Taylor & Francis
Taylor & Francis Group

Optimal designs for copula models

E. Perrone and W.G. Müller*

Department of Applied Statistics, Johannes Kepler University Linz, 4040 Linz, Austria

Copula modelling has in the past decade become a standard tool in many areas of applied statistics. However, a largely neglected aspect concerns the design of related experiments. Particularly the issue of whether the estimation of copula parameters can be enhanced by optimizing experimental conditions and how robust all the parameter estimates for the model are with respect to the type of copula employed. In this paper an equivalence theorem for (bivariate) copula models is provided that allows formulation of efficient design algorithms and quick checks of whether designs are optimal or at least efficient. Some examples illustrate that in practical situations considerable gains in design efficiency can be achieved. A natural comparison between different copula models with respect to design efficiency is provided as well.

1. Introduction

Due to their flexibility in describing dependencies and the possibility of separating marginal and joint effects copula models have become a popular device for coping with multivariate data. in many areas of applied statistics eg. for insurances,[1] econometrics,[2] medicine,[3] marketing,[4] spatial extreme events,[5] time series analysis,[6] even sports [7] and particularly in finance.[8]

The concept of copulas, however, has only been rarely employed in experimental design with notable exceptions of spatial design in [9,10], and sequential trials in [11]. The design question for copula parameter estimation has to our knowledge just been raised in [12], where a brute-force simulated annealing optimization was employed for the solution of a specific problem. By this paper we provide the necessary theory for fully embedding the situation into optimal design theory. Particularly we provide a Kiefer–Wolfowitz type equivalence theorem [13] in Section 3 as a basis for a substantial analysis of the arising issues in the example sections.

To be more concrete, let us consider a vector $\mathbf{x}^{\mathrm{T}} = (x_1, \ldots, x_r) \in \mathcal{X}$ of control variables, where $\mathcal{X} \subset \mathbb{R}^r$ is a compact set. The results of the observations and of the expectations in a regression

*Corresponding author. Email: werner.mueller@jku.at

experiments are the vectors:

$$\mathbf{y}(\mathbf{x}) = (y_1(\mathbf{x}), \ldots, y_m(\mathbf{x})),$$

$$\mathbf{E}[\mathbf{Y}(x)] = \mathbf{E}[(Y_1, \ldots, Y_m)] = \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\beta}) = (\eta_1(\mathbf{x}, \boldsymbol{\beta}), \ldots, \eta_m(\mathbf{x}, \boldsymbol{\beta})),$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ is a certain unknown vector of marginal parameters to be estimated and η_i $(i = 1, \ldots, m)$ are known functions. Let us call $F_{Y_i}(y_i(\mathbf{x}, \boldsymbol{\beta}))$ the marginal cumulative distributions of each Y_i for all $i = 1, \ldots, m$ and $f_{\mathbf{Y}}(\mathbf{y}(\mathbf{x}, \boldsymbol{\beta}), \boldsymbol{\alpha})$ the joint probability density function of the random vector $\mathbf{Y}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_l)$ are unknown (copula) parameters. In the remainder of the paper we will focus on the case $m = 2$, but generalizations of our results are possible.

DEFINITION 1.1   *Let $\mathcal{I} = [0, 1]$. A two-dimensional copula (or 2-copula) is a bivariate function $C : \mathcal{I} \times \mathcal{I} \longrightarrow \mathcal{I}$ with the following properties:*

(1) *for every $u_1, u_2 \in \mathcal{I}$*

$$C(u_1, 0) = 0, \quad C(u_1, 1) = u_1, \quad C(0, u_2) = 0, \quad C(1, u_2) = u_2; \tag{1}$$

(2) *for every $u_1, u_2, u_3, u_4 \in \mathcal{I}$ such that $u_1 \leq u_3$ and $u_2 \leq u_4$,*

$$C(u_3, u_4) - C(u_3, u_2) - C(u_1, u_4) + C(u_1, u_2) \geq 0.$$

Now let $F_{\mathbf{Y}}$ be a joint cumulative distribution function (cdf) with marginal cdfs F_{Y_1} and F_{Y_2}. According to Sklar's theorem [14] there exists then a 2-copula C such that

$$F_{\mathbf{Y}}(y_1, y_2) = C(F_{Y_1}(y_1), F_{Y_2}(y_2)) \tag{2}$$

for all reals y_1, y_2. If F_{Y_1} and F_{Y_2} are continuous, then C is unique; otherwise, C is uniquely defined on $\text{Ran}(F_{Y_1}) \times \text{Ran}(F_{Y_2})$. Conversely, if C is a 2-copula and F_{Y_1} and F_{Y_2} are distribution functions, then the function $F_{\mathbf{Y}}$ given by Equation (2) is a joint distribution with marginals F_{Y_1} and F_{Y_2}.

2.   Design issues

We need to quantify the amount of information on both (trend and copula) sets of parameters α and β respectively from the regression experiment embodied in the Fisher information matrix, which for an elemental information at a particular control x in the sense of Atkinson et al. [15] is a $(k + l) \times (k + l)$ matrix defined as

$$m(\mathbf{x}, \beta, \alpha) = \begin{pmatrix} m_{\beta\beta}(\mathbf{x}) & m_{\beta\alpha}(\mathbf{x}) \\ m_{\beta\alpha}^{\mathrm{T}}(\mathbf{x}) & m_{\alpha\alpha}(\mathbf{x}) \end{pmatrix}, \tag{3}$$

where the submatrix $m_{\beta\beta}(\mathbf{x})$ is the $(k \times k)$ matrix with the (i, j)th element defined as

$$\mathbf{E}\left(-\frac{\partial^2}{\partial \beta_i \partial \beta_j} \log[f_{\mathbf{Y}}(\mathbf{y}(\mathbf{x}, \boldsymbol{\beta}), \boldsymbol{\alpha})]\right), \tag{4}$$

