\mathbf{u})\{1 - \hat{C}_{ik}(\mathbf{u})\}, \tag{3}$$

which corresponds to the only term of the asymptotic variance of an empirical copula that does not involve a derivative of C_i. The value of $\boldsymbol{\lambda}$ minimizing the objective function $P_k(\boldsymbol{\lambda})$ defined in (2) with the substitution (3) is called the MAMSE weights.

Remark 3.4 *The MAMSE weights respect Assumption 3.2. Indeed, let* $\boldsymbol{\lambda} = [1, 0, \ldots, 0]^\mathsf{T}$ *be a possibly suboptimal choice of weights for* P_k*, and* $\boldsymbol{\mu}_k$ *denote the MAMSE weights, then*

$$\int \left\{\hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u})\right\}^2 \mathrm{d}M_k(\mathbf{u}) \leq P_k\{\boldsymbol{\mu}_k\} \leq P_k(\boldsymbol{\lambda}) = \int \widetilde{\mathrm{var}}\left\{\hat{C}_{1k}(\mathbf{u})\right\} \mathrm{d}M_k(\mathbf{u}) \leq \frac{1}{4n_{1k}}.$$

The algorithm for the MAMSE weights proposed by Plante (2008) can be used in the current context with copulas. Details specific to copulas may be found in Plante (2007).

Note that the data-dependence of the weights causes serious complications for the study of the asymptotic distribution of $\hat{C}_{\boldsymbol{\mu}_k}$. Although $\hat{C}_{\boldsymbol{\mu}_k}$ converges uniformly to the desired target, the weights $\boldsymbol{\mu}_k$ may remain random for an arbitrarily large k if a mixture of the true $C_2, \ldots, C_m$ is identical to C_1. This behaviour is observed with the MAMSE weights. A short discussion thereof is found in Plante (2008, 2009). Due to these complications, the study of the asymptotic distribution of $\hat{C}_{\boldsymbol{\mu}_k}$ is left to future work. For the time being, simulations or resampling methods can be used to determine confidence intervals or proceed with hypothesis testing.

4 Weighted Coefficients of Correlation

Due to its special formulation, we must treat Kendall's τ separately from the other coefficients of correlation who share a common linear form.

4.1 Linear forms

Most of the coefficients in Table 1 take the form

$$\hat{\kappa}_{ik} = a_{n_k} \int g(\mathbf{u}) \, \mathrm{d}\hat{C}_{ik}(\mathbf{u}) + b_{n_k} \tag{4}$$

that estimate $\kappa_i = a \int g(\mathbf{u}) \, \mathrm{d}C_i(\mathbf{u}) + b$ where $g(\mathbf{u})$ is a continuous bounded function on $[0,1]^2$. The coefficients $a_{n_k} \to a$ and $b_{n_k} \to b$ as $n_k \to \infty$ are chosen to ensure that $\hat{\kappa}_{ik} \in [-1, 1]$ for all sample sizes n_{ik}. Moreover, the values ± 1 occur only for perfect concordance or discordance, i.e. when the ranks are identical ($R^k_{ij1} = R^k_{ij2}$) or antithetic ($R^k_{ij1} = n_{ik} + 1 - R^k_{ij2}$).

Well known results by Ruymgaart, Shorack & van Zwet (1972) and Ruymgaart (1974) specify the asymptotic distribution of ρ, ν, $\bar{\nu}$, ξ and $\bar{\xi}$. See Genest & Plante (2003) for some illustrations.

Remark 4.1 *For scalar weights, it is clear that the random variable* $\sqrt{N_k/A_k}(\hat{\kappa}_{\boldsymbol{\lambda}_k} - \kappa)$ *converges weakly to a Normal variate with mean* 0 *and the same asymptotic variance as* $\sqrt{n_{ik}}(\hat{\kappa}_{ik} - \kappa)$ *when* $k \to \infty$*.*

Coefficients of correlation are often used as a test of independence. Suppose that the alternative hypothesis is expressed through a parameter θ for which $\theta = 0$ yields independence. The asymptotic relative efficiency (ARE) of the two tests represent the ratio of the sample sizes needed by both tests to achieve the same power. We find from Lehmann (1998), page 375, that

$$ARE(T_\kappa, T_{\kappa^*}) = \frac{n_{\kappa^*}}{n_\kappa} = \frac{\sigma^2_{\kappa^*}}{\sigma^2_\kappa} \left(\frac{\kappa_0'}{\kappa_0^{*\prime}}\right)^2$$

where T_κ is the independence test based on κ, σ^2_κ the asymptotic variance of $\hat{\kappa}$, $\kappa_0' = (\partial/\partial\theta)\kappa(\theta)|_{\theta=0}$ and similarly for κ^*.

Remark 4.2 *We have* $ARE(T_{\hat{\kappa}_{1k}}, T_{\hat{\kappa}_{\boldsymbol{\lambda}_k}}) = A_k n_{1k}/N_k$*. If the margins were comparable, we could pool the* N_k *data together, calculate* κ_{Pk} *and then* $ARE(T_{\hat{\kappa}_{Pk}}, T_{\hat{\kappa}_{\boldsymbol{\lambda}_k}}) = A_k$*. Recall that we have* $A_k = 1$ *when* $\lambda_i = n_{ik}/N_k$*, which means that there is no loss of power asymptotically in that case as far as the ARE is concerned.*

Replacing the copula by its empirical estimate in Table 1 yields an estimate of the corresponding coefficient. The expressions thus obtained typically differ slightly from the usual estimates based on ranks by an amount

that converges to 0 as the sample size goes to infinity. In particular, the bounds may not be exactly ± 1 for fixed sample sizes, but always converge to these limits. The coefficient thus obtained is asymptotically equivalent to its corresponding usual estimate. This combined with Theorem 3.1 implies Theorem 4.1.

