Copula-based Sensitivity Measures of Computer Experiments

Elmar Plischke
Institut für Endlagerforschung, Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany
Emanuele Borgonovo
Department of Decision Sciences, Bocconi University, 20136 Milano, Italy

ABSTRACT: We combine results from copula theory with results from global sensitivity analysis, on one hand investigating if classical measures of association and concordance can be used in a sensitivity context, while on the other hand embedding sensitivity measures in a copula context. Graphical tools for visual sensitivity analysis are presented. Empirical Bernstein copulas are used for numerical estimation of distribution-based sensitivity measures.

1 INTRODUCTION

Copulas are an indispensable tool for modelling dependence structures in mathematical finance and other branches of computational statistics. In global sensitivity analysis, one is interested in the converse: Gaining insight into the dependence structure between in- and output of a (black box simulation) model. Here distribution based sensitivity methods are gaining more and more attention to overcome the limitations of linear regression and variance-based sensitivity measures. The invariance of distribution based sensitivity measures under monotonic transformations Borgonovo et al. [2014] shows that these measures can be stated in terms of copula theory. In this article we combine results from copula theory with global sensitivity analysis. We limit our discussion to pairs (X, Y) of random variables. For a sensitivity analysis of model output one normally has multiple inputs available, so that such an analysis has to be carried out multiple times for obtaining a ranking list. Hence the X in the following discussion is a fixed input factor of interest. By ignoring all other inputs in the analysis, this implies that the approach is viable in a dependent input setting. However, one can not discriminate if the output uncertainty is then due to the input interdependence or the model interaction.

2 COPULAS

For a bivariate random vector (X, Y) the copula C is defined implicitly as the function which satisfies $F_{XY}(x, y) = C(F_X(x), F_Y(y))$. If the functional inverses of the marginal distributions exist, i.e. the distributions are strictly monotone and have no jump discontinuities, then $C(u, v) = F_{XY}(F_X^{-1}(u), F_Y^{-1}(v))$. In particular, $C(u, 0) = 0 = C(0, u)$ and $C(u, 1) = u$, $C(1, v) = v$. The copula density c is defined by $\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)} = c(F_X(x), F_Y(y))$ and satisfies $c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v)$. The density is nonnegative, $c(u, v) \geq 0$ for all $u, v \in [0, 1]$. If C is absolute continuous then $C(u, v) = \int_0^u \int_0^v c(s, t) \, \mathrm{d}t \, \mathrm{d}s$.

The conditional copula is given by $C(v|u) = \int_0^1 c(u, v) \, \mathrm{d}v = \frac{\partial C(u,v)}{\partial u}$, the conditional copula density is simply $c(u|v) = c(u, v) = c(v|u)$. Integration by parts gives $\int_0^1 vc(u, v) \, \mathrm{d}v = 1 - \int_0^1 C(v|u) \, \mathrm{d}v$ and $\iint_{[0,1]^2} uvc(u, v) \, \mathrm{d}u \, \mathrm{d}v = \iint_{[0,1]^2} C(u, v) \, \mathrm{d}u \, \mathrm{d}v$. More details can be found in Nelsen [2006].

Let us denote the uniform transformations of X and Y by $U = F_X(X)$ and $V = F_Y(Y)$, so that $U, V \sim U[0, 1]$. X and Y are called comonotone if $U = V$, they are called contramonotone if $U = -V$. Associated with these two configurations are the Fréchet-Hoeffding copulas $M(u, v) = \min\{u, v\}$ and $W(u, v) = \max\{u + v - 1, 0\}$. All bivariate copulas C are bounded by $W(u, v) \leq C(u, v) \leq M(u, v)$.

3 A VARIETY OF MEASURES

In the statistics literature there is a vivid discussion about measures of concordance/association, and coefficients of determination/goodness-of-fit. While the exact terms are not clearly defined, there are wishlists or axioms for the properties of the associated measures (e.g. Rényi [1959], Scarsini [1984], Taylor [2007], van der Linde and Tutz [2008]).

In sensitivity analysis one is interested in measures of association. Among practitioners, linear or rank linear regression methods are predominant, however, these are measures of concordance. First order effects, also called correlation ratios, are a notable exception as they measure association [Kruskal, 1958], see the discussion in Section 4.

As a notational reminder, the measures which are discussed in the following and are based on rank-transformed values are marked with an asterisk. This creates some nonstandard notation, however there seems not to be a commonly agreed notation to rely upon.

