

**Developments in Fuzzy Sets,
Intuitionistic Fuzzy Sets,
Generalized Nets and Related Topics.
Volume I: Foundations**

**New Developments in Fuzzy Sets,
Intuitionistic Fuzzy Sets,
Generalized Nets and Related Topics
Volume I: Foundations**

Editors

Editors
Krassimir T. Atanassov
Michał Baczyński
Józef Drewniak
Krassimir T. Atanassov
Janusz Kacprzyk
Michał Baczyński
Maciej Krawczyk
Józef Drewniak
Janusz Kacprzyk
Sławomir Zadrozny
Maciej Krawczyk
Eulalia Szmidt
Maciej Wygralak
Sławomir Zadrozny

SRI PAS



IBS PAN

**New Developments in Fuzzy Sets,
Intuitionistic Fuzzy Sets,
Generalized Nets and Related Topics
Volume I: Foundations**



**Systems Research Institute
Polish Academy of Sciences**

**New Developments in Fuzzy Sets,
Intuitionistic Fuzzy Sets,
Generalized Nets and Related Topics
Volume I: Foundations**

Editors

Krassimir T. Atanassov

Michał Baczyński

Józef Drewniak

Janusz Kacprzyk

Maciej Krawczak

Eulalia Szmidt

Maciej Wygralak

Sławomir Zadrozny

IBS PAN



SRI PAS

© **Copyright by Systems Research Institute**
Polish Academy of Sciences
Warsaw 2012

All rights reserved. No part of this publication may be reproduced, stored in retrieval system or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without permission in writing from publisher.

Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
www.ibspan.waw.pl
ISBN 83-894-7540-5

Dedicated to Professor Beloslav Riečan on his 75th anniversary

Computation of the measures of dependence for imprecise data

Olgierd Hryniewicz and Karol Opara

Systems Research Institute,

Newelska 6, 01-447 Warszawa, Poland

hryniewi@ibspan.waw.pl, Karol.Opara@ibspan.waw.pl

Abstract

Nonparametric measures of statistical dependence, such as Spearman's ρ or Kendall's τ , are very convenient when the type of dependence, described by a respective copula, and the type of marginal distributions are unknown. Computation of these statistics becomes difficult when available data are imprecise, and given as intervals of real numbers. In the paper we present the results of the research oriented on the algorithmic problems in the case of the computation of Kendall's τ .

Keywords: measures of dependence, imprecise data, Kendall's τ .

1 Introduction

Statistical dependencies between data have been investigated since the XIX century, First statistical measures of dependence were proposed by statisticians when the investigations of the probabilistic foundations of statistical dependence were on a very initial stage. For example, well known measures of dependence such as Pearson's coefficient of linear correlation r or Spearman's coefficient of rank correlation ρ were proposed without referring to particular probabilistic structures of dependence. Other measures of dependence, such as e.g. Kendall's τ have been proposed later on taking into account certain probabilistic properties of the analyzed data. However, the probabilistic properties of different statistical measures of dependence have been investigated much later, i.e. in the second half of the XX-th century.

Statistical tools for the computation of measures of dependence are available in many software packages. For example, Pearson's r can be

New Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume I: Foundations (K.T. Atanassov, M. Baczyński, J. Drewniak, J. Kacprzyk, M.Krawczak, E.Szmidt, M. Wygralak, S. Zadrozny, Eds.), IBS PAN - SRI PAS, Warsaw, 2012.

computed using all spreadsheets like MS Excel. More sophisticated measures based on ranks, such as Spearman's ρ or Kendall's τ , can be computed using statistical packages like SPSS, STATISTICA, SAS, and many others. Statistical routines for the computation of these measures, written in many programming languages, are also available. Therefore, for usual precise statistical data the computation of basic statistical measures of dependence does not create any problems.

Computation of statistical measures of dependence becomes much more difficult when the available data are presented in an imprecise form, e.g. as intervals of real numbers, or – in a more general setting – as fuzzy numbers. Denote by $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$ and $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_n)$ the components of n imprecise (interval) observations of the random (X, Y) . Let $s(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ be the observed value of a certain statistic S that estimates a certain measure of dependence. Minimal and maximal values of this measure that take into account imprecise character of statistical data we can found solving the following optimization problems:

$$s_L = \min_{\mathbf{x} \in \tilde{\mathbf{x}}, \mathbf{y} \in \tilde{\mathbf{y}}} s(\mathbf{x}, \mathbf{y}), \quad (1)$$

$$s_U = \max_{\mathbf{x} \in \tilde{\mathbf{x}}, \mathbf{y} \in \tilde{\mathbf{y}}} s(\mathbf{x}, \mathbf{y}). \quad (2)$$

When the statistic s is in the form of a continuous function of the observed data, as it is in the case of Pearson's r , the optimization problems (1) – (2) are the problems of the nonlinear mathematical programming, and may be solved using available optimization packages such as e.g. MATLAB. The only possible numerical problems may be related to the size of the optimization problem, $2n$, when the sample size n is large. However, when the statistic s is based on ranks, such as in the cases of Spearman's ρ or Kendall's τ , the optimization problems (1) – (2) become nonlinear integer programming problems of a combinatorial character. For such problems the amount of necessary computations increases exponentially with the sample size n . Denoeux et al. [2] have shown that exact optimization algorithms are effective for sample sizes smaller than 10, and approximate algorithms proposed in their paper are effective for sample sizes not exceeding 20-30. Therefore, in the case of large samples of imprecise data for the computation of the imprecise values of statistical measures of dependence there is a need to apply certain heuristic algorithms, such as e.g. the well known genetic algorithms.