Theorem 4.1 *Let κ represent any of ρ, γ, ν, $\bar{\nu}$, ξ or $\bar{\xi}$ and $\hat{\kappa}_{\boldsymbol{\mu}_k} = \sum_{i=1}^m \mu_{ik}\hat{\kappa}_{ik}$. Then, $\hat{\kappa}_{\boldsymbol{\mu}_k} \to \kappa_1$ almost surely as $k \to \infty$.*

4.2 Kendall's τ

Let us first consider the estimate $\hat{\tau}_{\boldsymbol{\lambda}_k} = \sum_{i=1}^m \lambda_i \hat{\tau}_{ik}$. Since $\hat{\tau}_{ik}$ is a U-statistics, $\sqrt{n_{ik}}(\hat{\tau}_{ik} - \tau)$ is asymptotically distributed as a centered Normal variable, hence $\sqrt{N_k/A_k}(\hat{\tau}_{\boldsymbol{\lambda}_k} - \tau)$ converges weakly to a Normal when the samples share a common copula and scalar weights are used.

Adaptive weights, however, are used when the underlying copula may differ between samples. As a consequence of the lack of linearity of τ, $\hat{\tau}_{\boldsymbol{\mu}_k}$ may not be consistent.

Remark 4.3 *Consider the Fréchet family of copula from Example 5.3 in Nelsen (1999), page 129. $C_1 = C_{\alpha,\beta} = \alpha M + (1-\alpha-\beta)\Pi + \beta W$ where $M = C_2$, $\Pi = C_3$ and $W = C_4$ represent respectively the Fréchet bounds of perfect concordance, independence and perfect discordance. In this situation, the adaptive weights will find $\boldsymbol{\mu}_k$ such that $C_{\boldsymbol{\mu}_k} \to C_1$, but the share of C_1 compared to C_2, C_3 and C_4 may remain random even for large k. Unfortunately, $\tau_1 = (\alpha-\beta)(\alpha+\beta+2)/3$ is not equal to $\alpha\tau_M + (1-\alpha-\beta)\tau_\Pi + \beta\tau_W = \alpha - \beta$, meaning that $\hat{\tau}_{\boldsymbol{\mu}_k}$ will not be consistent.*

We therefore construct an nonlinear definition for the weighted Kendall's τ that will be consistent. The empirical version of Kendall's τ in Table 1 can be rewritten

$$\hat{\tau}_{ik} = \frac{4n}{n-1}\int \hat{C}_{ik}(\mathbf{u})\,\mathrm{d}\hat{C}_{ik}(\mathbf{u}) - \left(1 + \frac{4}{n-1}\right), \tag{5}$$

which shows that Kendall's τ is asymptotically equivalent to replacing the copula by its empirical counterpart in the population value of τ. The functional form of τ is based on $\int C(\mathbf{u})\,\mathrm{d}C(\mathbf{u})$, hence the lack of linearity. If we estimate the copula by a weighted mixture of empirical copulas, we get

$$\hat{\tau}'_{\boldsymbol{\lambda}_k} \triangleq 4\int \hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u})\,\mathrm{d}\hat{C}_{\boldsymbol{\lambda}k}(\mathbf{u}) - 1 = \sum_{i=1}^m\sum_{j=1}^m \lambda_{ik}\lambda_{jk}\left\{4\int \hat{C}_{ik}(\mathbf{u})\,\mathrm{d}\hat{C}_{jk}(\mathbf{u}) - 1\right\} = \boldsymbol{\lambda}_k^\mathsf{T}\hat{\mathcal{T}}'_k\boldsymbol{\lambda}_k$$

where $\hat{\mathcal{T}}'_k$ is a $m \times m$ matrix with $[\hat{\mathcal{T}}'_k]_{ij} = 4\int \hat{C}_{ik}(\mathbf{u})\,\mathrm{d}\hat{C}_{jk}(\mathbf{u}) - 1$.

To facilitate their interpretation, coefficients of correlation are usually built to have a null expectation under the hypothesis of independence. In addition, under perfect negative or positive dependence, the coefficients take values -1 and 1 respectively. To preserve this property as best as possible, let us define the asymptotically equivalent expression

$$\begin{aligned}
[\hat{\mathcal{T}}^*_k]_{ij} &= \frac{1}{N_{ijk}}\sum_{s=1}^{n_{ik}}\sum_{t=1}^{n_{jk}} \operatorname{sign}\left(\frac{R^k_{is}}{n_{ik}} - \frac{R^k_{jt}}{n_{jk}}\right)\operatorname{sign}\left(\frac{S^k_{is}}{n_{ik}} - \frac{S^k_{jt}}{n_{jk}}\right) \\
&= 4\frac{n_{ik}n_{jk}}{N_{ijk}}\int \hat{C}_{jk}(\mathbf{u})\,\mathrm{d}\hat{C}_{ik}(\mathbf{u}) + O(n_{ik}) + O(n_{jk}) \\
&\qquad + \frac{n_{ik}n_{jk}}{N_{ijk}}\left[1 - 2\int\left\{\hat{C}_{jk}(u,1) + \hat{C}_{jk}(1,v)\right\}\mathrm{d}\hat{C}_{ik}(u,v)\right]
\end{aligned}$$

where

$$N_{ijk} = \sum_{s=1}^{n_{ik}}\sum_{t=1}^{n_{jk}} \mathbb{1}\left(\frac{R^k_{is}}{n_{ik}} \neq \frac{R^k_{jt}}{n_{jk}}\right)\mathbb{1}\left(\frac{S^k_{is}}{n_{ik}} \neq \frac{S^k_{jt}}{n_{jk}}\right)$$

is such that $n_{ik}n_{jk} - \min(n_{ik}, n_{jk}) \leq N_{ijk} \leq n_{ik}n_{jk}$, hence the ratio $n_{ik}n_{jk}/N_{ijk}$ is asymptotically equivalent to 1. The definition above is chosen so that $[\hat{\mathcal{T}}^*_k]_{ii} = \hat{\tau}_{ik}$.

