Ann Inst Stat Math (2014) 66:703–723
DOI 10.1007/s10463-013-0422-0

On the construction of minimum information bivariate copula families

Tim Bedford · Kevin J. Wilson

Received: 2 October 2012 / Revised: 21 June 2013 / Published online: 20 August 2013

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.

Keywords 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) (Jensen 1999) and copulas (Joe 1997; Kurowicka and Joe 2011).

T. Bedford · K. J. Wilson (✉)
Department of Management Science, University of Strathclyde, Glasgow G1 1QE, UK
e-mail: kevin.j.wilson@strath.ac.uk

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 overspecified distributions. In this paper, we investigate the issue of underspecification 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 underspecification 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 overspecified. 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, popularized by Jaynes (2003) and considered by Borwein and Lewis (2006).

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 normalizing 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 organized as follows. In Sect. 2, we review copulas, information and entropy. In Sect. 3, we outline the solution to the continuous optimization problem, initially considering the associated measurable optimization problem, and in Sect. 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 normalizing constant in Sect. 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 Sect. 6, we give some conclusions and areas for further work.

2 Copulas, information and entropy

2.1 Copulas

A copula (Nelsen 1999, 2006) 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

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

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 ρ_1 and ρ_2 on (x_i, y_j) for different values of n, ρ_1 is a *refinement* of ρ_2 if every element of ρ_1 is a subset of an element of ρ_2. Relative entropy is non-increasing under refinements (Uffink 1995), so that, if $p(\rho_k)$ is a probability distribution p under refinement of ρ_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 (Jaynes 2003) as the limit under increasing refinement of the discrete relative entropy,

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

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 $h_1(x, y), \ldots, h_m(x, y)$ which are assumed to be continuous on $[0, 1]^2$, satisfies $E[h_i(x, y)] = \alpha_i$, for some values α_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, underdetermined. 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

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

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,

$$\forall y \in [0, 1], \int_{[0,1]} f(x, y)\mathrm{d}x = 1,$$
$$\forall x \in [0, 1], \int_{[0,1]} f(x, y)\mathrm{d}y = 1.$$

We wish to solve the continuous optimization problem. However, 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

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

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 generalizing 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, $\alpha_1, \ldots, \alpha_m$, to lie in the interior of the convex hull of $h_1, \ldots, h_m$. In order for this to be true we impose the condition that the constraint functions, $h_1, \ldots, h_m$, are linearly independent modulo the constants. That is,

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

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

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

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

To solve the measurable optimization problem it shall be necessary to define a few quantities starting with the variation distance between two distributions.

Definition 1 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 σ-algebra B. Let $G_1 << G_3$ and $G_2 << G_3$. Then the *variation distance* between G_1 and G_2 is

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

Define Ω to be a convex set of probability distributions with respect to G_3. The set Ω is then said to be *variation closed* if Ω 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).

Theorem 1 *If Ω 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 Ω.*

We can use these results to show that the measurable optimization problem has a solution.

Proposition 1 *There is a solution to the measurable optimization problem.*

Proof In our case Ω is the set of copula densities satisfying the constraints which is clearly convex. If Ω is variation closed then, by Theorem 1, 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

$$\lim_{n\to\infty} \int \mathrm{d}x \int \phi(x, y) g^{(n)}(x, y)\mathrm{d}y \to \int \mathrm{d}x \int \phi(x, y) g(x, y)\mathrm{d}y,$$

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 $i = 1, \ldots, m$.

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

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

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

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

This implies that

$$\int \phi(x) \left[\int g(x, y)\mathrm{d}y\right]\mathrm{d}x = \int \phi(x)\mathrm{d}x,$$

and so

$$\int \left[1 - \int g(x, y)\mathrm{d}y\right] \phi(x)\mathrm{d}x = 0.$$

From this we can deduce that

$$\int g(x, y)\mathrm{d}y = 1$$

almost surely. Thus, $g(x, y)$ satisfies all of the constraints and the measurable optimization problem has a solution by Theorem 1. □

Now we know that a solution to the measurable optimization problem exists we wish to find a more explicit form. 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 $A : L_1(\mathcal{P}) \to Z$. The dual of A is denoted A^*. Suppose that the information expression we wish to minimize is $I(u) : L_1(\mathcal{P}) \to (-\infty, \infty)$ subject to the constraints satisfying the linear map $Au = b$.

