



## Strathprints Institutional Repository

**Bedford, Tim and Wilson, Kevin J. (2014) On the construction of minimum information bivariate copula families. *Annals of the Institute of Statistical Mathematics*, 66 (4). pp. 703-723. ISSN 0020-3157 , <http://dx.doi.org/10.1007/s10463-013-0422-0>**

This version is available at <http://strathprints.strath.ac.uk/44658/>

**Strathprints** is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: [strathprints@strath.ac.uk](mailto:strathprints@strath.ac.uk)

# On the Construction of Minimum Information Bivariate Copula Families

Tim Bedford, Kevin J. Wilson  
Department of Management Science, University of Strathclyde,  
Glasgow, G1 1QE, UK

## Abstract

Copulas have become very popular as modelling tools in probability applications. Given a finite number of expectation constraints for functions defined on the unit square, the minimum information copula is that copula which has minimum information (Kullback-Leibler divergence) from the uniform copula. This can be considered the most “independent” copula satisfying the constraints. We demonstrate the existence and uniqueness of such copulas, rigorously establish the relation with discrete approximations, and prove an unexpected relationship between constraint expectation values and the copula density formula. Bivariate copulas information uncertainty modelling expert judgement

## 1 Introduction

Uncertainty distributions are widely used in areas such as Operations Research and finance to represent the uncertainty inherent in any model of the real world. In such distributions it is important to include any dependencies between uncertain quantities. Common methods used to specify uncertainty distributions in the presence of dependency include Bayesian Belief Nets (BBNs) Jen99 and copulas Joe97,Joe11.

More generally, a vine is a nested set of trees in which each tree is made up of a series of bivariate copulas.

Uncertainty distributions can be specified either by fitting to data or by eliciting expert judgement. The data or the experts provide a number of specifications for functions of the problem variables which the uncertainty distribution must satisfy. These are known as constraints. In the case of bivariate copulas for use in vines, specification of the constraints will in general lead to either under- or over-specified distributions. In this paper we investigate the issue of under-specification of a copula. [Bedford *et al* (2013)] show how such methods can be applied in practice. [Bedford (2002)] and [Bedford *et al* (2012)] consider ways to avoid specifications which have no solution.

The approach taken to deal with the issue of under-specification is to use a quantity known as relative information. We seek the copula satisfying the

constraints which has minimum information relative to the uniform copula. This is referred to as the minimum information copula. We show rigorously that this copula exists and is unique, for problems which are not over-specified. Such a copula takes a form similar to that of an exponential family distribution. Minimum information methods have been used previously to specify uncertainty distributions, popularised by [Jaynes (2003)] and considered by [Borwein and Lewis (2006)].

In order to operationalize the use of such minimum information copulas we consider the discretized version of the problem. We show that we can approximate the continuous minimum information copula arbitrarily closely using the discrete copula with maximum entropy relative to the uniform copula. We give a result showing how it is possible to compute the expectations of the constraint functions using the derivative of the natural logarithm of the copula normalising constant.

There has been recent work considering maximum entropy copulas. [Pougaza and Djafari (2011)] considered the construction of maximum entropy copulas when only marginal distributions are specified, under different definitions of entropy. In the case of the Shannon entropy, due to the lack of constraints on the relationships between the variables, the resulting copula is simply the independent copula. [Piantadosi et al (2012)] looked at a simplified class of copulas that the authors called checkerboard. Within this class one has a finite number of regions on which probability is to be uniformly distributed - the problem of dealing with a continuous copula density is changed to one of dealing with a copula having a step function density. This means that one can apply many of the methods for a finite probability space.

The remainder of the paper is organised as follows. In Section 2 we review copulas, information and entropy. In Section 3 we outline the solution to the continuous optimization problem, initially considering the associated measurable optimization problem, and in Section 4 we solve the discretized problem and show that this converges to the continuous solution. We consider the expectations of the constraint functions in relation to the copula normalising constant in Section 5 and give an example of specifying a minimum information copula in two dimensions. We plot the feasible region of combinations of the constraints and indicate how we can parameterize this. In Section 6 we give some conclusions and areas for further work.

## 2 Copulas, Information and Entropy

### 2.1 Copulas

A copula Nels99,Nels06 is the restriction to the unit hypercube of a joint distribution function  $C$  with uniform marginals. In two dimensions, therefore, the copula is  $C(x, y)$ , where  $X, Y \sim U(0, 1)$ . The corresponding joint probability

density function, if it exists, is

$$c(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} C(x, y).$$

Copulas also obey certain properties. From above we see that  $C(0, 0) = 0$  and  $C(1, 1) = 1$ . It is also the case that  $C(x, y) = 0$  if either  $x$  or  $y = 0$ . The fact that each marginal is uniform means that  $C(x, 1) = x$  and  $C(1, y) = y$ . A copula must also be  $n$ -increasing, where  $n$  is the dimension of the copula.

Copulas can be used to define a joint distribution between variables with any marginal distributions we wish. If  $X \sim f_1$  and  $Y \sim f_2$  for densities  $f_1, f_2$ , then the copula of  $X, Y$  is the distribution of  $(F_1(X), F_2(Y))$ . If  $X$  and  $Y$  are independent their unique copula is  $\Pi(x, y) = xy$  whose density is uniform  $\pi(x, y) = 1$  on  $[0, 1]^2$ .

Bivariate copulas can also be used to construct more complex multivariate distributions by combining them in structures known as vines, see [Bedford and Cooke (2002)] and [Kurowicka and Cooke (2006)]. Vines provide a methodology to model any distribution with any general dependencies between the variables we wish.

## 2.2 Relative Information

For a vector quantity  $x$ , the relative information of a distribution  $g_1(x)$  to another  $g_2(x)$  measures the similarity of the two distributions. It is given by

$$I(g_1; g_2) = \int g_1(x) \log \left( \frac{g_1(x)}{g_2(x)} \right) dx.$$

Clearly if  $g_1(x) = g_2(x)$  then  $\log(g_1(x)/g_2(x)) = \log(1)$  and so the relative information of  $g_1(x)$  to  $g_2(x)$  is zero. A useful property of information is that it is invariant under monotone transformations. Thus if  $c_1(x)$  and  $c_2(x)$  are the copula densities associated with  $g_1(x)$  and  $g_2(x)$  respectively then

$$I(c_1; c_2) = I(g_1; g_2).$$

Therefore minimizing the relative information of  $g_1$  with respect to  $g_2$  is equivalent to minimizing the relative information of  $c_1$  with respect to  $c_2$ . This means that minimizing the relative information of a copula with respect to the uniform copula is equivalent to minimizing the relative information of the original density with respect to the independent distribution.

