



HAL
open science

On the estimation of Pareto fronts from the point of view of copula theory

Mickaël Binois, Didier Rullière, Olivier Roustant

► **To cite this version:**

Mickaël Binois, Didier Rullière, Olivier Roustant. On the estimation of Pareto fronts from the point of view of copula theory. 2014. hal-01097403v1

HAL Id: hal-01097403

<https://hal.science/hal-01097403v1>

Preprint submitted on 19 Dec 2014 (v1), last revised 23 Jun 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE ESTIMATION OF PARETO FRONTS FROM THE POINT OF VIEW OF COPULA THEORY

MICKAËL BINOIS ^{†‡}, DIDIER RULLIÈRE [§], AND OLIVIER ROUSTANT [‡]

Abstract. Given a first set of observations from a design of experiments sampled randomly in the design space, the corresponding set of non-dominated points usually does not give a good approximation of the Pareto front. We propose here to study this problem from the point of view of multivariate analysis, introducing a probabilistic framework with the use of copulas. This approach enables the expression of level lines in the objective space, giving an estimation of the position of the Pareto front when the level tends to zero. In particular, when it is possible to use Archimedean copulas, analytical expressions for Pareto front estimators are available. Several case studies illustrate the interest of the approach, which can be used at the beginning of the optimization when sampling randomly in the design space.

Key words. Multi-objective optimization, Pareto front, copulas, Archimedean copulas

AMS subject classifications. 62H12, 62H20, 90C26

1. Introduction. Multi-objective optimization (MOO) received a lot of attention recently, especially in evolutionary optimization [8], [42] and surrogate-based optimization [39]. Since no solution usually minimizes every objective at once, the definition of a solution for a multi-objective optimization problem is generally defined as a compromise: a solution is said to be optimal in the Pareto sense if there exists no other solution which is better for every component. All the optimal points in the objective space form the Pareto front. As a result, optimizers provide a set of non-dominated points to approximate the Pareto front. Methods are then designed to seek some properties for these sets, such as uniformity and coverage.

Usually an optimization process starts with random sampling, either to generate an initial population or as a basis to construct a metamodel. The current Pareto front estimated from this first sample may be highly variable, especially when only a small number of function evaluations is available, corresponding to time-consuming functions. This is illustrated in Figure 1 for the bi-objective problem ZDT1 [43], with five 50-points initial samples. However, the stochastic nature of sampling provides a probabilistic framework that can be exploited to quantify this variability and to give a better initial localization of the Pareto front. More precisely, if $\mathbf{X} = (X_1, \dots, X_d)$ is a d -dimensional random vector representing the inputs, and f_1, \dots, f_m the objective functions, then the Pareto front should be connected to the extreme level lines of the distribution of $\mathbf{Y} = (f_1(\mathbf{X}), \dots, f_m(\mathbf{X}))$. To investigate such connection is the aim of the paper.

In the mono-objective situation, a similar probabilistic connection is studied by [37] to estimate the value of the extremum. Considering a small sample of n observations (y_1, \dots, y_n) of \mathbf{Y} , the minimum of \mathbf{Y} is approximated using concepts from extreme order statistics. In multi-objective optimization, the connection seems to be

[†]Renault S.A.S., 78084 Guyancourt, France

[‡]Mines Saint-Étienne, UMR CNRS 6158, LIMOS, F-42023 Saint-Étienne, France, firstname.lastname@mines-stetienne.fr

[§]Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, EA2429, 50 avenue Tony Garnier, 69366 Lyon, France, didier.rulliere@univ-lyon1.fr

new. Uncertainty quantification around the Pareto front has been recently considered by [5], using conditional simulations of Kriging metamodels and concepts from random sets theory. Whereas such approach is relevant in a sequential algorithm, it may be inappropriate in the initial stage that we consider here, due to a potentially large model error in metamodeling.

In this paper, we give a theoretical framework in which the Pareto front appears as a zero level line of the multivariate distribution $F_{\mathbf{Y}}$ of \mathbf{Y} . This problem is known in the probabilistic literature as support curve estimation (see e.g. [20], [21], [18]). However, the existing methods rely on assumptions, such as domain of attraction or polynomial rate of decrease, that can hardly be checked in an optimization context. As an alternative, we propose to take advantage of copulas [32] which are multivariate probability distributions with uniform marginals, allowing to consider separately the estimation of the marginals and the dependence structure. This allows estimating extreme level lines, without making specific assumptions about domain of attractions. Copulas have already been used in optimization, mainly in the variable space to estimate distribution in evolutionary algorithms, see e.g. [40], [15], [14], while here we focus on the objective space. We propose a first estimation of the Pareto front relying on the empirical copula. Then, we consider the case where the copula belongs to the class of Archimedean copulas, parameterized by a function. This assumption can be checked visually or statistically with specific tests of the literature. If relevant, a better localization of the Pareto front is found. Furthermore, a parametric expression of the approximated Pareto front is available.

The paper is structured as follows. Section 2 proposes alternative definitions of the Pareto front from the point of view of the cumulative distribution function, presents some background about copulas and describes the estimation procedure in the Archimedean case. Section 3 discusses the applicability of the model and more specifically the consequences of the Archimedean copula model. Section 4 illustrates in several configurations the application of the proposed approach to Pareto front localization. Section 5 concludes and describes possibilities for further improvements.

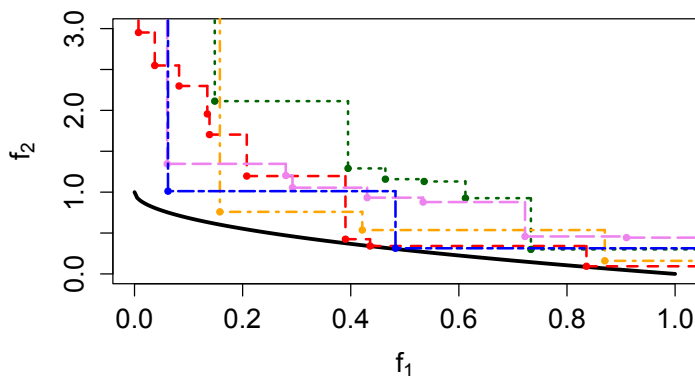


Figure 1: Non-dominated points obtained with 5 different random samples (one color and type of line per sample) of 50 points for the bi-objective problem ZDT1. The true Pareto front is the black solid line.

2. Methodology. The present section describes the interest of using a probabilistic framework in multi-objective optimization by establishing the link between both domains. Based on the resulting theorem, the expression of level lines of the multivariate cumulative distribution functions $F_{\mathbf{Y}}$ using copulas is described as well as a procedure for their estimation. Empirical and parametric model are discussed, with emphasis on Archimedean models.

2.1. Link between Pareto front and level curves. For a variety of methods ranging from evolutionary optimization [8] to surrogate-based methods [36], optimization starts with random sampling in the design space, with uniform sampling or with a random Latin Hypercube. In this case, it is possible to study the resulting observations in the objective space as a set of points. Specifically, assuming that the outputs can be considered as independent and identically distributed (i.i.d.) random variables, they enter the scope of multivariate analysis.

Let us start with a definition of Pareto front. For two points $\mathbf{y} = (y_1, \dots, y_m)$ and $\mathbf{z} = (z_1, \dots, z_m)$ of \mathbb{R}^m , $m \geq 2$, we first define the respective weak and strict dominance operators \preceq and \prec as:

$$\begin{cases} \mathbf{z} \preceq \mathbf{y} \Leftrightarrow \forall i = 1, \dots, m, z_i \leq y_i, \\ \mathbf{z} \prec \mathbf{y} \Leftrightarrow \forall i = 1, \dots, m, z_i < y_i. \end{cases}$$

Consider a subset \mathbf{G} of \mathbb{R}^m . We define here the Pareto front \mathcal{P} of the set \mathbf{G} as the subset of \mathbf{G} of all points that are weakly dominated only by themselves:

$$(2.1) \quad \mathbf{y} \in \mathcal{P} \Leftrightarrow \{\mathbf{z} \in \mathbf{G}, \mathbf{z} \preceq \mathbf{y}\} = \{\mathbf{y}\}$$

Now assume that $\mathbf{G} = \mathbf{f}(\mathbf{E})$ is the image of a set $\mathbf{E} \subseteq \mathbb{R}^d$ by a vector-valued objective function $\mathbf{f} : \mathbf{E} \rightarrow \mathbb{R}^m$, with $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$, $\mathbf{x} \in \mathbf{E}$. Then, the Pareto front of \mathbf{f} is defined as the Pareto front of the image set \mathbf{G} . In this case we retrieve the usual interpretation that a solution in the objective space is Pareto-optimal if there exists no other solution which is better in every component: for $\mathbf{y} \in \mathbf{G}$, there exists no $\mathbf{z} \in \mathbf{G}$ such that $\mathbf{z} \preceq \mathbf{y}$ and $\mathbf{z} \neq \mathbf{y}$.

Assume that \mathbf{X} is a random vector with values in \mathbf{E} and denote $\mathbf{Y} = (Y_1, \dots, Y_m)$ with $Y_i = f_i(\mathbf{X})$. Then if \mathbf{Y} has an absolutely continuous distribution with respect to the Lebesgue measure in \mathbf{G} , one easily gets $\mathbf{y} \in \mathcal{P} \Rightarrow \mathbb{P}[\mathbf{Y} \in \{\mathbf{z} \in \mathbf{G}, \mathbf{z} \preceq \mathbf{y}\}] = 0$. As a direct consequence, denoting $F_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}[\mathbf{Y} \preceq \mathbf{y}]$ the multivariate cumulative distribution function of \mathbf{Y} ,

$$(2.2) \quad \mathbf{y} \in \mathcal{P} \Rightarrow F_{\mathbf{Y}}(\mathbf{y}) = 0.$$

The Pareto front thus belongs to the zero level set of $F_{\mathbf{Y}}$, $\{\mathbf{y} \in \mathbf{G}, F_{\mathbf{Y}}(\mathbf{y}) = 0\}$, which enlightens the connection between Pareto front and level sets of $F_{\mathbf{Y}}$.

