Maximum Entropies Copulas

Doriano-Boris Pougaza, Ali Mohammad-Djafari

Maximum Entropies Copulas

Doriano-Boris Pougaza and Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes
UMR 8506 (CNRS-SUPELEC-UNIV PARIS SUD 11)
Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette Cedex, France

Abstract. New families of copulas are obtained in a two-step process : first considering the inverse problem which consists of finding a joint distribution from its given marginals as the constrained maximization of some entropies (Shannon, Rényi, Burg, Tsallis-Havrda-Charvát), and then using Sklar's theorem, to define the corresponding copula.

Keywords: Copula, Maximum entropy, Shannon, Rényi, Burg, Tsallis-Havrda-Charvát entropies
PACS: 02.30.Gp,02.50.Cw, 02.50.Sk

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* [4, 5] 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 $\mathbf{I} = [0,1]$ and given correlation.

Here the partial information is the knowledge of the marginal distributions. 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 $\mathbf{I}^2$ to $\mathbf{I}$ with the following properties:

1. $\forall u,v \in \mathbf{I}$, $C(u,0) = 0 = C(0,v)$,
2. $\forall u,v \in \mathbf{I}$, $C(u,1) = u$ 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 $u_1 \leq u_2$, $v_1 \leq v_2$.

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:

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

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. $J_1(f) = -\iint f(x,y)\ln f(x,y)\,dx\,dy$ (Shannon);
2. $J_2(f) = \dfrac{1}{1-q}\ln\left(\iint f^q(x,y)\,dx\,dy\right)$, $q \geqslant 0$ and $q \neq 1$ (Rényi);
3. $J_3(f) = \iint \ln f(x,y)\,dx\,dy$ (Burg);
4. $J_4(f) = \dfrac{1}{1-q}\left(1 - \iint f^q(x,y)\,dx\,dy\right)$ $q \geqslant 0$ 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

$$\begin{aligned}\mathscr{L}_g(f,\lambda_0,\lambda_1,\lambda_2) = J_i(f) &+ \lambda_0\left(1 - \iint f(x,y)dxdy\right)\\ &+ \int \lambda_1(x)\left(f_1(x) - \int f(x,y)dy\right)dx\\ &+ \int \lambda_2(y)\left(f_2(y) - \int f(x,y)dx\right)dy,\end{aligned}$$

and the critical point of $\mathscr{L}_g$ holds for the following system of equations:

$$\frac{\partial \mathscr{L}_g(f,\lambda_0,\lambda_1,\lambda_2)}{\partial f} = 0, \frac{\partial \mathscr{L}_g(f,\lambda_0,\lambda_1,\lambda_2)}{\partial \lambda_i} = 0.$$

Assuming that the integrals converge within the interval $\mathbf{I}$, this system of equations yields:

$$\begin{aligned}
f(x,y) &= \exp(-\lambda_1(x) - \lambda_2(y) - \lambda_0) && \text{(Shannon's entropy);}\\
\frac{f^{q-1}(x,y)}{\iint_{\mathbf{I}^2} f^q(x,y)\,dx\,dy} &= \frac{1-q}{q}(\lambda_1(x) + \lambda_2(y) + \lambda_0) && \text{(Rényi's entropy);}\\
f(x,y) &= (\lambda_1(x) + \lambda_2(y) + \lambda_0)^{-1} && \text{(Burg's entropy);}\\
f(x,y) &= \frac{1-q}{q}(\lambda_1(x) + \lambda_2(y) + \lambda_0)^{\frac{1}{q-1}} && \text{(Tsallis-Havrda-Charvát's entropy),}
\end{aligned}$$

where $\lambda_1(x)$, $\lambda_2(y)$ and λ_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:

$$\lambda_1(x) = -\ln\left(f_1(x) \int_{\mathbf{I}} \lambda_1(x)\, dx \right),$$

$$\lambda_2(y) = -\ln\left(f_2(y) \int_{\mathbf{I}} \lambda_2(y)\, dx \right)$$

and

$$\lambda_0 = \ln\left(\int_{\mathbf{I}} \lambda_1(x)\, dx \int_{\mathbf{I}} \lambda_2(y)\, dx \right),$$

and the joint distribution becomes

$$f(x,y) = f_1(x) f_2(y). \tag{2}$$

Unfortunately, in the cases of Rényi, Burg and Tsallis-Havrda-Charvát entropies, it is not possible, to find general solutions for λ_0, λ_1, and λ_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 λ_0:

$$\lambda_1(x) = -2 f_1(x) + \int_{\mathbf{I}} \lambda_1(x)\, dx + 2,$$

$$\lambda_2(y) = -2 f_2(y) + \int_{\mathbf{I}} \lambda_2(y)\, dy + 2,$$

$$\lambda_0 = -2 - \int_{\mathbf{I}} \lambda_1(x)\, dx - \int_{\mathbf{I}} \lambda_2(y)\, dy.$$

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

$$f(x,y) = f_1(x) + f_2(y) - 1. \tag{3}$$

Assuming $\|f\|_2^2 = \iint_{\mathbf{I}^2} f^2(x,y)\, dx\, dy = 1$, 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 $\mathbf{I}^n$ follows

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i) - n + 1. \tag{4}$$

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 :

$$\begin{aligned} F(x,y) &= \int_0^x \int_0^y f(s,t)\,ds\,dt \\ &= \int_0^x \int_0^y f_1(s) f_2(t)\,ds\,dt \\ &= \int_0^x f_1(s)\,ds \int_0^y f_2(t)\,dt. \end{aligned}$$