and the submatrices $m_{\beta\alpha}(\mathbf{x})$ $(k \times l)$ and $m_{\alpha\alpha}(\mathbf{x})$ $(l \times l)$ are defined accordingly. Here we model the dependence between Y_1 and Y_2 with a copula function $C_\alpha(F_{Y_1}(y_1(\mathbf{x}, \boldsymbol{\beta})), F_{Y_2}(y_2(\mathbf{x}, \boldsymbol{\beta})))$ and find the joint density of the random variables from

$$f_{\mathbf{Y}}(\mathbf{y}(\mathbf{x}, \boldsymbol{\beta}), \boldsymbol{\alpha}) = \frac{\partial^2}{\partial y_1 \partial y_2} C_\alpha(F_{Y_1}(y_1(\mathbf{x}, \boldsymbol{\beta})), F_{Y_2}(y_2(\mathbf{x}, \boldsymbol{\beta}))).$$

DEFINITION 2.1 *For a concrete (discrete) experiment with N independent observations at $n \leq N$ support points $\mathbf{x_1}, \ldots, \mathbf{x_n}$, the corresponding total information matrix is*

$$M(\xi, \beta, \alpha) = N^{-1} \sum_{i=1}^{n} w_i m(\mathbf{x_i}, \beta, \alpha), \quad \sum_{i=1}^{n} w_i = 1, \quad \xi = \begin{Bmatrix} \mathbf{x_1} & \cdots & \mathbf{x_n} \\ w_1 & \cdots & w_n \end{Bmatrix},$$

with so-called design weights w_i.

The aim of approximate optimal design theory is concerned with finding an optimal design measure $\xi^*(\beta, \alpha)$, such that it maximizes some scalar function $\phi(M(\xi, \beta, \alpha))$, the so-called design criterion. In the following we will consider only *D-optimality*, that is, the criterion $\phi(M) = \log \det M$, if M is non-singular. There exist several well written monographs on optimal design theory and its application, but in this paper we follow mainly the style and notation of Silvey.[16]

3. Equivalence theory

The cornerstone of a theoretical investigation into optimal design is usually the formulation of a Kiefer–Wolfowitz type equivalence relation, which is given in the following theorem. It is a generalized version of a theorem given without proof in [17] and follows from a multivariate version of the basic theorem given in [16], its full proof can be found in the Appendix.

THEOREM 3.1 *Denote by $(\bar{\beta}, \bar{\alpha})$ fixed values (local guesses) for the parameter vector. Then, the following properties are equivalent:*

- ξ^* *is D-optimal;*
- $\mathrm{tr}[M(\xi^*, \bar{\beta}, \bar{\alpha})^{-1} m(x, \bar{\beta}, \bar{\alpha})] \leq (k + l), \forall x \in \mathcal{X}$;
- ξ^* *minimize $\max_{x \in \mathcal{X}} \mathrm{tr}[M(\xi^*, \bar{\beta}, \bar{\alpha})^{-1} m(x, \bar{\beta}, \bar{\alpha})]$, over all $\xi \in \Xi$.*

This theorem provides simple checks for D-optimality through the maxima of

$$d(x, \xi^*) = \mathrm{tr}[M(\xi^*, \bar{\beta}, \bar{\alpha})^{-1} m(x, \bar{\beta}, \bar{\alpha})],$$

which is usually called *sensitivity function*. It also allows us the use of standard design algorithms such as of the Fedorov-Wynn-type,[18,19] which will yield an optimal approximate design $\xi^* = \begin{Bmatrix} \mathbf{x_1}^* & \cdots & \mathbf{x_n}^* \\ w_1^* & \cdots & w_n^* \end{Bmatrix}$.

Note that these resulting optimal designs will now depend not only upon the marginal model structure, but also upon the chosen copula and through the induced nonlinearities potentially also on the unknown parameter values for α and β, which is why we are resorting to localized designs around the values $(\bar{\beta}, \bar{\alpha})$. A sensitivity analysis with respect to the effect of these choice on a particular example can be found in [20].

DEFINITION 3.2 *For the comparison of designs define D-Efficiency of the design ξ with respect to the design ξ^* as the ratio*

$$\left(\frac{|M(\xi, \bar{\beta}, \bar{\alpha})|}{|M(\xi^*, \bar{\beta}, \bar{\alpha})|} \right)^{1/(k+l)}, \tag{5}$$

where $(k + l)$ is the number of the model parameters. We will report all our findings in percentage losses of these D-efficiencies.

4. Examples

A main question now of course concerns whether ignorance or wrong guesses of copula function and/or parameters may lead to inefficiencies of the designs.

4.1. *Tools*

For that purpose let us here give the list of copulas used in our examples (for more details see, eg. [21] or [22]). We provide the copula function along with the so-called Kendall's τ, which is a dependence measure that allows us to conveniently relate different copulas (for a definition and a more exhaustive comparison see [23]).

DEFINITION 4.1   (1) *Product Copula, which represents the independence case:*

$$C(u_1, u_2) = u_1 u_2,$$

with $\tau = 0$.
(2) *Gaussian Copula:*

$$C_\alpha(u_1, u_2) = \frac{1}{2\Pi\sqrt{1 - \alpha^2}} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left(-\frac{z_1{}^2 - 2\alpha z_1 z_2 + z_2{}^2}{2(1 - \alpha^2)}\right) dz_1\, dz_2,$$

with $\alpha \in [-1, 1]$ and $\tau = (2/\Pi)\arcsin(\alpha)$.
(3) *Farlie–Gumbel–Morgenstern (FGM):*

$$C_\alpha(u_1, u_2) = u_1 u_2[1 + \alpha(1 - u_1)(1 - u_2)],$$

with $\alpha \in [-1, 1]$ and $\tau = \frac{2}{9}\alpha$.
(4) *Clayton:*

$$C_\alpha(u_1, u_2) = [\max(u_1^{-\alpha} + u_2^{-\alpha} - 1, 0)]^{-(1/\alpha)},$$

with $\alpha \in (0, +\infty)$ and $\tau = \alpha/(\alpha + 2)$.
(5) *Frank:*

$$C_\alpha(u_1, u_2) = -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha u_1} - 1)(e^{-\alpha u_2} - 1)}{e^{-\alpha} - 1}\right),$$

with $\alpha \in (-\infty, +\infty)$, and $\tau = 1 - (4/\alpha)(1 - (1/\alpha)\int_0^\alpha (t/(e^t - 1))\, dt)$.
(6) *Gumbel:*

$$C_\alpha(u_1, u_2) = \exp(-[(-\ln u_1)^\alpha + (-\ln u_2)^\alpha]^{1/\alpha}),$$

with $\alpha \in [1, +\infty)$ and $\tau = (\alpha - 1)/\alpha$.