## 2.3 Relative Entropy

Suppose we have a discrete probability distribution over two dimensions defined by  $P(x_i, x_j) = p_{ij}$ . We can define the relative entropy of one distribution to another as we did above for information. The relative entropy of the discrete distribution  $p$  to a second distribution  $q$  is

$$H(p; q) = - \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$

Note that  $H(p; q) \neq H(q; p)$  in general as the above definition is not symmetric in  $p$  and  $q$ . The relative entropy  $H(p; q) \geq 0$ .

The notions of entropy and information are closely linked. First recall that for two partitions  $\rho_1$  and  $\rho_2$  on  $(x_i, y_j)$  for different values of  $n$ ,  $\rho_1$  is a *refinement* of  $\rho_2$  if every element of  $\rho_1$  is a subset of an element of  $\rho_2$ . Relative entropy is non-increasing under refinements Uff95, so that, if  $p(\rho_k)$  is a the probability distribution  $p$  under refinement of  $\rho_k$ ,

$$-\sum_i \sum_j p(\rho_1) \log \frac{p(\rho_1)}{q(\rho_1)} \geq -\sum_i \sum_j p(\rho_2) \log \frac{p(\rho_2)}{q(\rho_2)}.$$

Thus, if  $f, g$  are continuous two dimensional densities being approximated by  $p_{ij}, q_{ij}$  respectively, the continuous relative entropy can be defined Jay03 as the limit under increasing refinement of the discrete relative entropy,

$$\begin{aligned} \lim_{n \rightarrow \infty} -\sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} &= -\int_I dx \int_I f(x, y) \log \frac{f(x, y)}{g(x, y)} dy \\ &= \int_I \int_I \frac{f(x, y)}{g(x, y)} \log \frac{f(x, y)}{g(x, y)} dg(x, y). \end{aligned}$$

Hence we see that the minimum information distribution can be approximated by an equivalent discrete distribution with maximum entropy.

### 3 The Continuous Optimization Problem

Suppose we have uniform variables  $x, y$  and the copula density we wish to find is  $f(x, y)$ . Further suppose that we wish to find a copula which, for some functions of the uniform variables  $h_1(x, y), \dots, h_m(x, y)$  which are assumed to be continuous on  $[0, 1]^2$ , satisfies  $E[h_i(x, y)] = \alpha_i$ , for some values  $\alpha_i$ . We call these the constraints of the problem. If we make the assumption that a copula satisfying the constraints exists then this problem is, in general, under-determined. In order to select a unique distribution we wish to find the copula with minimum information with respect to the uniform copula satisfying these expectations.

The relative information of  $f(x, y)$  with respect to the uniform copula is

$$\int_{[0,1]} dx \int_{[0,1]} f(x, y) \log f(x, y) dy.$$

The requirement that  $f(x, y)$  is to be a copula density introduces the further constraints that the marginal distributions for  $x$  and  $y$  are uniform. That is,

$$\begin{aligned} \forall y \in [0, 1], \quad \int_{[0,1]} f(x, y) dx &= 1, \\ \forall x \in [0, 1], \quad \int_{[0,1]} f(x, y) dy &= 1. \end{aligned}$$

We wish to solve the continuous optimization problem. However, in order to do so, we shall first consider the associated measurable optimization problem. We

can then use this to give a solution in the continuous case. Thus, the measurable optimization problem we wish to solve is

$$\begin{aligned}
& \text{Minimize } \int_{[0,1]} dx \int_{[0,1]} f(x,y) \log(f(x,y)) dy, \\
& \text{Subject to } \int_{[0,1]} f(x,y) dx = 1, \quad a.e. y \in [0,1], \\
& \int_{[0,1]} f(x,y) dy = 1, \quad a.e. x \in [0,1], \\
& \int_{[0,1]} dx \int_{[0,1]} h_i(x,y) f(x,y) dy = \alpha_i, \quad i = 1, \dots, m, \\
& f(x,y) \geq 0, \\
& f(x,y) \in L_1([0,1]^2),
\end{aligned}$$

where a.e. means “for almost every”, with respect to the uniform measure. We shall determine the unique solution to this measurable optimization problem. We do so by generalising the work of [Bedford and Meeuwissen (1997)], who considered just a single constraint on the rank correlation. Their derivation was based on the work in [Nussbaum (1989)] and [Borwein *et al* (1994)].

The continuous problem we shall use to solve this is the measurable optimization problem above but with each “for almost every” replaced with a “for all” and  $f(x,y)$  constrained to being a continuous function rather than in  $L_1$ .

We shall impose a further condition on both the measurable and continuous optimization problems. This will allow us to use theoretical results developed in [Nussbaum (1989)], [Borwein *et al* (1994)] and [Lanford (1973)] and, in practice, does not unduly restrict the use of the developed approach.

The minimum information solution shall require the constraints on the expectations,  $\alpha_1, \dots, \alpha_m$ , to lie in the interior of the convex hull of  $h_1, \dots, h_m$ . In order for this to be true we impose the condition that the constraint functions,  $h_1, \dots, h_m$ , are linearly independent modulo the constants. That is,

$$\sum_i \lambda_i h_i \neq c,$$

for any constant  $c$ . This implies that  $CH(h_1, \dots, h_m)$  has an interior, where  $CH(\cdot)$  denotes convex hull. To see this consider  $CH(\text{Range}(h_1, \dots, h_m))$ . Suppose that  $CH(\text{Range}(h_1, \dots, h_m))$  does not have an interior in  $\mathbb{R}^m$ . Then  $\text{Range}(h_1, \dots, h_m)$  is restricted to a linear subspace of  $\mathbb{R}^m$ , which means there exist  $\lambda_1, \dots, \lambda_m$  and  $c$  such that

$$\sum_i \lambda_i h_i = c,$$

which would mean that  $h_1, \dots, h_m$  were linearly dependent modulo the constraints. This contradicts our condition and so  $CH(h_1, \dots, h_m)$  has an interior.

In order to solve the measurable optimization problem it shall be necessary to define a few quantities starting with the variation distance between two distributions. Let  $G_1, G_2$  and  $G_3$  be probability distributions, with corresponding densities  $g_1, g_2$  and  $g_3$ , on a measurable space  $S = (M, B)$  for the set  $M$  over the  $\sigma$ -algebra  $B$ . Let  $G_1 \ll G_3$  and  $G_2 \ll G_3$ . Then the *variation distance* between  $G_1$  and  $G_2$  is

$$|G_1 - G_2| = \int_M |g_1 - g_2| dG_3.$$

Define  $\Omega$  to be a convex set of probability distributions with respect to  $G_3$ . The set  $\Omega$  is then said to be *variation closed* if  $\Omega$  is closed in the topology of the variation distance.

We are now almost in a position to prove that there is a solution to the measurable optimization problem. First we shall quote a theorem from [Nussbaum (1989)] in the form given in [Bedford and Meeuwissen (1997)]. If  $\Omega$  is variation closed, and if there exists some  $G_1 \in \Omega$  with  $I(G_1; G_3) < \infty$ , then  $\inf_{G_2 \in \Omega} I(G_2; G_3)$  is found in  $\Omega$ .