Using this notation, the optimization problem given at the beginning of Sect. 3 takes the form

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

where

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

We can now express the Theorem from Borwein et al. (1994) as follows.

Theorem 2 *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 $\mu_0, \mu_1, \ldots \in Z^*$ with*

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

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

Corollary 1 *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 optimization problem is given by $A : L_1([0, 1]^2) \to L_1([0, 1]) \times L_1([0, 1]) \times \mathbb{R} \times \cdots \times \mathbb{R}$. 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

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

The remaining elements are the constraints on the expectations, namely

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

$i = 1, \ldots, m$. We can now prove the following theorem giving the form of the solution to the measurable optimization problem.

Theorem 3 *The solution, $f(x, y)$, to the measurable optimization problem can be written in the form*

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

where the kernel is given by

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

for Lagrange multipliers $\lambda_1 \ldots, \lambda_m$ and measurable functions $\mathrm{d}^{(1)}(x), \mathrm{d}^{(2)}(y) : [0, 1] \to \mathbb{R}$.

Proof We can appeal to Theorem 2 to state that there is a sequence of vectors $\mu_1, \mu_2, \ldots$ in the dual of A, for which

$$\| f(x, y)(A^*\mu_n - \log f(x, y)) \|_1 \to 0. \tag{1}$$

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

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

which, by Fubini's Theorem, gives

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

for all $(u, a, b, c_1, \ldots, c_m)$, 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,

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

almost everywhere. We now wish to use Corollary 1 to make the link between these asymptotic multipliers and multipliers $a, b, c_1, \ldots, c_m$. To do so, we need to show that $R(A^*)$ is closed. Consider the space associated with A,

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

This is finite dimensional and hence $R(A^*)$ is closed. Thus, by Corollary 1,

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

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

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

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.

Theorem 4 *There is a unique solution to the continuous optimization problem of the form*

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

where $\mathrm{d}^{(1)}(x), \mathrm{d}^{(2)}(y) : [0,1] \to \mathbb{R}$ *are continuous.*

Proof The proof follows one of Nussbaum in Nussbaum (1989). First, we show that $\mathrm{d}^{(1)}(x)$ and $\mathrm{d}^{(2)}(y)$ 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 $\mathrm{d}^{(1)}\mathrm{d}^{(2)}$ belongs to $L_1([0,1]^2)$, and so individual functions $\mathrm{d}^{(1)}, \mathrm{d}^{(2)}$ belong to $L_1([0,1])$. Now consider the marginal constraint on Y. This is

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

and so

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

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

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

4 Approximation using discrete densities

Suppose that the input space has been discretized into the points (x_i, y_j) for $i, j = 1, \ldots, n$, 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 maximizes 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

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

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

Write $R = (h_1, \ldots, h_m)(\Omega)$ and $R_n = (h_1, \ldots, h_m)(\Omega_n)$ for the ranges of $(h_1, \ldots, h_m)$ 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 $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$.

Proposition 2 *If* $\boldsymbol{\alpha} \in \text{int(CH)}$ *then for all large enough* n, $\boldsymbol{\alpha} \in \text{int}(\text{CH}_n)$.

Proof The first step is to show that in a small neighbourhood of $\boldsymbol{\alpha}$, 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 $\boldsymbol{\alpha} \in \text{int}(\text{CH})$ then we can find $m+1$ points $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{m+1} \in \text{int}(\text{CH})$ close to $\boldsymbol{\alpha}$ so that $\boldsymbol{\alpha}$ is in the convex hull of $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{m+1}$. 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 $\boldsymbol{\alpha}$ 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 $\boldsymbol{\beta} \in B$ there is a point $(x, y) \in \Omega$ so that $(h_1, \ldots, h_m)(x, y) = \boldsymbol{\beta}$. By taking a sequence $(x_n, y_n) \in \Omega$ converging to (x, y), and using continuity of the $(h_1, \ldots, h_m)$ we see that $(h_1, \ldots, h_m)(x_n, y_n)$ is a sequence in R_n converging to $\boldsymbol{\beta}$. Define B_n to be the set of points constructed in this way at the nth 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 $\boldsymbol{\alpha}$ when n is large enough. Hence, U is also contained in CH_n, which implies that $\boldsymbol{\alpha} \in \text{int}(\text{CH}_n)$ for large enough n. □