In the second section of this paper we present the most popular measures of dependence, and their relationship to the most important probabilistic models of dependence, described in the form of copulas. Then, in the third section of the paper, we present the results of preliminary investigations on the possibility to use general purpose optimization algorithms for the computation of imprecise Kendall's τ statistic.

2 Measures of statistical dependence

Analysis of statistical dependence is one of the most important areas of mathematical statistics. When the data are described by a p -dimensional random vector (X_1, X_2, \dots, X_p) the full knowledge about possible statistical dependencies between the components of this vector is equivalent to the knowledge of its p -dimensional cumulative distribution function $F(x_1, x_2, \dots, x_p)$.

There exist infinitely many multidimensional probability distributions, but for many years it was only the multivariate normal distribution that was used in practice. The analysis of real multivariate statistical data have shown, however, that there exist other multivariate probability distributions, even with normal marginal distributions, that can be used for the description of mutually dependent statistical data. This phenomenon was noticed by many statisticians, like Fréchet or Gumbel, but the break-through in the analysis of dependent data began with the publication of the paper by Sklar [17] on copulas.

Sklar [17] has proved that for every two-dimensional cumulative probability distribution function $H(x, y)$ with one-dimensional marginal cumulative probability functions denoted by $F(x)$ i $G(y)$, respectively, there exists a unique function C , called a *copula*, such that $H(x, y) = C(F(x), G(y))$. Later on, the concept of the copula has been generalized for the case of any p -dimensional probability distribution. The formal definition of the copula can be found in many sources, such as e.g. the monograph by Nelsen [14].

Let $u = F(x)$, and $v = G(y)$. The simplest copula, the product copula $\Pi(u, v) = uv$, describes *independent* random variables. All other bivariate copulas fulfill the Fréchet-Hoeffding inequalities

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v). \quad (3)$$

The most popular bivariate probability distribution, the bivariate normal distribution, is the particular case (with normal marginal) of the normal copula defined as

$$C_{\rho}(u, v) = \Phi_{\mathcal{N}}(\Phi^{-1}(u), \Phi^{-1}(v); r), \quad (4)$$

where $\Phi_{\mathcal{N}}(x, y; \rho)$ is the cumulative distribution function of the bivariate standardized normal distribution with the correlation coefficient r , and $\Phi^{-1}(u)$ is the inverse of the cdf of the univariate standardized normal distribution (the quantile function). Other popular copulas are the Fairlie-Gumbel-Morgenstern (FGM) copula defined in the paper by Fairlie [3] or the Marshall-Olkin family of copulas defined in the work of Marshall and Olkin [13].

Genest and MacKay [6] introduced the family of the Archimedean copulas defined by

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \quad (5)$$

where φ^{-1} is a pseudo-inverse of the continuous and strictly decreasing function $\varphi: [0, 1] \rightarrow [0, \infty]$, called copula's generator such that $\varphi(1) = 0$. To this family belong well known copulas such as the Clayton copula (defined in [1]), the Gumbel copula (defined in [9]) or the Frank copula (defined in [4]). The comprehensive information about copulas and their properties can be found in the book by Nelsen [14].

Statistical dependence between random variables is encoded in the form of the respective copula. Therefore, in order to evaluate its strength it is necessary to estimate the parameters of the copula. This is not, at least theoretically, difficult task if we know the type of the copula and the type of the marginal distribution, as it is in the case of the classical multivariate normal distribution. In such cases the maximum likelihood methodology can be used. When marginal distributions are not known we can use a semi-parametric methods proposed by Shih and Louis [18]. However, in many cases usually encountered in practice we do not have all necessary information, and we have to use nonparametric statistical methods.

Before we present the main nonparametric measures of dependence, let us comment on the popular usage of the Pearson's linear correlation r . It can be proved that this statistic is an unbiased estimator of the parameter r of the normal copula (5) only in the case of the normal marginal distributions, i.e. in the case of the classical multivariate normal distribution. Genest and Verret [8] have shown that tests based on Pearson's r are equivalent to certain optimal nonparametric tests only for specially chosen copulas and marginal distributions. Therefore, the linear correlation r is not a good general measure of dependence, as its properties depend on the type of copula and the type of marginal distributions.