We look at the values of $[\hat{\mathcal{I}}_k^*]_{ij}$ in some special cases. Under the assumption of independence, $\mathrm{E}([\hat{\mathcal{I}}_k^*]_{ij}) = (n_{ik}n_{jk}/N_{ijk})(1/n_{jk} - 1/n_{ik})^2$. When the ranks are in perfect positive dependence, $[\hat{\mathcal{I}}_k^*]_{ij} = 1$. When the ranks are antithetic, we rather get

$$-1 + \frac{2}{N_{ijk}} \sum_{s=1}^{n_{ik}} \sum_{t=1}^{n_{jk}} \mathbb{1}\left\{ \frac{sn_{jk}}{n_{ik}} < t < \frac{(s+1)n_{jk}}{n_{ik}} - 1 \right\} + \mathbb{1}\left\{ \frac{(s+1)n_{jk}}{n_{ik}} - 1 < t < \frac{sn_{jk}}{n_{ik}} \right\}.$$

These three values depend only on the values of n_{ik} and n_{jk}. If $n_{ik} = n_{jk}$, we get the usual -1, 0, 1 values. Unfortunately however, $\mathrm{E}([\hat{\mathcal{I}}_k^*]_{ij})$ is not in general the mid-point of the values of $\hat{\mathcal{I}}_k^*$ under perfect dependence. As a consequence, even a linear transformation cannot make $\hat{\tau}^*_{\boldsymbol{\lambda}_k} = \boldsymbol{\lambda}_k^\mathsf{T} \hat{\mathcal{I}}_k^* \boldsymbol{\lambda}_k$ fit the magical values of -1, 0 and 1 appropriately, contrarily to the inconsistent $\hat{\tau}_{\boldsymbol{\lambda}_k}$ who preserves this property.

Theorem 4.2 *$\hat{\tau}^*_{\boldsymbol{\lambda}_k} \to \tau$ and $\hat{\tau}^*_{\boldsymbol{\mu}_k} \to \tau_1$ almost surely as $k \to \infty$.*

The asymptotic normality of Kendall's τ can be derived from the theory on U-statistics. Unfortunately $\tau^*_{\boldsymbol{\lambda}_k}$ does not fall within this paradigm and its asymptotic distribution remains to be derived.

5 Simulations

Simulations are used to explore the finite-sample performances of the weighted methods we propose under different scenarios. Scalar weights that are proportional to the sample sizes are considered as well as the adaptive MAMSE weights. Note that a function to compute the MAMSE weights is available in the library MAMSE available from the Comprehensive R Archive Network.

For each table, a small bootstrap study is used to evaluate the error due to simulation. Unless it is otherwise stated, the standard deviation of that error is less than the last significant digit shown in the tables.

5.1 Common copula

We first consider a scenario where data comes from a common bivariate copula: a Clayton distribution (see Clayton (1978) or Nelsen (1999)) whose parameter is set to yield a Spearman correlation of $\rho \in \{0.1, 0.5, 0.9\}$.

A total of $5n$ data points are available as samples of equal sizes from 5 populations. It is assumed that their marginal distributions are not comparable, hence the data cannot be pooled. The weighted methods allow using all of the data, rather than relying only on one sample of size n. For each value of ρ and $n \in \{10, 20, 50\}$, 10000 sets of samples are simulated.

To evaluate the precision of the weighted empirical copula $\hat{C}^*$, Table 2 shows the ratio $100 \int |\hat{C}_1(\mathbf{u}) - C(\mathbf{u})| \, \mathrm{d}\mathbf{u} / \int |\hat{C}^*(\mathbf{u}) - C(\mathbf{u})| \, \mathrm{d}\mathbf{u}$ for both the scalar and the MAMSE weights. Values over 100 favor the weighted method. The empirical copula based on one population of size n serves as the basis of comparison and is denoted $\hat{C}_1(\mathbf{u})$. In practice, pooling the data would not be possible, but in a simulation framework, these results are shown for comparison purposes. Note that the standard error due to simulation can reach 1.3 units in Table 2

The pooled estimate provides estimates about 2.25 times more accurate than using a single of the five populations. This number corresponds approximately to $\sqrt{5}$, the value that would be expected theoretically. Using scalar weights yields exactly the same performance, hence when the data cannot be pooled, using $\hat{C}_{\boldsymbol{\lambda}}(\mathbf{u})$ allows to use all of the data with virtually no loss. Not surprisingly, the MAMSE weights perform slightly worse: estimating weights from the data has a cost. The magnitude of the difference is however rather small.

Weighted coefficients of correlation are calculated on the same samples described above. Their performance is measured by ratios such as $100\mathrm{MSE}(\hat{\rho}_1)/\mathrm{MSE}(\hat{\rho}_{\boldsymbol{\lambda}})$. The results appear in Table 3; values above 100 favor the weighted methods. For the pooled estimates, the standard deviation of the error due to simulation can reach 30, for all other cases, it is less than 10, hence the rounding.

Table 2: Performance of the weighted methods as measured by $100 \int |\hat{C}_1(\mathbf{u}) - C(\mathbf{u})| \, \mathrm{d}\mathbf{u} / \int |\hat{C}^*(\mathbf{u}) - C(\mathbf{u})| \, \mathrm{d}\mathbf{u}$. Five samples of size n are simulated from a Clayton distribution with Spearman's correlation ρ. Each scenario is repeated 10000 times.

	$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
	$n = 10$	20	50	$n = 10$	20	50	$n = 10$	20	50
Pooled	227	225	224	228	224	224	226	226	224
Scalar weights	227	225	224	228	224	224	226	226	224
MAMSE weights	210	209	208	214	212	212	223	223	222

Table 3: Performance of different weighted coefficients of correlation as measured by a ratio of the kind $100\,\mathrm{MSE}(\hat{\rho}_1)/\mathrm{MSE}(\hat{\rho}_{\boldsymbol{\lambda}})$. Five samples of size n are simulated from a Clayton distribution with Spearman's correlation ρ. Each scenario is repeated 10000 times.

	$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
	$n = 10$	20	50	$n = 10$	20	50	$n = 10$	20	50
ρ pooled	540	510	490	600	550	510	910	750	580
$\hat{\rho}_{\boldsymbol{\lambda}}$	500	480	480	470	480	480	320	360	400
$\hat{\rho}_{\boldsymbol{\mu}}$	320	330	330	320	350	370	300	350	390
τ pooled	630	560	510	670	580	520	750	650	550
$\hat{\tau}_{\boldsymbol{\lambda}}$	500	480	490	510	500	500	490	500	490
$\hat{\tau}_{\boldsymbol{\mu}}$	330	330	330	360	370	380	460	480	480
$\hat{\tau}^*_{\boldsymbol{\lambda}}$	430	450	470	420	450	470	300	340	400
$\hat{\tau}^*_{\boldsymbol{\mu}}$	290	310	320	310	340	370	290	330	390

The pooled methods should theoretically yield a mean squared error five times smaller than using one population. The deviance from that number may be explained in part by the larger simulation error, but also by the structure implied by the use of ranks. As a matter of fact, the improvement is closer to 500 for larger sample size where the effect of the ranks start fading.