Define the upper level set $L_{\alpha}^F = \{\mathbf{y} \in \mathbb{R}^m, F_{\mathbf{Y}}(\mathbf{y}) \geq \alpha\}$ with $\alpha \in (0, 1)$, and the corresponding level line $\partial L_{\alpha}^F = \{\mathbf{y} \in \mathbb{R}^m, F_{\mathbf{Y}}(\mathbf{y}) = \alpha\}$. The following main result is that the upper level set L_{α}^F converges towards the area dominated by the Pareto front, when α tends to 0. This seems quite natural, as illustrated in Figure 2. However, the rigorous proof involves topological arguments that are different in the continuous and discrete case. Some pitfalls are that for $\alpha > 0$, L_{α}^F is not necessarily included in \mathcal{Y} ,

as illustrated in Figure 2, and that all points of \mathcal{Y} are not necessarily dominated by the Pareto front, as when $\mathcal{Y} = \mathbb{R}^m$ for some unbounded objective functions.

THEOREM 2.1. *Consider a random vector \mathbf{Y} which admits a probability density function $f_{\mathbf{Y}}$ with respect to the Lebesgue measure on \mathbb{R}^m , and denote by \mathcal{Y} its support (i.e. the essential support of the function $f_{\mathbf{Y}}$). Let \mathcal{P} be the Pareto front of the set \mathcal{Y} . Define the respective weak and strict dominated sets:*

$$\mathcal{P}^{\succ} = \bigcup_{\mathbf{y} \in \mathcal{P}} \{\mathbf{z} \in \mathbb{R}^m, \mathbf{y} \preceq \mathbf{z}\} \text{ and } \mathcal{P}^{\succ} = \bigcup_{\mathbf{y} \in \mathcal{P}} \{\mathbf{z} \in \mathbb{R}^m, \mathbf{y} \prec \mathbf{z}\} .$$

If all points of \mathcal{Y} are dominated by the Pareto front, i.e. $\mathcal{Y} \subseteq \mathcal{P}^{\succ}$, then the dominated area is obtained as the union of all upper level sets :

$$\mathcal{P}^{\succ} = \bigcup_{\alpha > 0} L_{\alpha}^F .$$

As a consequence we have $\lim_{\alpha \rightarrow 0} \mathbb{P}[\mathbf{Y} \in L_{\alpha}^F] = \mathbb{P}[\mathbf{Y} \in \mathcal{P}^{\succ}] = 1$.

Proof. We want to prove that the dominated area is equal to the area dominated by the set $L_0 = \bigcup_{\alpha > 0} L_{\alpha}^F$.

- $L_0 \subseteq \mathcal{P}^{\succ}$: It is sufficient to prove that if $\mathbf{y} \notin \mathcal{P}^{\succ}$, then $F_{\mathbf{Y}}(\mathbf{y}) = 0$.
 - Assume first that $\mathbf{y} \notin \mathcal{P}^{\succ}$, $F_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}[\mathbf{Y} \preceq \mathbf{y}] = \mathbb{P}[\mathbf{Y} \in \{\mathbf{z} \in \mathcal{Y}, \mathbf{z} \preceq \mathbf{y}\}]$. One can show that if $\mathbf{y} \notin \mathcal{P}^{\succ}$ and if $\mathbf{z} \preceq \mathbf{y}$ then $\mathbf{z} \notin \mathcal{P}^{\succ}$, so that $\{\mathbf{z} \in \mathbb{R}^m, \mathbf{z} \preceq \mathbf{y}\} \cap \mathcal{P}^{\succ} = \emptyset$. Finally $\{\mathbf{z} \in \mathbb{R}^m, \mathbf{z} \preceq \mathbf{y}\} \cap \mathcal{Y} = \emptyset$, since by assumption $\mathcal{Y} \subseteq \mathcal{P}^{\succ}$, and $F_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}[\mathbf{Y} \in \emptyset] = 0$.
 - Now assume that $\mathbf{y} \in \mathcal{P}^{\succ} \setminus \mathcal{P}^{\succ}$. One can show that the Lebesgue measure $\mu(\mathcal{P}^{\succ} \setminus \mathcal{P}^{\succ}) = 0$: Otherwise, there would exist an hypercube $\prod_{i=1}^m [a_i, b_i]$ included in $\mathcal{P}^{\succ} \setminus \mathcal{P}^{\succ}$ such that for all $i = 1, \dots, m$, $b_i > a_i$; This would be in contradiction with $\mathbf{b} = (b_1, \dots, b_m) \in \mathcal{P}^{\succ}$. Now, $F_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}[\mathbf{Y} \preceq \mathbf{y}] = \mathbb{P}[\mathbf{Y} \preceq \mathbf{y} \text{ and } \mathbf{Y} \in \mathcal{P}^{\succ}]$ since by assumption $\mathcal{Y} \subseteq \mathcal{P}^{\succ}$. This probability is equal to $\mathbb{P}[\mathbf{Y} \preceq \mathbf{y} \text{ and } \mathbf{Y} \in \mathcal{P}^{\succ} \setminus \mathcal{P}^{\succ}]$ because $\mathbf{y} \notin \mathcal{P}^{\succ}$ and $\mathbf{Y} \preceq \mathbf{y} \Rightarrow \mathbf{Y} \notin \mathcal{P}^{\succ}$. Thus this probability is 0 by absolute continuity of \mathbf{Y} since $\mu(\mathcal{P}^{\succ} \setminus \mathcal{P}^{\succ}) = 0$.
- $\mathcal{P}^{\succ} \subseteq L_0$: Recall that the complementary set \mathcal{Y}^C of the support \mathcal{Y} is defined as the union of all open sets Ω such that $f_{\mathbf{Y}}(\cdot) = 0$ almost everywhere on Ω . Let $\mathbf{y} \in \mathcal{P}^{\succ}$, $\exists \mathbf{y}^* \in \mathcal{P}$ such that $\mathbf{y}^* \prec \mathbf{y}$. Denote $D_{\mathbf{y}} = \{\mathbf{z} \in \mathbb{R}^m, \mathbf{z} \preceq \mathbf{y}\}$. There exists an open set $B_{\mathbf{y}^*} \subseteq D_{\mathbf{y}}$ which contains \mathbf{y}^* . Now, we show that we cannot have $\mathbb{P}[\mathbf{Y} \preceq \mathbf{y}] = 0$. Otherwise, by assumption of absolute continuity of \mathbf{Y} , this would imply that almost everywhere $f_{\mathbf{Y}}(\cdot) = 0$ on $D_{\mathbf{y}}$. Then $B_{\mathbf{y}^*}$ would be an open set belonging to \mathcal{Y}^C . This would be in contradiction with $\mathcal{P} \subseteq \mathcal{Y}$, by definition of \mathcal{P} , which implies that the non-empty set $\mathcal{P} \cap B_{\mathbf{y}^*} \subseteq \mathcal{Y}$. Thus necessarily $\mathbb{P}[\mathbf{Y} \preceq \mathbf{y}] > 0$, and there exists $\alpha > 0$ such that $\mathbf{y} \in L_{\alpha}^F$. Therefore $\mathcal{P}^{\succ} \subseteq L_0$.

Given that $\mathbb{P}[\mathbf{Y} \in \mathcal{P}^{\succ}] = 1$ and $\mathbb{P}[\mathbf{Y} \in \mathcal{P}^{\succ} \setminus \mathcal{P}^{\succ}] = 0$, the last part of the proposition is obtained by considering a decreasing sequence α_n and $L_{\alpha_n}^F$ and using Proposition 1.27 in [6]. \square

The case when \mathbf{G} is discrete is also of practical interest and the corresponding result is detailed in Remark 1. Notice that the result is slightly different than the previous Theorem 2.1.

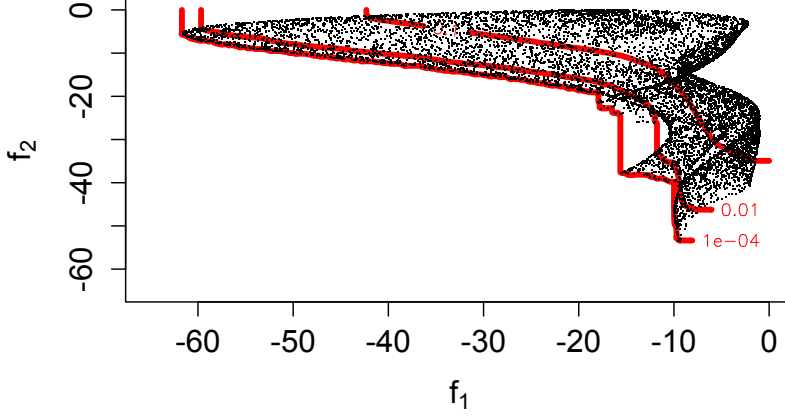


Figure 2: Level lines ∂L_α^F with $\alpha = 0.0001, 0.01, 0.1$ of the empirical cumulative distribution function of $\mathbf{f}(\mathbf{X})$ obtained with sampled points (in black), showing the link between the level line of level α and the Pareto front \mathcal{P} (apart from the vertical and horizontal components), as α tends to zero.

REMARK 1. Let \mathbf{Y} be a discrete random vector with support \mathcal{Y} . Let \mathcal{P} be the Pareto front of the set \mathcal{Y} . Assume that all points of the support \mathcal{Y} are dominated, i.e., $\mathcal{Y} \subseteq \mathcal{P}^\succ$. Then

$$\mathcal{P}^\succ = \bigcup_{\alpha > 0} L_\alpha^F.$$

Proof. Denote $L_0 = \bigcup_{\alpha > 0} L_\alpha^F$. As \mathbf{Y} is a discrete random vector, for any $\mathbf{y} \in \mathcal{Y}$, by definition of the support $\mathbb{P}[\mathbf{Y} = \mathbf{y}] > 0$. Now let us show that $\mathcal{P}^\succ = L_0$.