Instead of Pearson's r we can use two popular nonparametric measures of monotonic dependence, namely Spearman's ρ and Kendall's τ . Schweizer and Wolff [16] have shown that for any copula $C(u, v)$ it is possible to define theoretical (population) equivalents of these two statistics. Denote by \mathbf{I}^2 the unit square. The population version of Spearman's ρ is given by

$$\rho = 12 \iint_{\mathbf{I}^2} C(u, v) du dv = 12 \iint_{\mathbf{I}^2} uv dC(u, v). \quad (6)$$

Similarly, the population version of Kendall's τ is given by

$$\tau = 4 \iint_{\mathbf{I}^2} C(u, v) dC(u, v). \quad (7)$$

Relationship between these two characteristics have been investigated by Fredricks and Nelsen [5].

Denote by $K(t)$ the cdf of the random variable $T = C(U_1, U_2)$, where $C(*, *)$ is the function describing the copula, and, U_1 and U_2 are the random variables uniformly distributed on the unit interval $[0, 1]$. It can be proved, see Nelsen [14], that the following relation holds

$$\tau = 3 - 4 \int_0^1 K(t) dt \quad (8)$$

In the case of the family of Archimedean copulas defined by (5), Genest and Rivest [7] have found that for the given generator $\phi(v)$ we have

$$\tau = 4 \int_0^1 \frac{\phi(v)}{\phi'(v)} dv. \quad (9)$$

Formulae (8) and (9) for Kendall's τ are much more convenient than the general formula (6) Spearman's ρ . Thus, Kendall's τ is much often used in the analysis of copulas than Spearman's ρ . Therefore, in the remaining part of this paper we restrict our attention to this particular statistic.

For the estimation of τ Genest and Rivest [7] propose to use the following statistic

$$V_i = \text{card} \left\{ (X_j, Y_j) : X_j < X_i, Y_j < Y_i \right\} / (n-1), i = 1, \dots, n. \quad (10)$$

The estimator of Kendall's τ is thus given by

$$\tau_n = \frac{4}{n} \sum_{i=1}^n V_i - 1. \quad (11)$$

Similarly, if we use the following statistics

$$W_i = \text{card} \left\{ (X_j, Y_j) : X_j > X_i, Y_j > Y_i \right\} / (n-1), i = 1, \dots, n, \quad (12)$$

we have

$$\tau_n = \frac{4}{n} \sum_{i=1}^n W_i - 1. \quad (13)$$

Moreover, Genest and Rivest [7] have shown that for

$$S_\tau^2 = \sum_{i=1}^n \left(V_i + W_i - \frac{2}{n} \sum_{i=1}^n V_i \right)^2 / (n-1), \quad (14)$$

The statistic $\sqrt{n}(\tau_n - \tau) / S_\tau$ is asymptotically distributed according to the standardized normal distribution.

3 Computing fuzzy rank-based statistics

3.1 Motivation for research

The problem of computing fuzzy rank-based statistics considered in this paper arises from the analysis of time series in the autoregressive model AR(1):

$S_t = \varphi S_{t-1} + \varepsilon$. Let y denote a vector of the last N elements of the time series $y = [S_{t-N}, S_{t-(N-1)}, \dots, S_{t-1}, S_t]$ and x a vector of the last N lagged observations $x = [S_{t-N-1}, S_{t-N}, \dots, S_{t-2}, S_{t-1}]$. Dependence between elements x_i and y_i (i.e. time series elements S_{i-1} and S_i) can be measured using standard statistical tools. However, this issue becomes more complex when the exact values of x and y are unknown but must fulfill inequalities $x^{\min} \leq x \leq x^{\max}$ and $y^{\min} \leq y \leq y^{\max}$ where values x^{\min} , x^{\max} , y^{\min} , and y^{\max} are known. Such problem occurs when only lower and upper bounds of the time series S_t are known.

The case when one knows only ranges which considered variables (e.g. response and explanatory) belong to can occur also in problems of other nature than those of the time series analysis. For this reason, we treat such data from a more general point of view dropping the assumption that i^{th} constraint of vector x must be equal to the $i-1^{\text{st}}$ constraint of vector y . An example of two-dimensional range data is presented in Fig. 1. The i^{th} rectangle (for $i = 1, 2, \dots, N$) shows constraints for element (x_i, y_i) which is plotted with a dot. In practice, location of the dot within the rectangle is unknown.