Proposition 3 *If $h_1, \ldots, h_m$ are linearly independent, modulo the constants as functions on Ω, then they also have that property as functions on Ω_n for large enough n.*

Proof If not then there is a set of constants $c_{1,n}, \ldots, c_{m,n}$ 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 ($i = 0, \ldots, m$), with $c_{i,n} \to c_i$ 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 $\boldsymbol{g}$. If we define

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

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

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

We see that $Z(\boldsymbol{\theta})$ is the normalizing 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 (Lanford 1973)

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

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 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ satisfying the constraints which give a maximum entropy distribution of the form of Eq. (2). It states

Lemma 1 *If $\boldsymbol{u} = (u_1, \ldots, u_m)$ is in the interior of the convex hull of the essential range of $\mathbf{g}$ then there is a unique $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{u}) \in \mathbb{R}$ such that*

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

Lanford gives a further Theorem, A4.7, which gives further results concerning the maximum entropy distribution and which will be useful to us.

Theorem 5 *Let $\mathbf{g}$ be a bounded measurable function on Ω with values in $\mathbb{R}^t$; assume that the components of $\mathbf{g}$ are linearly independent modulo the constants. For $\boldsymbol{u}$ in the interior of the convex hull of the range of $\mathbf{g}$, let $\boldsymbol{\theta}(\boldsymbol{u})$ be the unique solution of*

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

Then

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

where $s(\boldsymbol{g}, \boldsymbol{u})$ is a real-analytic and strictly concave function of $\boldsymbol{u}$.

The quantity $s(\boldsymbol{g}, \boldsymbol{u})$ is the maximum possible entropy for a probability vector $\boldsymbol{g}$ giving expectation $\boldsymbol{u}$.

We wish to apply the above results 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 $\boldsymbol{u}$ for our problem is in the interior of the convex hull of the essential range of the relevant $\boldsymbol{g}$.

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,

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

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

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

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 $\boldsymbol{u}$.

(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,

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

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 $\boldsymbol{u}$ to be the reduced vector of expectations associated with these constraints.

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

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

That is, we must show that $\boldsymbol{u}$ 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.

Theorem 6 *There are functions $h_l(x_i, y_j)$ and further functions $\mathrm{d}^{(1)}(x_i), \mathrm{d}^{(2)}(y_j)$ 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 $l = 1, \ldots, m$ has the form

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

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

Proof 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 θ_l is an analytic function of the corresponding α_l. Define $\lambda_l = -\theta_l$. Then, by Theorem 5, we have the relation

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

If we take the derivative of λ_l with respect to α_l,

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

everywhere as $s(\boldsymbol{g}, \boldsymbol{u})$ is a strictly concave function. Thus, λ_l is an analytic function of α_l and θ_l is an analytic function of λ_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

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

To show that these discrete distributions converge to the continuous distribution consider the sequences of functions $\mathrm{d}_{(t)}^{(1)}(x), \mathrm{d}_{(t)}^{(2)}(x) : [0, 1] \to \mathbb{R}$, $t = 1, 2, \ldots$ for fixed $\lambda_1, \ldots, \lambda_m$ given by

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

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

5 Calculating the expectations of the constraint functions

We saw in Sect. 4 that the normalizing 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 normalizing constant

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

the expectation of $h_l(x, y)$ is found to be (Kullback 1959; Lanford 1973),

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

In the case of the minimum information copulas which are our interest in this paper, the role of the normalizing constant has been fulfilled by $\mathrm{d}^{(1)}(\cdot)$ and $\mathrm{d}^{(2)}(\cdot)$. We can show how these two quantities can be used in an equivalent manner to calculate the expectations of the constraint functions in the copula case. We give the following theorem.

Theorem 7 *For a bivariate minimum information copula* $f : [0,1]^2 \to \mathbb{R}$ *of the form*

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

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

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

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

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

If we take logarithms then

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where $E[h_l]$ 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

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

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 7 to calculate all of the expectations. We use the discretized version of the density and discretize over 50×50 points.

Clearly when α_1 changes this will have an effect on the values which α_2 can take. Thus, we can map out the two-dimensional feasible region for the constraints (α_1, α_2). This is given in Fig. 1.