4.2. *The linear case*

Let us first consider a simple example reported in [18]. For each design point $x \in [0, 1]$, we may observe an independent pair of random variables Y_1 and Y_2, such that

$$E[Y_1(x)] = \beta_1 + \beta_2 x + \beta_3 x^2,$$
$$E[Y_2(x)] = \beta_4 x + \beta_5 x^3 + \beta_6 x^4,$$

which is linear in β and has dependence described by the product copula with Gaussian margins. Since this case is covered by Theorem 1, we were able to compute the optimal design ξ^* by a
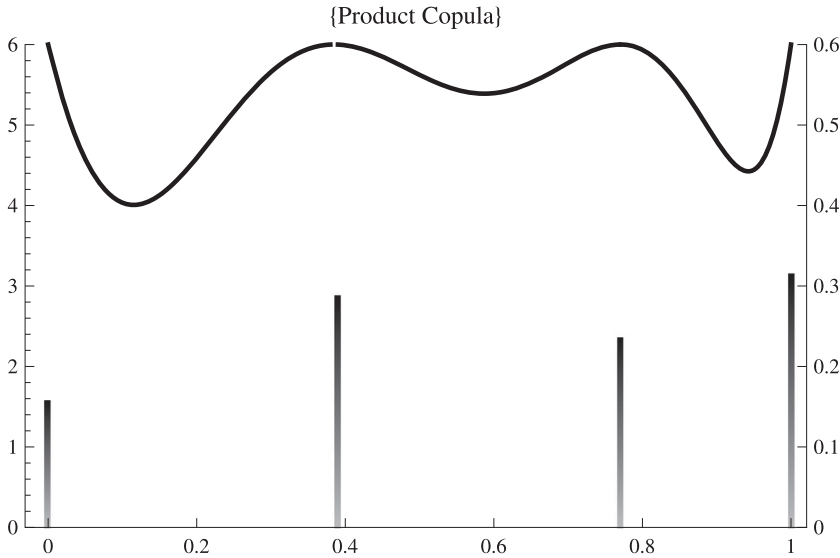
Figure 1.   Sensitivity function (left axis) and optimal design (right axis) for the Fedorov example.

standard algorithm and we display it in Figure 1 along with its sensitivity function. As rather typical ξ^* is supported on only a small number (here four) of design points, smaller than the number of parameters. From the sensitivity function we can see that it is indeed optimum as it reaches (and not exceeds) the number of parameters at all design points. Furthermore our optimal design coincides with the one reported in [18], namely

$$\xi^* = \begin{pmatrix} x_i^* \\ w_i^* \end{pmatrix} = \begin{pmatrix} 0 & 0.38 & 0.76 & 1.0 \\ 0.16 & 0.28 & 0.23 & 0.33 \end{pmatrix}. \tag{6}$$

Let us consider a more general case, for which the joint distribution is described by a Gaussian copula and we thus allow the random variables Y_1 and Y_2 to be dependent. In this case the joint probability function of the random vector $\mathbf{Y} = (Y_1, Y_2)$ is simply

$$\begin{aligned} F_{\mathbf{Y}}(y_1, y_2) &= C_\alpha(\Phi(y_1 - \eta_1(x, \beta)), \Phi(y_2 - \eta_2(x, \beta))) \\ &= \Phi_2(y_1 - \eta_1(x, \beta), y_2 - \eta_2(x, \beta); \alpha), \end{aligned} \tag{7}$$

where $\Phi_2(\cdot, \cdot; \alpha)$ denotes the bivariate normal cdf with correlation $\alpha \in (-1, 1)$ and Φ denotes the cdf of the standard normal distribution $N(0, 1)$ (see [24]).

Our computations gave rise to the following

COROLLARY 4.2   *For different values of α the optimal design is the same as for the independence case, which is the Gaussian case with $\alpha = 0$.*

Note, that the sensitivity function now has a different scaling (with a maximum at 7) as we have an additional copula parameter. This corollary, however, is hardly surprising as this fact coincides with the classic findings for the multivariate Gaussian distribution by Krafft and Schaefer.[25]

Table 1. Losses in D-efficiency (in bold) by ignoring the dependence in per cent.