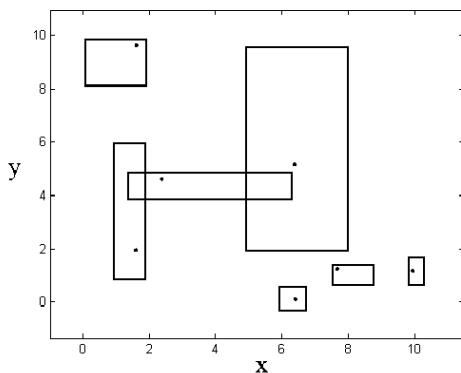


Figure 1: Graphical interpretation of two-dimensional range data

3.2 Fuzzy rank-based statistics

The analysis of monotonic dependence between two variables can be conducted by means of rank-based correlation coefficients such as Spearman's ρ or Kendall's τ . In case when variables x and y are only known to lie within certain ranges their correlation is usually not uniquely defined. Interval character of data results in many possible linear orderings of elements of x and y , which in turn causes many possible values of correlation coefficients. Intuitively speaking, one can imagine such placing of dots within their rectangles in Fig. 1,

that their order with respect to each coordinate changes and so do their correlation coefficient. Dencoux et al. [2] proposed to use in this case a generalized rank-based Kendall statistic $[\tau^{\min}, \tau^{\max}]$ and appropriate hypothesis tests based on such interval values.

3.3 Optimization problem

To apply nonparametric fuzzy rank-based statistics one needs to calculate values of τ^{\min} and τ^{\max} . This requires solving an optimization problem of finding such ordering of elements of vectors x and y within their ranges which minimizes or maximizes the value of Kendall's τ .

$$(x_*, y_*) = \arg \min \{ \tau(x, y) : x^{\min} \leq x \leq x^{\max} \wedge y^{\min} \leq y \leq y^{\max} \} \quad (15)$$

$$(x^*, y^*) = \arg \max \{ \tau(x, y) : x^{\min} \leq x \leq x^{\max} \wedge y^{\min} \leq y \leq y^{\max} \} \quad (16)$$

To solve this problem we decided to use global optimization algorithms. We tried to find a heuristic rule which would allow us to generate good starting points for optimization methods, so that process of finding maximal and minimal values of τ becomes faster and more robust.

Calculating τ^{\min} and τ^{\max} can be also stated as a discrete optimization problem in a space of linear extensions of partial orders with respect to each variable [2]. Heuristic solutions to discrete optimization problems are usually bound to certain problems, since they are strictly dependent on their encoding. For this reason, it might be convenient to solve such problems using some elaborated and reliable "general purpose" continuous optimization algorithms. In this approach the problem-specific knowledge (in this case its rank-based character) should be incorporated into the objective function. For this reason we decided to substitute the original problems (15) and (16) with their rank-based counterparts.

$$(x_*, y_*) = \arg \min \{ \tau(x, y) : r(x^{\min}) \leq x \leq r(x^{\max}) \wedge r(y^{\min}) \leq y \leq r(y^{\max}) \} \quad (17)$$

$$(x^*, y^*) = \arg \max \{ \tau(x, y) : r(x^{\min}) \leq x \leq r(x^{\max}) \wedge r(y^{\min}) \leq y \leq r(y^{\max}) \} \quad (18)$$

where $r(x^{\min}) = [r_1^{\min}, r_2^{\min}, \dots, r_N^{\min}]$ denotes a vector of minimal possible ranks of each element and $r(x^{\max}) = [r_1^{\max}, r_2^{\max}, \dots, r_N^{\max}]$ denotes a vector of maximal ranks. Each minimal rank r_i^{\min} can be calculated as a number of these ranges which lie below the i^{th} range and have an empty intersect with it $r_i^{\min} = \# \{ j \in \{1, 2, \dots, N\} : x_j^{\max} \leq x_i^{\min} \}$. Similarly, maximal rank may be computed as $r_i^{\max} = N - \# \{ j \in \{1, 2, \dots, N\} : x_i^{\max} \leq x_j^{\min} \}$. Ranges of ranks for variable y are obtained analogously.

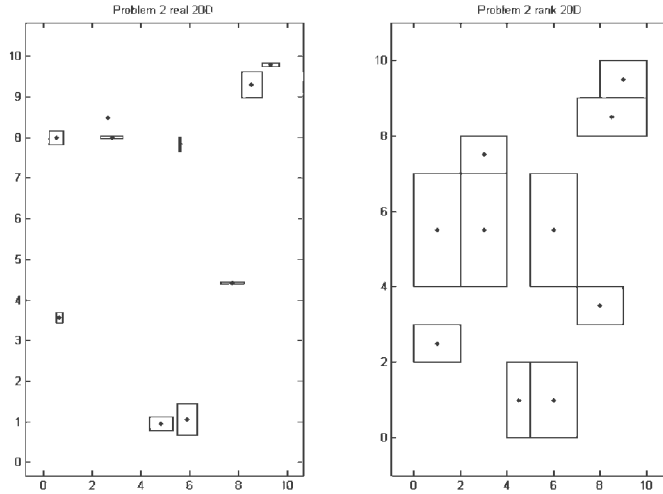


Figure 2: Example of an optimization problem in its natural (left) and rank-based (right) forms