Let us discuss some of the measures found in the literature. Kendall τ compares the concordance probability to the discordance probability [Nelson, 1991],

$$\tau^* = \iint_{\mathbb{R}^2} \big(\mathbb{P}(X{<}x, Y{<}y \vee X{>}x, Y{>}y)$$

$$-\mathbb{P}(X{<}x, Y{>}y \vee X{>}x, Y{<}y)\big) f_{XY}(x,y)\mathrm{d}x\mathrm{d}y \quad (1)$$

$$= 4\iint_{[0,1]^2} C(u,v)c(u,v)\,\mathrm{d}u\,\mathrm{d}v - 1.$$

Spearman ϱ^* is the Pearson correlation coefficient of U and V. Using the Hoeffding formula for the covariance of two variables, which is given by

$$\mathrm{Cov}(X,Y) = \iint_{\mathbb{R}^2} (F_{XY}(x,y) - F_X(x)F_Y(y))\mathrm{d}x\mathrm{d}y$$

$$= \iint_{[0,1]^2} (C(u,v) - uv)\mathrm{d}F_X^{-1}(u)\mathrm{d}F_Y^{-1}(v), \quad (2)$$

on $\mathrm{Cov}(U,V)$ we obtain a formula for Spearman ϱ^* in terms of copula,

$$\varrho^* = \frac{\mathrm{Cov}(U,V)}{\sqrt{\mathbb{V}[U]\,\mathbb{V}[V]}}$$

$$= 12\iint_{[0,1]^2} (C(u,v) - uv)\,\mathrm{d}u\,\mathrm{d}v \quad (3)$$

$$= 12\iint_{[0,1]^2} C(u,v)\,\mathrm{d}u\,\mathrm{d}v - 3$$

where $\mathbb{E}[U] = \mathbb{E}[V] = \frac{1}{2}$ and $\mathbb{V}[U] = \mathbb{V}[V] = \frac{1}{12}$. Scarsini [1984] uses for Spearman ϱ^* the expression

$$\varrho^* = 3\iint_{[0,1]^2} (2u-1)(2v-1)c(u,v)\,\mathrm{d}u\,\mathrm{d}v \quad (4)$$

which, after some manipulations, can be identified with (3). Spearman ϱ^* is a measure of correlation, but it may not be suited to detect stochastic dependence. Based on this observation, refined indicators have been investigated. Schweizer and Wolff [1981] discuss the sigma index using the L^1 norm

$$s^* = 12\iint_{[0,1]^2} |C(u,v) - uv|\,\mathrm{d}u\,\mathrm{d}v, \quad (5)$$

its L^2 variant (this Cramér/von Mises ω^2-statistics was introduced in Hoeffding [1948], see also Blum et al. [1961], Schmid et al. [2010])

$$w^* = \sqrt{90\iint_{[0,1]^2} (C(u,v) - uv)^2\,\mathrm{d}u\,\mathrm{d}v} \quad (6)$$

and its L^∞ variant (normalized star discrepancy, Niederreiter [1992])

$$D^* = 4\max_{[0,1]^2} |C(u,v) - uv|\,. \quad (7)$$

These are all special cases of the following approach.

Proposition 1. *Suppose that the distance between copulas C and D is measured by an L^p norm,*

$$d_p(C,D) = \sqrt[p]{\iint_{[0,1]^2} |C(u,v) - D(u,v)|^p\,\mathrm{d}u\,\mathrm{d}v},$$

if $p \geq 1$ and $d_\infty = \max_{[0,1]^2} |C(u,v) - D(u,v)|$. Then a normalized measure of association for a copula C is given by

$$\gamma_p^* = \frac{d_p(C,\Pi)}{d_p(M,\Pi)} \quad (8)$$

where $\Pi(u,v) = uv$ is the product (independence) copula.

For all copulas, one has $d_p(M,\Pi) \geq d_p(C,\Pi)$ and $d_p(M,\Pi) = d_p(W,\Pi)$. For $p < \infty$ the distance of the independent copula to the co/contramonotone copulas is given by $d_p(M,\Pi) = d_p(W,\Pi) = (p+1)^{-2/p}\binom{2p+1}{p}^{-1/p}$ and $\lim_{p\to\infty} d_p(M,\Pi) = \frac{1}{4} = d_\infty(M,\Pi)$.

The measures of (8) are 0 if and only if the associated random variables X and Y are independent, they yield 1 in case of co- or contramonotone data.

4 RANK CORRELATION RATIO

In global sensitivity analysis the correlation ratio [Pearson, 1905]

$$\eta = \frac{\mathbb{V}[\mathbb{E}[Y|X]]}{\mathbb{V}[Y]} \quad (9)$$

which is also called main effect, Sobol' first order index, etc., is a well-studied object [Saltelli et al., 2000]. However, it is only invariant with respect to affine linear output transformations. Let us replace X with U and Y with V in order to create a version which is invariant with respect to general monotonic transformations [Saltelli and Sobol', 1995]. Note that the nonparametric regression curve of V conditional to $U = u$

is given by

$$\mathbb{E}[V|U=u] = \int_0^1 v c(u,v)\,\mathrm{d}v$$

$$= \frac{\partial C}{\partial u}(u,1) - \int_0^1 \frac{\partial C}{\partial u}(u,v)\,\mathrm{d}v$$

$$= 1 - \int_0^1 C(v|u)\,\mathrm{d}v. \quad (10)$$

Analogously to Spearman ρ^* the rank correlation ratio is then defined by

$$\eta^* = \frac{\mathbb{V}[\mathbb{E}[V|U]]}{\mathbb{V}[V]} = 12\,\mathbb{E}[(\mathbb{E}[V|U]-\mathbb{E}[V])^2]$$
$$= 3\int_0^1 \left(1-2\int_0^1 C(v|u)\,\mathrm{d}v\right)^2 \mathrm{d}u. \quad (11)$$