We can use these results to show that the measurable optimization problem has a solution. There is a solution to the measurable optimization problem. In our case  $\Omega$  is the set of copula densities satisfying the constraints which is clearly convex. If  $\Omega$  is variation closed then, by Theorem 3, the measurable optimization problem has a solution.

Consider  $g^{(n)}(x, y)$ , a sequence of densities converging in variation to  $g(x, y)$ . We show that if  $g^{(n)}(x, y)$  satisfy the constraints then so does  $g(x, y)$ .

As  $g^{(n)}(x, y)$  converge in variation to  $g(x, y)$  this means that

$$\lim_{n \rightarrow \infty} \int dx \int \phi(x, y) g^{(n)}(x, y) dy \rightarrow \int \int \phi(x, y) g(x, y) dy,$$

for any functions  $\phi(x, y) \in L_\infty$ , the dual of  $L_1$ . Clearly,  $g(x, y)$  satisfies the expectation constraints by setting  $\phi(x, y) = h_i(x, y)$  for  $i = 1, \dots, m$ .

Now consider the constraints associated with the uniform marginals for  $X, Y$ . By Fubini's Theorem

$$\int dx \int \phi(x, y) g^{(n)}(x, y) dy = \int dy \int \phi(x, y) g^{(n)}(x, y) dx.$$

We shall consider the marginal constraint for  $Y$ . The argument for  $X$  is similar. Take  $\phi$  of the form  $\phi(x, y) = \phi(x)$ . Thus, if  $g^{(n)}(x, y)$  satisfies the marginal constraint for  $Y$ ,  $\int g^{(n)}(x, y) dy = 1$ , then the following holds,

$$\int dx \int \phi(x) g^{(n)}(x, y) dy = \int \phi(x) \left[ \int g^{(n)}(x, y) dy \right] dx = \int \phi(x) dx.$$

This implies that

$$\int \phi(x) \left[ \int g(x, y) dy \right] dx = \int \phi(x) dx,$$

and so

$$\int \left[ 1 - \int g(x, y) dy \right] \phi(x) dx = 0.$$

From this we can deduce that

$$\int g(x, y) dy = 1$$

almost surely. Thus  $g(x, y)$  satisfies all of the constraints and the measurable optimization problem has a solution by Theorem 3. Now we know that a solution to the measurable optimization problem exists we wish to find a more explicit form. In order to do so we shall consider a property of the dual space of the linear map associated with the constraints given in the below theorem from [Borwein *et al* (1994)].

First, however, let  $\mathcal{P}$  be the support of a probability space,  $Z$  be an arbitrary local convex topological vector space whose topological dual is denoted  $Z^*$  and  $A$  be a linear map such that  $A : L_1(\mathcal{P}) \rightarrow Z$ . The dual of  $A$  is denoted  $A^*$ . Suppose that the information expression we wish to minimize is  $I(u) : L_1(\mathcal{P}) \rightarrow (-\infty, \infty)$  subject to the constraints satisfying the linear map  $Au = b$ .

Using this notation the optimization problem given at the beginning of Section 3 takes the form

$$\text{minimize } I(u) \text{ subject to } Au = b,$$

where

$$\begin{aligned} I(u) &= \int_{[0,1]} dx \int_{[0,1]} u(x, y) \log u(x, y) dy \\ \mathcal{P} &: [0, 1]^2 \\ A &: L_1([0, 1]^2) \rightarrow L_1(0, 1) \times L_1(0, 1) \times \mathbb{R}^m \\ b &= (1, 1, \alpha_1, \dots, \alpha_m) \\ u &\in L_1([0, 1]^2). \end{aligned}$$

We can now express the Theorem from [Borwein *et al* (1994)] as follows. Suppose that a feasible solution to the above problem,  $\hat{u}$ , exists. Then there exists a unique optimal solution,  $u_0$ . Furthermore,  $u_0 > 0$  almost everywhere and there exists a sequence  $\mu_0, \mu_1, \dots \in Z^*$  with

$$\| u_0(A^* \mu_n - \log u_0) \|_1 \rightarrow 0.$$

In order to use this to make explicit statements about the form of the desired densities we shall also require Corollary 2.13 from [Borwein *et al* (1994)]. This is If  $u_0$  is a feasible solution as in the previous Theorem and  $R(A^*)$ , the range of  $A^*$ , is closed as a subspace of  $L_1(\mathcal{P})$  then  $u_0$  is optimal if and only if there exists  $\mu \in Z^*$  with

$$A^* \mu = \log u_0.$$



The linear map associated with our optimisation problem is given by  $A : L_1([0, 1]^2) \rightarrow L_1([0, 1]) \times L_1([0, 1]) \times \mathbb{R} \times \dots \times \mathbb{R}$ . For a two dimensional density  $u(x, y) \in L_1([0, 1]^2)$  the linear constraints in this space form a vector of length  $m + 2$ , the first two elements of which are

$$\int_{[0,1]} u(x, y) dy, \quad \int_{[0,1]} u(x, y) dx.$$

The remaining elements are the constraints on the expectations, namely

$$\int_{[0,1]} dx \int_{[0,1]} u(x, y) h_i(x, y) dy,$$

$i = 1, \dots, m$ . We can now prove the following theorem giving the form of the solution to the measurable optimization problem. The solution,  $f(x, y)$ , to the measurable optimization problem can be written in the form

$$f(x, y) = d^{(1)}(x) d^{(2)}(y) K(x, y),$$

where the kernel is given by

$$K(x, y) = \exp\{\lambda_1 h_1(x, y) + \dots + \lambda_m h_m(x, y)\},$$

for Lagrange multipliers  $\lambda_1, \dots, \lambda_m$  and measurable functions  $d^{(1)}(x), d^{(2)}(y) : [0, 1] \rightarrow \mathbb{R}$ . We can appeal to Theorem 3 to state that there is a sequence of vectors  $\mu_1, \mu_2, \dots$  in the dual of  $A$ , for which

$$\| f(x, y) (A^* \mu_n - \log f(x, y)) \|_1 \rightarrow 0. \quad (1)$$

Thus, in order to find the form of  $f(x, y)$ , we shall calculate the dual  $A^*$ . We determine this by calculating

$$\begin{aligned} \langle u, A^*(a, b, c_1, \dots, c_m) \rangle &= \langle Au, (a, b, c_1, \dots, c_m) \rangle \\ &= \int a(x) \int u(x, y) dy dx \\ &\quad + \int b(y) \int u(x, y) dx dy \\ &\quad + \sum_i c_i \int \int h_i(x, y) u(x, y) dx dy \\ &= \int dx \int u(x, y) \left[ a(x) + b(y) + \sum_i c_i h_i(x, y) \right] dy, \end{aligned}$$

which, by Fubini's Theorem, gives

$$(A^*(a, b, c_1, \dots, c_m))(x, y) = a(x) + b(y) + \sum_{i=1}^m c_i h_i(x, y),$$

for all  $(u, a, b, c_1, \dots, c_m)$ , and a.e.  $(x, y)$ . Combining this with Eq. (1) gives us an equation which indicates how sequences of this form converge to the desired density,

$$a_n(x) + b_n(y) + \sum_{i=1}^k c_{i,n} h_i(x, y) \rightarrow \log f(x, y),$$

almost everywhere. We now wish to use Corollary 3 to make the link between these asymptotic multipliers and multipliers  $a, b, c_1, \dots, c_k$ . In order to do so we need to show that  $R(A^*)$  is closed. Consider the space associated with  $A$ ,

$$L_1([0, 1]) \times L_1([0, 1]) \times \mathbb{R} \times \dots \times \mathbb{R}.$$