Using scalar weights allows to use the information from all the samples when their margins are not comparable. In most cases, the improvement is by a factor of 5. A notable exception arises for $\hat{\rho}_{\boldsymbol{\lambda}}$ when $\rho = 0.9$ and could be caused by the collapsing of the data points to a line as we approach the limit case of $\rho = 1$: n points may not align as nicely as $5n$ points. The quadratic form $\hat{\tau}^*_{\boldsymbol{\lambda}}$ does not perform as well as expected. The complexity of its estimation which involves estimating many terms of the type $\int \hat{C}_i \, \mathrm{d}\hat{C}_j$ may have an impact.

The coefficients that use the MAMSE weights feature a mean squared error about three times smaller than that obtained when using only one sample of size n. The cost of estimating the weights seems more important compared to the results obtained in Table 2, but the improvement is still substantial.

Overall, all weighted methods perform very well, allowing to recover information that would otherwise be lost.

5.2 Different correlations

We now turn to a situation where the hypothesis on the similarities between the copulas fails. We propose two scenarios where copulas from a same family have different correlations. Equal samples of size $n \in \{20, 50, 100, 250\}$ are drawn from each.

Scenario A: $\rho = \textit{0.35}$, 0.25, 0.30, 0.40 and 0.45 respectively
Scenario B: $\rho = \textit{0.25}$, 0.30, 0.35, 0.40 and 0.45 respectively

Table 4: Performance of the weighted Spearman's ρ with scalar or adaptive weights as measured by a ratio of the kind $100\,\mathrm{MSE}(\hat{\rho}_1)/\mathrm{MSE}(\hat{\rho}_{\boldsymbol{\lambda}})$. Five samples of size n are simulated from a Clayton, a Gumbel-Hougaard, a Frank or a Normal distribution. Their true correlations correspond to one of two scenarios. Averages over obtained from 10000 repetitions.

		Scalar Weights				MAMSE Weights			
	Family	$n=20$	50	100	250	$n=20$	50	100	250
Scenario A	Normal	530	490	500	490	330	330	310	270
	Clayton	530	490	500	490	330	330	310	270
	Frank	500	500	500	500	340	330	310	270
	Gumbel	490	510	510	500	330	330	310	270
Scenario B	Normal	310	160	80	30	280	200	130	70
	Clayton	310	150	80	40	280	200	130	80
	Frank	310	150	80	30	280	200	130	80
	Gumbel	310	160	80	30	290	200	140	80

Population 1 (whose true correlation is in italic above) is deemed of prime interest. Different families of copula are used: Clayton, Gumbel-Hougaard, Frank and Normal. In each case, 10000 repetitions are produced. Table 4 displays the ratio of the mean squared error for estimating the correlation of Population 1.

Under Scenario A, the correlation of the population of interest sits squarely within the range of correlations. It is thus not surprising to see improvements of a magnitude similar to that of Section 5.1.

Scenario B represents a case where intuition suggests to use adaptive weights rather than scalar weights. While the scalar weights perform 10% better than the MAMSE weights for $n = 20$, the MAMSE weights are to be favored for all other sample sizes. Note however that for $n = 250$, the weighted methods perform worse than the estimate based only on the population of interest. Under that scenario, the improved variance does not compensate for the bias caused by using populations with the wrong correlation.

Overall, the use of a weighted method allows to use information which would otherwise be discarded, yielding once more appreciable gains under different circumstances.

5.3 Tests of independence

Finally, we use the weighted coefficients of correlation as statistics to test independence and study the power of such tests.

Five populations of equal size $n \in \{10, 20, 50\}$ are simulated from a Clayton copula. Different values of ρ are used to produce smooth power plots. For each such values, 5000 repetitions are produced.

Some coefficients are asymptotically Normal. Their test of independence is then based on that approximation since this is the approach that would likely be used in practice. More specifically, with a sample of size n, Spearman's ρ is compared to a centered Normal with variance $1/(n-1)$ and Kendall's τ to a centered Normal with variance $(4n+10)/\{9n(n-1)\}$. For a weighted coefficient based on five populations of equal sizes n, the variances above are simply divided by 5.

The asymptotic distribution of the MAMSE weighted coefficients of correlation are not known. That of $\hat{\tau}^*_{\boldsymbol{\lambda}}$ remains to be determined as well. In these cases, a simulation study based on 100000 repetitions was used to determine the sampling distributions under the hypothesis of independence. Critical values appear in Table 5: the hypothesis of independence is rejected at the 5% level when an estimated coefficient does not belong to the interval $[LB, UL]$.

Figure 1 shows the power of a test of independence based on different coefficients of correlation. The dashed line shows the power of a test based only on one population of size n. Even though such an operation would not be possible in practice, the coefficients are also calculated on the pooled data and drawn as a plain line. The dotted line shows the power of a test based on a coefficient with scalar weights proportional to the

Table 5: Critical values for a 5% level test of independence using different weighted coefficients of correlation whose asymptotic distribution is unknown. These values are based on a large simulation with 100000 repetitions and are valid when 5 samples of equal size n are used. Other situations can be simulated easily in the same fashion.

	$n = 10$		$n = 20$		$n = 50$	
Coefficient	LB	UB	LB	UB	LB	UB
$\hat{\tau}^*_{\boldsymbol{\lambda}}$	-0.233	0.233	-0.147	0.147	-0.087	0.087
$\hat{\rho}_{\boldsymbol{\mu}}$	-0.384	0.336	-0.262	0.235	-0.160	0.150
$\hat{\tau}^*_{\boldsymbol{\mu}}$	-0.304	0.268	-0.191	0.171	-0.110	0.104
$\hat{\tau}_{\boldsymbol{\mu}}$	-0.286	0.251	-0.186	0.165	-0.109	0.102

sample sizes. The mixed (dashes and dots) line gives the power of a test based on the MAMSE-weighted coefficients.