As $\iint_{[0,1]^2} C(v|u)\,\mathrm{d}u\,\mathrm{d}v = \int_0^1 (C(1,v)-C(0,v))\,\mathrm{d}v = \int_0^1 v\,\mathrm{d}v = \frac{1}{2}$ we can further simplify (11)

$$\eta^* = 3\int_0^1 \left(1-4\int_0^1 C(v|u)\mathrm{d}v + 4\left(\int_0^1 C(v|u)\mathrm{d}v\right)^2\right)\mathrm{d}u \quad (12)$$
$$= 12\int_0^1 \left(\int_0^1 C(v|u)\,\mathrm{d}v\right)^2 \mathrm{d}u - 3.$$

Contrary to the indicators discussed in Section 3, (12) is not symmetric. As it is a measure of the explanatory power of the nonlinear regression curve symmetry is not intended.

The reader may have guessed by now that η^* is by no means a newly discovered quantity. There is a direct link to the Kruskal-Wallis test [Kleijnen and Helton, 1999]. Given a sample $(u_i, v_i) \in \{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}^{n\times 2}$ (scaled ranks, assuming no ties in the data) and partition size M the associated statistic is given by

$$H = (n-1)\frac{\sum_{m=1}^{M} n_m\left(\bar{v}_m - \bar{v}\right)^2}{\sum_{i=1}^{n}\left(v_i - \bar{v}\right)^2}, \text{where}$$

$$n_m = \sum_{i:Mu_i\in(m-1,m]} 1, \qquad \bar{v}_m = n_m^{-1}\sum_{i:Mu_i\in(m-1,m]} v_i. \quad (13)$$

and $\bar{v} = 0.5$ is the average of V. Now, $(n-1)^{-1}H$ is the estimator for η^* of the pair (U,V) already suggested by Pearson [1905] using piecewise constant step functions as regression model.

From Fréchet [1934] it follows that $\eta^* = \rho^2(V, \mathbb{E}[V|U])$. As a consequence, if the nonparametric regression curve is linear (i.e., the dependence in the original data is monotonic) then $\eta^* = \rho^{*2}$.

5 DISTRIBUTION-BASED SENSITIVITY METHODS

One drawback of the measures of Prop. 1 lies in the fact that functional dependences between X and Y other than co- and contra-monotonic are not well-detected. For functional dependence we suggest to use the rank correlation ratio (12), however this measure is not uncovering stochastic dependences. Hence the ideal measure should be positive-definite in the sense that it vanishes if and only if X and Y are independent and it should be equal to one if Y is a function of X. But stochastic dependence is broader than functional dependence and hence the measure might be unity even if there is no functional dependence of Y on X.

Instead of considering the copula directly as in Section 3, one can compare the copula density $c(u,v)$ with the constant 1 [Wei et al., 2014], using L^p distances,

$$\delta_p^*(c) = \iint_{[0,1]^2} |c(u,v)-1|^p\,\mathrm{d}u\,\mathrm{d}v. \quad (14)$$

As the density of the independent copula $\Pi(u,v) = uv$ is the constant 1 we have $\delta_p^*(\pi) = 0$. For the comonotone copula the density m is singular and we obtain a Dirac δ measure along the diagonal $m(u,v) = \delta(u-v)$, $\delta_p^*(m) = \iint_{[0,1]^2} |\delta(u-v)-1|^p\,\mathrm{d}u\,\mathrm{d}v = 2$

Conjecture 2. For all copula densities c, $\delta_p^*(c) \le 2$.

Replacing the underlying L^p norm in (14) by the Kullback-Leibler discrepancy one obtains a so-called negentropy/mutual information measure [Joe, 1989, Blumentritt and Schmid, 2012].

$$\theta^* = \iint_{[0,1]^2} c(u,v)\log c(u,v)\,\mathrm{d}u\,\mathrm{d}v \quad (15)$$

using the convention $0\log(0) = 0$. The nonnegativity $\theta^* \ge 0$ follows from Jensen's inequality.

As discussed in Borgonovo et al. [2013], distribution based sensitivity measures are invariant under monotonic transformations. As copula theory is the study of properties invariant under increasing transformations we may consider distribution based measures in a copula framework. When rewriting the Kolmogorov-Smirnov distance, the Kuiper discrepancy and the Borgonovo measure [Plischke and Borgonovo, 2015] in terms of U and V, they can be expressed as

$$\beta = \int_0^1 \max_{v\in[0,1]} |C(v|u) - v| \,\mathrm{d}u,$$

$$\kappa = \int_0^1 \Bigg[\max_{v\in[0,1]} (C(v|u) - v) + \max_{v\in[0,1]} (v - C(v|u)) \Bigg] \,\mathrm{d}u,$$

$$\delta = \frac{1}{2} \iint_{[0,1]^2} |c(u,v) - 1| \,\mathrm{d}u\,\mathrm{d}v.$$