- $\mathcal{P}^\succ \subseteq L_0$: For any $\mathbf{z} \in \mathcal{P}^\succ$, there exists $\mathbf{y} \in \mathcal{P}$ such that $\mathbf{y} \preceq \mathbf{z}$. Now for any $\mathbf{y} \in \mathcal{P}$, as $\mathcal{P} \subseteq \mathcal{Y}$ by definition of the Pareto front, then $\mathbb{P}[\mathbf{Y} = \mathbf{y}] > 0$. Then $\mathbb{P}[\mathbf{Y} \preceq \mathbf{z}] \geq \mathbb{P}[\mathbf{Y} = \mathbf{y}] > 0$ and $\mathcal{P}^\succ \subseteq L_0$.
- $L_0 \subseteq \mathcal{P}^\succ$: Let $\mathbf{y} \in L_\alpha^F$, $\alpha > 0$, then $\mathbb{P}[\mathbf{Y} \preceq \mathbf{y}] \geq \alpha$. As $\mathbb{P}[\mathbf{Y} \preceq \mathbf{y}] = \sum_{\mathbf{y}_0 \in \mathcal{Y}, \mathbf{y}_0 \preceq \mathbf{y}} \mathbb{P}[\mathbf{Y} = \mathbf{y}_0] > 0$, there exists $\mathbf{y}_0 \in \mathcal{Y}$ such that $\mathbf{y}_0 \preceq \mathbf{y}$. Since by assumption $\mathcal{Y} \subseteq \mathcal{P}^\succ$, $\mathbf{y}_0 \in \mathcal{P}^\succ$ and since $\mathbf{y}_0 \preceq \mathbf{y}$, one gets $\mathbf{y} \in \mathcal{P}^\succ$.

□

2.2. Expression of level curves using copulas. The m -dimensional distribution function $F_{\mathbf{Y}}$ contains all the information about the problem at hand, in particular about the Pareto front. The copula framework offers the possibility to study the dependence on the level lines separately from the univariate marginal distributions. Furthermore, under the Archimedean hypothesis, the level lines have a parametric expression. To distinguish between the objective space and the copula space, we denote by $\mathbf{y} = (y_1, \dots, y_m) \in \mathbb{R}^m$ vectors in the objective space and by $\mathbf{u} = (u_1, \dots, u_m) \in [0, 1]^m$ vectors in the copula space.

2.2.1. General case. Supposing that the multivariate distribution function $F_{\mathbf{Y}}$ is continuous, from Sklar theorem [38], there is a unique copula function C such that:

$$F_{\mathbf{Y}}(y_1, \dots, y_m) = C(F_1(y_1), \dots, F_m(y_m)),$$

where the F_i , $1 \leq i \leq m$ are the univariate marginals, i.e. the cumulative distribution function (cdf) of $Y_i = f_i(\mathbf{X})$. Note that C is a cdf defined on $[0, 1]^m$ with uniform marginals.

From now on, we consider that the F_i 's are continuous and invertible functions. Thus we can write for $\mathbf{u} \in [0, 1]^m$:

$$C(u_1, \dots, u_m) = F_{\mathbf{Y}}(F_1^{-1}(u_1), \dots, F_m^{-1}(u_m)).$$

Let $\alpha \in (0, 1)$. The α -level lines of C , i.e. $\{\mathbf{u} \in [0, 1]^m, C(u_1, \dots, u_m) = \alpha\}$ are denoted ∂L_{α}^C . They are connected to the level lines ∂L_{α}^F of $F_{\mathbf{Y}}$ by the following relationship:

$$\partial L_{\alpha}^F = \{(y_1, \dots, y_m) = (F_1^{-1}(u_1), \dots, F_m^{-1}(u_m)) \in \mathbb{R}^m, \mathbf{u} \in \partial L_{\alpha}^C\}.$$

It follows that given a model of the copula and given the marginals, the levels lines of $F_{\mathbf{Y}}$ are obtained without additional effort. We describe next a specific model of copula which allows a parametric expression of those level lines. The methods to estimate both the copula model and the marginals will be detailed in Section 2.3.

2.2.2. Parametric form in the Archimedean case. A parametrization of the Pareto front has sometimes been proposed based on a metamodel of one output in function of the others [19] or using B-splines [3]. It seems that in both cases the results do not necessarily follow the Pareto dominance, which might cause problems when dealing with Pareto fronts. Here we propose a method usable with any number of points, after sampling randomly in the design space, and respecting (weak) Pareto dominance for the proposed results.

Among other available parametric models of copulas (see e.g. [33]), a practical class of copula is the class of Archimedean copula, see e.g. [31]. The family of Archimedean copulas is a flexible family that depends on a real function $\phi : \mathbb{R}^+ \rightarrow [0, 1]$, called the generator of the copula. An Archimedean copula is defined by

$$C_{\phi}(u_1, \dots, u_m) = \phi(\phi^{-1}(u_1) + \dots + \phi^{-1}(u_m)),$$

where the function ϕ^{-1} is the generalized inverse of the generator ϕ :

$$\phi^{-1}(t) = \inf \{x \in \mathbb{R}^+, \phi(x) \leq t\}.$$

The generator ϕ is supposed to be continuous, m -monotone (see [31], which implies convexity), strictly decreasing on $[0, \phi^{-1}(0)]$ with $\phi(0) = 1$ and $\lim_{x \rightarrow +\infty} \phi(x) = 0$. If $\phi(x) > 0$ for all $x \in \mathbb{R}^+$, the generator and the corresponding Archimedean copula are said to be *strict*, otherwise they are called *non-strict*.

Also ϕ can be seen as a particular univariate survival function, so that in the following we will say that $\psi_0 = \phi^{-1}(0)$ is the *end-point* of the generator, with $\psi_0 < +\infty$ for non-strict generators, and $\psi_0 = +\infty$ for strict generators. A list of bivariate Archimedean copulas with parametric generator functions is given in [32], with strict and non-strict examples.

The interest of representing C with an Archimedean copula (or a transformed copula [2]) is that we know how to express parametrically the level curves of such copulas, and consequently those of $F_{\mathbf{Y}}$.

PROPOSITION 2.1 (Level curves for an Archimedean copula). *Let \mathcal{S} denotes the simplex $\mathcal{S} = \{\mathbf{s} \in [0, 1]^m, s_1 + \dots + s_m = 1\}$. If C_ϕ is an Archimedean copula with generator ϕ then for all $\alpha \in (0, \psi_0)$, we have*

$$\partial L_\alpha^{C_\phi} = \{\mathbf{u} \in [0, 1]^m, u_i = \phi(s_i \phi^{-1}(\alpha)), 1 \leq i \leq m, \mathbf{s} \in \mathcal{S}\},$$

and the level lines of $F_{\mathbf{Y}}$ are expressed as:

$$\partial L_\alpha^F = \{\mathbf{y} \in \mathbb{R}^m, y_i = F_i^{-1}(u_i), u_i = \phi(s_i \phi^{-1}(\alpha)), 1 \leq i \leq m, \mathbf{s} \in \mathcal{S}\}.$$

Proof. For an Archimedean copula with generator ϕ , the level curve of level $\alpha > 0$ is $\partial L_\alpha^{C_\phi} = \{\mathbf{u} \in [0, 1]^m, C_\phi(u_1, \dots, u_m) = \alpha\}$. Let $\mathbf{u} \in [0, 1]^m$, $\mathbf{u} \in \partial L_\alpha^{C_\phi} \Leftrightarrow C_\phi(u_1, \dots, u_m) = \alpha \Leftrightarrow \phi(\phi^{-1}(u_1) + \dots + \phi^{-1}(u_m)) = \alpha$.

Suppose that in addition $\alpha \in (0, \psi_0)$, then $\mathbf{u} \in \partial L_\alpha^{C_\phi} \Leftrightarrow \frac{\phi^{-1}(u_1) + \dots + \phi^{-1}(u_m)}{\phi^{-1}(\alpha)} = 1$.

By re-parameterizing with $s_i = \phi^{-1}(u_i)/\phi^{-1}(\alpha)$, $1 \leq i \leq m$ (equivalent to $u_i = \phi(s_i \phi^{-1}(\alpha))$), we obtain that those s_i belongs to the simplex \mathcal{S} . Hence $\partial L_\alpha^{C_\phi} = \{\mathbf{u} \in [0, 1]^m, u_i = \phi(s_i \phi^{-1}(\alpha)), 1 \leq i \leq m, \mathbf{s} \in \mathcal{S}\}$.

The expression of ∂L_α^F follows from the connection between ∂L_α^F and $\partial L_\alpha^{C_\phi}$.

□

Other parameterizations of level curves of $F_{\mathbf{Y}}$ can be found in the literature (see e.g. [2], Proposition 2.4.).

A difference between strict and non-strict generators lies in the behaviour of the level lines when α tends to 0:

DEFINITION 2.1 (zero set, from [32, 27] and extended in the general case). *The zero set of a copula C is the set*

$$S_0 = \{\mathbf{u} \in [0, 1]^m, C(u_1, \dots, u_m) = 0\}.$$

The Lebesgue measure on \mathbb{R}^m of this zero set S_0 will be denoted m_{S_0} .

As recalled in [27], based on [32], the zero set is of Lebesgue measure zero if and only if the copula is strict. In the other case, for non-strict generators, the boundary of the zero set, $\{\mathbf{u} \in [0, 1]^m, \phi^{-1}(u_1) + \dots + \phi^{-1}(u_m) = \psi_0\}$ is called the *zero curve* of C_ϕ . For such a non-strict Archimedean copula and with $m = 2$, the zero curve can be expressed with $\partial L_0^{C_\phi} = \{(u_1, u_2) \in [0, 1]^2, u_2 = \phi(\phi^{-1}(0) - \phi^{-1}(u_1))\}$. This form can be extended to any dimension m by writing the m^{th} output as a function of the $m - 1$ first ones. Still when $m = 2$, setting $\psi = \phi^{-1}$, the probability mass of the zero curve is equal to $-\frac{\psi(0)}{\psi'(0^+)}$ (cf. i.e. Theorem 4.3.3. in [32]).