This is finite dimensional and hence  $R(A^*)$  is closed. Thus, by Corollary 3,

$$a(x) + b(y) + \sum_{i=1}^k c_i h_i(x, y) = \log f(x, y),$$

almost everywhere. If we take  $d^{(1)}(x) = e^{a(x)}$ ,  $d^{(2)}(y) = e^{b(y)}$  and  $\lambda_i = c_i$  then rearranging gives

$$f(x, y) = d^{(1)}(x) d^{(2)}(y) K(x, y).$$

This concludes the proof. Now that we have an explicit form for the solution to the measurable optimization problem we can use this to give the solution to the continuous optimization problem. There is a unique solution to the continuous optimization problem of the form

$$f(x, y) = d^{(1)}(x) d^{(2)}(y) K(x, y),$$

where  $d^{(1)}(x), d^{(2)}(y) : [0, 1] \rightarrow \mathbb{R}$  are continuous. The proof follows one of Nussbaum in [Nussbaum (1989)]. First we show that  $d^{(1)}(x)$  and  $d^{(2)}(y)$  are in  $L_1([0, 1])$  and then use this fact to determine that they are continuous.

We know that  $f(x, y)$  is measurable and in  $L_1([0, 1]^2)$  and that  $K(x, y)$  is bounded. This implies that the product  $d^{(1)} d^{(2)}$  belongs to  $L_1([0, 1]^2)$ , and so individual functions  $d^{(1)}, d^{(2)}$  belong to  $L_1([0, 1])$ . Now consider the marginal constraint on  $Y$ . This is

$$\int_{[0,1]} d^{(1)}(x) d^{(2)}(y) K(x, y) dy = 1,$$

and so

$$\frac{1}{d^{(1)}(x)} = \int_{[0,1]} d^{(2)}(y) K(x, y) dy.$$

Now, as  $d^{(2)}(y)$  belongs to  $L_1([0, 1])$  and  $K(x, y)$  is continuous this implies that  $1/d^{(1)}(x)$  is continuous. Combining this with the fact that  $1/d^{(1)}(x) > 0$ , we can conclude that  $d^{(1)}(x)$  exists and is continuous.

The same argument can be applied to the marginal constraint for  $X$  to show that  $d^{(2)}(y)$  is continuous.

## 4 Approximation Using Discrete Densities

Suppose that the input space has been discretized into the points  $(x_i, y_j)$  for  $i, j = 1, \dots, n$ , so that the different combinations of  $i, j$  make up points on the unit square. This forms a partition of  $[0, 1]^2$ .

The solution to the continuous minimum information problem can then be approximated by the distribution,  $P(x_i, y_j) = p_{ij}$ , which maximises the Shannon entropy

$$-\sum_i \sum_j p_{ij} \log p_{ij},$$

and which satisfies the constraints on the marginal distributions and expectations. The requirement of uniform marginals in the discrete case brings about the constraints

$$\sum_{i=1}^n p_{ij} = \frac{1}{n}, \quad \sum_{j=1}^n p_{ij} = \frac{1}{n},$$

so that all of the rows and columns sum to one. Thus the discrete optimization problem we wish to solve is

$$\begin{aligned} & \text{maximise } -\sum_{i=1}^n \sum_{j=1}^n p_{ij} \log(p_{ij}), \\ & \text{subject to } \sum_{j=1}^n p_{ij} = \frac{1}{n}, \quad i = 1, \dots, n, \\ & \quad \sum_{i=1}^n p_{ij} = \frac{1}{n}, \quad j = 1, \dots, n, \\ & \quad \sum_{i=1}^n \sum_{j=1}^n h_l(x_i, y_j) p_{ij} = \alpha_l, \quad l = 1, \dots, m, \\ & \quad \text{and } p_{ij} \geq 0. \end{aligned}$$

Having moved from the continuous to the discrete case it is no longer the case that  $\alpha_1, \dots, \alpha_m$  are necessarily in the convex hull for the discrete problem. It is also no longer necessarily the case that  $h_1, \dots, h_m$  are linearly independent modulo the constants for the discrete problem. Thus, in order to make the link between the continuous and discrete problems we provide the following two propositions.

Write  $R = (h_1, \dots, h_m)(\Omega)$  and  $R_n = (h_1, \dots, h_m)(\Omega_n)$  for the ranges of  $(h_1, \dots, h_m)$  in the continuous and discrete cases respectively. We also write  $CH$  to be the convex hull of  $R$ , and  $CH_n$  the convex hull of  $R_n$ . Also define  $\alpha = (\alpha_1, \dots, \alpha_m)$ .

If  $\alpha \in \text{int}(CH)$  then for all large enough  $n$ ,  $\alpha \in \text{int}(CH_n)$ .

The first step is to show that in a small neighbourhood of  $\alpha$ , all the points can be obtained as a convex combination of a fixed finite collection of points in  $R$ . To see this, note that if  $\alpha \in \text{int}(CH)$  then we can find  $m+1$  points  $\alpha_1, \dots, \alpha_{m+1} \in$

$\text{int}(CH)$  close to  $\alpha$  so that  $\alpha$  is in the convex hull of  $\alpha_1, \dots, \alpha_{m+1}$ . But each of these points is in  $CH$  so can be written as a convex combination of  $m+1$  points in  $R$ . Hence any point in a small neighbourhood of  $\alpha$  can be written as a convex combination of the set  $B$  of  $(m+1)^2$  points from  $R$ .

The next step is to show that each point of  $B$  can be arbitrarily well approximated by points from  $R_n$ . Given  $\beta \in B$  there is a point  $(x, y) \in \Omega$  so that  $(h_1, \dots, h_m)(x, y) = \beta$ . By taking a sequence  $(x_n, y_n) \in \Omega$  converging to  $(x, y)$ , and using continuity of the  $(h_1, \dots, h_m)$  we see that  $(h_1, \dots, h_m)(x_n, y_n)$  is a sequence in  $R_n$  converging to  $\beta$ . Define  $B_n$  to be the set of points constructed in this way at the  $n$ th step.