Here, only δ is symmetric with respect to u and v. Both β and κ are closely related to the star discrepancy (7). While D^* is constructed from a two-dimensional L^∞ measure, β mixes one-dimensional L^∞ and L^1 measures. Furthermore, κ uses a range as a distance measure instead of the maximum of the absolute values. These measures are independent under monotone transformations of inputs or outputs. Hence their rank-based starred versions coincide with the originals. With respect to (14) we have $\delta = 0.5\delta_1^*$. For each $u \in [0,1]$, $c(u,v)$ is a density in v. Hence we can apply a theorem of Scheffé [Devroye and Györfi, 1985] which considers differences of densities,

$$\delta = \frac{1}{2} \iint_{[0,1]^2} |c(u,v) - 1| \,\mathrm{d}v\,\mathrm{d}u = \int_0^1 \int_{v: c(u,v)\geq 1} (c(u,v) - 1) \,\mathrm{d}v\,\mathrm{d}u. \tag{16}$$

Now, suppose that the set $\{v : c(u,v) \geq 1\}$ is an interval (or the complement of an interval) with boundaries $v_{\text{bottom}}(u)$ and $v_{\text{top}}(u)$ then these boundary points satisfy $c(u,v) - 1 = 0$. Therefore they are (inner) extreme points of $v \mapsto C(v|u) - v$. If this interval case holds for all $u \in [0,1]$ then

$$\delta = \int_0^1 (C(v|u) - v)\Big|_{v_{\text{bottom}}(u)}^{v_{\text{top}}(u)} \,\mathrm{d}u = \int_0^1 \left(\max_{v\in[0,1]} (C(v|u) - v) - \min_{v\in[0,1]} (C(v|u) - v) \right) \mathrm{d}u = \kappa. \tag{17}$$

If for all u the function $v \mapsto C(v|u) - v$ has constant sign then also $\kappa = \beta$.

6 SAMPLE-BASED INDICATORS AND GRAPHICAL METHODS

For some of the presented indicators there are sample-based estimators available. Especially, the Monte-Carlo technique yields $\iint G(u,v)c(u,v)\,\mathrm{d}u\,\mathrm{d}v \approx n^{-1}\sum_{i=1}^n G(u_i, v_i)$ for bivariate functions G and samples $(u_i, v_i)_{i=1,\ldots,n}$ which follow the C copula distribution. Hence (1) and (4) are well suited for this approach. Following Wei et al. [2013], we may write (16) as $\frac{1}{2}\iint |1 - c^{-1}(u,v)|\, c(u,v)\mathrm{d}u\mathrm{d}v$ which also shows the possibility of a MC estimator for δ. The empirical copula $\widehat{C}$ for the sample $(u_i, v_i)_{i=1,\ldots,n}$ is given by

$$\widehat{C}(u,v) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{u_i \leq u\}\mathbb{1}\{v_i \leq v\}. \tag{18}$$

With no ties, the order statistics of u and v satisfy $u_{(i)} = v_{(i)} = \frac{i}{n}$. A MC estimator of Kendall τ^* is

$$\hat{\tau}^* = \frac{1}{n-1}\left(4\sum_{i=1}^n \hat{C}(u_i, v_i) - (n+3)\right). \tag{19}$$

The constants are chosen in such way that $\hat{\tau}^* = \frac{C-D}{C+D}$ where C is the number of concordant pairs and D is the number of discordant pairs.

With computed values for the sample mean $\mu_n = \frac{1}{n}\sum_{i=n}^n \frac{i}{n} = \frac{n+1}{2n}$ and the empirical variance $s_n^2 = \frac{1}{n-1}\sum_{i=n}^n \left(\frac{i}{n} - \mu_n\right)^2 = \frac{n+1}{12n}$ we obtain

$$\hat{\varrho}^*_{\text{MC}} = \frac{1}{ns_n^2}\sum_{i=n}^n (u_i - \mu_n)(v_i - \mu_n) = \frac{12}{n+1}\sum_{i=1}^n \left(u_i - \frac{n+1}{2n}\right)\left(v_i - \frac{n+1}{2n}\right).$$

Using (18) to estimate ϱ^* from (4) gives

$$\hat{\varrho}^* = 12\iint_{[0,1]^2} \hat{C}(s,t)\,\mathrm{d}s\,\mathrm{d}t - 3 = -3 + \frac{12}{n}\sum_{i=1}^n (1 - u_i)(1 - v_i).$$

Following the same line, Gaißer [2010] reports for the square of (6) the so-called Warnock formula,

$$\hat{w}^{*2} = \frac{90}{n^2}\sum_{j=1}^n\sum_{k=1}^n (1 - \max\{u_j, u_k\}) \cdot (1 - \max\{v_j, v_k\}) - \frac{45}{n}\sum_{j=1}^n (1 - u_j^2)(1 - v_j^2) + 10. \tag{20}$$

In Genest and Boies [2003] a graphical method is suggested for the classification of empirical copulas by comparing the area $\mathbb{P}(UV < \omega)$ given by the ω contour of the independent copula with an estimate