In all cases, the weighted methods offer an appreciable improvement over the use of only one sample of size n. The scalar weights offer performances very close to the pooled data, which is expected since their ARE is 1. Very little information is lost. The MAMSE weighted coefficient offers a slightly less powerful test which is nonetheless clearly superior to the option of using only n data.

Let us now consider two scenario where the copulas underlying each distribution differ. Samples of size 20 come from 5 populations. The x-axis of the power graphs below correspond to Spearman's correlation in Population 1. Under Scenario A, the other four populations have a correlation of 0.1. Under Scenario B, Population 2 has a correlation of 0.1, but the other 3 are independent ($\rho = 0$). The power plots are shown in the last two rows of Figure 1.

Scenarios A and B suppose that we use data that does not follow the same distribution as the target. The tests based on pooled data or on scalar weights do not behave well under such situations, displaying very low power. On the other hand, the test based on the MAMSE weights adapts to the discrepancies and features a power similar to a test based on Population 1 only. It thus offers a robust alternative akin to the best available choice, even though we do not know what that choice is in a given situation.

6 Conclusion

Suppose that data comes from m populations whose distributions share a common copula. Statistics based on ranks can be calculated on each sample even if they have different margins. Mixing the samples with scalar weights yields empirical estimates of the copula and weighted coefficients of correlation that were shown consistent and asymptotically Normal.

In the case where copulas may differ from a population to another, we propose adaptive weights, such as the MAMSE weights, that use the data to evaluate the discrepancies between the populations. The weighted empirical copula and weighted coefficients of correlation based on adaptive weights are consistent, but we do not know their limiting distribution. The MAMSE weights are not optimized for performance, but they prove that it is possible to weight copulas adaptively to obtain consistent estimates that offer improved performance.

Simulations shown that when the true underlying copula is the same in all populations, using scalar weights yields performances nearly as good as pooling the data. The MAMSE weights are slightly worse, but feature nonetheless an appreciable improvement compared to the dismissing of some data. When the true copula in the m populations differ, the scalar weights can fail badly while the adaptive MAMSE weights perform similarly to using the only reliable sample. The adaptive weights thus offer improved performance as well as robustness to discrepancies between populations.

Now that the concept is shown to work well, optimizing the performance of MAMSE-like weights could lead to better estimates. We leave this to future research and are indeed looking forward to applications of the weighted methods in case studies and its development in a time series context.

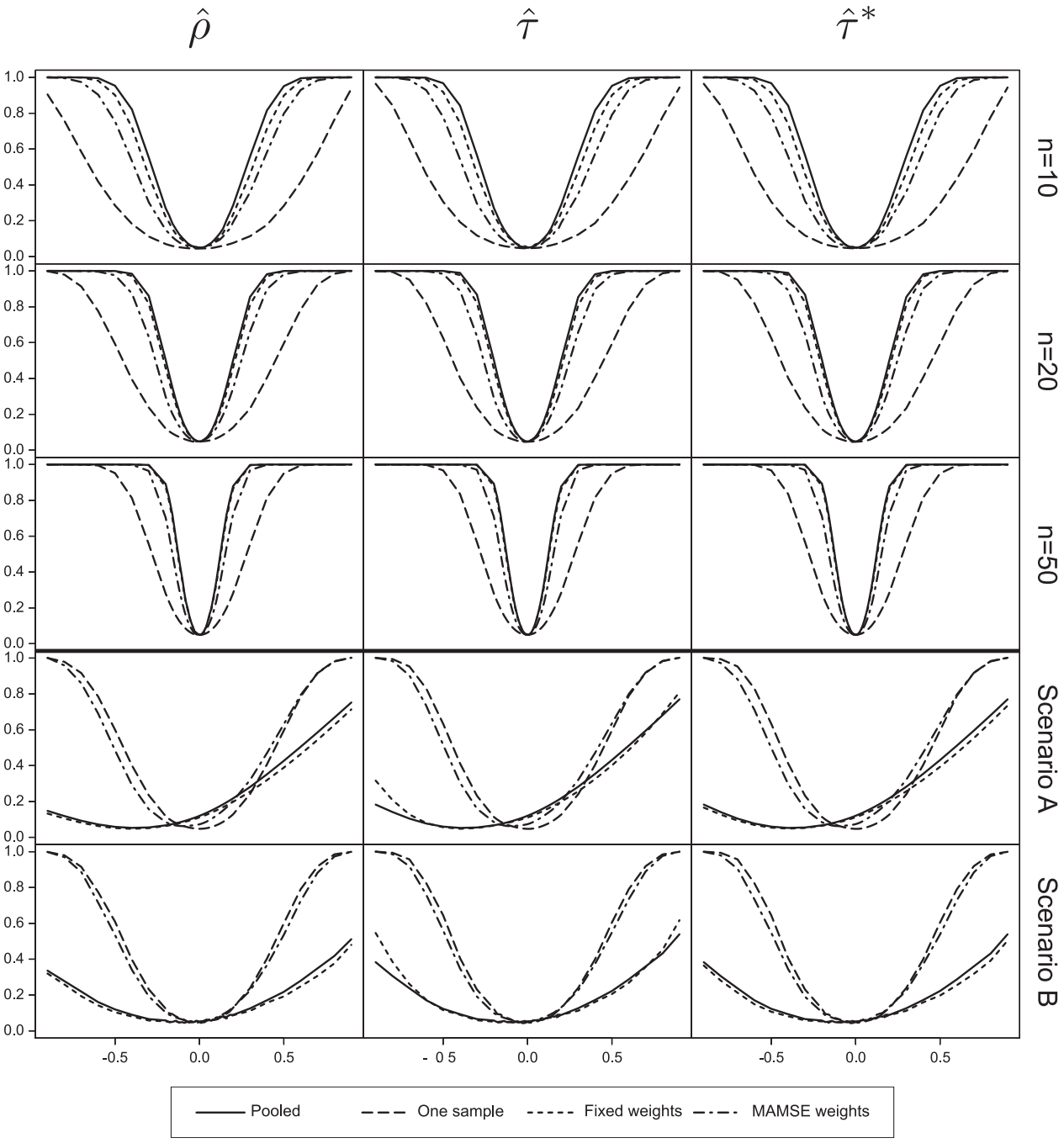

Figure 1: Power of a test of independence based on different coefficient of correlations. The three columns of plots are respectively for ρ, τ and τ^*. Equal samples of size n are drawn from 5 populations distributed as a Clayton distribution with correlation ρ. For the last two rows, Equal samples of size $n = 20$ are drawn from 5 populations distributed as a Clayton distribution with correlations based on two different scenarios.