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 Fig. 2.

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

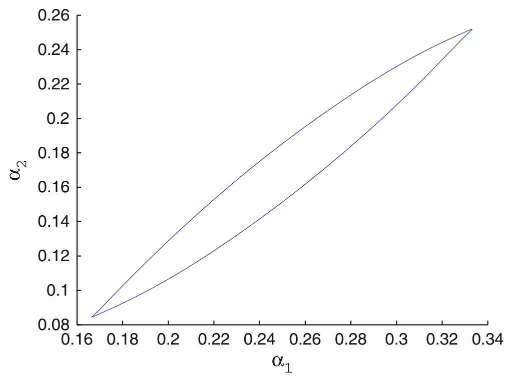

Fig. 1 The feasible region for α_1 and α_2

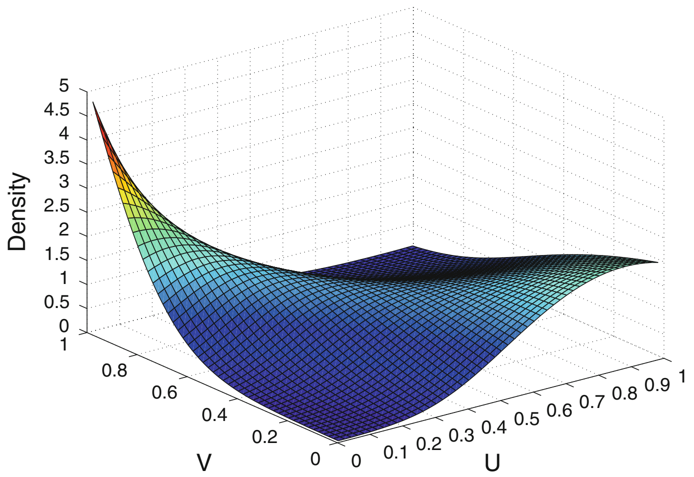

Fig. 2 The minimum information copula satisfying the constraints

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 underspecification 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 normalizing functions in the copula case. This generalizes a similar result in the non-copula case.

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

References

Bedford, T. (2002). Interactive expert assessment of minimally informative copulae. In B. Lindqvist (Ed.) *Proceedings of mathematical methods in reliability*.

Bedford, T., Cooke, R. M. (2002). Vines: a new graphical model for dependent random variables. *The Annals of Statistics*, *30*, 1031–1068.

Bedford, T., Daneshkhah, A., Wilson, K. J. (2013). Approximate uncertainty modelling with vine copulas. *European Journal of Operational Research* (submitted).

Bedford, T., 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, T., Wilson, K.J., Daneshkhah, A. (2012). Uncertainty modelling on coupled models using minimum information methods. In T. Aven (Ed.) *Proceedings of PSAM 11 & ESREL 2012*.
Borwein, J., Lewis, A. (2006). *Convex analysis and nonlinear optimization: theory and examples*. Montreal: Canadian Mathematical Society.
Borwein, J., Lewis, A., Nussbaum, R. (1994). Entropy minimization, DAD problems, and doubly stochastic kernels. *Journal of Functional Analysis*, *123*, 264–307.
Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge: Cambridge.
Jensen, F. V. (1999). *An introduction to Bayesian belief networks*. London: UCL Press.
Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman and Hall.
Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.
Kurowicka, D., Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*. New York: Wiley.
Kurowicka, D., Joe, H. (2011). *Dependence modeling: Vine copula handbook*. Singapore: World Scientific.
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, R. (1999). *An introduction to copulas*. New York: Springer.
Nelsen, R. (2006). *An introduction to copulas* (2nd ed.). New York: Springer.
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, R. D. (1989). *Iterated nonlinear maps and Hilbert projective metric, II*. Providence: American Mathematical Society.
Piantadosi, J., Howlett, P., Borwein, J. (2012). Copulas with maximum entropy. *Optimization Letters*, *6*, 99–125.
Pougaza, D. B., Djafari, A. M. (2011). Maximum entropy copulas. In A. Mohammad-Djafari, J.-F. Bercher, P. Bessiere (Eds.) *Bayesian inference and maximum entropy methods in science and engineering* (pp. 329–336) New York: American Institute of Physics.
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.