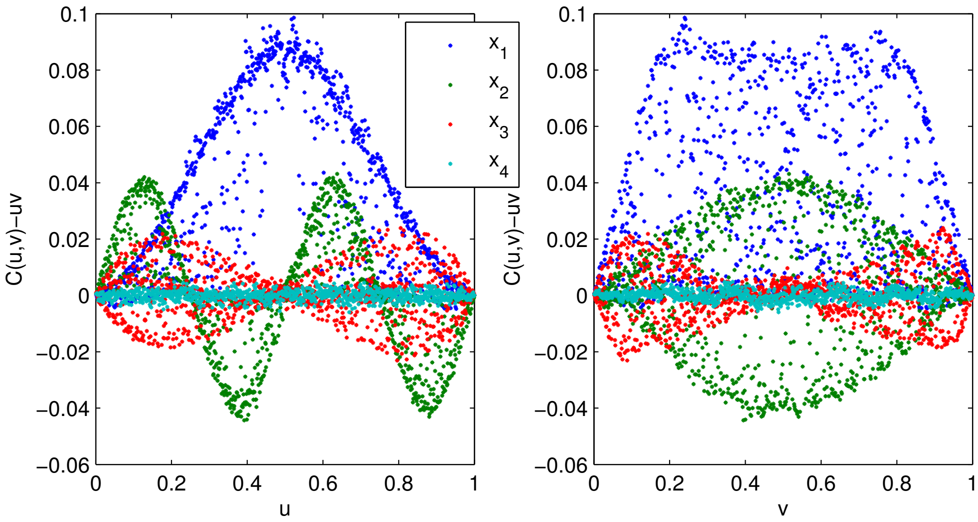

Figure 1: Copula distance plots, 1023 QMC realisations.

from the empirical copula. Here, we choose a different approach for a visualisation technique. In particular, note that the unifying term in the indicators of Spearman (3), Schweizer-Wolff (5), Hoeffding (6) and the discrepancy (7) is $C(u,v)-uv$. In particular, $\hat{C}(u,v)$ is a Monte-Carlo integral counting the relative frequency of hits of the box area $u\cdot v$. Hence comparing the empirical copula and the product copula therefore has an obvious graphical interpretation.

Given a sample of X and Y the associated pairs from the empirical marginal cdfs (ignoring tied data) populate a Latin hypercube. Hence the pairs $(u_i, \hat{C}(u_i,v_i)-u_iv_i)$ and $(v_i, \hat{C}(u_i,v_i)-u_iv_i)$ are graphs of functions φ and ψ mapping $\{\frac{i}{n}; i=1,\ldots,n\}$ into $[-0.25, 0.25]$. The average of (either of) these functions yields $\frac{1}{12}\varrho^*$, the absolute average $\frac{1}{12}s^*$, the absolute maximum $\frac{1}{4}D^*$, etc.

Figure 1 shows these copula distance plots $\varphi: u_i \mapsto \hat{C}(u_i,v_i)-u_iv_i$ and $\psi: v_i \mapsto \hat{C}(u_i,v_i)-u_iv_i$ for the Ishigami test function,

$$Y=\sin X_1+7\sin^2 X_2+0.1X_3^4\sin X_1, \qquad (21)$$

with $X_j \sim U(-\pi,\pi)$ independently distributed. Here parameter 4 is a dummy parameter which does not enter into the calculations. It is immediately clear that Spearman $\varrho^*=12\int_0^1\varphi(u)\mathrm{d}u\approx 0$ for parameters 2, 3 and 4. The discrepancy (7) may also be read directly off the graph using $D^*=4\max_{u=0}^1|\varphi(u)|$, yielding $D_1^*\approx 4\cdot 0.10=0.40$, $D_2^*\approx 4\cdot 0.045=0.18$, $D_3^*\approx 4\cdot 0.025=0.10$ and $D_4^*\approx 0$ for the dummy parameter.

We investigate the relation of the copula distance to the cumulative sum of reordered normalized outputs [Plischke, 2012] which is the absolute concentration curve [Yitzhaki, 2003] of the normalized output $Z=\frac{Y-\mathbb{E}[Y]}{\sqrt{\mathbb{V}[Y]}}$, $\mathrm{ACC}_{X,Z}:\theta\mapsto\int_{-\infty}^{F_X^{-1}(\theta)}\mathbb{E}[Z|X=x]dF_X(x)$. We obtain a copula version by exchanging X and Z with U and $\sqrt{3}(2V-1)$. Hence the empirical rank-based cusunoro curve reduces to

$$\mathrm{csn}^*:\quad [0,1]\to\mathbb{R},\quad u\mapsto\frac{\sqrt{3}}{n}\sum_{i:\,u_i\le u}(2v_i-1)$$
$$=\frac{\sqrt{3}}{n}\sum_{i=1}^n \mathbb{1}\{u_i\le u\}(2v_i-1). \qquad (22)$$

Here u_i controls which of the output realizations are to be considered in the sum. Note that $\int_0^1\hat{C}(u,v)\,\mathrm{d}v=\frac{1}{n}\sum_{i=1}^n\mathbb{1}\{u_i\le u\}(1-v_i)$ and $\frac{1}{n}\sum_{i=1}^n\mathbb{1}\{u_i\le u\}\approx u$ which implies that $\mathrm{csn}^*(u)\approx\sqrt{12}\int_0^1(uv-\hat{C}(u,v))\,\mathrm{d}v$. Hence the cusunoro curve is obtained from the copula distance plot by changing sign and scale and averaging out the dependence on V.