	FGM		Clayton		Frank	
τ	$\bar{\alpha}$	D-eff	$\bar{\alpha}$	D-eff	$\bar{\alpha}$	D-eff
−0.15	−0.67	**0.29**	n.d.	–	− 1.37	**0.10**
−0.10	−0.45	**0.23**	n.d.	–	− 0.90	**0.10**
−0.05	−0.22	**0.59**	n.d.	–	− 0.45	**0.10**
0.05	0.22	**0.68**	0.10	**0.16**	0.45	**0.10**
0.10	0.45	**0.39**	0.22	**0.13**	0.90	**0.10**
0.15	0.67	**0.28**	0.35	**0.34**	1.37	**0.10**
0.35	n.d.	–	1.08	**0.11**	3.51	**0.11**
0.75	n.d.	–	6.00	**0.27**	14.13	**0.16**

But now for a contrast consider the FGM copula. Following our approach, we must calculate the density corresponding to the function:

$$C_\alpha(\Phi(Y_1(x;\boldsymbol{\beta})), \Phi(Y_2(x;\boldsymbol{\beta}))) = \Phi(Y_1(x;\boldsymbol{\beta}))\Phi(Y_2(x;\boldsymbol{\beta}))$$
$$\times [1 + \alpha(1 - \Phi(Y_1(x;\boldsymbol{\beta})))(1 - \Phi(Y_2(x;\boldsymbol{\beta})))],$$

which eventually leads to expressions like

$$\mathbf{E}\left(-\frac{\partial^2}{\partial\beta_i\partial\beta_j}\log\left[\frac{\partial^2}{\partial y_1\partial y_2}C_\alpha(\Phi(Y_1(x;\boldsymbol{\beta})), \Phi(Y_2(x;\boldsymbol{\beta})))\right]\right)$$

for the information matrix. These integrals are not analytically solvable, but we can evaluate them numerically and we can use the algorithm in order to find the optimum designs.

Not surprisingly those optimal designs do depend upon the choice of $\bar{\alpha}$ (i.e. the assumed dependence) – and similar calculations can be performed for other copula functions as well. Some results are subsumed in Table 1, which displays the loss in D-efficiency that occurs by using the optimal design ξ^* from Equation (6) compared to the respective optimal designs for various copula models and Kendall's τ. It can be seen that these losses are generally quite small for all considered copulas.

4.3. A binary bivariate model

In order to better investigate the role of the copula parameter, we analyse a more elaborate example with potential applications in clinical trials. Let us formally introduce the model. We consider a bivariate binary response (Y_{i1}, Y_{i2}), $i = 1,\ldots,n$ with four possible outcomes $\{(0,0), (0,1), (1,0), (1,1)\}$ where 1 usually represents a success and 0 a failure (of eg. a drug treatment). For a single observation denote the joint probabilities of Y_1 and Y_2 by $p_{y_1,y_2} = \text{pr}(Y_1 = y_1, Y_2 = y_2)$ for $(y_1, y_2 = 0, 1)$. In a clinical trial context Y_1 and Y_2 could represent efficacy and toxicity of a tested drug.

Now, define

$$p_{11} = C_\alpha(\pi_1, \pi_2), \quad p_{10} = \pi_1 - p_{11},$$
$$p_{01} = \pi_2 - p_{11}, \quad p_{00} = 1 - \pi_1 - \pi_2 + p_{11}. \tag{8}$$

The complete log-likelihood for the bivariate binary model is then given by

$$l(\boldsymbol{\theta};\mathbf{y}) = \sum_{i=1}^{n} w_i l_i(\boldsymbol{\theta};\mathbf{y}), \quad \boldsymbol{\theta} = (\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \alpha), \tag{9}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the parameters associated with the respective margins and the log-likelihood for a single observation is given by

$$l_i(\boldsymbol{\theta}; \mathbf{y}) = y_1 y_2 \log p_{11} + y_1(1 - y_2) \log p_{10} + (1 - y_1)y_2 \log p_{01} + (1 - y_1)(1 - y_2) \log p_{00}. \tag{10}$$

As shown in [26] the Fisher information matrix for a single observation can then be written as

$$M(\boldsymbol{\theta}, \xi_i) = \frac{\partial \mathbf{p}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \left(P^{-1} + \frac{1}{1 - p_{11} - p_{10} - p_{01}} \mathbf{e} \mathbf{e}^{\mathrm{T}} \right) \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}}, \tag{11}$$

where $\mathbf{p} = (p_{11}, p_{10}, p_{01})$, $P = \mathrm{diag}(\mathbf{p})$ and $\mathbf{e} = (1, 1, 1)^{\mathrm{T}}$. Some useful formulae for calculating information matrices in copula models can also be found in [27].

A particular case of the introduced model has already been analysed in [17]. In that work, the authors assume the marginal probabilities of success given by the models

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_{i1} + \beta_{i2} x, \quad i = 1, 2 \tag{12}$$

with $x \in [0, 10]$ and 'localized' parameters $\bar{\boldsymbol{\beta}}_1 = [-1, 1]$ and $\bar{\boldsymbol{\beta}}_2 = [-2, 0.5]$. The considered joint cdf is the Gumbel cdf, which corresponds to the following choice for the probability of success p_{11}:

$$p_{11} = F_{Y_1, Y_2}(\pi_1, \pi_2) = \pi_1 \pi_2 (1 + \alpha(1 - \pi_1)(1 - \pi_2)),$$

that is, in terms of copulas, the FGM copula.

In [17], the choice of this model is highlighted by arguing that it allows for dependence between efficacy and toxicity and it is claimed that including estimation of α rather then independently analysing efficacy and toxicity is preferable. We reanalyse this example studying both the role of the copula parameter and the impact of a particular type of dependence structure on the D-optimal designs obtained.

To get a clearer idea of the role played by the copula parameter, let us first focus on the benchmark case of independence, described both by the Product copula and the FGM copula with $\alpha = 0$. However, using these two copulas has substantially different interpretation and effect. In the one case (Product copula) we completely ignore potential dependence, whereas in the other case (FGM copula) we allow for its estimation, but assume it inexistent.

In a second step, to examine the impact of the dependence structure, we compare the model analysed in [17] with the more general ones proposed in [12]. Those have the same assumptions on the marginals probabilities as in [17], as well as same design space and initial parameter for the betas, but the dependencies are instead represented by using the copulas Frank, Gumbel, and Clayton. Note that in [12] the authors employed a brute-force simulated annealing algorithm for their calculations and had no means for checking definitive optimality, which is now possible through the equivalence theorem (Theorem 5.1) provided.

Now, using the D-optimal designs for the FGM copula and for the Product copula as benchmarks, we note the losses in D-efficiency in per cent as reported respectively in Table 2 (Product copula) and in Table 3 (FGM). In both cases, the losses are much stronger than in the previous example.