Figure 3 illustrates the different cases described on the level lines of the copulas. With strict generators, the level lines converge towards the axis $[0, \infty) \times \{0\}$ and $\{0\} \times [0, \infty)$ as α tends to zero. This is not the case for non-strict generators, where zero sets have a strictly positive Lebesgue measure m_{S_0} , as visible on lower left

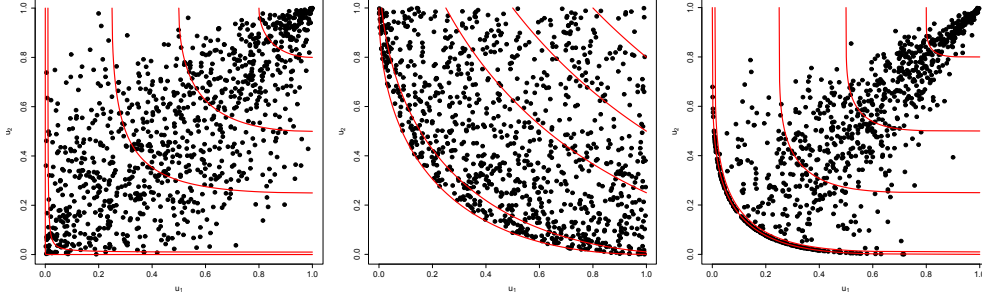


Figure 3: Scatterplots of samples of a thousand points $\mathbf{U}^1, \dots, \mathbf{U}^{1000}$ from Archimedean copulas with different generators and level lines with $\alpha = \{0, 0.01, 0.25, 0.5, 0.8\}$. Left : strict generator (Gumbel copula with $\theta = 2$). Center : non-strict generator (Clayton copula with $\theta = -0.8$). Right: non-strict generator with a probability mass on the zero curve (copula No.2 from [32] with $\theta = 5$).

corners of center and right panels of Figure 3.

As a summary, the Archimedean family of copulas has the advantage to be very flexible (it is indexed by a whole real function), to provide simple parametric expressions of the level curves, and to distinguish naturally degenerated or non degenerated Pareto front (via strict or non-strict generators). In the next section, we explain how to get an approximation of the Pareto front \mathcal{P} from this parametric expression of $\partial L_\alpha^{C_\phi}$. The relevance of this Archimedean model in practice is the subject of Section 3.

2.3. Estimation of the level lines. When working with black-box functions in order to find Pareto optimal solutions, the marginal distribution functions and copulas of the output \mathbf{Y} must be estimated from the data. In the general case, only empirical estimation is possible while supposing that the copula is Archimedean gives parametric expressions for the level lines.

We aim here at proposing estimators of the level lines ∂L_α^F for small values of α . In particular, when α tends to 0, ∂L_α^F is directly related to the Pareto front \mathcal{P} (see Theorem 2.1). As shown in Section 2.2.1, ∂L_α^F can be expressed as a function of ∂L_α^C and F_1, \dots, F_m . For $\alpha \in (0, 1)$, the proposed plug-in estimators of the α -level lines are thus of the form

$$\widehat{\partial L}_\alpha^F = \left\{ (y_1, \dots, y_m) = (\hat{F}_1^{-1}(u_1), \dots, \hat{F}_m^{-1}(u_m)) \in \mathbb{R}^m, \mathbf{u} \in \partial L_\alpha^{\hat{C}} \right\},$$

where \hat{C} and $\hat{F}_1, \dots, \hat{F}_m$ are respective estimators of C and F_1, \dots, F_m , and where $\hat{F}_1^{-1}, \dots, \hat{F}_m^{-1}$ are generalized pseudo inverse of $\hat{F}_1, \dots, \hat{F}_m$.

The proposed estimator of the Pareto front will be

$$(2.3) \quad \widehat{\mathcal{P}} = \widehat{\partial L}_{\alpha^*}^F$$

where $\alpha^* \in [0, 1)$ is a small level value whose choice will be discussed hereafter. In the following, we first investigate the case where \hat{C} is an empirical copula, and then the case where \hat{C} is an Archimedean copula with generator ϕ .

2.3.1. Empirical copula. Several estimators of an empirical copula can be proposed, see for example [9] and [34]. Consider a set of n observations in \mathbb{R}^m : $\left\{ \mathbf{Y}^k = (Y_1^k, \dots, Y_m^k) \right\}_{k=1, \dots, n}$. The corresponding pseudo-observations are defined as $\left\{ \mathbf{U}^k = (U_1^k, \dots, U_m^k) \right\}_{k=1, \dots, n}$, with

$$U_i^k = \frac{1}{n+1} \sum_{j=1}^n \mathbb{1}_{\{Y_i^j \leq Y_i^k\}}, \quad i \in \{1, \dots, m\},$$

where $\mathbb{1}$ is the indicator function.

The empirical copula can be estimated using the following formula:

$$\hat{C}_n(u_1, \dots, u_m) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{U_1^k \leq u_1, \dots, U_m^k \leq u_m\}}.$$

This is in fact the empirical distribution of the (normalized) ranks of the data. More details can be found in [9] and [34].

The empirical copula \hat{C}_n being a step function, we mostly consider its level sets: $L_\alpha^{\hat{C}_n} = \left\{ \mathbf{u} \in [0, 1]^m, \hat{C}_n(u_1, \dots, u_m) \geq \alpha \right\}$. In this case, different values of α may lead to the same level sets. An estimator of the level lines can be obtained by considering the frontiers of these upper level sets. This operation may be computationally costly, especially in large dimension. Furthermore, no simple analytical expression is available for these frontiers, justifying the use of parametric models when they are relevant.

2.3.2. Estimation in the Archimedean case. We consider here the case where the copula C is estimated by an Archimedean copula \hat{C}_ϕ , having a generator ϕ . There exists a vast literature on the estimation of Archimedean copulas, see for example [17] or [25]. In the case of parametric estimation, methods to fit an Archimedean copula C_ϕ rely for instance on Maximum Likelihood estimation or on dependence measures. A review of these methods can be found e.g. in [26].

In the case of non-parametric estimation, among other different possible estimation procedures, one can cite [11] or [16]. Under what is called *Frank's condition* (see [13]), the Archimedean copula is uniquely determined by its diagonal section $\delta(u) = C(u, \dots, u)$, $u \in [0, 1]$. In this paper, for strict generators, we use a non-parametric estimator of the generator ϕ , based on an initial estimator of the diagonal section of the empirical copula, as detailed in Algorithm 2 in [10].

Another important choice is about non-strict Archimedean copula, for which one has to estimate the *end-point* of the generator $\psi_0 = \phi^{-1}(0) = \inf \{x \in \mathbb{R}^+, \phi(x) = 0\}$. In parametric estimation, a recent method has been proposed in [27]. Among admissible parameters leading the zero curve to dominate all pseudo-observations, the choice is based on the functional form of the zero curve of the copula. The selected parameter is the one giving the closest zero curve to the pseudo-observations, under the assumption that the Lebesgue measure of the zero set is monotone with respect to the parameter. More formally, considering that the generator depends on ψ_0 and other parameters $\theta \in \Theta$, selected parameters are:

$$(2.4) \quad (\psi_0^*, \theta^*) = \underset{(\psi_0, \theta) \in \mathbb{R}^{++} \times \Theta}{\operatorname{argmax}} m_{S_0}(\psi_0, \theta) \text{ s.t. } \mathbf{U}^k \notin S_0, 1 \leq k \leq n,$$

where $m_{S_0}(\psi_0, \theta)$ represents the Lebesgue measure of the zero set S_0 of the copula (see Definition 2.1).

2.3.3. Choice of the level α^* . Depending on the copula model: empirical copula or Archimedean with strict/non-strict generator, the behavior of L_α^C when α tends to zero differs. Consider an estimator \hat{C} of the copula C . Notice that any admissible level α^* for the Pareto front estimator \hat{P} should dominate all pseudo-observations (i.e. be such that $\forall k \in \{1, \dots, n\}, \mathbf{U}^k \in L_{\alpha^*}^{\hat{C}}$). Otherwise, pseudo-observations of the data would have a zero likelihood. Inspired by the method in [27], we also want to select the level α giving the closest zero curve to the pseudo-observations:

$$\alpha^* = \sup \left\{ \alpha \in [0, 1] : \forall k \in \{1, \dots, n\}, \mathbf{U}^k \in L_\alpha^{\hat{C}} \right\} .$$

It follows directly that:

LEMMA 2.1 (conservative threshold α^*).

If $\alpha^* = \sup \left\{ \alpha \in [0, 1] : \forall k \in \{1, \dots, n\}, \mathbf{U}^k \in L_\alpha^{\hat{C}} \right\}$, then $\alpha^* = \min_{k=1, \dots, n} \hat{C}(\mathbf{U}^k)$.

Proof. Let $\alpha_1 \leq \alpha_2$, for any $\mathbf{u} \in L_{\alpha_2}^{\hat{C}}$, \mathbf{u} is also in $L_{\alpha_1}^{\hat{C}}$ by definition of the upper level sets. Hence by taking $\alpha = \min_{i=1, \dots, n} \hat{C}(\mathbf{U}^i)$, all $\mathbf{U}^k \in L_\alpha^{\hat{C}}$, $k \in \{1, \dots, n\}$. Furthermore, there exists $k^* \in \{1, \dots, n\}$ such that $\hat{C}(\mathbf{U}^{k^*}) = \alpha$, so that for any $\alpha' > \alpha$, $\mathbf{U}^{k^*} \notin L_{\alpha'}^{\hat{C}}$. \square

We discuss here consequences of this choice of the level α^* on the estimated copulas considered in this paper:

- For empirical copulas, the conservative threshold is almost surely $\alpha^* = \frac{1}{n}$ since any inferior value results in a zero set included in the axis $[0, \infty) \times \{0\}$ and $\{0\} \times [0, \infty)$.
- For strict Archimedean copulas, this choice leads to $\alpha^* > 0$ as soon as pseudo-observations are all strictly positive. It thus avoids setting $\alpha^* = 0$ which would lead to a degenerated zero set included in the axis $[0, \infty) \times \{0\}$ and $\{0\} \times [0, \infty)$.
- For non-strict Archimedean copulas, the choice of ψ_0^* as in Equation (2.4) leads to $\alpha^* = 0$ by construction. It would be possible to set smaller values of ψ_0^* leading to admissible parameter $\alpha^* \geq 0$, but for the sake of simplicity, we have considered here only the case where ψ_0^* was given by Equation (2.4).

