Maximum Entropies Copulas
Doriano-Boris Pougaza, Ali Mohammad-Djafari

To cite this version:

HAL Id: hal-00591062
https://hal.archives-ouvertes.fr/hal-00591062
Submitted on 6 May 2011

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

Copulas have been proved useful for modelling the dependence structure between variables in the presence of partial information: the knowledge of marginal distributions. For example, recently we pointed out how to use the notion of copula in tomography [1, 2]. The problem in which we are interested in the present paper is to find the bivariate distribution when we know only its marginals. This problem is an ill-posed inverse problem [3] in the sense that it does not have a unique solution (existence, uniqueness and stability of the solution being the three necessary conditions of well-posedness). One possible way to select a unique solution to this problem is to choose an appropriate copula and then use Sklar’s theorem according to which there exists a copula which relates the marginal distributions yielding to the joint distribution. The problem then becomes the choice of a copula. Note that there are many other ways to derive families of continuous multivariate distributions with given univariate marginals (e.g. [6, 7, 8], and references therein).

Two years before Sklar’s theorem was published, Edwin Jaynes proposed, in two seminal papers [9, 10], the Principle of Maximum Entropy (PME) which defines probability distributions given only partial information. PME has been used in many areas and originally when the partial information is in the form of knowledge of some geometric or harmonic moments (e.g. [11, 12]).

Entropy maximization of a joint distribution subject to given marginals has been studied in statistical and probabilistic literature since the 1930s [13]. The condition for existence of the solution has also been known [14]. This problem was also considered in [15] and [16]. The case where the entropy considered is the Shannon entropy on a measurable space was discussed more rigorously in [17], and this idea was later used in [18], where the authors derive the joint distribution with given uniform marginals on \( I = [0, 1] \) and given correlation.

Here the partial information is the knowledge of the marginal distributions. The
The main result is that we can determine a multivariate distribution with given marginals and which maximizes an entropy. Many types of entropies have been proposed. A consequence of this is that we can now, depending on the entropy expression used, obtain different multivariate distributions, and hence different families of new copulas. The main contribution of this paper is to consider the cases where we can obtain explicit expressions for the maximum entropy problem and so the copula families. To our knowledge, these families have not been discussed before in the literature.

**MAXIMUM ENTROPIES COPULAS**

Denote by $F(x, y)$ an absolutely continuous bivariate cumulative distribution function (cdf), and $f(x, y)$ its bivariate probability density function (pdf). Let $F_1(x), F_2(y)$ be the marginal cdf’s and $f_1(x), f_2(y)$ their respective pdf’s.

A bivariate copula $C$ is a function from $I^2$ to $I$ with the following properties:

1. $\forall u, v \in I, C(u, 0) = 0 = C(0, v)$,
2. $\forall u, v \in I, C(u, 1) = u$ and $C(1, v) = v$ and
3. $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$

for all $u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq u_2$,

One can construct copulas $C$ from joint distribution functions by

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)),$$

where the quantile function is $F_i^{-1}(t) = \inf\{u : F_i(u) \geq t\}$. For further details see [19].

**Problem’s formulation**

In order to find the bivariate maximum entropy pdf $f(x, y)$, the marginal distributions become the constraints:

$$\begin{cases} 
C_1 : \int f(x, y) \, dy = f_1(x), & \forall x \\
C_2 : \int f(x, y) \, dx = f_2(y), & \forall y \\
C_3 : \iint f(x, y) \, dx \, dy = 1.
\end{cases}$$

Hence, the goal is then to find the bivariate density distribution $f(x, y)$, compatible with available information in the PME sense. Among all possible $f(x, y)$ satisfying the constraints (1), PME selects the one which optimizes an entropy $J(f)$, i.e. :

$$\hat{f} := \text{maximize } J(f) \text{ subject to (1)}.$$

Because the constraints are linear, the choice of a concave objective function $J$ guarantees the existence of a unique solution to the problem. Many entropy functionals
can serve as concave objective functions. We focus on the Shannon entropy [20], Rényi entropy [21], Burg entropy [22], and Tsallis-Havrda-Charvát entropy [23, 24] respectively given by:

1. \( J_1(f) = -\int \int f(x,y) \ln f(x,y) \, dx \, dy \) (Shannon);
2. \( J_2(f) = \frac{1}{1-q} \ln \left( \int \int f^q(x,y) \, dx \, dy \right), \quad q \geq 0 \text{ and } q \neq 1 \) (Rényi);
3. \( J_3(f) = \int \int \ln f(x,y) \, dx \, dy \) (Burg);
4. \( J_4(f) = \frac{1}{1-q} \left( 1 - \int \int f^q(x,y) \, dx \, dy \right), \quad q \geq 0 \text{ and } q \neq 1 \) (Tsallis-Havrda-Charvát),

One can get a continuum of entropy measures by choosing different values of parameter \( q \neq 1 \). Shannon entropy is the special limit of \( J_2(f) \) and \( J_4(f) \) as \( q \to 1 \).

**Method and parametric solution**

The main tool is to define the following Lagrange multipliers technique. When solving the Lagrangian functional equation, we assume that there exists only one feasible \( f > 0 \) with finite entropy satisfying

\[
\mathcal{L}_g(f, \lambda_0, \lambda_1, \lambda_2) = J_1(f) + \lambda_0 \left( 1 - \int \int f(x,y) \, dx \, dy \right) + \int \lambda_1(x) \left( f_1(x) - \int f(x,y) \, dy \right) \, dx \\
+ \int \lambda_2(y) \left( f_2(y) - \int f(x,y) \, dx \right) \, dy,
\]

and the critical point of \( \mathcal{L}_g \) holds for the following system of equations:

\[
\frac{\partial \mathcal{L}_g(f, \lambda_0, \lambda_1, \lambda_2)}{\partial f} = 0, \quad \frac{\partial \mathcal{L}_g(f, \lambda_0, \lambda_1, \lambda_2)}{\partial \lambda_i} = 0.
\]

Assuming that the integrals converge within the interval \( I \), this system of equations yields:

\[
f(x,y) = \exp(-\lambda_1(x) - \lambda_2(y) - \lambda_0) \quad \text{(Shannon’s entropy)};
\]

\[
\frac{\int_{\Omega^2} f^{q-1}(x,y) \, dx \, dy}{\int_{\Omega^2} f^q(x,y) \, dx \, dy} = \frac{1-q}{q} (\lambda_1(x) + \lambda_2(y) + \lambda_0) \quad \text{(Rényi’s entropy)};
\]

\[
f(x,y) = (\lambda_1(x) + \lambda_2(y) + \lambda_0)^{-1} \quad \text{(Burg’s entropy)};
\]

\[
f(x,y) = \frac{1-q}{q} (\lambda_1(x) + \lambda_2(y) + \lambda_0)^{\frac{q-1}{q}} \quad \text{(Tsallis-Havrda-Charvát’s entropy)}.
\]
where $\lambda_1(x, y)$ and $\lambda_0$ are obtained by replacing these expressions in the constraints (1) and solving the resulting system of equations.

For the Shannon entropy, the constraints can be solved analytically:

$$
\lambda_1(x) = -\ln \left( f_1(x) \int_I \lambda_1(x) \, dx \right),
$$

$$
\lambda_2(y) = -\ln \left( f_2(y) \int_I \lambda_2(y) \, dx \right)
$$

and

$$
\lambda_0 = \ln \left( \int_I \lambda_1(x) \, dx \int_I \lambda_2(y) \, dy \right),
$$

and the joint distribution becomes

$$
f(x, y) = f_1(x) f_2(y).
$$

Unfortunately, in the cases of Rényi, Burg and Tsallis-Havrda-Charvát entropies, it is not possible to find general solutions for $\lambda_0, \lambda_1, \lambda_2$ as explicit functions of $f_1$ and $f_2$, and numerical approaches become necessary.