Analysing these results by focusing on the Frank and the Gumbel copulas, one can notice that lower losses in Table 3 correspond to the lower values of τ. Conversely, the losses already become much higher in Table 3 for a moderate level of the association τ. Moreover, looking at the results for the Clayton copula, we even have lower losses by ignoring the dependence for almost all the levels of the association τ. This suggests, at first glance, that it is not generally

Table 2. Losses in D-efficiency (in bold) by ignoring the dependence in per cent (product copula).

	Frank		Clayton		Gumbel	
τ	$\bar{\alpha}$	D-eff	$\bar{\alpha}$	D-eff	$\bar{\alpha}$	D-eff
0.11	1.00	**1.72**	0.24	**1.75**	1.12	**0.95**
0.45	5.00	**1.31**	1.68	**1.49**	1.84	**1.29**
0.66	10.00	**1.87**	3.98	**0.71**	3.00	**2.31**
0.76	15.00	**2.89**	6.42	**2.84**	4.21	**2.99**
0.82	20.00	**3.10**	8.89	**9.48**	5.45	**3.25**

Table 3. Losses in D-efficiency (in bold) in per cent with respect to the FGM copula ($\alpha = 0$).

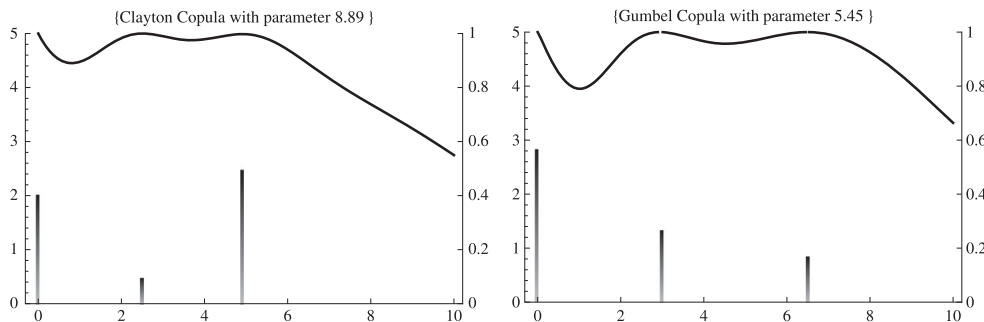|	Frank		Clayton		Gumbel	
τ	$\bar{\alpha}$	D-eff	$\bar{\alpha}$	D-eff	$\bar{\alpha}$	D-eff
0.11	1.00	**0.01**	0.24	**2.42**	1.12	**0.87**
0.45	5.00	**0.36**	1.68	**1.13**	1.84	**0.5**
0.66	10.00	**3.18**	3.98	**1.34**	3.00	**2.84**
0.76	15.00	**5.63**	6.42	**5.54**	4.21	**5.13**
0.82	20.00	**6.24**	8.89	**13.94**	5.45	**6.12**

Figure 2.   The optimal designs and the sensitivity functions for the binary example (Clayton left, Gumbel right). The copula parameters chosen correspond to Kendall's $\tau = 0.816$.

preferable to insert a dependence parameter to be estimated, since this might increase the losses when the model is chosen badly.

Interestingly, if the copula parameter is not estimated, that is, if the model is just a four-parameter model, the optimal designs found are almost the same for all the investigated copulas, and can be represented by

$$\xi^* = \begin{pmatrix} x_i^* \\ w_i^* \end{pmatrix} = \begin{pmatrix} \sim 0 & 2.80 & 6.79 \\ 0.42 & 0.36 & 0.22 \end{pmatrix}.$$

Evidently, the structure of the dependence has an impact as soon as its parameter requires estimation. In Figure 2 we display the designs and sensitivity functions for a representative case contrasting the very different optimal designs for a Clayton and a Gumbel copula with identical Kendall's τ. It will therefore be of great practical value to compare different copulas with respect to their optimal design properties (as well as eventually to be able to efficiently discriminate between them).

A first such step of comparing different copula models was taken in [12] where the authors evaluate designs for various copula choices against each other (in their Table 8). However, they have been using the same parameter values for all the copulas without considering the different

Table 4. Losses in D-efficiency (in bold) by comparing the true copula model with the assumed one for a fixed Kendall's τ.

True Copula	Frank		Clayton		Gumbel	
Assumed Copula	Clayton	Gumbel	Frank	Gumbel	Frank	Clayton
$\tau = 0.11$	**2.24**	**0.67**	**1.99**	**2.70**	**0.82**	**2.75**
$\tau = 0.45$	**0.26**	**0.03**	**0.26**	**0.11**	**0.03**	**0.15**
$\tau = 0.66$	**1.09**	**0.11**	**1.04**	**1.28**	**0.14**	**1.57**
$\tau = 0.76$	**4.27**	**0.02**	**3.87**	**4.08**	**0.01**	**4.73**
$\tau = 0.82$	**8.24**	**0.01**	**10.91**	**10.96**	**0.01**	**8.43**

meaning of the copula parameter for various copula families. Therefore, we instead provide in Table 4 an improved comparison between different dependence structures along the same Kendall's τ values by exploiting the relationship between the copula parameter to the measure of concordance τ. Thanks to this comparison, now the pure impact of the choice of the copula is highlighted. It turns out, that even in extreme case the efficiency losses are only small to moderate. They are greatest if Frank or Gumbel are used instead of Clayton, which may be explained by their opposing representations of tail dependencies.

4.4. *A more flexible model*

Let us now allow the strength of the dependence itself be dependent upon the regressors x, a situation completely covered by our equivalence theorem. Thus here the copula parameters themselves become model-dependent such as, for example, in [28]. As in our context only positive associations (between efficacy and toxicity) make sense we consider in the following the τ modelled by a logistic:

$$\tau(x, \alpha_1) = \frac{e^{\alpha_1 x}}{1 + e^{\alpha_1 x}}, \tag{13}$$

which takes values in $[0, 1]$ for $\alpha_1 \in [0, 1]$.