It now follows that the convex hull of  $B_n$  converges to that of  $B$  and, in particular, that it contains any given small neighbourhood  $U$  of  $\alpha$  when  $n$  is large enough. Hence  $U$  is also contained in  $CH_n$ , which implies that  $\alpha \in \text{int}(CH_n)$  for large enough  $n$ .

If  $h_1, \dots, h_m$  are linearly independent, modulo the constants as functions on  $\Omega$ , then they also have that property as functions on  $\Omega_n$  for large enough  $n$ .

If not then there is a set of constants  $c_{1,n}, \dots, c_{m,n}$  and  $c_{0,n}$  (not all zero) so that

$$\sum_i c_{i,n} h_i(x, y) + c_{0,n} = 0,$$

for all  $(x, y) \in \Omega_n$ .

Without loss of generality we can assume that the constants are normalized so that  $\sum_{i=0}^m c_{i,n}^2 = 1$ . This means that there is a subsequence along which the  $c_{i,n}$  simultaneously converge ( $i = 0, \dots, m$ ), with  $c_{i,n} \rightarrow c_i$  say.

This implies that

$$\sum_i c_i h_i(x, y) + c_0 = 0,$$

for all  $(x, y) \in \Omega_n$ , and therefore by continuity that it also holds for all  $(x, y) \in \Omega$ .

[Lanford (1973)] considered the problem of finding the maximum entropy distribution satisfying the constraints  $E[g_l(x_i, y_j)] = u_l$ , for bounded vector  $\mathbf{g}$ . If we define

$$Z(\boldsymbol{\theta}) = \sum_i \sum_j \exp \left\{ - \sum_k \theta_k g_k(x_i, y_j) \right\},$$

where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$  is the vector of Lagrange multipliers, then the probability distribution with maximum entropy is of the form

$$p_{ij} = \frac{\exp\{\sum_k \theta_k g_k(x_i, y_j)\}}{Z(\boldsymbol{\theta})}. \quad (2)$$

We see that  $Z(\boldsymbol{\theta})$  is the normalising constant constraining the  $p_{ij}$  to sum to one. This  $Z$ -function has useful properties. If we take logs and differentiate

with respect to the Lagrange multipliers we obtain Lan73

$$\begin{aligned}\frac{\partial}{\partial \theta_l} \log Z(\boldsymbol{\theta}) &= -\frac{\sum_i \sum_j g_l(x_i, y_j) \exp\{\sum_k \theta_k g_k(x_i, y_j)\}}{\sum_i \sum_j \exp\{-\sum_k \theta_k g_k(x_i, y_j)\}} \\ &= -E[g_l].\end{aligned}$$

Thus we can find the expectations associated with the constraints easily using the  $Z$ -function. Lemma A4.6 of [Lanford (1973)] gives conditions for when there is a unique vector of parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$  satisfying the constraints which give a maximum entropy distribution of the form of Eq. (2). It states If  $\mathbf{u} = (u_1, \dots, u_m)$  is in the interior of the convex hull of the essential range of  $\mathbf{g}$  then there is a unique  $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{u}) \in \mathbb{R}$  such that

$$\mathbf{u} = -\text{grad}_{\boldsymbol{\theta}}(\log Z(\boldsymbol{\theta})).$$

Lanford gives a further Theorem, A4.7, which gives further results concerning the maximum entropy distribution and which will be useful to us. Let  $\mathbf{g}$  be a bounded measurable function on  $\Omega$  with values in  $\mathbb{R}^t$ ; assume that the components of  $\mathbf{g}$  are linearly independent modulo the constants. For  $\mathbf{u}$  in the interior of the convex hull of the range of  $\mathbf{g}$ , let  $\boldsymbol{\theta}(\mathbf{u})$  be the unique solution of

$$\mathbf{u} = -\text{grad}_{\boldsymbol{\theta}} \log Z(\boldsymbol{\theta}).$$

Then

$$s(\mathbf{g}, \mathbf{u}) = \log Z(\boldsymbol{\theta}(\mathbf{u})) + \mathbf{u} \cdot \boldsymbol{\theta}(\mathbf{u}),$$

where  $s(\mathbf{g}, \mathbf{u})$  is a real-analytic and strictly concave function of  $\mathbf{u}$ . The quantity  $s(\mathbf{g}, \mathbf{u})$  is the maximum possible entropy for a probability vector  $\mathbf{g}$  giving expectation  $\mathbf{u}$ .

We wish to apply the above results in order to solve our discrete optimization problem. To do so we need to show that

- (i) we can represent all of the constraints in our problem, and in particular the constraints on the marginal distributions, as expectations,
- (ii) all of the constraints are linearly independent modulo the constraints, and
- (iii) the expectation vector  $\mathbf{u}$  for our problem is in the interior of the convex hull of the essential range of the relevant  $\mathbf{g}$ .

We now consider each of these conditions in turn.

(i) The constraints on the functions of the discretized variables,  $E[h_l] = \alpha_l$ ,  $l = 1, \dots, m$  are already expressed as expectations and so we need only to consider the constraints on the marginals,

$$\sum_{j=1}^n p_{ij} = \frac{1}{n}, \quad \sum_{i=1}^n p_{ij} = \frac{1}{n}.$$

Define the Kronecker deltas

$$\delta_q^{(r)}(i, j) = \begin{cases} 1, & \text{if } i = q, \\ 0, & \text{if } i \neq q, \end{cases}$$

$$\delta_q^{(c)}(i, j) = \begin{cases} 1, & \text{if } j = q, \\ 0, & \text{if } j \neq q, \end{cases}$$

which indicate whether we are in the  $q$ 'th row and  $q$ 'th column respectively. The marginal constraints are then the expectations of these indicator functions. That is

$$\mathbb{E} \left[ \delta_q^{(r)} \right] = \sum_i \sum_j \delta_q^{(r)}(i, j) p_{i,j} = \frac{1}{n}, \quad \mathbb{E} \left[ \delta_q^{(c)} \right] = \sum_i \sum_j \delta_q^{(c)}(i, j) p_{i,j} = \frac{1}{n},$$

for  $q = 1, \dots, n$ . Thus the set of constraints for the discrete problem can be represented by a vector which has length  $2n + m$ , with the first  $2n$  elements of the expectation vector being  $1/n$  and the final  $m$  being  $\alpha_1, \dots, \alpha_m$ . We shall denote this vector  $\mathbf{u}$ .