**Special case $q=2$**

The special case when Tsallis-Havrda-Charvát’s entropy index $q$ is equal to 2, is known as the Simpson’s diversity index [25]. Here the probability density function has the form $f(x, y) = -\frac{1}{2} (\lambda_1(x) + \lambda_2(y) + \lambda_0)$ and we can obtain explicit expressions for $\lambda_1(x), \lambda_2(x)$ and $\lambda_0$:

$$
\lambda_1(x) = -2 f_1(x) + \int_I \lambda_1(x) \, dx + 2,
$$

$$
\lambda_2(y) = -2 f_2(y) + \int_I \lambda_2(y) \, dy + 2,
$$

$$
\lambda_0 = -2 - \int_I \lambda_1(x) \, dx - \int_I \lambda_2(y) \, dy.
$$

Substituting these expressions gives the following probability density function on the bounded interval $I$ (where $f_1$ and $f_2$ are chosen properly):

$$
f(x, y) = f_1(x) + f_2(y) - 1.
$$

Assuming $\|f\|_2^2 = \int_I \int_I f^2(x, y) \, dx \, dy = 1$, the resulting pdf obtained when maximizing Rényi’s entropy is the same as the pdf (3). General form of the pdf over any bounded interval is obtained by substituting $x$ and $y$ respectively with \(\frac{x-x_{\min}}{x_{\max}-x_{\min}}\) and \(\frac{y-y_{\min}}{y_{\max}-y_{\min}}\).

The multivariate case of (3) over $I^n$ follows

$$
f(x_1, \ldots, x_n) = \sum_{i=1}^n f_i(x_i) - n + 1.
$$
FAMILIES OF COPULAS

With the bivariate density obtained from the maximum entropy principle, we can immediately find the corresponding bivariate copula.

For the case of the Shannon entropy, (2), we have:

$$F(x, y) = \int_0^x \int_0^y f(s, t) \, ds \, dt$$
$$= \int_0^x \int_0^y f_1(s) f_2(t) \, ds \, dt$$
$$= \int_0^x f_1(s) \, ds \int_0^y f_2(t) \, dt.$$  

The cdf becomes

$$F(x, y) = F_1(x) F_2(y),$$

and the copula is

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)) = uv.$$  \hspace{1cm} (5)

The maximum copula obtained from the Shannon entropy is the well-known independent copula which describes independence between two random variables.

In the particular case ($q = 2$) of the Tsallis-Havrda-Charvát entropy, (3)

$$F(x, y) = \int_0^x \int_0^y f(s, t) \, ds \, dt$$
$$= \int_0^x \int_0^y (f_1(s) + f_2(t) - 1) \, ds \, dt$$
$$F(x, y) = y \int_0^x f_1(s) \, ds + x \int_0^y f_2(t) \, dt - xy;$$

with the cdf

$$F(x, y) = y F_1(x) + x F_2(y) - xy, \quad 0 \leq x, y \leq 1$$  \hspace{1cm} (6)

and the associated copula

$$C(u, v) = u F_2^{-1}(v) + v F_1^{-1}(u) - F_1^{-1}(u) F_2^{-1}(v).$$  \hspace{1cm} (7)

In the multivariate case (4), the cdf is :

$$F(x_1, \ldots, x_n) = \int_0^{x_1} \cdots \int_0^{x_n} f(s_1, \ldots, s_n) \prod_{i=1}^n ds_i$$
$$= \int_0^{x_1} \cdots \int_0^{x_n} \left( \sum_{i=1}^n f_i(s_i) - n + 1 \right) \prod_{i=1}^n ds_i$$
$$= \sum_{i=1}^n F_i(x_i) \prod_{j=1}^n x_j + (1 - n) \prod_{i=1}^n x_i, \quad 0 \leq x_i \leq 1$$  \hspace{1cm} (8)
and the associated multivariate copula, depending on $F^{-1}_i$ will have the following form

$$C(u_1, \ldots, u_n) = \sum_{i=1}^n u_i \prod_{j=1 \atop j \neq i}^n F^{-1}_j(u_j) + (1 - n) \prod_{i=1}^n F^{-1}_i(u_i).$$  \hspace{1cm} (9)

One has to verify that (9) satisfies the properties of a copula or equivalently that (8) is a cdf on $\mathbb{I}^n$. The first two properties of copula are easily proven (since $F^{-1}_i(0) = 0$ and $F^{-1}_i(1) = 1$).