Considering the Archimedian copulas Clayton C_1 and Gumbel C_2, the following relationships between the Kendall's τ and the copula parameters hold:

$$\tau_{C_1}(x, \alpha_1) = \frac{\gamma_1(x, \alpha_1)}{\gamma_1(x, \alpha_1) + 2} \quad \text{andso } \gamma_1(x, \alpha_1) = \frac{2\tau_{C_1}(x, \alpha_1)}{1 - \tau_{C_1}(x, \alpha_1)} \quad \text{for the Clayton family;}$$

$$\tau_{C_2}(x, \alpha_1) = \frac{\gamma_2(x, \alpha_1) - 1}{\gamma_2(x, \alpha_1)} \quad \text{andso } \gamma_2(x, \alpha_1) = \frac{1}{1 - \tau_{C_2}(x, \alpha_1)} \quad \text{for the Gumbel family.}$$

Then, we model the current probability of success by a convex combination of the Clayton and the Gumbel copulas

$$C(\pi_1, \pi_2; \alpha_1, \alpha_2) = \alpha_2 C_1(\pi_1, \pi_2; \gamma_1(x, \alpha_1)) + (1 - \alpha_2) C_2(\pi_1, \pi_2; \gamma_2(x, \alpha_1)),$$

and when we link them at the same τ values we end up with

$$C(\pi_1, \pi_2; \alpha_1, \alpha_2) = \alpha_2 C_1(\pi_1, \pi_2; 2 e^{\alpha_1 x}) + (1 - \alpha_2) C_2(\pi_1, \pi_2; 1 + e^{\alpha_1 x}).$$

The added flexibility of such a model is that the impact of the dependence structure and the association level is reflected by two different parameters. While the α_2 parameter is strictly related to the structure of the dependence, the α_1 parameter is only related to the measure of association Kendall's τ.

Table 5. Losses in D-efficiency for the convex combination model in per cent (in bold).

	$\tau \in [0.5, 0.95]$	$\tau \in [0.5, 0.99]$	$\tau \in [0.5, 0.995]$
$\bar{\alpha_2}$	Loss in D-eff.	Loss in D-eff.	Loss in D-eff.
0.1	**11.15**	**31.97**	**47.22**
0.3	**6.51**	**25.11**	**40.33**
0.6	**2.50**	**17.20**	**32.27**
0.9	**1.11**	**10.90**	**24.96**

In Table 5 we report the efficiency losses with respect to the independence case (Product Copula). By fixing three localized values $\bar{\alpha_1}$, we assume three different intervals for τ. Corresponding to these intervals we fixed four localized values $\bar{\alpha_2}$. From the table it is clear that when the range of the τ increases, the losses in terms of D-efficiency can become quite substantial. By focusing on the various localized values for α_2, it is evident that also the structure of the dependence plays a big role in the design obtained. Here, we can see that when the highest weight in the convex combination is on the Clayton copula, the efficiency losses are lowest.

5. Discussion

In general, our theory forms the basis to investigate further showcase examples from the literature, like, for example, in [29] or eventually treat mixed discrete/continuous type models like in [30]. Particularly for the latter, but also quite generally the methods provided in this paper can thus be expected to be valuable for real applications from clinical trials, environmental sampling, industrial experiments, etc.

Although here we provide only examples on limited types of copulas, we might expect similar or greater effects for some more non-symmetric copulae (see eg. [31]), which are subject to our current investigations.

Note that in the convex combination example by focusing on α_2 we could find designs with the sole purpose of efficiently discriminating between different copula models, which we plan to do future research on.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

W.G. Müller http://orcid.org/0000-0002-3564-766X

References

[1] Valdez EA. Understanding relationships using copulas. N Am Actuar J. 1998;2(1):1–25.

[2] Trivedi PK, Zimmer DM. Copula modeling: an introduction for practitioners. Found Trends Econ. 2006;1(1):1–111.

[3] Nikoloulopoulos AK, Karlis D. Multivariate logit copula model with an application to dental data. Stat Med. 2008;27(30):6393–6406.

[4] Danaher PJ, Smith MS. Modeling multivariate distributions using copulas: applications in marketing. Mark Sci. 2011;30(1):4–21.

[5] Wadsworth JL, Tawn JA. Dependence modelling for spatial extremes. Biometrika. 2012;99(2):253–272.

[6] Patton AJ. A review of copula models for economic time series. J Multivariate Anal. 2012;110:4–18.

[7] McHale I, Scarf P. Modelling the dependence of goals scored by opposing teams in international soccer matches. Stat Model. 2011;11(3):219–236.

[8] Cherubini U, Luciano E, Vecchiato W. Copula methods in finance. Chichester: Wiley; 2004.

[9] Li J, Bárdossy A, Guenni L, Liu M. A copula based observation network design approach. Environ Model Softw. 2011;26(11):1349–1357.

[10] Pilz J, Kazianka H, Spöck G. Some advances in Bayesian spatial prediction and sampling design. Spat Statist. 2012;1:65–81.

[11] Schmidt R, Faldum A, Witt O, Gerß J. Adaptive designs with arbitrary dependence structure. Biom J. 2014;56(1):86–106.

[12] Denman NG, McGree JM, Eccleston JA, Duffull SB. Design of experiments for bivariate binary responses modelled by Copula functions. Comput Statist Data Anal. 2011;55(4):1509–1520.

[13] Kiefer J, Wolfowitz J. The equivalence of two extremum problems. Canad J Math. 1960;12:363–366.

[14] Sklar A. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de Paris. 1959;8:229–231.

[15] Atkinson AC, Fedorov VV, Herzberg AM, Zhang R. Elemental information matrices and optimal experimental design for generalized regression models. J Statist Plann Inference. 2014;144:81–91.