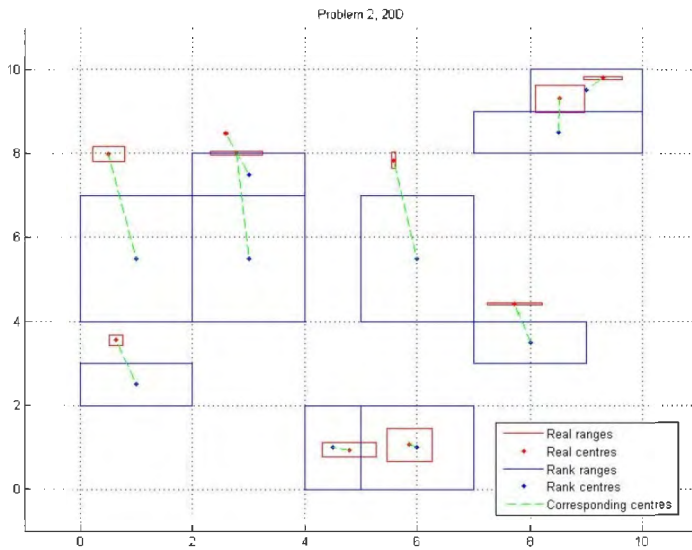


Figure 3: Transformation of natural problem to rank-based one; dashed lines connect centers of corresponding rectangles

Transformations of optimization task from its natural form (15) and (16) to rank-based one (17) and (18) are illustrated in Fig. 2 and 3. Numerical simulations proved that this transformation considerably improved the

conditioning of optimization task, and hence the time of obtaining results and their quality. Moreover, there exists an inverse transformation, which means that there is no loss of generality.

3.4 Benchmarking procedure

We wanted to check whether the process of finding τ_{\min} and τ_{\max} can be improved by initializing optimization methods with values obtained through heuristic procedure. To verify this hypothesis we needed to compare performance of optimization algorithms initialized with and without use of each heuristic. Benchmarking procedures for optimization algorithms require comparing performance for a set of test functions [11, 15]. In this study test problems were created by generating vectors x and y from Frank's copula, which allowed for controlling the strength of dependence between them. Next, for each point (x_i, y_i) a rectangle $[x_i^{\min}, x_i^{\max}] \times [y_i^{\min}, y_i^{\max}]$ was chosen according to the following formulae:

$$\begin{aligned} x_i^{\min} &= x - \xi_1 \xi_2 u \\ x_i^{\max} &= x + (1 - \xi_1) \xi_2 u \\ y_i^{\min} &= y - \xi_3 \xi_4 u \\ y_i^{\max} &= y + (1 - \xi_3) \xi_4 u \end{aligned}$$

where ξ_1, ξ_2, ξ_3 and ξ_4 are IID realizations from the uniform distribution $U(0,1)$, while u is a parameter used to control the expected size of rectangles. Figure 4 presents nine test functions obtained for variables x and y generated for expected value $\tau \in \{0, 0.5, 1\}$ and maximal relative uncertainty $u \in \{0.1, 0.2, 0.5\}$. In Fig. 4 there are 15 rectangles which means that corresponding optimization problem is thirty-dimensional and its feasible set is a hypercube. Having transformed rectangles to its rank-based form the ranges (and plot) were scaled down to a unit square $[0,1] \times [0,1]$.

Among common performance measures used for comparing optimization algorithms the widest interpretation of results is provided by the expected runtime, see [11, 15] for details. In case of benchmark used in this paper a problem emerged that the real values τ_{\min} and τ_{\max} are unknown. This issue was resolved by substituting them with their estimates $\hat{\tau}_{\min}$ and $\hat{\tau}_{\max}$ obtained through running for a long time (a week) independently restarted optimizers: Monte Carlo (MC), Covariance Matrix Adaptation (CMA-ES, [10]), as well as simulated annealing (SA) and genetic algorithm (GA) taken from Matlab Optimization Toolbox. Next, the stopping conditions for benchmarking

simulations (each lasting a few hours) were set for achieving a solution for which the optimization error $\tau - \hat{\tau}_{\min}$ of estimating Kendall's correlation coefficient is lower than $\Delta\tau = 0.2(\hat{\tau}_{\max} - \hat{\tau}_{\min})$. The expected runtime required to solve the test problem with accuracy $\Delta\tau$ is used as the performance measure. This time is counted as the number of objective function evaluations (FEs), in order to measure the performance of an algorithm rather than the hardware and implementation. Details concerning definition and interpretation of expected runtime can be found in papers [11, 15].