(ii) The second condition is that all of the constraints must be independent modulo the constants. This is not the case when we consider the full complement of  $2n + m$  constraints. We can see this as within a column or row the Kronecker deltas will all be zero apart from where  $i = q$  or  $j = q$  respectively. That is,

$$\sum_{i=1}^q \delta_q^{(r)}(i, j) = 1, \quad \sum_{j=1}^q \delta_q^{(c)}(i, j) = 1. \quad (3)$$

However, let us instead consider the  $2n + m - 2$  functions given by  $\delta_q^{(r)}(i, j)$  and  $\delta_q^{(c)}(i, j)$  for  $q = 1, \dots, n - 1$  and  $h_l(x_i, y_j)$  for  $l = 1, \dots, m$ . We no longer have the restriction given in Eq. (3) and all of the constraints are now linearly independent modulo the constants. Thus we redefine  $\mathbf{u}$  to be the reduced vector of expectations associated with these constraints.

(iii) The final condition is to show that the vector  $\mathbf{u}$  is in the interior of the convex hull of  $\mathbf{g}$ , where

$$\mathbf{g} = (\delta_1^{(r)}, \dots, \delta_{n-1}^{(r)}, \delta_1^{(c)}, \dots, \delta_{n-1}^{(c)}, \alpha_1, \dots, \alpha_m).$$

That is, we must show that  $\mathbf{u}$  is in the space of all possible expectation specifications. In fact, the proof of this follows immediately from Proposition 2 of [Bedford and Meeuwissen (1997)] for a single function.

We are now in a position to bring all of the results of this section together and, using Lanford's results, give the form of the discrete copula density which solves the discrete optimization problem. We can also then link this to the solution of the continuous optimization problem. There are functions  $h_l(x_i, y_j)$  and further functions  $d^{(1)}(x_i), d^{(2)}(y_j)$  such that the probability distribution on  $\{(x_i, y_j) : 1 \leq i, j \leq n\}$  with maximum entropy under the constraints

$$\sum_i p_{ij} = \sum_j p_{ij} = \frac{1}{n}, \quad \sum_{i,j} h_l(x_i, y_j) p_{ij} = \alpha_l,$$

for  $l = 1 \dots, m$  has the form

$$p_{ij} = \frac{1}{n^2} d^{(1)}(x_i) d^{(2)}(y_j) \exp \left\{ \sum_k \lambda_k h_k(x_i, y_j) \right\}.$$

These discrete probability distributions converge pointwise to the solution of the continuous optimization problem.

We have shown that conditions (i), (ii) and (iii) are satisfied and so we can apply the results of Lanford to our problem. Inserting our constraints into the form of the maximum entropy distribution given in Eq. (2),

$$p_{ij} \propto \exp \left\{ - \sum_q \left( \theta_q^{(r)} \delta_q^{(r)}(i, j) + \theta_q^{(c)} \delta_q^{(c)}(i, j) \right) - \sum_k \theta_k h_k(x_i, y_j) \right\}.$$

Now, each of the Kronecker deltas will be equal to one exactly once, when  $i = q$  and  $j = q$  respectively, and so the discrete distribution becomes

$$p_{ij} = \frac{e^{-\theta_i^{(r)}} e^{-\theta_j^{(c)}}}{Z(\boldsymbol{\theta})} \exp \left\{ - \sum_k \theta_k h_k(x_i, y_j) \right\}.$$

We see that each  $\theta_l$  is an analytic function of the corresponding  $\alpha_l$ . Define  $\lambda_l = -\theta_l$ . Then, by Theorem 4, we have the relation

$$\lambda_l = -\theta_l = - \frac{\partial s(\mathbf{g}, \mathbf{u})}{\partial \alpha_l}.$$

If we take the derivative of  $\lambda_l$  with respect to  $\alpha_l$ ,

$$\frac{\partial \lambda_l}{\partial \alpha_l} = - \frac{\partial^2 s(\mathbf{g}, \mathbf{u})}{\partial \alpha_l^2} > 0,$$

everywhere as  $s(\mathbf{g}, \mathbf{u})$  is a strictly concave function. Thus  $\lambda_l$  is an analytic function of  $\alpha_l$  and  $\theta_l$  is an analytic function of  $\lambda_l$ . This means we can now write the maximum entropy distribution as

$$p_{ij} = \frac{e^{-\theta_i^{(r)}} e^{-\theta_j^{(c)}}}{Z(\boldsymbol{\theta})} \exp \left\{ \sum_k \lambda_k h_k(x_i, y_j) \right\}.$$

The form of the discrete density follows by setting

$$d^{(1)}(x_i) = \frac{n e^{-\theta_i^{(r)}}}{\sqrt{Z(\boldsymbol{\theta})}}, \quad d^{(2)}(y_j) = \frac{n e^{-\theta_j^{(c)}}}{\sqrt{Z(\boldsymbol{\theta})}}.$$

In order to show that these discrete distributions converge to the continuous distribution consider the sequences of functions  $d_{(t)}^{(1)}(x), d_{(t)}^{(2)}(x) : [0, 1] \rightarrow \mathbb{R}$ ,  $t = 1, 2, \dots$  for fixed  $\lambda_1, \dots, \lambda_m$  given by

$$\begin{aligned} x_i &\mapsto d_{i(t)}^{(1)} \quad \text{for } x_i \in I_i \\ y_j &\mapsto d_{j(t)}^{(2)} \quad \text{for } y_j \in I_j. \end{aligned}$$

The proof of [Nosowad (1966)] states that these sequences converge pointwise to the continuous functions  $d^{(1)}(x)$  and  $d^{(2)}(y)$  in the solution of the continuous optimization problem. The result follows from this.

## 5 Calculating the expectations of the constraint functions

We saw in Section 4 that the normalising constant in the maximum entropy distribution,  $Z(\cdot)$ , has useful properties associated with calculating the means of the constraint functions. This is also true in the continuous case. That is, for the bivariate minimum information distribution  $g(x, y)$  with normalising constant

$$Z(\boldsymbol{\theta}) = \int_{[0,1]} dx \int_{[0,1]} \exp \left\{ - \sum_k \theta_k h_k(x, y) \right\} dy,$$

the expectation of  $h_l(x, y)$  is found to be Kul59,Lan73,

$$-\frac{\partial}{\partial \theta_l} \log Z(\boldsymbol{\theta}).$$

In the case of the minimum information copulas which are our interest in this paper the role of the normalising constant has been fulfilled by  $d^{(1)}(\cdot)$  and  $d^{(2)}(\cdot)$ . We can show how these two quantities can be used in an equivalent manner in order to calculate the expectations of the constraint functions in the copula case. We give the following theorem. For a bivariate minimum information copula  $f : [0, 1]^2 \rightarrow \mathbb{R}$  of the form

$$f(x, y) = d^{(1)}(x)d^{(2)}(y)K(x, y),$$

where  $K(x, y) = \exp \{ \lambda_k \sum_k h_k(x, y) \}$ , the mean of the constraint functions can be calculated as

$$\mathbb{E}[h_l(x, y)] = - \int_{[0,1]} dx \int_{[0,1]} \frac{\partial}{\partial \lambda_l} \log \{ d^{(1)}(x)d^{(2)}(y) \} dy.$$