[16] Silvey SD. Optimal design (science paperbacks). London: Chapman & Hall; 1980.

[17] Heise MA, Myers RH. Optimal designs for bivariate logistic regression. Biometrics. 1996;52(2):613–624.

[18] Fedorov VV. The design of experiments in the multiresponse case. Theory Probab Appl. 1971;16(2):323–332.

[19] Wynn HP. The sequential generation of D-optimum experimental designs. Ann Math Statist. 1970;41(5):1655–1664.

[20] Perrone E. A study on robustness in the optimal design of experiments for copula models. In: Steland A, Rafajłowicz E, Szajowski K, editors. Stochastic models, statistics and their applications. Springer Proceedings in Mathematics & Statistics, Vol. 122. Wrocław, Poland: Springer International Publishing; 2015. p. 335–342.

[21] Nelsen RB. An introduction to copulas (Springer series in statistics). 2nd ed. New York: Springer; 2007.

[22] Durante F, Sempi C. Copula theory: an introduction. In: Bickel P, Diggle P, Fienberg S, Gather U, Olkin I, Zeger S, Jaworski P, Durante F, Härdle WK, Rychlik T, editors. Copula theory and its applications. Lecture Notes in Statistics, chapter 1, Vol. 198. Berlin, Heidelberg: Springer; 2010. p. 3–31.

[23] Michiels F, De Schepper A. A copula test space model: how to avoid the wrong copula choice. Kybernetika. 2008;44(6):864–878.

[24] Meyer C. The bivariate normal copula. Comm Statist Theory Methods. 2013;42(13):2402–2422.

[25] Krafft O, Schaefer M. D-optimal designs for a multivariate regression model. J Multivariate Anal. 1992;42(1):130–140.

[26] Dragalin V, Fedorov V. Adaptive designs for dose-finding based on efficacy–toxicity response. J Statist Plann Inference. 2006;136(6):1800–1823.

[27] Schepsmeier U, Stöber J. Derivatives and Fisher information of bivariate copulas. Stat. Papers. 2014;55(2):525–542.

[28] Noh H, Ghouch AE, Bouezmarni T. Copula-based regression estimation and inference. J Amer Statist Assoc. 2013;108(502):676–688.

[29] Oakes D, Ritz J. Regression in a bivariate copula model. Biometrika. 2000;87(2):345–352.

[30] de Leon AR, Wu B. Copula-based regression models for a bivariate mixed discrete and continuous outcome. Stat Med. 2011;30(2):175–185.

[31] Klement EP, Mesiar R. How non-symmetric can a copula be? Comment Math Univ Carolin. 2006;47(1):141–148.

Appendix

Equivalence theorem

For all the basics in what follows cf.[16] For a given vector of parameters (β, α), let $\mathcal{M}_{(\beta,\alpha)}$ be the set of the information matrices generated as ξ ranges over the class of all set of probability distribution on $\mathcal{X}$. Then $\mathcal{M}_{(\beta,\alpha)}$ is the convex hull of $\{m(x, \beta, \alpha) : x \in \mathcal{X}\}$.

Let us now recall the definition of two derivatives that will play an important role in our theory.

DEFINITION A.1   *Considering two elements M_1 and M_2 in $\mathcal{M}$, the Gâteaux derivative of ϕ at M_1 in the direction of M_2 is:*

$$G_\phi(M_1, M_2) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \{\phi(M_1 + \varepsilon M_2) - \phi(M_1)\},$$

the Fréchet derivative of ϕ at M_1 in the direction of M_2 is:

$$F_\phi(M_1, M_2) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \{\phi\{(1 - \varepsilon)M_1 + \varepsilon M_2\} - \phi(M_1)\}.$$

The following are the properties of the derivatives that we defined before: the concavity of ϕ implies that

$$\frac{1}{\varepsilon}[\phi\{(1 - \varepsilon)M_1 + \varepsilon M_2\} - \phi(M_1)]$$

is a non-increasing function of I_t in $0 < \varepsilon \leq 1$. Hence when ϕ is concave, $F_\phi(M_1, M_2)$ exists if we allow the value $+\infty$.

It is clear that if we put $\varepsilon = 1$ in the previous equation, we obtain: $F_\phi(M_1, M_2) \geq \phi(M_2) + \phi(M_1)$.

According to the definitions of Fréchet and Gâteaux derivatives, we can stress the following relationship between them: $F_\phi(M_1, M_2) = G_\phi(M_1, M_2 - M_1)$. Then, if we assume the differentiability of ϕ it is clear that for scalars a_i

$$F_\phi\left(M_1, \sum a_i M_i\right) = \sum a_i F_\phi(M_1, M_i).$$

THEOREM A.2   *Suppose to have a fixed parameters vector $(\bar{\beta}, \bar{\alpha})$, a concave function ϕ on $\mathcal{M}_{(\bar{\beta},\bar{\alpha})}$ which is also differentiable at all points of $\mathcal{M}_{(\bar{\beta},\bar{\alpha})}$ where $\phi(M) < -\infty$, so where a ϕ optimal measure exists.*
   Then the following are equivalent:

- ξ^* is ϕ-optimal;
- $F_\phi(M(\xi^*, \bar{\beta}, \bar{\alpha}), M(\xi, \bar{\beta}, \bar{\alpha})) \leq 0, \forall \xi \in \Xi$;
- $F_\phi(M(\xi^*, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})) \leq 0, \forall x \in \mathcal{X}$;
- $\max_{x \in \mathcal{X}} F_\phi(M(\xi^*, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})) = \min_{\xi \in \Xi} \max_{x \in \mathcal{X}} F_\phi(M(\xi, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha}))$.

Proof   Let us prove the theorem by double implications.