Appendix

Mathematical proofs are presented below. Different Lemmas appear in the Appendix as well to make the more technical proofs more readable. Proofs of trivial results are not provided.

Lemma 6.1 *Let* $\mathbf{u}, \mathbf{v} \in [0,1]^p$ *be such that* $v_\ell \leq u_\ell$ *for* $\ell = 1, \ldots, p$. *Then*

$$0 \leq \hat{C}_{ik}(\mathbf{u}) - \hat{C}_{ik}(\mathbf{v}) \leq \sum_{\ell=1}^{p} \frac{\lceil n_{ik}(u_\ell - v_\ell) \rceil}{n_{ik}}$$

where $\lceil x \rceil$ *denotes the smallest integer greater or equal to* x.

Proof of Lemma 6.1. The lower bound is a consequence of the monotone properties of distribution functions. The upper bound can be derived from developing the probability represented by $\hat{C}_{ik}(\mathbf{u}) - \hat{C}_{ik}(\mathbf{v})$ and using the uniformity of the margins on the points of the form $\{a/n_{ik} : a = 1, \ldots, n_{ik}\}$. ■

Let $\mathcal{G}_k^*$ be an extended grid that includes the axes:

$$\mathcal{G}_k^* = \left\{0, \frac{1}{n_{1k}}, \frac{2}{n_{1k}}, \ldots, 1\right\} \times \cdots \times \left\{0, \frac{1}{n_{1k}}, \frac{2}{n_{1k}}, \ldots, 1\right\}.$$

Lemma 6.2

$$\sup_{\mathbf{u} \in [0,1]^p} \left| \hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) \right| \leq \frac{p}{n_{1k}} + \sup_{\mathbf{u} \in \mathcal{G}_k^*} \left| \hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) \right|.$$

Proof of Lemma 6.2. For a fixed k, $|\hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u})|$ is a bounded function on the compact set $[0,1]^p$ and hence its maximum is attained. Let $\mathbf{v} \in [0,1]^p$ be a point where that maximum is achieved. We treat two cases.

Case 1: $\hat{C}_{1k}(\mathbf{v}) \geq \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v})$.

Let $\mathbf{v}^* = [v_1^*, \ldots, v_p^*]^\mathsf{T}$ be defined by $v_\ell^* = \lfloor n_{1k} v_\ell \rfloor / n_{1k}$, where $\lfloor x \rfloor$ denotes the largest integer smaller or equal to x. Then $\mathbf{v}^* \in \mathcal{G}_k^*$ is on the same "plateau" of the multivariate step function $\hat{C}_{1k}(\mathbf{u})$ as $\mathbf{v}$, meaning that $\hat{C}_{1k}(\mathbf{v}^*) = \hat{C}_{1k}(\mathbf{v})$ and $\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v}^*) \leq \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v})$.

Recalling that $\mathbf{v}$ is the point where the difference between $\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u})$ and $\hat{C}_{1k}(\mathbf{u})$ is maximized, we can write

$$\begin{aligned} |\hat{C}_{1k}(\mathbf{v}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v})| &= \hat{C}_{1k}(\mathbf{v}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v}) \leq \hat{C}_{1k}(\mathbf{v}^*) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v}^*) \\ &\leq \sup_{\mathbf{u} \in \mathcal{G}_k^*} \left| \hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) \right| \leq \frac{p}{n_{1k}} + \sup_{\mathbf{u} \in \mathcal{G}_k^*} \left| \hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) \right|, \end{aligned}$$

meaning that the maximum occurs at a point of the grid $\mathcal{G}_k^*$.

Case 2: $\hat{C}_{1k}(\mathbf{v}) \leq \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v})$.

Let $\mathbf{v}^* = [v_1^*, \ldots, v_p^*]^\mathsf{T}$ be defined by $v_\ell^* = \lceil n_{1k} v_\ell \rceil / n_{1k}$, where $\lceil x \rceil$ denotes the smallest integer greater or equal to x. Then, $\mathbf{v}^* \in \mathcal{G}_k^*$ and $\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v}^*) \geq \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v})$ since $C_{\boldsymbol{\mu}_k}(\mathbf{u})$ is a nondecreasing function and $\mathbf{v}^* \geq \mathbf{v}$. By Lemma 6.1, $\hat{C}_{1k}(\mathbf{v}^*) - \hat{C}_{1k}(\mathbf{v}) \leq p/n_{1k}$. Recalling that $\mathbf{v}$ maximizes the difference between $\hat{C}_{1k}(\mathbf{u})$ and $C_{\boldsymbol{\mu}_k}(\mathbf{u})$, we can write

$$\begin{aligned} |\hat{C}_{1k}(\mathbf{v}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v})| &= \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v}) - \hat{C}_{1k}(\mathbf{v}) \leq \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{v}^*) - \hat{C}_{1k}(\mathbf{v}^*) + \frac{p}{n_{1k}} \\ &\leq \frac{p}{n_{1k}} + \sup_{\mathbf{u} \in \mathcal{G}_k^*} \left| \hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) \right|. \end{aligned}$$

Combining Cases 1 and 2 yields the desired result. ■

Lemma 6.3 *We have* $\sup_{\mathbf{u} \in \mathcal{G}_k^*} \left| \hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) \right| \to 0$ *almost surely as* $k \to \infty$.