2.3.4. Estimation of the marginals. The univariate marginals and their inverses also need to be estimated. This can be performed with the empirical quantiles or any method using truncated or non-truncated kernel density estimation. The choice of the method relates to the need to smooth and/or extrapolate the result. In some experiments with scarce data, we use the method proposed in [37] to estimate the support of the cumulative distribution function and its inverse, based on a catalog of beta distributions. Increasing the number of objectives usually implies to sample more points in the variable space to cover the objective space, providing more points to estimate each of the univariate marginals.

At last, the expression of the estimated level lines of the multivariate distribution also depends on the inverse functions of the marginal distributions, see Equation (2.3). Some parametric methods has been proposed in order to fit univariate distributions

and to obtain straightforward simple expressions for their inverse functions, see e.g. [4].

We recapitulate the general procedure for estimating level lines in Algorithm 1.

Algorithm 1 Estimation of the level curve of $F_{\mathbf{Y}}$

- 1: Get the univariate marginal distributions $\hat{F}_1, \dots, \hat{F}_m$ of $\mathbf{Y}_1, \dots, \mathbf{Y}_m$ empirically, with kernel density estimation and/or as in [37] and select the best solution.
 - 2: Compute the pseudo-observations $\left\{ \mathbf{U}^k \right\}_{k=1, \dots, n}$ from the data and compute the empirical copula \hat{C} .
 - 3: **if** Archimedean model too inaccurate or Archimedean hypothesis rejected (See Section 3 for more details.) **then**
 - 4: Estimate $\partial L_{\alpha}^{\hat{C}}$ with the empirical copula.
 - 5: **else**
 - 6: Select the generator ϕ giving the best result, based on prior knowledge or depending on a comparison with the empirical copula.
 - 7: Estimate $\partial L_{\alpha}^{C_{\phi}}$ with the best generator.
 - 8: **end if**
 - 9: Estimate the level lines of $F_{\mathbf{Y}}$ with Proposition 2.1.
 - 10: Select the level α^* depending on the chosen copula as prescribed in Section 2.3.3.
-

3. Pertinence of the Archimedean model. The interest of such a model, if appropriate, lies in the fact that if the dependency is accurately modeled, every observation gives information about the whole Pareto front, providing a continuous and smooth estimation. The parametric expression for the level curve of $F_{\mathbf{Y}}$, written in Proposition 2.1, requires the assumption that the copula describing the dependency structure can be approximated by an Archimedean copula. This section provides a discussion of the associated restrictions in practice and about the choice of an Archimedean copula model from the alternatives presented in Section 2.3.2.

3.1. Properties of Archimedean copulas: convexity, symmetry and associativity. The Archimedean model is convenient and tractable even with many objectives, but it imposes symmetry and associativity, e.g. with $m = 2$, $(u_1, u_2, u_3) \in [0, 1]^3$, $C(u_1, u_2) = C(u_2, u_1)$ and $C(C(u_1, u_2), u_3) = C(u_1, C(u_2, u_3))$. In addition the level lines of the copula are convex.

PROPOSITION 3.1 (Convexity of $\partial L_{\alpha}^{C_{\phi}}$). *The level curves of an Archimedean copula of dimension m are convex.*

Proof. This proposition is demonstrated in the case $m = 2$ in [32]. In the case $m > 2$, the result is still valid.

Given $\mathbf{u} = (u_1, \dots, u_m)$ and $\mathbf{v} = (v_1, \dots, v_m)$ two points of $\partial L_{\alpha}^{C_{\phi}}$. Given $\lambda \in [0, 1]$, we denote $\mathbf{w} = \lambda \mathbf{u} + (1 - \lambda) \mathbf{v}$. In dimension m , the generator ϕ is a m -monotone function, implying in particular that ϕ^{-1} is a decreasing convex function. Hence for all $i \in \{1, \dots, m\}$, $\phi^{-1}(w_i) = \phi^{-1}(\lambda u_i + (1 - \lambda) v_i) \leq \lambda \phi^{-1}(u_i) + (1 - \lambda) \phi^{-1}(v_i)$. Then $\phi^{-1}(w_1) + \dots + \phi^{-1}(w_m) \leq \lambda (\phi^{-1}(u_1) + \dots + \phi^{-1}(u_m)) + (1 - \lambda) (\phi^{-1}(v_1) + \dots + \phi^{-1}(v_m))$. Since \mathbf{u} and \mathbf{v} belongs to $\partial L_{\alpha}^{C_{\phi}}$,

$$(\phi^{-1}(u_1) + \dots + \phi^{-1}(u_m)) = (\phi^{-1}(v_1) + \dots + \phi^{-1}(v_m)) = \phi^{-1}(\alpha).$$

Then $\phi^{-1}(w_1) + \dots + \phi^{-1}(w_m) \leq \phi^{-1}(\alpha)$, which is equivalent to $\mathbf{w} \in L_{\alpha}^{C_{\phi}}$. \square

Note that having convex level lines does not imply that the level lines in the objective space will also be convex since it depends on the marginals. In the case when it is known that the Pareto front is convex, a sufficient condition to ensure the convexity of the Pareto front is to have concave marginals with an Archimedean copula.

PROPOSITION 3.2 (Convexity of ∂L_α^F). *If the marginals F_1, \dots, F_d are concave, then the level lines of ∂L_α^F are also convex.*

Proof. Given $\mathbf{y} = (y_1, \dots, y_m)$ and $\mathbf{z} = (z_1, \dots, z_m)$ two points of ∂L_α^F . Given $\lambda \in [0, 1]$, we denote $\mathbf{w} = \lambda\mathbf{y} + (1 - \lambda)\mathbf{z}$. The F_i 's are concave, hence for all $i \in (1, \dots, m)$, $\lambda F_i(y_i) + (1 - \lambda)F_i(z_i) \leq F_i(w_i)$. Since the generator ϕ^{-1} is a decreasing convex function, $\lambda\phi^{-1}(F_i(y_i)) + (1 - \lambda)\phi^{-1}(F_i(z_i)) \geq \phi^{-1}(F_i(w_i))$. And thus by summation $\lambda(\phi^{-1}(F_1(y_1)) + \dots + \phi^{-1}(F_m(y_m))) + (1 - \lambda)(\phi^{-1}(F_1(z_1)) + \dots + \phi^{-1}(F_m(z_m))) \geq \phi^{-1}(F_1(w_1)) + \dots + \phi^{-1}(F_m(w_m))$. Now, ϕ is a decreasing function: $\phi(\lambda(\phi^{-1}(F_1(y_1)) + \dots + \phi^{-1}(F_m(y_m))) + (1 - \lambda)(\phi^{-1}(F_1(z_1)) + \dots + \phi^{-1}(F_m(z_m)))) \leq \phi(\phi^{-1}(F_1(w_1)) + \dots + \phi^{-1}(F_m(w_m))) = F_\mathbf{Y}(\mathbf{w})$. Since \mathbf{y} and \mathbf{z} are in ∂L_α^F , $\phi^{-1}(F_1(y_1)) + \dots + \phi^{-1}(F_m(y_m)) = \phi^{-1}(F_1(z_1)) + \dots + \phi^{-1}(F_m(z_m)) = \phi^{-1}(\alpha)$. Then $F_\mathbf{Y}(\mathbf{w}) \geq \alpha$, which means that $\mathbf{w} \in \partial L_\alpha^F$. \square

If the level curves must be concave, then the use of survival copulas (associated with $1 - F_\mathbf{Y}$) can be a solution.

It is important to mention that even if the hypothesis of Archimedeanity is restrictive, it can still cover a great variety of situations, as illustrated in Figure 4 with varying copulas and marginals. The assessment of this hypothesis is detailed in the next paragraph.

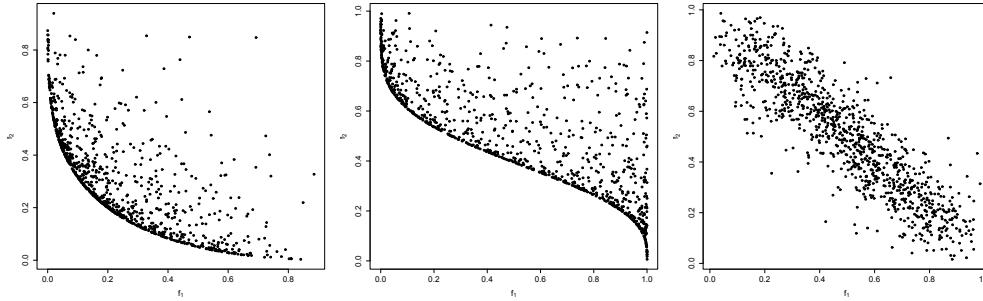


Figure 4: Scatterplots with a thousand of sample points $\mathbf{Y}^1, \dots, \mathbf{Y}^{1000}$ generated from Archimedean copulas models and further applying inverse of beta distribution functions as univariate marginals. Left: Clayton copula, $\theta = -0.8$, $F_1 = \text{Beta}(1, 3)$ and $F_2 = \text{Beta}(1.5, 3.5)$. Center: Clayton copula, $\theta = -0.8$, $F_1 = \text{Beta}(0.5, 0.5)$ and $F_2 = \text{Beta}(2.5, 2.5)$. Right: Frank copula, $\theta = -12$, $F_1 = \text{Beta}(2, 2)$ and $F_2 = \text{Beta}(2, 2)$.

3.2. Archimedeanity tests - Choosing between the different options.

An immediate solution is to test whether the hypothesis of Archimedeanity holds or not. Recent work exist in the bivariate case, see e.g. [7]. Otherwise a simple test is to compare visually the level curves of the empirical copula with those of the fitted

Archimedean copula in the same spirit as the normal probability plot in dimension one.

It remains to decide which Archimedean model is the best to estimate the Pareto front, by trying the different possibilities: parametric strict and non-strict models or non-parametric strict models. Non-strict parametric models seem best suited to estimate Pareto fronts due to the presence of the zero set but in certain circumstances non-parametric strict models perform better. For parametric families with analytical strict generator function, one can mention for instance the Clayton family ($\theta > 0$), Gumbel family or Frank families of Archimedean copulas. The parameters are evaluated using Maximum Likelihood.