The further examination shows that the cusunoro curve is related to the correlation ratio and therefore to the Kruskal–Wallis test (13) by taking the mean squared gradient. For rank correlation ratios we therefore have $\eta^*=\int_0^1(\frac{d}{du}\mathrm{cusunoro}^*(u))^2\,\mathrm{d}u$. From the discussion following (22) we have the sample-based representation

$$\hat{\eta}^*=3\int_0^1\left(2\frac{d}{du}\int_0^1(uv-\hat{C}(u,v))\,\mathrm{d}v\right)^2\mathrm{d}u \qquad (23)$$

which can be immediately identified with (11).

7 NUMERICAL EXPERIMENTS

Normally, copulas are used to model dependence structures. In this section we investigate if empirical copulas can be used to obtain estimators of the measures discussed above. Starting with a sample $(X,Y)\in\mathbb{R}^{2\times n}$ and passing over to the marginal empirical cumulative distribution functions $(U,V)\in$

$[0,1]^{2\times n}$ (with no ties: $u, v \in \{\frac{1}{n}, \frac{2}{n}, \dots, 1\}$) the empirical copula is given by (18). More advanced methods of estimating $\hat{C}$ are discussed in Schmid et al. [2010]. In order to obtain a copula density estimate we also use the Bernstein copula approach [Sancetta and Satchell, 2004],

$$C_m(u,v) = \sum_{i=1}^{m}\sum_{j=1}^{m} C(\tfrac{i}{m}, \tfrac{j}{m}) P_{mi}(u) P_{mj}(v) \quad (24)$$

with Bernstein polynomials $P_{mk}(t) = \binom{m}{k} t^k (1-t)^{m-k}$, $t \in [0,1]$. If C is a copula then C_m is also copula. However, for the empirical copula $\hat{C}(1,t) \neq t \neq \hat{C}(t,1)$ holds, hence $\hat{C}_m$ is only an approximate copula. The Bernstein conditional copula is $\hat{C}_m(v|u) = \sum_i \sum_j \hat{C}(\frac{i}{m}, \frac{j}{m}) P'_{mi}(u) P_{mj}(v)$ and the Bernstein copula density is given by $\hat{c}_m(u,v) = \sum_i \sum_j \hat{C}(\frac{i}{m}, \frac{j}{m}) P'_{mi}(u) P'_{mj}(v)$. As the Bernstein polynomial base is closed under differentiation, $P'_{mk}(u) = m\left(P_{m-1,k-1}(u) - P_{m-1,k}(u)\right)$ holds. The Bernstein copula density has no singular part. This causes problems if the original copula has a singular density as then the convergence of the Bernstein copula is slow, e.g., to obtain at least ± 0.9 indicator levels for M and W given 4096 simple random samples a polynomial order of $m \approx 200$ is needed. For computing such high polynomial degrees the transformation from the power-basis $\{1, t, t^2, t^3, \dots\}$ is ill-conditioned and it is advantageous to use the recursion formula $P_{mk}(t) = (1-t)P_{m-1,k}(t) + tP_{m-1,k-1}(t)$ known as De Casteljau's algorithm.

We can now compare $\hat{C}$ with the product copula on a regular grid. Figure 2 shows the empirical copulas based upon a 32×32 regular grid, the product copula (in red) and the underlying dependogram (scatterplot of ranks) for all parameters of the Ishigami test function. Parameter 4 is a dummy parameter, note that empirical copula and independent copula are almost coincident. For parameters 2 and 3 Spearman ϱ^* reports 0 in Table 1, η^* sees no functional dependence in parameter 3 while the discrepancy, SW and δ register a dependence for both parameters. Note that for some measures sample based formulas are available so that the discretisation step suggested above is not needed, see Section 6.

For L^p measures of Prop. 1 a direct Monte Carlo estimation on a regular grid $\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\widehat{C}\left(\frac{i}{m}, \frac{j}{m}\right) - \frac{ij}{m^2}\right)^p$ is possible. Here $\widehat{C}(u,v)$ is given by (18). Differentiation of the Bernstein copula yields estimates for η^* and δ.

8 CONCLUSIONS

This paper offers a refreshingly new look on established statistical facts and embeds them in a sensitivity analysis context. Some of the theoretical results and estimators show their usefulness in a SA context. They augment the toolbox of cheap methods/given data methodology for global sensitivity analysis considerably. Table 2 synthesizes our findings. It is clear to see that there is no all-in-one method that provides all information in just one number.

REFERENCES

J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, 32(2):485–498, 1961.