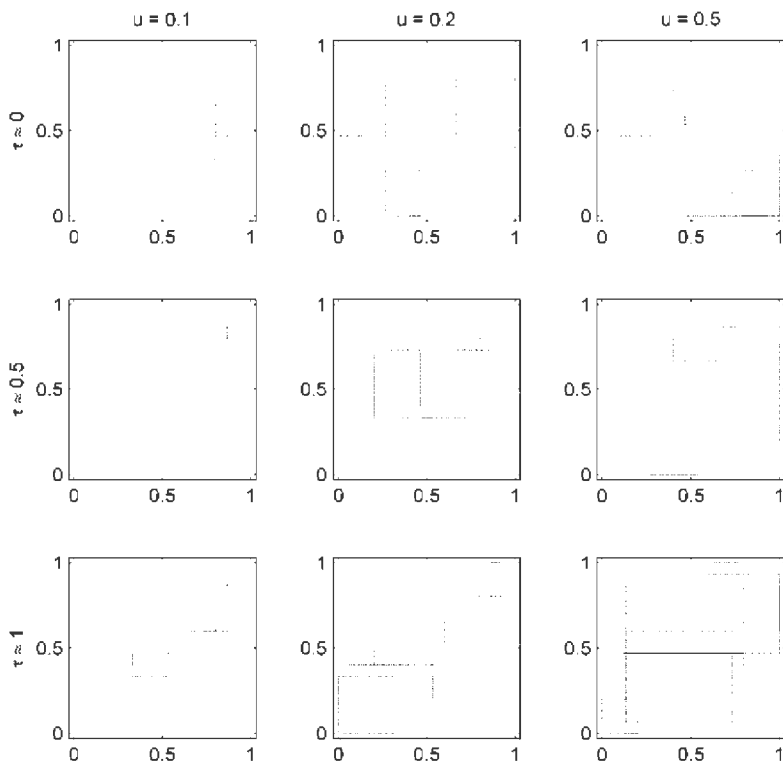


Figure 4: Thirty-dimensional test functions created from Frank copula

3.5 Investigated heuristics

This study was confined to investigate only one heuristic. It was inspired by geometric properties of solutions, which give $\hat{\tau}_{\min}$ and $\hat{\tau}_{\max}$. These solutions - obtained for problems shown in Fig. 4 - are presented in Fig. 5 and 6. Decreasing the size of each plot and increasing markers helps to notice some

general tendencies within location of points: they seem to align along one or a few parallel lines running in the same direction.

Our heuristic was setting the points along one of the diagonals of the plot. Since each point must lie within a rectangle, the points were projected to the nearest empty rectangle. Results of this procedure are plotted in Fig. 7.

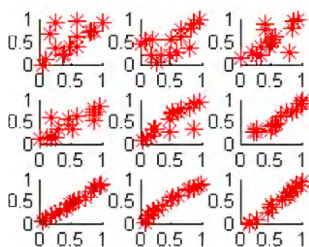


Figure 5: Location of point for $\hat{\tau}_{\max}$

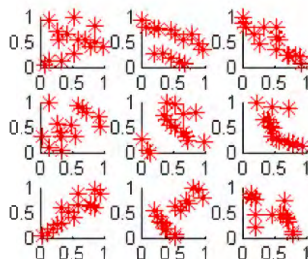


Figure 6: Location of points for $\hat{\tau}_{\min}$

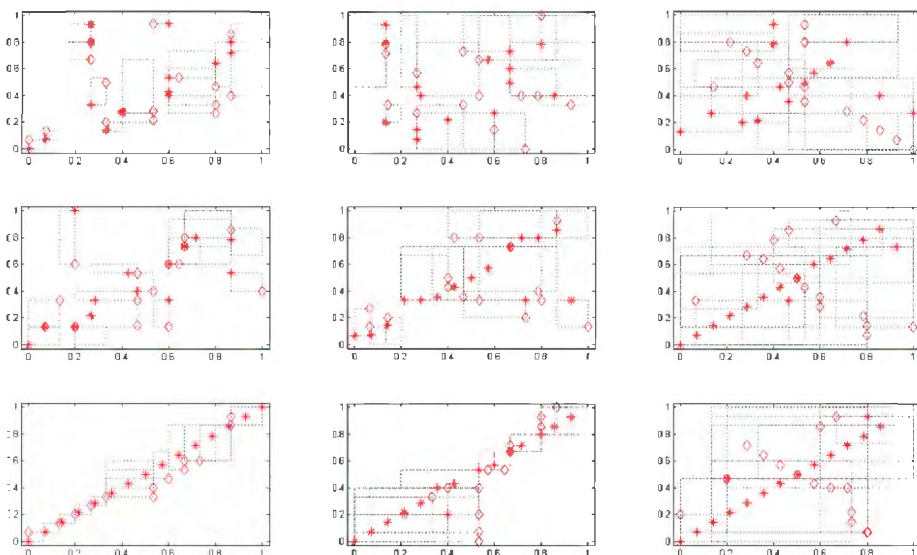


Figure 7: Location of points generated by the heuristic; asterisks denote location for $\hat{\tau}_{\max}$ while diamonds for $\hat{\tau}_{\min}$

3.6 Initial results

Some initial simulations were run with use of the aforementioned algorithms: MC, CMA-ES, GA, and SA. In Table 1 and 2 we give the expected runtimes for each algorithm and test problem for the task of seeking maximal and minimal

value of coefficient τ . The upper part of each cell gives result obtained with use of heuristic while the lower part gives results obtained with uniform random initialization of an optimizer. Expected runtimes were estimated for a sample of 20 runs for each test problem and both variants of an algorithms. Value *Inf* means that none of those runs succeeded, i.e. found a solution for which its coefficient τ fulfills condition $\tau - \hat{\tau}_{\min} \leq \Delta\tau = 0.2(\hat{\tau}_{\max} - \hat{\tau}_{\min})$ for minimization or condition $\hat{\tau}_{\max} - \tau \leq \Delta\tau = 0.2(\hat{\tau}_{\max} - \hat{\tau}_{\min})$ for maximization.