Estimating a non-parametric generator from the data gives more flexibility when the Archimedean hypothesis is too strong, as illustrated in the applications. Even if it cannot capture the dissymmetry of the empirical copula, the fitted model is often more accurate with this non-parametric generator.

Apart from visual assessment, it is possible to estimate a distance between the empirical copula \hat{C} and a fitted copula C_ϕ , based on an integrated mean squared error (IMSE): $\int_{[0,1]^m} (\hat{C}(\mathbf{u}) - C_\phi(\mathbf{u}))^2 du_1 \dots du_m$ or with a root mean squared error (RMSE): $\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{C}(\mathbf{U}^i) - C_\phi(\mathbf{U}^i))^2}$.

Non-strict Archimedean copulas play a particular role for modeling the Pareto front, due to their non degenerated zero-sets. A generator of such a copula can be linked to a non-observable univariate random variable (e.g. the radial part of the copula, see [31]). The maximum value of such random variable is directly related to the location of the Pareto front, and using *end-point* probabilistic literature would be an interesting perspective (see e.g. [29], [22], [28], [18], and references therein).

4. Applications. To illustrate the benefits of the approach proposed in Algorithm 1 we take three classical bi-objective f_1, f_2 problems from the MOO literature: the ZDT1, ZDT6 [43] and Poloni [35] test problems. They have respectively convex, concave and disconnected Pareto fronts. The variable dimension is two in all the examples, but it could be much higher since the estimation procedure only deals with the objective space. Note that with an increasing variable dimension, it becomes necessary to increase the sample size. We use R packages *copula* [41, 26, 30, 23] for estimating strict Archimedean copulas and *ks* [12] for kernel density estimation.

4.1. Estimation of the Pareto front for the ZDT1 test problem. The first test problem, ZDT1 is a relatively simple benchmark problem:

ZDT1. Let $\mathbf{x} \in [0, 1]^d$ and $g(\mathbf{x}) = 1 + \frac{9}{d-1} \sum_{i=2}^d x_i$. Consider:

$$f_1(\mathbf{x}) = x_1, \quad f_2(\mathbf{x}) = g(\mathbf{x}) \left(1 - \sqrt{\frac{f_1(\mathbf{x})}{g(\mathbf{x})}} \right).$$

Here we choose $d = 2$ and draw a sample of size $n = 100$, uniformly in $[0, 1]^2$.

The first step is to estimate the marginals. As one can see from Figure 5, the parametric estimation based on beta distribution gives a good fit of the empirical

inverse of the marginals while non-parametric estimation is clearly too optimistic on the range of the ZDT1 test problem : $[0, 1]$ for f_1 and $[0, 10]$ for f_2 . Then we select the model with the best fit for the copula, which is the non-parametric copula model in this case, based on Figure 6. Here several models would be acceptable, since all the other Archimedean models look close to the empirical copula, except the non-strict model No. 2. However, the RMSE error on the pseudo-observations is the lowest with the non-parametric generator.

Finally we obtain the estimation of the position of the Pareto front, cf. Figure 7. While being slightly too optimistic on the right side, it is more accurate than the Pareto front approximation from the non-dominated points of the observations. Also a comparison with what would have been obtained using only the empirical copula illustrates that the Archimedean hypothesis brings a smoother and better localization.

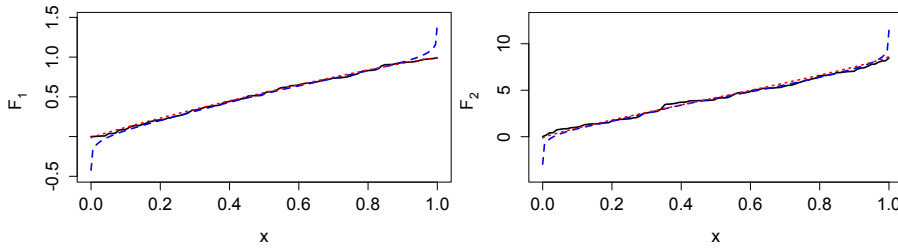


Figure 5: ZDT6 test problem: comparison between three estimation methods of the marginals F_1 and F_2 – empirical (black solid line), kernel density (blue dashed line) and fit of a generalized beta distribution (red dotted line) – for the objectives f_1 (left) and f_2 (right).

4.2. Estimation of the Pareto front for the ZDT6 test problem. The second test problem has a concave Pareto front and is harder due to a very low density around the Pareto optimal area:

ZDT6. Let $\mathbf{x} \in [0, 1]^d$ and $g(\mathbf{x}) = 1 + 9 \left(\sum_{i=2}^d \frac{x_i}{i} \right)^{1/4}$. Consider:

$$f_1(\mathbf{x}) = 1 - \exp(-4x_1) \sin^6(6\pi x_1), \quad f_2(\mathbf{x}) = g(\mathbf{x}) \left(1 - \left(\frac{f_1(\mathbf{x})}{g(\mathbf{x})} \right)^2 \right).$$

Again, we choose $d = 2$ and we draw a sample of size $n = 100$ uniformly in $[0, 1]^2$, giving observations farther away from the true Pareto front.

This time non-parametric estimation gives the best fit of the marginal distributions, see Figure 8. The best copula model is given by the non-parametric copula model, see Figure 9. Here again only the non-strict model No.2 is clearly not relevant. For all the models the level lines with α^* closely approximate the corresponding level line of the empirical copula, indicating that the Archimedean hypothesis is acceptable. Note that the lowest RMSE error is also in this case obtained with the non-parametric generator, taking advantage of the higher flexibility offered by this option.

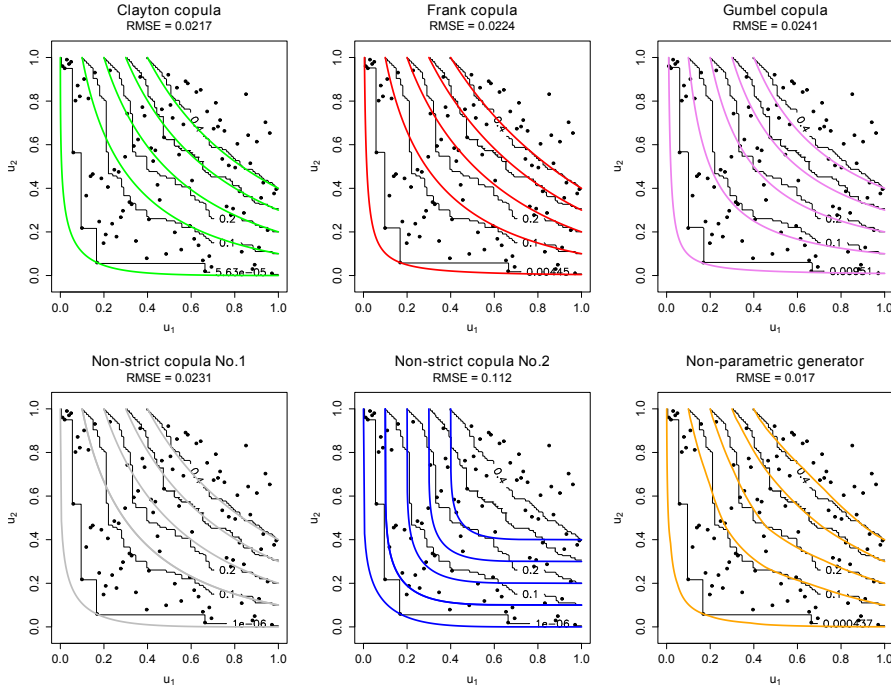


Figure 6: Levels lines $\partial L_{\alpha}^{C_{\phi}}$ of the different fitted Archimedean models based on the pseudo-data \mathbf{U}^k , $k = 1, \dots, n$, from test problem ZDT1. The level lines correspond in each case to α^* , 0.1, 0.2, 0.3 and 0.4.

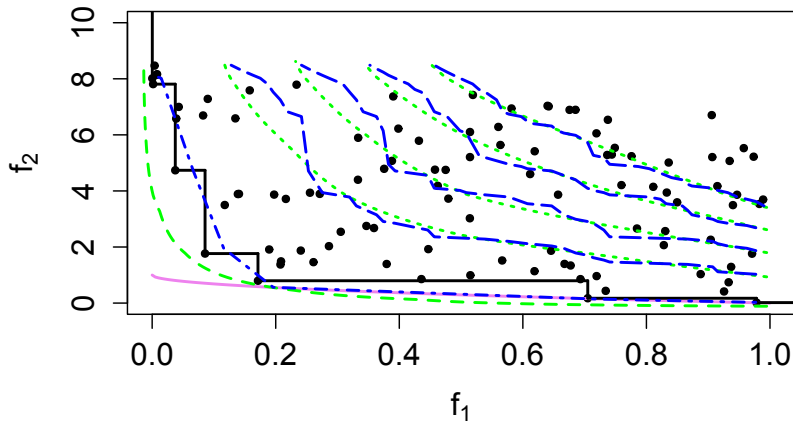


Figure 7: Estimated level line $\partial L_{\alpha^*}^F$ with the best C_{ϕ} for the ZDT1 test problem (green dashed line), compared to the Pareto front approximation from the observations \mathcal{P}_n (black line), the result with the empirical copula \hat{C} (blue dashed-dotted line) and the true Pareto front \mathcal{P} (violet solid line). Other level lines with levels 0.1, 0.2, 0.3 and 0.4 are also displayed with thinner lines.

The estimation of the position of the Pareto front is presented in Figure 10, showing that the model can extend the information of the extremal observations to improve estimation in the center of the attainable image space, where no observations are available. In this case, knowing the range of the objectives, for instance by minimizing each objective separately would help selecting the best estimation of marginals.