All integrals in the proof are over  $[0, 1]$ . We begin by considering the marginal constraints. We can use them to deduce that

$$d^{(1)}(x)d^{(2)}(y) = \frac{1}{\int d^{(1)}(x)K(x, y)dx \int d^{(2)}(y)K(x, y)dy}.$$

If we take logarithms then

$$\log \{ d^{(1)}(x)d^{(2)}(y) \} = - \log \int d^{(1)}(x)K(x, y)dx - \log \int d^{(2)}(y)K(x, y)dy.$$



We differentiate this with respect to the Lagrange multipliers. This gives

$$\begin{aligned} \frac{\partial}{\partial \lambda_l} \log d^{(1)}(x)d^{(2)}(y) \\ = -\frac{\frac{\partial}{\partial \lambda_l} \int d^{(1)}(x)K(x,y)dx}{\int d^{(1)}(x)K(x,y)dx} - \frac{\frac{\partial}{\partial \lambda_l} \int d^{(2)}(y)K(x,y)dy}{\int d^{(2)}(y)K(x,y)dy}. \end{aligned} \quad (4)$$

As  $d^{(1)}(x)$ ,  $d^{(2)}(y)$  and  $K(x,y)$  are all functions of the Lagrange multipliers,  $\lambda_1, \dots, \lambda_m$ , it is necessary to evaluate the derivatives above using the product rule. That is, in the case of the first derivative,

$$\begin{aligned} \frac{\partial}{\partial \lambda_l} \left[ \int d^{(1)}(x)K(x,y)dx \right] \\ = \int h_l(x,y)d^{(1)}(x)K(x,y)dx + \int K(x,y)\frac{\partial}{\partial \lambda_l}d^{(1)}(x)dx. \end{aligned} \quad (5)$$

To proceed further we re-express the differential of  $d^{(1)}(x)$  in terms of  $y$ . This gives

$$\begin{aligned} \frac{\partial}{\partial \lambda_l} d^{(1)}(x) &= \frac{\partial}{\partial \lambda_l} \left[ \frac{1}{\int d^{(2)}(y)K(x,y)dy} \right] \\ &= -d^{(1)}(x) \left[ \int h_l(x,y)d^{(1)}(x)d^{(2)}(y)K(x,y)dy \right. \\ &\quad \left. + \int K(x,y)d^{(1)}(x)\frac{\partial}{\partial \lambda_l}d^{(2)}(y)dy \right], \end{aligned}$$

after some simple manipulation. We first substitute this back into Eq. (5) and then substitute this back into Eq. (4). We also tidy the denominators of Eq. (4). The resulting equation, again after some basic calculations, is

$$\begin{aligned} \frac{\partial}{\partial \lambda_l} \log \left[ d^{(1)}(x)d^{(2)}(y) \right] &= - \int h_l(x,y)d^{(1)}(x)d^{(2)}(y)K(x,y)dx \\ &\quad + \int d^{(1)}(x)d^{(2)}(y)K(x,y) \left[ \int h_l(x,y)d^{(1)}(x)d^{(2)}(y)K(x,y)dy \right. \\ &\quad \left. + \int d^{(1)}(x)K(x,y)\frac{\partial}{\partial \lambda_l}d^{(2)}(y)dy \right] dx \\ &\quad - \int h_l(x,y)d^{(1)}(x)d^{(2)}(y)K(x,y)dy - \int d^{(1)}(x)K(x,y)\frac{\partial}{\partial \lambda_l}d^{(2)}(y)dy. \end{aligned}$$

The final stage is to integrate over  $x$  and  $y$ . After some straightforward computations we find that

$$\begin{aligned} \int dx \int \frac{\partial}{\partial \lambda_l} \log [d^{(1)}(x)d^{(2)}(y)] dy &= -\mathbb{E}[h_l(x, y)] \\ &+ \int d^{(1)}(x) \int d^{(2)}(y) K(x, y) dy \left[ \int h_l(x, y) d^{(1)}(x) d^{(2)}(y) K(x, y) dy \right. \\ &\quad \left. + \int d^{(1)}(x) K(x, y) \frac{\partial}{\partial \lambda_l} d^{(2)}(y) dy \right] dx \\ &- \mathbb{E}[h_l(x, y)] - \int dx \int d^{(1)}(x) K(x, y) \frac{\partial}{\partial \lambda_l} d^{(2)}(y) dy, \end{aligned}$$

and, since we can again use the first marginal constraint to cancel the  $d^{(1)}(x)$  outside of the square brackets, the result in the continuous case follows immediately from this.

If the problem is discretized by taking  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$  then we obtain an  $n$  on the denominator when substituting back for  $d^{(1)}(x_i)$  and  $d^{(2)}(y_j)$  in the discrete equivalent of Eq. (4). A similar derivation shows that

$$\mathbb{E}[h_l] = -\frac{1}{n} \sum_i \sum_j \frac{\partial}{\partial \lambda_l} \log d^{(1)}(x_i) d^{(2)}(y_j),$$

where  $\mathbb{E}[h_l]$  is a discrete expectation. We can use this discrete form in the following example.

## 5.1 Example

Suppose that the two unknowns in our analysis are  $X, Y \sim U(0, 1)$  and that we wish to specify the minimum information copula between them subject to the constraints

$$\mathbb{E}[h_1(x, y)] = \alpha_1, \quad \mathbb{E}[h_2(x, y)] = \alpha_2, \quad (6)$$

for constraint functions  $h_1(x, y) = xy$ , and  $h_2(x, y) = xy^2$ . If we wished to specify a copula with minimum information for non-uniform variables  $V, W$  then the relevant constraint functions would be  $h_1(v, w) = vw$  and  $h_2(v, w) = vw^2$  and the copula would be specified using

$$h_1'(x, y) = F_V^{-1}(x)F_W^{-1}(y), \quad h_2'(x, y) = F_V^{-1}(x)[F_W^{-1}(y)]^2,$$

where  $F(\cdot)$  denotes the relevant distribution function.

We can investigate the range of possible values the two constraints in Eq. (6) can take. To do so we use Theorem 5 to calculate all of the expectations. We use the discretized version of the density and discretize over  $50 \times 50$  points.

Clearly when  $\alpha_1$  changes this will have an effect on the values which  $\alpha_2$  can take. Thus we can map out the two dimensional feasible region for the constraints  $(\alpha_1, \alpha_2)$ . This is given in Figure 1.

[Figure 1 about here]

We see from the figure that the feasible region is clearly convex. Let us suppose that the expectations we wish to satisfy are

$$E[XY] = 0.2, E[XY^2] = 0.12.$$

We can find the resulting minimum information copula. The Lagrange multipliers are found to be  $\lambda_1 = -25.489, \lambda_2 = 14.306$ . A plot of the copula with these parameter values is given in Figure 2.