Table 1: Expected runtimes for each minimization test problem (in FEs)

ERT _{Min}	Number of test problem								
	1	2	3	4	5	6	7	8	9
MC	1,1E+04	2,1E+05	2,7E+05	9,3E+03	6,6E+02	1,0E+00	3,9E+04	1,0E+05	<i>Inf</i>
	1,3E+04	2,5E+05	3,1E+05	7,7E+03	8,7E+04	<i>Inf</i>	5,7E+04	1,1E+05	<i>Inf</i>
CMA-ES	6,3E+01	1,2E+02	2,2E+01	8,4E+02	2,6E+01	5,4E+01	3,3E+02	1,4E+02	1,4E+02
	1,3E+02	1,8E+02	1,1E+02	1,4E+02	2,3E+02	8,5E+02	2,1E+02	2,9E+02	<i>Inf</i>
GA	4,0E+02	7,3E+02	4,0E+01	3,3E+02	4,0E+01	4,0E+01	<i>Inf</i>	1,9E+03	4,3E+02
	6,2E+02	1,3E+03	1,7E+02	3,5E+02	5,0E+02	2,0E+04	9,6E+03	1,2E+03	1,0E+04
SA	2,2E+03	1,2E+04	1,0E+00	1,9E+03	1,0E+00	1,0E+00	1,2E+04	3,5E+04	<i>Inf</i>
	2,0E+03	1,1E+04	2,0E+03	2,6E+03	1,3E+04	5,2E+05	7,7E+03	4,0E+04	<i>Inf</i>

Table 2: Expected runtimes for each maximization test problem (in FEs)

ERT _{Max}	Number of test problem								
	1	2	3	4	5	6	7	8	9
MC	1,8E+04	3,4E+05	<i>Inf</i>	1,2E+04	1,2E+04	2,0E+00	2,0E+00	3,6E+04	1,3E+05
	3,2E+04	4,1E+05	<i>Inf</i>	2,1E+04	2,2E+04	2,4E+05	2,9E+02	3,0E+04	2,4E+05
CMA-ES	5,3E+01	1,7E+02	7,4E+01	8,9E+01	3,4E+01	1,6E+01	1,6E+01	1,6E+01	1,6E+01
	1,1E+02	3,5E+02	8,4E+02	2,6E+02	1,1E+02	6,0E+01	2,6E+01	4,1E+01	4,1E+01
GA	2,1E+02	1,2E+03	5,5E+02	2,4E+02	1,4E+02	4,0E+01	4,0E+01	4,0E+01	4,0E+01
	7,6E+02	9,6E+02	3,5E+03	4,5E+02	3,6E+02	8,5E+01	4,0E+01	4,2E+01	5,5E+01
SA	3,7E+03	2,5E+04	<i>Inf</i>	9,0E+03	5,3E+03	1,0E+00	1,0E+00	1,0E+00	1,0E+00
	4,0E+03	4,1E+04	5,0E+05	8,5E+03	4,3E+03	2,6E+02	5,6E+00	1,9E+01	2,7E+01

Table 3 contains medians and interquartile ranges of speed increase of each optimizer owing to heuristic initialization. For instance, for minimization problem algorithm CMA-ES works approximately two times faster when initialized with heuristic, while SA works at the same median speed. The analysis of IQR values show that the speed advantage ranges a lot between test problems. Nevertheless, the use of heuristic initialization seems to improve the results of optimizers.

Table 3: Relative decrease of optimization time (speedup) obtained by using heuristic initialization for mimimization (left) and maximization (right) problems

ERT _{Min}	Speedup		ERT _{Max}	Speedup	
	Median	IQR		Median	IQR
MC	1,2	0,4	MC	1,7	0,7
CMA-ES	2,1	4,8	CMA-ES	2,6	1,4
GA	3,0	11,5	GA	1,9	1,6
SA	1,1	2048,4	SA	1,6	18,4

4 Conclusions and further study

Results reported in this paper suggest that proposed heuristic may be a good choice for initialization of optimization methods. One can use, for example, heuristics proposed in the paper by Hryniewicz and Szewdz [12]. However, to confirm or reject this hypothesis a more elaborated statistical analysis is required. In particular, we plan to check whether time differences between algorithms initialized randomly and with heuristic are significantly different both for single problems and for the whole benchmark.