**SOME FAMILIES OF COPULAS**

Beta distributions are very interesting and general continuous distribution on the finite interval $[0, 1]$. This is the main reason for choosing this family as a first example for our development. We consider then:

$$f_1(x) = \frac{1}{B(a_1, b_1)} x^{a_1-1}(1-x)^{b_1-1}$$
and
$$f_2(y) = \frac{1}{B(a_2, b_2)} y^{a_2-1}(1-y)^{b_2-1},$$
where

$$B(a_i, b_j) = \int_0^1 t^{a_i-1}(1-t)^{b_j-1} \, dt, \quad 0 \leq x, y \leq 1 \text{ and } a_i, b_j > 0.$$  

We consider the inverse of the Beta cumulative distribution function in some particular and interesting values of the parameters $a_i$ and $b_j$.

**Case 1**: $a_i = 1, b_j = 1$ which corresponds to uniform marginals $f_1$ and $f_2$

$$f_1(x) = 1 \rightarrow F_1(x) = x \rightarrow F_1^{-1}(u) = u$$
$$f_2(y) = 1 \rightarrow F_2(y) = y \rightarrow F_2^{-1}(v) = v$$
$$F(x, y) = xy,$$
gives the well known independent copula:

$$C(u, v) = uv.$$  \hspace{1cm} (10)

**Case 2**: $a_i > 0, b_j = 1$

$$f_1(x) = a_1 x^{a_1-1} \rightarrow F_1(x) = x^{a_1} \rightarrow F_1^{-1}(u) = u^{1/a_1}$$
$$f_2(y) = a_2 y^{a_2-1} \rightarrow F_2(y) = y^{a_2} \rightarrow F_2^{-1}(v) = v^{1/a_2}$$
$$F(x, y; a_1, a_2) = xy^{a_1} + xy^{a_2} - xy,$$
using this in (7) gives:

$$C(u, v; a_1, a_2) = uv^{1/a_2} + vu^{1/a_1} - u^{1/a_1} v^{1/a_2}.$$  \hspace{1cm} (11)
which is a well defined copula for appropriate values of $a_1$, $a_2$ and for almost $u,v$ in $I$. If $a_1 = a_2 = \frac{1}{a}$, we notice that (11) can be rewritten as

$$C(u,v;a) = (uv)^a(u^{1-a} \otimes_1 v^{1-a}),$$

(12)

where $a \geq 1$ and $u \otimes_a v = [u^a + v^a - 1]^\frac{1}{a}$ is the generalized product [26].

**Case 3:** $a_i = b_j = 1/2$ which corresponds to the density of the arcsine distribution:

$$f_1(x) = \frac{1}{\pi \sqrt{x(1-x)}} \rightarrow F_1(x) = \frac{2}{\pi} \arcsin(\sqrt{x}) \rightarrow F_1^{-1}(u) = \sin^2(\frac{\pi}{2}u)$$

$$f_2(y) = \frac{1}{\pi \sqrt{y(1-y)}} \rightarrow F_2(y) = \frac{2}{\pi} \arcsin(\sqrt{y}) \rightarrow F_2^{-1}(v) = \sin^2(\frac{\pi}{2}v)$$

$$F(x,y) = \frac{2y}{\pi} \arcsin(\sqrt{x}) + \frac{2x}{\pi} \arcsin(\sqrt{y}) - xy, \quad 0 \leq x, y \leq 1.$$  

The corresponding copula:

$$C(u,v) = u \sin^2(\frac{\pi v}{2}) + v \sin^2(\frac{\pi u}{2}) - \sin^2(\frac{\pi u}{2}) \sin^2(\frac{\pi v}{2}).$$

There are also other bounded distributions beyond the Beta distribution [27] which have explicit quantile functions and the procedure of construction we have discussed can be extended to obtain other new families of copulas.

**CONCLUSION**

In this paper we have proposed a new way to derive families of copulas using the principle of maximum entropy. PME is used for finding a joint distribution given its marginals as the linear constraints, and Sklar’s theorem to obtain the corresponding copula. We considered only some particular cases for which we could obtain explicit expressions, but we are now investigating other cases of entropy expressions as well as other cases of marginals in an effort to obtain either analytical or numerical representations of other new families of continuous and discrete copulas.

**ACKNOWLEDGMENTS**

The authors thank Christian Genest, Fabrizio Durante, Christophe Vignat and Jean-François Bercher for theirs comments and suggestions on the first version of this paper. We thank Michael J. Betancourt and Fábio Macêdo Mendes for their careful and critical reading.
REFERENCES

23. J. Havrda, and F. Charvát, Kybernetika 3, 0–3 (1967).