[Figure 2 about here]

The copula which results is a smooth function of  $X$  and  $Y$ .

## 6 Conclusions

We have considered the specification of a copula under a number of constraints on the expectations of functions of the variables. In particular, we have considered the issue of under-specification of such a copula in which there are multiple possible copulas which satisfy all of the constraints imposed. We have operationalized such modelling by proposing to use the copula with minimum information satisfying the constraints.

We have shown that such a problem has a unique solution and found an explicit function form for this. We achieved this by initially deriving the unique solution for the measurable version of the problem. We then considered discretization of the continuous problem so that such a process can be carried out in practice. We showed that we can approximate the continuous minimum information copula arbitrarily closely using a series of discrete densities.

Finally, we proved a theorem linking the expectations of the constraint functions to the normalising functions in the copula case. This generalizes a similar result in the non-copula case.

We are grateful to reviewers and the editor for comments and suggestions which improved the paper.

## References

- [Bedford (2002)] Bedford, T. (2002). Interactive expert assessment of minimally informative copulae. In B. Lindqvist (Ed.) *Proceedings of Mathematical Methods in Reliability 2002*.

- [Bedford and Cooke (2002)] Bedford, T. and Cooke, R.M. (2002). Vines: a new graphical model for dependent random variables. *The Annals of Statistics*, 30, 1031-1068.
- [Bedford *et al* (2013)] Bedford, T., Daneshkhah, A., and Wilson, K.J. (2013). Approximate uncertainty modelling with vine copulas. *Submitted to the European Journal of Operational Research*.
- [Bedford and Meeuwissen (1997)] Bedford, T. and Meeuwissen, A.M.H. (1997). Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, 57, 143-174.
- [Bedford *et al* (2012)] Bedford, T., Wilson, K.J., and Daneshkhah, A. (2012). Uncertainty modelling on coupled models using minimum information methods. In T. Aven (Ed.) *Proceedings of PSAM 11 & ESREL 2012*.
- [Borwein and Lewis (2006)] Borwein, J., and Lewis, A. (2006). *Convex analysis and nonlinear optimization: theory and examples*. Canadian Mathematical Society.
- [Borwein *et al* (1994)] Borwein, J., Lewis, A., and Nussbaum, R. (1994). Entropy minimization, DAD problems, and doubly stochastic kernels. *Journal of Functional Analysis*, 123, 264-307.
- [Jaynes (2003)] Jaynes, E.T. (2003). *Probability theory: the logic of science*. Cambridge: Cambridge.
- [Jensen (1999)] Jensen, F.V. (1999). *An introduction to Bayesian Belief Networks*. London: UCL Press.
- [Joe (1997)] Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman and Hall.
- [Kullback (1959)] Kullback, S. (1959). *Information Theory and Statistics*. New York: John Wiley & Sons.
- [Kurowicka and Cooke (2006)] Kurowicka, D., and Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*. New York: Wiley.
- [Kurowicka and Joe (2011)] Kurowicka, D., and Joe, H. (2011). *Dependence modeling: vine copula handbook*. Singapore: World Scientific.
- [Lanford (1973)] Lanford, O.E. (1973). Entropy and equilibrium states in classical statistical mechanics. In A. Lenard (Ed.) *Statistical Mechanics and Mathematical Problems* (pp. 1-113). New York: Springer.
- [Nelsen (1999)] Nelsen, R. (1999). *An introduction to copulas*. New York: Springer.

- [Nelsen (2006)] Nelsen, R. (2006). *An introduction to copulas: second edition*. New York: Springer.
- [Nosowad (1966)] Nosowad, P. (1966). On the integral equation  $kf = 1/f$  arising from a problem in communication. *Journal of Mathematical Analysis and Applications*, 14, 484-492.
- [Nussbaum (1989)] Nussbaum, R.D. (1989). *Iterated nonlinear maps and Hilbert projective metric, II*. Providence: American Mathematical Society.
- [Piantadosi et al (2012)] Piantadosi, J., Howlett, P., and Borwein, J. (2012). Copulas with maximum entropy *Optimization Letters*, 6, 99-125.
- [Pougaza and Djafari (2011)] Pougaza, D.B., and Djafari, A.M. (2011). Maximum entropy copulas. In A. Mohammad-Djafari, J.-F. Bercher and P. Bessiere (Eds.) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (p.p. 329-336.) New York: American Institute of Physics.
- [Uffink (1995)] Uffink, J. (1995). Can the maximum entropy principle be explained as a consistency requirement? *Studies in the History and Philosophy of Modern Physics*, 26, 223-261.

## Figures

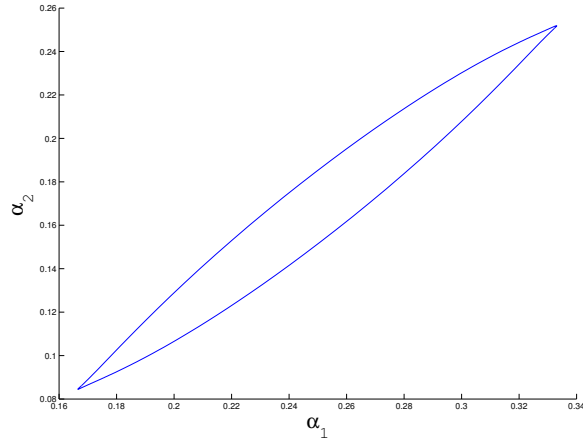


Figure 1: The feasible region for  $\alpha_1$  and  $\alpha_2$ .

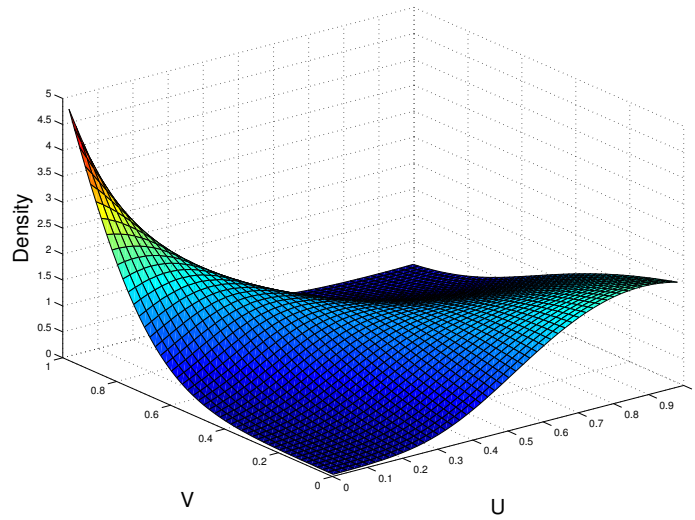


Figure 2: The minimum information copula satisfying the constraints.