T. Blumentritt and F. Schmid. Mutual information as a measure of multivariate association: analytical properties and statistical estimation. *Journal of Statistical Computation and Simulation*, 82(9): 1257–1274, 2012.

E. Borgonovo, S. Tarantola, E. Plischke, and M. D. Morris. Transformations and invariance in the sensitivity analysis of computer experiments. *Journal of the Royal Statistical Society, Series B*, 2013. DOI:10.1111/rssb.12052.

E. Borgonovo, S. Tarantola, E. Plischke, and M. D. Morris. Transformations and invariance in the sensitivity analysis of computer experiments. *Journal of the Royal Statistical Society, Series B*, 76:925–947, 2014.

L. Devroye and L. Györfi. *Nonparametric Density Estimation: The L^1 View*. John Wiley&Sons, New York, NY, 1985.

M. Fréchet. Sur le coefficient, dit de corrélation et sur la corrélation en géneral. *Revue de l'Institut International de Statistique*, 1(4):16–23, 1934.

S. C. Gaißer. *Statistics for Copula-based Measures of Multivariate Association. Theory and Applications to Financial Data*. PhD thesis, Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität zu Köln, 2010.

C. Genest and J.-C. Boies. Detecting dependence with Kendall plots. *The American Statistician*, 57(4): 275–284, 2003.

W. Hoeffding. A non-parametric test of independence. *The Annals of Mathematical Statistics*, 19 (4):546–557, 1948.

H. Joe. Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, 84(405):157–164, 1989.

J. P. C. Kleijnen and J. C. Helton. Statistical analyses of scatterplots to identify important factors

Table 1: Copula-based sensitivity measures for the Ishigami model. (D) direct, (C) empirical copula, (B) Bernstein estimates with $m = 48$.

Factor	Spearman			Discrepancy		Schweizer–Wolf		Kendall
	D	C	B	C	B	C	B	D
1	0.4391	0.4355	0.4067	0.3759	0.3384	0.4377	0.4078	0.2923
2	0.0008	-0.0024	-0.0047	0.1708	0.1271	0.1773	0.1300	0.0003
3	0.0015	-0.0015	-0.0039	0.0883	0.0722	0.0755	0.0676	0.0006
4	0.0000	-0.0031	-0.0054	0.0160	0.0060	0.0091	0.0059	-0.0001

Factor	Rank Correlation Ratio				Borgonovo			Kuiper
	D	C	B	B	D	D	B	D
1	0.3033	0.3097	0.2579	0.2796	0.25787	0.2630	0.1995	0.26287
2	0.5096	0.4987	0.2843	0.3130	0.43932	0.4059	0.2480	0.44391
3	0.0014	0.0018	0.0001	0.0019	0.22713	0.1813	0.1552	0.21134
4	0.0013	0.0010	0.0000	0.0004	0.08449	0.0684	0.0153	0.10852

Table 2: Synthesis of the results for the discussed copula-based sensitivity methods.

Method	Monotone Dep.	Functional Dep.	Independence	Estimation
Spearman ρ^*	✓			Direct
Rank correlation ratio η^*	✓	✓		Nonlinear regression
Copula distance plot	✓	(✓)	(✓)	Empirical copula
L^p Copula distance	✓		✓	Empirical copula
L^p Copula density distance		(✓)	✓	Bernstein copula density
Moment-independent: δ, β, κ		(✓)	✓	Conditional CDFs/PDFs

in large-scale simulations, 1: Review and comparison of techniques. *Reliability Engineering&System Safety*, 65(2):147–185, 1999.

W. H. Kruskal. Ordinal measures of association. *Journal of the American Statistical Association*, 53 (284):814–861, 1958.

R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer Verlag, 2nd edition, 2006.

R. B. Nelson. Copulas and association. In *Advances in Probability Distributions With Given Marginals: Beyond the Copulas*, pages 51–74. Kluwer Academic Publisher, Dordrecht, 1991.

H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.

K. Pearson. *On the General Theory of Skew Correlation and Non-linear Regression*, volume XIV of *Mathematical Contributions to the Theory of Evolution, Drapers' Company Research Memoirs*. Dulau & Co., London, 1905.

E. Plischke. An adaptive correlation ratio method using the cumulative sum of the reordered output. *Reliability Engineering&System Safety*, 107:149–156, 2012.

E. Plischke and E. Borgonovo. Probabilistic sensitivity measures from cumulative distribution functions. 2015. In preparation.

A. Rényi. On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungaricae*, 10: 441–451, 1959.

A. Saltelli and I. M. Sobol'. About the use of rank transformation in the sensitivity analysis of model output. *Reliability Engineering&System Safety*, 50 (3):225–239, 1995.

A. Saltelli, K. Chan, and E. M. Scott. *Sensitivity Analysis*. John Wiley&Sons, Chichester, 2000.

A. Sancetta and S. Satchell. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20 (3):535–562, 2004.

M. Scarsini. Strong measures of concordance and convergence in probability. *Rivista di matematica per le scienze economiche e sociali*, 7(l–2):39–44, 1984.