Proof of Lemma 6.3. Let $\epsilon > 0$. For any given $k \in \mathbb{N}$, let $\mathbf{u}_k = [u_{k1}, \ldots, u_{kp}]^\mathsf{T}$ be the point of the grid $\mathcal{G}_k^*$ where $|\hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u})|$ is maximized. Consider the events

$$\begin{aligned} A_k &= \left\{\hat{C}_{1k}(\mathbf{u}_k) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}_k) > \epsilon\right\}, \\ B_k &= \left\{\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}_k) - \hat{C}_{1k}(\mathbf{u}_k) > \epsilon\right\}, \\ C_k &= \left\{\mathbf{u}_k \in \left[\frac{\epsilon}{2}, 1\right]^p\right\}. \end{aligned}$$

The negation of Lemma 6.3 is $\{A_k \cup B_k\}$ *i.o.* which will happen if and only if $\{\left(A_k \cup B_k \cap C_k^C\right) \cup (A_k \cup B_k \cap C_k)\}$*i.o.* We will show that neither of the two events in this decomposition can occur infinitely often.

Case 1: $A_k \cup B_k \cap C_k^C$.

We have

$$\begin{aligned} \left|\hat{C}_{1k}(\mathbf{u}_k) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}_k)\right| &\leq \left|\hat{C}_{1k}(\mathbf{u}_k)\right| + \left|\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}_k)\right| \\ &= \hat{C}_{1k}(\mathbf{u}_k) + \sum_{i=1}^{m} \mu_{ik} \hat{C}_{ik}(\mathbf{u}_k) \leq 2 \min_{\ell \in \{1,\ldots,p\}} u_{k\ell} \leq \epsilon \end{aligned}$$

because $\hat{C}_{ik}(\mathbf{u}_k) > \min_{\ell \in \{1,\ldots,p\}} u_{k\ell}$ is incompatible with uniform univariate margins and the MAMSE weights sum to 1. Consequently, $A_k \cup B_k \cap C_k^C = \emptyset$ for all k.

Case 2: $A_k \cup B_k \cap C_k$.

Let $\mathbf{v}$ be a vector of integers such that $\mathbf{v}/n_{1k} = \mathbf{u}_k$; we temporarily omit the index k for notational simplicity. Let also $\mathbf{w} = [w_1, \ldots, w_p]^\mathsf{T}$ be a point from the set

$$\mathcal{W} = \left\{0, 1, \ldots, \left\lfloor \frac{n_{1k}}{2p} \epsilon \right\rfloor\right\} \times \cdots \times \left\{0, 1, \ldots, \left\lfloor \frac{n_{1k}}{2p} \epsilon \right\rfloor\right\}.$$

The points $(\mathbf{v} - \mathbf{w})/n_{1k}$ belong to $\mathcal{G}_k$ since $\mathbf{u}_k \in [\epsilon/2, 1]^p$. Next, we show that

$$\left|\hat{C}_{\boldsymbol{\mu}_k}\left(\frac{\mathbf{v} - \mathbf{w}}{n_{1k}}\right) - \hat{C}_{1k}\left(\frac{\mathbf{v} - \mathbf{w}}{n_{1k}}\right)\right| \geq \frac{\epsilon}{2} - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} \geq 0$$

by treating two subcases. Note that the last inequality holds because

$$\frac{1}{n_{1k}} \sum_{\ell=1}^{p} w_\ell \leq \frac{p}{n_{1k}} \left\lfloor \frac{n_{1k}}{2p} \epsilon \right\rfloor \leq \frac{\epsilon}{2}.$$

Subcase A: $A_k \cap C_k$.

From the fact monotonicity of copulas and Lemma 6.1, we have

$$\begin{aligned} \hat{C}_{1k}\left(\frac{\mathbf{v} - \mathbf{w}}{n_{1k}}\right) - \hat{C}_{\boldsymbol{\mu}_k}\left(\frac{\mathbf{v} - \mathbf{w}}{n_{1k}}\right) &\geq \hat{C}_{1k}(\mathbf{u}_k) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}_k) - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} \\ &\geq \frac{\epsilon}{2} - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} \geq 0. \end{aligned}$$

Subcase B: $B_k \cap C_k$.

By Lemma 6.1, we have

$$\hat{C}_{\boldsymbol{\mu}_k}\left(\frac{\mathbf{v}}{n_{1k}}\right) - \hat{C}_{\boldsymbol{\mu}_k}\left(\frac{\mathbf{v} - \mathbf{w}}{n_{1k}}\right) = \sum_{i=1}^{m} \mu_{ik} \left\{\hat{C}_{ik}\left(\frac{\mathbf{v}}{n_{1k}}\right) - \hat{C}_{ik}\left(\frac{\mathbf{v} - \mathbf{w}}{n_{1k}}\right)\right\}$$

$$\leq \sum_{i=1}^{m} \frac{1}{n_{ik}} \sum_{\ell=1}^{p} \left\lceil \frac{n_{ik} w_\ell}{n_{1k}} \right\rceil \leq \sum_{i=1}^{m} \frac{1}{n_{ik}} \sum_{\ell=1}^{p} \left(\frac{n_{ik} w_\ell}{n_{1k}} + 1 \right) = \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} + \sum_{i=1}^{m} \frac{p}{n_{ik}}.$$

Hence,

$$\hat{C}_{\boldsymbol{\mu}_k}\left(\frac{\mathbf{v}-\mathbf{w}}{n_{1k}}\right) \geq \hat{C}_{\boldsymbol{\mu}_k}\left(\frac{\mathbf{v}}{n_{1k}}\right) - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} - \sum_{i=1}^{m} \frac{p}{n_{ik}} = \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}_k) - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} - \sum_{i=1}^{m} \frac{p}{n_{ik}}.$$

Let us consider only k that are large enough to make $\sum_{i=1}^{m} p/n_{ik} < \epsilon/2$. From the previous inequality and the monotonicity of $\hat{C}_{1k}(\mathbf{u})$, we obtain

$$\begin{aligned} \hat{C}_{\boldsymbol{\mu}_k}\left(\frac{\mathbf{v}-\mathbf{w}}{n_{1k}}\right) - \hat{C}_{1k}\left(\frac{\mathbf{v}-\mathbf{w}}{n_{1k}}\right) &\geq \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}_k) - \hat{C}_{1k}(\mathbf{u}_k) - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} - \sum_{i=1}^{m} \frac{p}{n_{ik}} \\ &\geq \epsilon - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} - \sum_{i=1}^{m} \frac{p}{n_{ik}} \geq \frac{\epsilon}{2} - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} \geq 0. \end{aligned}$$