(i) $\Rightarrow$ (ii) ξ^* is ϕ-optimal. This means that $\phi(M(\xi^*, \bar{\beta}, \bar{\alpha}))$ is maximal.
   For the properties of the function ϕ, the following relation holds:

$$\phi\{(1 - \varepsilon)M(\xi^*, \bar{\beta}, \bar{\alpha}) + \varepsilon M(\xi, \bar{\beta}, \bar{\alpha})\} - \phi\{M(\xi^*, \bar{\beta}, \bar{\alpha})\} \leq 0$$

for $\varepsilon \in [0, 1]$ and all $\xi \in \Xi$.
   For all the elements of $\mathcal{M}_{(\bar{\beta},\bar{\alpha})}$ holds that

$$(1 - \varepsilon)M(\xi^*, \bar{\beta}, \bar{\alpha}) + \varepsilon M(\xi, \bar{\beta}, \bar{\alpha}) = M\{(1 - \varepsilon)\xi^* + \varepsilon \xi\}$$

and this means, from the definition of the Fréchet derivative, that

$$F_\phi\{M(\xi^*, \bar{\beta}, \bar{\alpha}), M(\xi, \bar{\beta}, \bar{\alpha})\} \leq 0$$

for all $\xi \in \Xi$.

(ii) $\Rightarrow$ (iii)   Since $m(x, \bar{\beta}, \bar{\alpha})$ are elements of the convex hull $\mathcal{M}_{(\bar{\beta},\bar{\alpha})}$, the condition (iii) follows directly from the hypothesis.

(iii) $\Rightarrow$ (iv)   For a particular control x and the elemental informations m, since $F_\phi\{M(\xi, \bar{\beta}, \bar{\alpha}), M(\xi, \bar{\beta}, \bar{\alpha})\} = 0$, it must be that:

$$\max_{x \in \mathcal{X}} F_\phi\{M(\xi, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})\} \geq 0$$

But, according to the hypothesis, we have that for the design ξ^*

$$\max_{x \in \mathcal{X}} F_\phi\{M(\xi^*, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})\} \leq 0.$$

Hence

$$\max_{x \in \mathcal{X}} F_\phi\{M(\xi^*, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})\} = 0$$

$$= \min_{\xi} \max_{x \in \mathcal{X}} F_\phi\{M(\xi, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})\}.$$

(iv) $\Rightarrow$ (i) Suppose now that ξ^* satisfies the hypothesis, then

$$\max_{x \in \mathcal{X}} F_\phi \{M(\xi^*, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha}))\} = 0,$$

that means that $F_\phi \{M(\xi^*, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})\} \leq 0, \forall x \in \mathcal{X}$. According to the definition of the matrices $M \in \mathcal{M}$, any M can be written as $M(\xi, \bar{\beta}, \bar{\alpha}) = \sum_{i=n}^{r} w_i m(x_i, \bar{\beta}, \bar{\alpha})$, where $\sum_{i=n}^{r} w_i = 1$ and $w_i > 0$ for every $i = 1, \ldots, n$. Then, since ϕ is differentiable at $M(\xi, \bar{\beta}, \bar{\alpha})$, it holds that:

$$F_\phi \{M(\xi^*, \bar{\beta}, \bar{\alpha}), M(\xi, \bar{\beta}, \bar{\alpha})\} = \sum_{i=1}^{n} \lambda_i F_\phi \{M(\xi^*, \bar{\beta}, \bar{\alpha}), m(x, \bar{\beta}, \bar{\alpha})\} \leq 0$$

for every $\xi \in \Xi$.
This means, clearly, that

$$\phi(M(\xi, \bar{\beta}, \bar{\alpha})) - \phi(M(\xi^*, \bar{\beta}, \bar{\alpha})) \leq 0$$

for every $\xi \in \Xi$, then ξ^* is ϕ-optimal. ∎

D-optimality

Let consider now as design criterion the following function:

$$\phi(M) = \begin{cases} \log \det M & \text{if } M \text{ is non-singular} \\ -\infty & \text{otherwise} \end{cases}$$

A design that maximizes such a ϕ function is called *D-optimal design*.
In the case of D-optimality the Fréchet and the Gâteaux derivatives have the following expression:
Gâteaux derivative

$$\log \det(M_1 + \varepsilon M_2) - \log \det M_1 = \log \det(I + \varepsilon M_2 M_1^{-1})$$

$$= \log\{1 + \varepsilon \operatorname{tr}(M_2 M_1^{-1})\} + O(\varepsilon^2) = \varepsilon \operatorname{tr}(M_2 M_1^{-1}) + O(\varepsilon^2)$$

Hence, $G_\phi(M_1, M_2) = \operatorname{tr}(M_2 M_1^{-1})$.
Fréchet derivative

$$F_\phi(M_1, M_2) = G_\phi(M_1, M_2 - M_1) = \operatorname{tr}((M_2 - M_1) M_1^{-1})$$

$$= \operatorname{tr}(M_2 M_1^{-1} - I_{(k+l)}) = \operatorname{tr}(M_2 M_1^{-1}) - (k + l)$$

where $(k + l)$ is the number of the model parameters.
We are ready now to give an equivalence theorem which holds in the particular case of the D-criterion.

THEOREM A.3   *For a fixed parameters vector $(\bar{\beta}, \bar{\alpha})$, the following properties are equivalent:*

- ξ^* *is D-optimal;*
- $\operatorname{tr}(M(\xi^*, \bar{\beta}, \bar{\alpha})^{-1} m(x, \bar{\beta}, \bar{\alpha})) \leq (k + l), \forall x \in \mathcal{X};$
- ξ^* *minimize* $\max_{x \in \mathcal{X}} \operatorname{tr}(M(\xi^*, \bar{\beta}, \bar{\alpha})^{-1} m(x, \bar{\beta}, \bar{\alpha})),$ *over all $\xi \in \Xi$.*

Proof   The proof comes directly from the Theorem A.1 by imputing the Fréchet derivative for the D-criterion. ∎