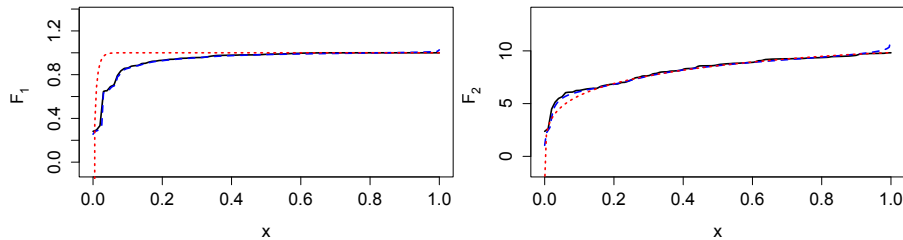


Figure 8: ZDT6 test problem: comparison between three estimation methods of the marginals F_1 and F_2 – empirical (black solid line), kernel density (blue dashed line) and fit of a generalized beta distribution (red dotted line) – for the objectives f_1 (left) and f_2 (right).

4.3. Estimation of the Pareto front for the Poloni test problem. This last problem has a disconnected Pareto front with concave and convex parts. A mathematical description of the problem can be found in [35].

The estimation of marginals suggest the use of non-parametric estimation for f_1 and the estimation from the catalog of beta distribution for f_2 , as visible in Figure 11. Concerning the copula model, it appears in Figure 12 that the level lines α^* from the Archimedean models do not approximate well the shape of the Pareto front. In particular, the lowest level of the empirical level lines are highly non-symmetric. Thus we discard the Archimedean assumption and we keep the empirical copula, getting the extrapolation from the marginals.

The estimation of the position of the Pareto front is shown in Figure 13, showing that the proposed approach is also suited when the Archimedean model hypothesis does not hold. Even if the approximation cannot be improved on the lowest part of the Pareto front due to the absence of observations in this area, it effectively gives a better estimation of the Pareto front in the other parts.

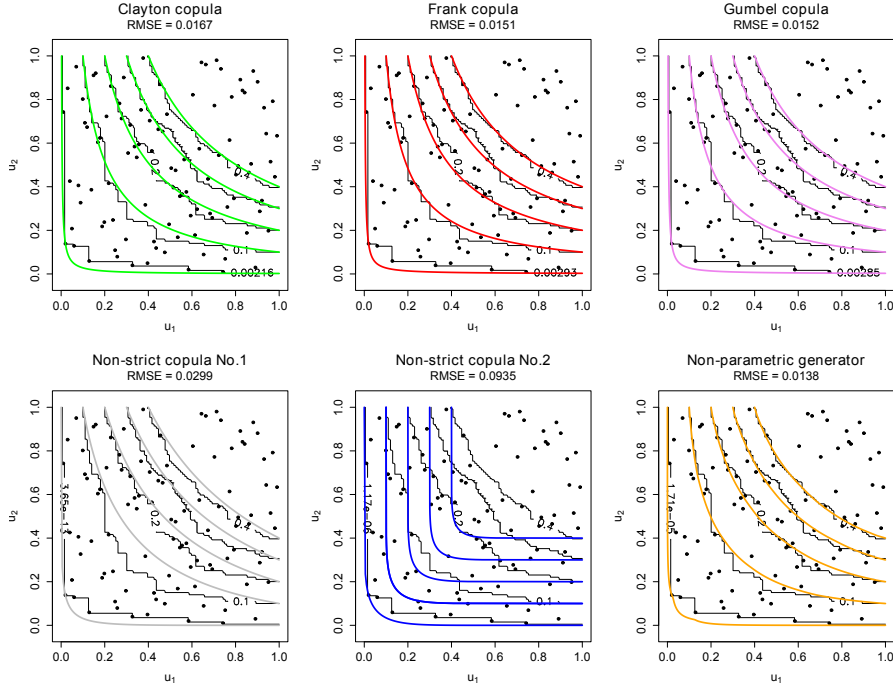


Figure 9: Levels lines $\partial L_{\alpha^*}^{C_\phi}$ of the different fitted Archimedean models based on the pseudo-data \mathbf{U}^k , $k = 1, \dots, n$, from test problem ZDT6. The level lines correspond in each case to α^* , 0.1, 0.2, 0.3 and 0.4

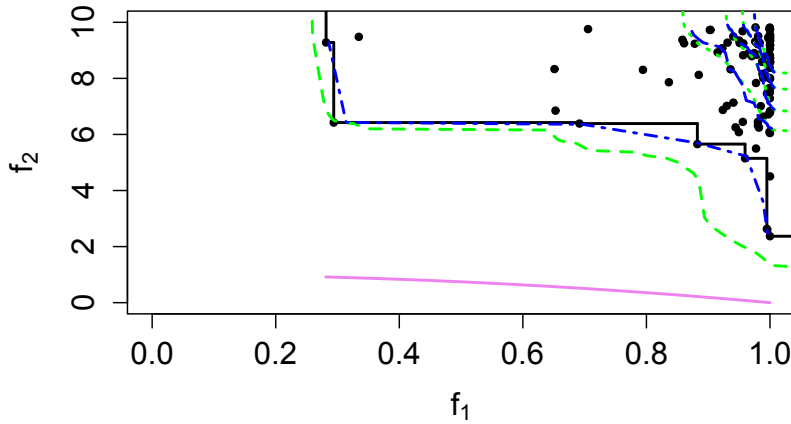


Figure 10: Estimated level line $\partial L_{\alpha^*}^F$ with the best C_ϕ for the ZDT6 test problem (green dashed line), compared to the Pareto front approximation from the observations \mathcal{P}_n (black line), the result with the empirical copula \hat{C} (blue dashed-dotted line) and the true Pareto front \mathcal{P} (violet solid line). Other level lines with levels 0.1, 0.2, 0.3 and 0.4 are also displayed with thinner lines.

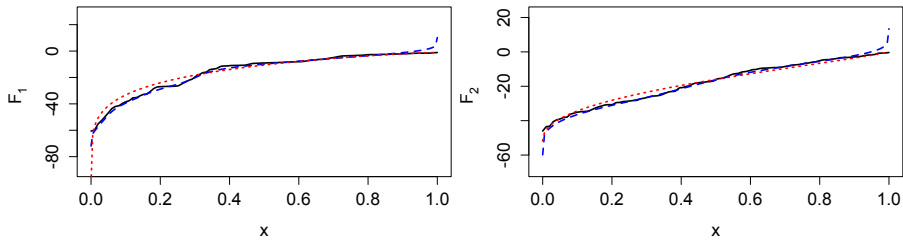


Figure 11: Poloni test problem: comparison between three estimation methods of the marginals F_1 and F_2 – empirical (black solid line), kernel density (blue dashed line) and fit of a generalized beta distribution (red dotted line) – for the objectives f_1 (left) and f_2 (right).

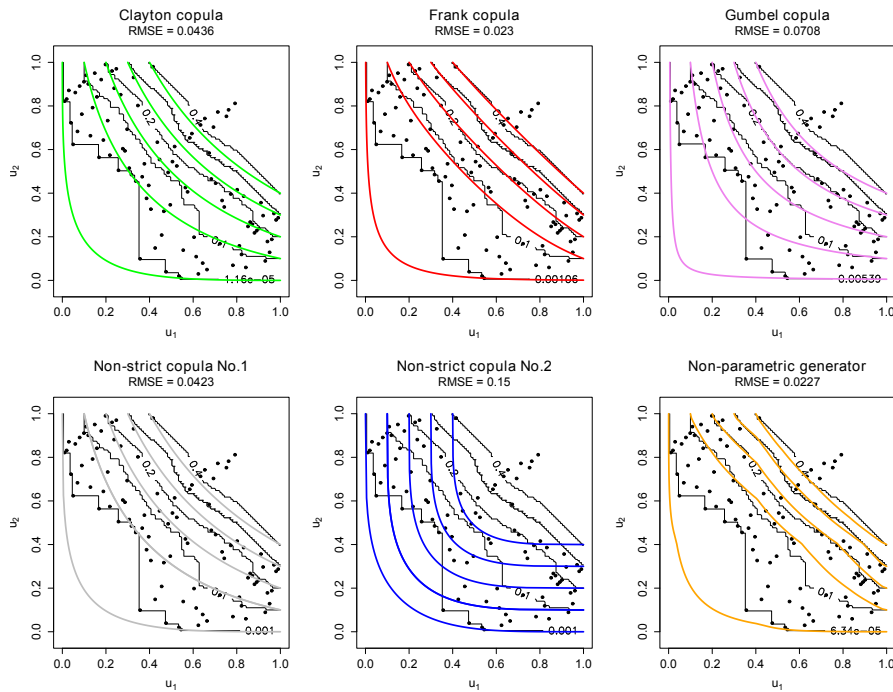


Figure 12: Levels lines $\partial L_{\alpha}^{C_{\phi}}$ of the different fitted Archimedean models based on the pseudo-data \mathbf{U}^k , $k = 1, \dots, n$, from test problem Poloni. The level lines correspond in each case to α^* , 0.1, 0.2, 0.3 and 0.4.

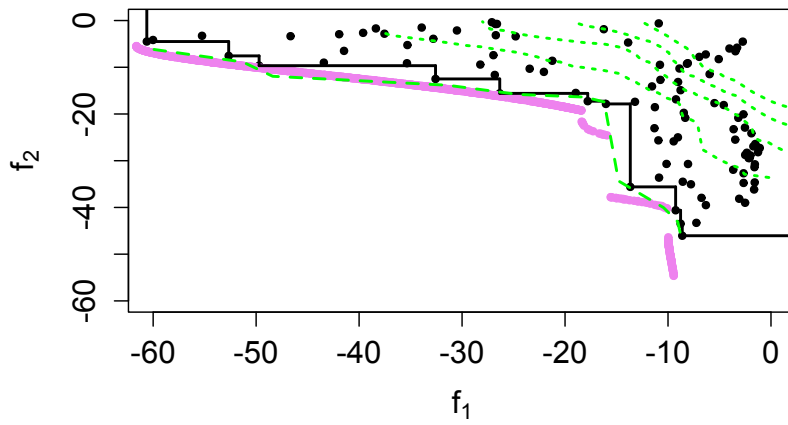


Figure 13: Estimated level line $\partial L_{\alpha^*}^F$ for the Poloni test problem (green dashed line), compared to the Pareto front approximation from the observation \mathcal{P}_n (black line) and the true Pareto front \mathcal{P} (violet solid line). Other level lines with levels 0.1, 0.2, 0.3 and 0.4 are also displayed with thinner lines.