Combining subcases A and B yields

$$P_k(\boldsymbol{\mu}_k) \geq \int \{\hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) - \hat{C}_{1k}(\mathbf{u})\}^2 \, \mathrm{d}M_k(\mathbf{u}) \geq \frac{1}{n_{1k}^p} \sum_{\mathbf{w} \in \mathcal{W}} \left(\frac{\epsilon}{2} - \frac{\sum_{\ell=1}^{p} w_\ell}{n_{1k}} \right)^2.$$

The sum above corresponds to a Riemann sum for the multiple integral

$$\int_0^{\frac{\epsilon}{2p}} \cdots \int_0^{\frac{\epsilon}{2p}} \left(\frac{\epsilon}{2} - \sum_{\ell=1}^{p} y_\ell \right)^2 \mathrm{d}y_1 \cdots \mathrm{d}y_p = K_p.$$

The number K_p is a fixed positive constant for any fixed p.

As a consequence, there exists a k_0 such that for all $k \geq k_0$, $P_k(\boldsymbol{\mu}_k) > K_p/2 > 0$, a contradiction with Assumption 3.2. We must thus conclude that $A_k \cup B_k \cap C_k$ occurs at most a finite number of times.

Hence, $A_k \cup B_k$ occurs at most a finite number of times and $\sup_{\mathbf{u} \in \mathcal{G}_k^*} \left| \hat{C}_{1k}(\mathbf{u}) - \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) \right| \to 0$ almost surely as $k \to \infty$. ■

Proof of Theorem 3.3. Consider the decomposition

$$\sup_{\mathbf{u} \in [0,1]^p} \left| \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) - C_1(\mathbf{u}) \right| \leq \sup_{\mathbf{u} \in [0,1]^p} \left| \hat{C}_{\boldsymbol{\mu}_k}(\mathbf{u}) - \hat{C}_{1k}(\mathbf{u}) \right| + \sup_{\mathbf{u} \in [0,1]^p} \left| \hat{C}_{1k}(\mathbf{u}) - C_1(\mathbf{u}) \right|.$$

The first term on the right-hand side goes to 0 almost surely by Lemma 6.2 and Lemma 6.3, the second term does likewise by Equation 1. ■

Proof of Lemma 3.1. The uniform convergence in Theorem 3.1 implies that a sequence of random vectors with distributions $\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u})$ will converge weakly to a random vector with distribution $C(\mathbf{u})$. As a consequence, expectations of continuous bounded functions of these variables converge almost surely. ■

Proof of Theorem 4.2. It is sufficient to show that $|\int \hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) \, \mathrm{d}\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) - \int C(\mathbf{u}) \, \mathrm{d}C(\mathbf{u})| \to 0$ almost surely as $k \to \infty$ and similarly with $\boldsymbol{\mu}_k$. This expression is bounded by

$$\left| \int \hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) \, \mathrm{d}\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) - \int C(\mathbf{u}) \, \mathrm{d}\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) \right| + \left| \int C(\mathbf{u}) \, \mathrm{d}\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) - \int C(\mathbf{u}) \, \mathrm{d}C(\mathbf{u}) \right|.$$

The first term is bounded by $\sup_{\mathbf{u} \in [0,1]^p} |\hat{C}_{\boldsymbol{\lambda}_k}(\mathbf{u}) - C(\mathbf{u})|$ which converges to 0 almost surely by Theorem 3.1 or Theorem 3.3. The second converges to zero by the argument of Lemma 3.1. ■

References

D. Blest (2000). Rank correlation—an alternative measure, *Australian and New Zealand Journal of Statistics*, **42**, 101–111.

D. G. Clayton (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, **65**, 141–151.

P. Deheuvels (1979). La fonction de dépendance empirique et ses propriétés: un test non paramétrique d'indépendance, *Académie royale de Belgique–Bulletin de la classe des sciences*, **65**, 274–292.

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

C. Genest & J.-F. Plante (2003). On Blest's measure of rank correlation, *The Canadian Journal of Statistics*, **31**, 35–52.

E. L. Lehmann (1998). *Nonparametrics: Statistical Methods Based on Ranks*. Revised First Edition. Prentice Hall, Upper Saddle River, New Jersey.

D. G. Luenberger (2003). *Linear and nonlinear programming, second edition*, Kluwer, Norwell, 491 pp.

R. B. Nelsen (1999). *An Introduction to Copulas*, Lecture Notes in Statistics No. 139, Springer, Berlin.

J. Pinto da Costa & C. Soares (2005). A weighted rank measure of correlation, *Australian and New Zealand Journal of Statistics*, **47**, 515–529.

J.-F. Plante (2007). *Adaptive likelihood weights and mixtures of empirical distributions*, unpublished doctoral dissertation, Department of Statistics, The University of British Columbia, 171 pp.

J.-F. Plante (2008). Nonparametric adaptive likelihood weights, *The Canadian Journal of Statistics*, **36**, 443–461.

J.-F. Plante (2009). Asymptotic properties of the MAMSE adaptive likelihood weights. *Journal of Statistical Planning and Inference*, **139**, 2147–2161.

F. H. Ruymgaart, G. R. Shorack & W. R. van Zwet (1972). Asymptotic normality of nonparametric tests for independence, *The Annals of Mathematical Statistics*, **43**, 1122–1135.

F. H. Ruymgaart (1974). Asymptotic normality of nonparametric tests for independence, *The Annals of Statistics*, **2**, 892–910.

A. Sklar (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, **8**, 229–231.

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

X. Wang (2001). *Maximum weighted likelihood estimation*, unpublished doctoral dissertation, Department of Statistics, The University of British Columbia, 151 pp.

X. Wang & J. V. Zidek (2005). Selecting likelihood weights by cross-validation, *The Annals of Statistics*, **33**, 463–501.