F. Schmid, R. Schmidt, T. Blumentritt, S. Gaißer, and M. Ruppert. Copula-based measures of multivariate association. In *Copula Theory and Its Applications*, Lecture Notes in Statistics, pages 209–236. Springer Verlag, Berlin, 2010.

B. Schweizer and E. F. Wolff. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879–885, 1981.

M. D. Taylor. Multivariate measures of concordance. *AISM*, 59:789–806, 2007.

A. van der Linde and G. Tutz. On association in regression: The coefficient of determination revisited. *Statistics*, 42(1):1–24, 2008.

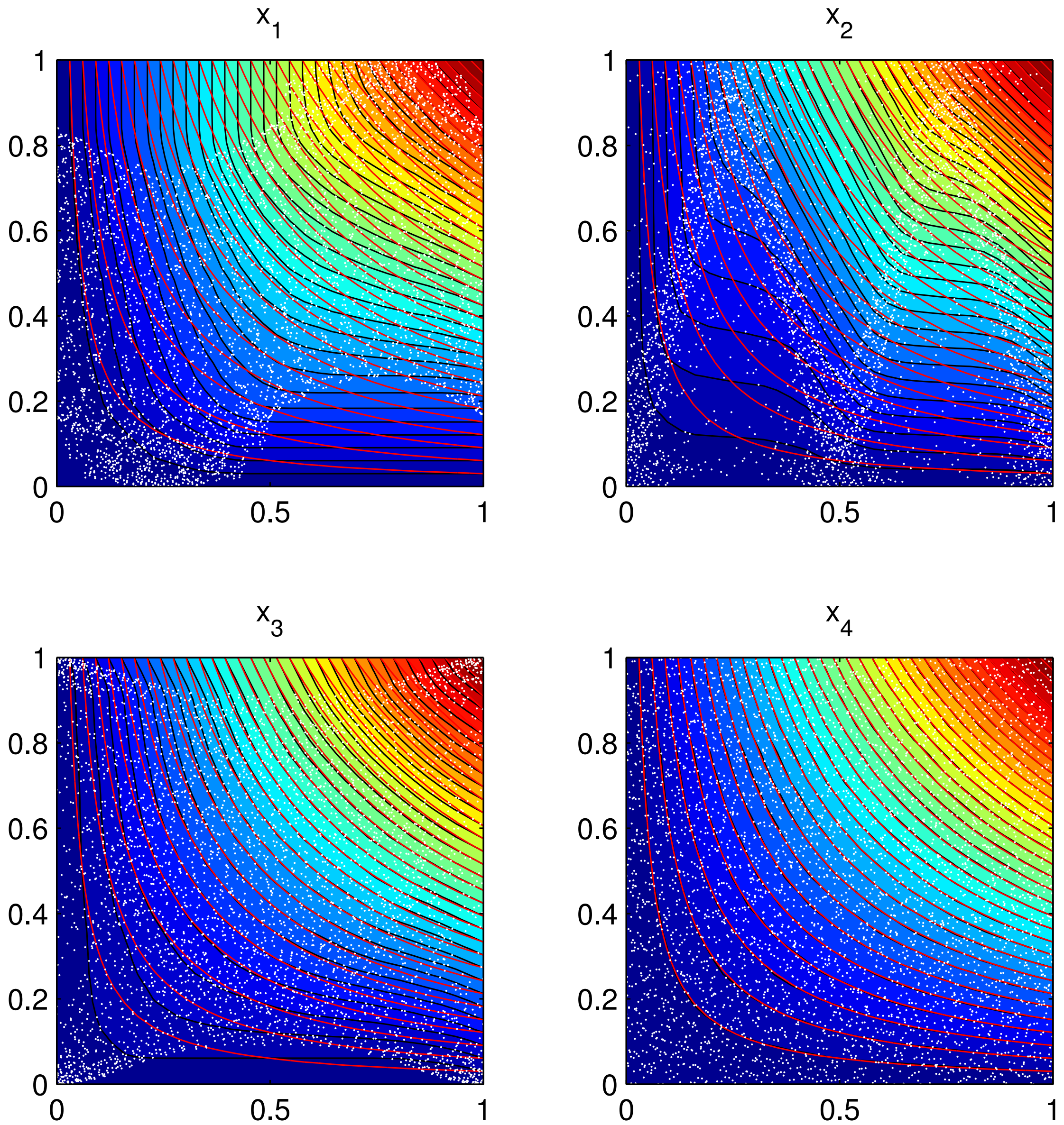

Figure 2: Empirical copulas, 4096 QMC realisations, 32 partitions

P. Wei, Z. Lu, and X. Yuan. Monte Carlo simulation for moment-independent sensitivity analysis. *Reliability Engineering&System Safety*, 110: 60–67, 2013.

P. Wei, Z. Lu, and J. Song. Moment-independent sensitivity analysis using copula. *Risk Analysis*, 34(2): 210–222, 2014.

S. Yitzhaki. Gini's mean difference: a superior measure of variability for non-normal distributions. *Metron*, 61(2):285–316, 2003.