5. Conclusions and perspectives. In this paper, we addressed the problem of estimating the Pareto front in an initial phase of multiobjective problems when an i.i.d. sample is available.

At the theoretical level, we established a connection between Pareto fronts and upper level lines of the outputs sample. The approximation of these level lines can be done with very few natural assumptions by using the theory of copulas. An interesting particular case is for Archimedean copulas, for which analytical expressions are available. This assumption can be checked visually or statistically with specific tests of the literature.

The benefits of this methodology are illustrated on some common bi-objective problems from multi-objective optimization literature.

There are several perspectives of this research. Though the Archimedean assumption corresponds to a large range of copulas, it is sometimes inappropriate. As an intermediate solution to the general alternative proposed here – i.e. usage of empirical copula –, it may be interesting to consider nested Archimedean copulas, see e.g. [30] and references therein, or other families of copulas. Further developments about non-strict generators have also been evocated in Section 3.

Secondly, the restriction to i.i.d. samples can be mitigated since in some particular cases random search performs relatively well, see e.g. [1]. However, it might be possible to extend the approach of [24] to deal with non independent observations.

Finally, this methodology relies on the estimated distribution of the outputs. In the context of time-consuming objective functions, such estimation could be improved by using surrogate models.

Acknowledgements. Part of this work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments.

REFERENCES

- [1] J. BERGSTRA AND Y. BENGIO, *Random search for hyper-parameter optimization*, The Journal of Machine Learning Research, 13 (2012), pp. 281–305.
- [2] E. DI BERNARDINO AND D. RULLIÈRE, *Distortions of multivariate distribution functions and associated level curves: Applications in multivariate risk theory*, Insurance: Mathematics and Economics, 53 (2013), pp. 190 – 205.
- [3] P. BHARDWAJ, B. DASGUPTA, AND K. DEB, *Modeling Pareto-optimal set using B-spline basis functions*, tech. report, Tech. Rep. 2012008, Indian Institute of Technology Kanpur, Kanpur, India, 2012.
- [4] A. BIENVENÛE AND D. RULLIÈRE, *Iterative adjustment of survival functions by composed probability distortions*, The Geneva Risk and Insurance Review, 37 (2012), pp. 156–179.
- [5] M. BINOIS, D. GINSBOURGER, AND O. ROUSTANT, *Quantifying uncertainty on Pareto fronts with Gaussian process conditional simulations*, to appear in the European Journal of Operational Research, (+2014).
- [6] L. BREIMAN, *Probability*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
- [7] A. BÜCHER, H. DETTE, AND S. VOLGUSHEV, *A test for Archimedeanity in bivariate copula models*, Journal of Multivariate Analysis, 110 (2012), pp. 121–132.
- [8] KALYANMOY DEB, *Introduction to evolutionary multiobjective optimization*, in Multiobjective Optimization, J. Branke et al., eds., vol. 5252 of LNCS, Springer, 2008, p. 59–96.
- [9] P. DEHEUVELS, *La fonction de dépendance empirique et ses propriétés*, Acad. Roy. Belg. Bull. Cl. Sci., 65 (1979), pp. 274–292.

- [10] E. DI BERNARDINO AND D. RULLIÈRE, *On certain transformations of Archimedean copulas : Application to the non-parametric estimation of their generators*, Dependence Modeling, 1 (2013), pp. 1–36. doi:10.2478/demo-2013-0001.
- [11] D. S. DIMITROVA, V. K. KAISHEV, AND S. I. PENEV, *Ged spline estimation of multivariate archimedean copulas*, Computational Statistics & Data Analysis, 52 (2008), pp. 3570–3582.
- [12] T. DUONG, *ks: Kernel smoothing*, 2014. R package version 1.9.2.
- [13] A. ERDELY, J. M. GONZÁLEZ-BARRIOS, AND M. M. HERNÁNDEZ-CEDILLO, *Frank’s condition for multivariate Archimedean copulas*, Fuzzy Sets and Systems, In press (2013).
- [14] Y. GAO, L. PENG, F. LI, M. LIU, AND X. HU, *Estimation of distribution algorithm with multivariate t-copulas for multi-objective optimization*, Intelligent Control and Automation, 4 (2013), p. 63.
- [15] Y. GAO, L. PENG, F. LI, M. LIU, AND W. LIU, *Archimedean copula-based estimation of distribution algorithm for multi-objective optimisation*, International Journal of Trust Management in Computing and Communications, 1 (2013), pp. 200–211.
- [16] C. GENEST, J. NEŠLEHOVÁ, AND J. ZIEGEL, *Inference in multivariate Archimedean copula models*, TEST, 20 (2011), pp. 223–256.
- [17] C. GENEST AND L.-P. RIVEST, *Statistical inference procedures for bivariate Archimedean copulas*, Journal of the American Statistical Association, 88 (1993), pp. 1034–1043.
- [18] S. GIRARD, A. GUILLOU, AND G. STUPFLER, *Estimating an endpoint with high-order moments*, test, 21 (2012), pp. 697–729.
- [19] T. GOEL, R. VAIDYANATHAN, R. T. HAFTKA, W. SHYY, N. V. QUEIPO, AND K. TUCKER, *Response surface approximation of Pareto optimal front in multi-objective optimization*, Computer Methods in Applied Mechanics and Engineering, 196 (2007), pp. 879–893.
- [20] P. HALL, *On estimating the endpoint of a distribution*, The Annals of Statistics, (1982), pp. 556–568.
- [21] P. HALL, M. NUSSBAUM, AND S. E. STERN, *On the estimation of a support curve of indeterminate sharpness*, Journal of Multivariate Analysis, 62 (1997), pp. 204–232.
- [22] P. HALL AND J. Z. WANG, *Estimating the end-point of a probability distribution using minimum-distance methods*, Bernoulli, (1999), pp. 177–189.
- [23] M. HOFERT, I. KOJADINOVIC, M. MAECHLER, AND J. YAN, *copula: Multivariate Dependence with Copulas*, 2014. R package version 0.999-10.
- [24] J. HÜSLER, *On applications of extreme value theory in optimization*, in Experimental Methods for the Analysis of Optimization Algorithms, Springer, 2010, pp. 185–207.
- [25] G. KIM, M. J. SILVAPULLE, AND P. SILVAPULLE, *Comparison of semiparametric and parametric methods for estimating copulas*, Computational Statistics & Data Analysis, 51 (2007), pp. 2836–2850.
- [26] I. KOJADINOVIC AND J. YAN, *Comparison of three semiparametric methods for estimating dependence parameters in copula models*, Insurance: Mathematics and Economics, 47 (2010), pp. 52–63.
- [27] S. KÖNIG, H. KAZIANKA, J. PILZ, AND J. TEMME, *Estimation of nonstrict Archimedean copulas and its application to quantum networks*, Applied Stochastic Models in Business and Industry, (2014).
- [28] D. LI, L. PENG, AND X. XU, *Bias reduction for endpoint estimation*, Extremes, 14 (2011), pp. 393–412.
- [29] W.-Y. LOH, *Estimating an endpoint of a distribution with resampling methods*, The Annals of Statistics, (1984), pp. 1543–1550.
- [30] M. HOFERT AND M. MÄCHLER, *Nested Archimedean copulas meet R: The nacopula package*, Journal of Statistical Software, 39 (2011), pp. 1–20.
- [31] A.J. MCNEIL AND J. NEŠLEHOVÁ, *Multivariate Archimedean copulas, d-monotone functions and l_1 -norm symmetric distributions*, The Annals of Statistics, 37 (2009), pp. 3059–3097.
- [32] R. B. NELSEN, *An introduction to copulas*, vol. 139 of Lecture Notes in Statistics, Springer-Verlag, New York, 1999.
- [33] R. B. NELSEN, J. J. QUESADA-MOLINAB, J. A. RODRIGUEZ-LALLENAC, AND M. ÚBEDA-FLORESC, *Kendall distribution functions*, Statistics and Probability Letters, 65 (2003), pp. 263–268.
- [34] M. OMELKA, I. GIJBELS, AND N. VEREVERBEKE, *Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing*, The Annals of Statistics, 37 (2009), pp. 3023–3058.
- [35] C. POLONI, A. GIURGEVICH, L. ONESTI, AND V. PEDIRODA, *Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimizer for a complex design problem in fluid dynamics*, Computer Methods in Applied Mechanics and Engineering, 186 (2000), pp. 403–420.
- [36] W. PONWEISER, T. WAGNER, D. BIERMANN, AND M. VINCZE, *Multiobjective optimization on a limited budget of evaluations using model-assisted S-metric selection*, in PPSN X,

- G. Rudolph et al., eds., vol. 5199 of LNCS, Springer, 2008, p. 784–794.
- [37] N. QUEIPO, S. PINTOS, E. NAVA, AND A. VERDE, *Setting targets for surrogate-based optimization*, Journal of Global Optimization, (2010), pp. 1–19.
 - [38] M. SKLAR, *Fonctions de répartition à n dimensions et leurs marges*, Université Paris 8, 1959.
 - [39] I. VOUTCHKOV AND A. KEANE, *Multi-objective optimization using surrogates*, Computational Intelligence in Optimization, (2010), pp. 155–175.
 - [40] L.-F. WANG AND J.-C. ZENG, *Estimation of distribution algorithm based on copula theory*, in Exploitation of linkage learning in evolutionary algorithms, Springer, 2010, pp. 139–162.
 - [41] J. YAN, *Enjoy the joy of copulas: with a package copula*, Journal of Statistical Software, 21 (2007), pp. 1–21.
 - [42] A. ZHOU, B.-Y. QU, H. LI, S.-Z. ZHAO, P. N. SUGANTHAN, AND Q. ZHANG, *Multiobjective evolutionary algorithms: A survey of the state of the art*, Swarm and Evolutionary Computation, 1 (2011), pp. 32–49.
 - [43] E. ZITZLER, K. DEB, AND L. THIELE, *Comparison of multiobjective evolutionary algorithms: Empirical results*, Evolutionary computation, 8 (2000), pp. 173–195.