It seems interesting to perform analogous simulations for other copulae, e.g. normal (Gaussian), Clayton or Gumbel and for various types of marginal distributions. Results obtained in this way would be more general.

We also plan to check other kinds of geometric heuristics exploiting concave, convex and saw-like shapes as well as heuristics based on other inspirations. Using many heuristics may require running multiple instances of an optimizer initialized differently, which may be not very convenient in practice. For this reason, in further study we plan to concentrate on population-based optimizers such as evolutionary algorithms.

References

- [1] Clayton G.G (1978) A model for Association in Bivariate Life Tables and its Applications in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, **65**, 141 – 151.
- [2] Denceux T, Masson M.-H., Hebert P.-A. (2005) Nonparametric rank-based statistics and significance tests for fuzzy data. *Fuzzy Sets and Systems*, **153**, 1-28.
- [3] Farlie D.J.G. (1960) The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307 – 323.

- [4] Frank M.J. (1979) On the Simultaneous Associativity of $F(x,y)$ and $x+y-F(x,y)$. *Æquationes Mathematicæ*, **19**, 194 – 226.
- [5] Fredricks G.A., Nelsen R.B. (2007) On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, **137**, 2143-2150.
- [6] Genest C., MacKay J. (1986) Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, **14**, 145 – 159.
- [7] Genest C., Rivest L-P. (1993) Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, **88**, 1034 – 1043.
- [8] Genest C., Verret F. (2005) Locally most powerful rank tests for independence for copula models. *Nonparametric Statistics*, **17**, 521-539.
- [9] Gumbel E.J. (1960) Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut statistique de l'Université de Paris*, **9**, 171 – 173.
- [10] Hansen, N. (2006) The CMA evolution strategy: a comparing review. In: Lozano, J. (Ed.) *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*, Springer, Berlin, 75-102.
- [11] Hansen, N., Auger, A., Finck, S., Ros, R. (2009), *Real-parameter black-box optimization benchmarking 2009: Experimental setup*, Techn. Rep. INRIA, <http://coco.gforge.inria.fr/>.
- [12] Hryniewicz O., Szediw A. (2008) Fuzzy Kendall τ statistic for Autocorrelated data. In : D.Dubois, M.A. Lubiano, H.Prade, M.A.Gil, P.Grzegorzewski, O.Hryniewicz (Eds.), *Soft Methods for Handling Variability and Imprecision*. Springer, Heidelberg, 155-162.
- [13] Marshall A.W., Olkin I. (1967) A generalized bivariate exponential distribution. *Journal of Applied Probability*, **4**, 291 – 302.
- [14] Nelsen R.B. (1999) *Introduction to Copulas*, Springer, New York.
- [15] Opara, K., Arabas, J. (2011) Benchmarking Procedures for Continuous Optimization Algorithms, *Journal of Telecommunications and Information Technology* (in press), no. 4, 2011.
- [16] Schweizer B., Wolff E.F (1981) On nonparametric measures of dependence for random variables. *Annals of Statistics*, **9**, 879-885.
- [17] Sklar A. (1959) Fonctions de répartitions à n dimensions et leur marges. *Publications de l'Institut de statistique de l'Université de Paris*, **8**, 229-231.
- [18] Shih J.H., Louis T.A. (1995) Inferences on the Association Parameters in Copula Models for Bivariate Survival Data. *Biometrics*, **51**, 1384 – 1399.

The papers presented in this Volume 1 constitute a collection of contributions, both of a foundational and applied type, by both well-known experts and young researchers in various fields of broadly perceived intelligent systems.

It may be viewed as a result of fruitful discussions held during the Tenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2011) organized in Warsaw on September 30, 2011 by the Systems Research Institute, Polish Academy of Sciences, in Warsaw, Poland, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences in Sofia, Bulgaria, and WIT - Warsaw School of Information Technology in Warsaw, Poland, and co-organized by: the Matej Bel University, Banska Bystrica, Slovakia, Universidad Publica de Navarra, Pamplona, Spain, Universidade de Tras-Os-Montes e Alto Douro, Vila Real, Portugal, and the University of Westminster, Harrow, UK:

[Http://www.ibspan.waw.pl/ifs2011](http://www.ibspan.waw.pl/ifs2011)

The consecutive International Workshops on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGNs) have been meant to provide a forum for the presentation of new results and for scientific discussion on new developments in foundations and applications of intuitionistic fuzzy sets and generalized nets pioneered by Professor Krassimir T. Atanassov. Other topics related to broadly perceived representation and processing of uncertain and imprecise information and intelligent systems have also been included. The Tenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2011) is a continuation of this undertaking, and provides many new ideas and results in the areas concerned.

We hope that a collection of main contributions presented at the Workshop, completed with many papers by leading experts who have not been able to participate, will provide a source of much needed information on recent trends in the topics considered.

ISBN-13 9788389475404
ISBN 838947540-5



9 788389 475404