The cdf becomes

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

and the copula is

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

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)

$$\begin{aligned} F(x,y) &= \int_0^x \int_0^y f(s,t)\,ds\,dt \\ &= \int_0^x \int_0^y (f_1(s) + f_2(t) - 1)\,ds\,dt \\ F(x,y) &= y\int_0^x f_1(s)\,ds + x\int_0^y f_2(t)\,dt - xy, \end{aligned}$$

with the cdf

$$F(x,y) = yF_1(x) + xF_2(y) - xy, \quad 0 \le x,y \le 1 \tag{6}$$

and the associated copula

$$C(u,v) = uF_2^{-1}(v) + vF_1^{-1}(u) - F_1^{-1}(u)F_2^{-1}(v). \tag{7}$$

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

$$\begin{aligned} F(x_1,\ldots,x_n) &= \int_0^{x_1} \cdots \int_0^{x_n} f(s_1,\ldots,s_n) \prod_{i=1}^n ds_i \\ &= \int_0^{x_1} \cdots \int_0^{x_n} \left(\sum_{i=1}^n f_i(s_i) - n + 1 \right) \prod_{i=1}^n ds_i \\ &= \sum_{i=1}^n F_i(x_i) \prod_{\substack{j=1 \\ j\neq i}}^n x_j + (1-n)\prod_{i=1}^n x_i, \quad 0 \le x_i \le 1 \end{aligned} \tag{8}$$

and the associated multivariate copula, depending on F_i^{-1} will have the following form

$$C(u_1,\ldots,u_n)=\sum_{i=1}^{n}u_i\prod_{\substack{j=1\\ j\neq i}}^{n}F_j^{-1}(u_j)+(1-n)\prod_{i=1}^{n}F_i^{-1}(u_i). \tag{9}$$

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

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 :

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

where

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

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

<u>Case 1</u>: $a_i=1\,,\,b_j=1$ 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 :

$$C(u,v)=uv. \tag{10}$$

<u>Case 2</u>: $a_i>0\,,\,b_j=1$

$$f_1(x)=a_1x^{a_1-1}\rightarrow F_1(x)=x^{a_1}\rightarrow F_1^{-1}(u)=u^{\frac{1}{a_1}}$$
$$f_2(y)=a_2y^{a_2-1}\rightarrow F_2(y)=y^{a_2}\rightarrow F_2^{-1}(v)=v^{\frac{1}{a_2}}$$
$$F(x,y;a_1,a_2)=yx^{a_1}+xy^{a_2}-xy,$$

using this in (7) gives :

$$C(u,v;a_1,a_2)=uv^{\frac{1}{a_2}}+vu^{\frac{1}{a_1}}-u^{\frac{1}{a_1}}v^{\frac{1}{a_2}}, \tag{11}$$

which is a well defined copula for appropriate values of a_1, a_2 and for almost u, v in $\mathbf{I}$. If $a_1 = a_2 = \dfrac{1}{a}$, we notice that (11) can be rewritten as

$$C(u,v;a) = (uv)^a(u^{1-a} \otimes_1 v^{1-a}), \tag{12}$$

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:

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

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

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

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

1. D.-B. Pougaza, A. Mohammad-Djafari, and J.-F. Bercher, "Utilisation de la notion de copule en tomographie," in *XXIIe colloque GRETSI*, Dijon, France, 2009.
2. D.-B. Pougaza, A. Mohammad-Djafari, and J.-F. Bercher, *Pattern Recognition Letters* **31**, 2258–2264 (2010).
3. J. Hadamard, *Princeton University Bulletin* **13**, 49–52 (1902).
4. A. Sklar, *Publications de l'Institut de Statistique de L'Université de Paris 8* pp. 229–231 (1959).
5. B. Schweizer, and A. Sklar, *Probabilistic Metric Spaces*, North Holland New York, 1983.
6. A. Marshall, and I. Olkin, *Journal of the American Statistical Association* pp. 30–44 (1967).
7. C. Genest, and J. MacKay, *The American Statistician* **40**, 280–283 (1986).
8. H. Joe, *Journal of multivariate analysis* **46**, 262–282 (1993).
9. E. Jaynes, *Physical Review* **106**, 620–630 (1957).
10. E. Jaynes, *Physical Review* **108**, 171–190 (1957).
11. A. Mohammad-Djafari, *A Matlab Program to Calculate the Maximum Entropy Distributions*, Kluwer Academic Publ., 1991, T.W. Grandy edn.
12. A. Mohammad-Djafari, *Traitement du signal* **11**, 87–116 (1994).
13. H. Cramér, and H. Wold, *J.London Math. Soc.* **11**, 290–294 (1936).
14. V. Strassen, *The Annals of Mathematical Statistics* **36**, 423–439 (1965).
15. S. Kullback, *The Annals of Mathematical Statistics* pp. 1236–1243 (1968).
16. I. Csiszar, *Ann. Probab* **3**, 146–158 (1975).
17. J. Borwein, A. Lewis, and R. Nussbaum, *Journal of Functional Analysis* **123**, 264–307 (1994).
18. A. Meeuwissen, and T. Bedford, *Journal of Statistical Computation and Simulation* **57**, 143–174 (1997).
19. R. Nelsen, *An introduction to copulas*, Springer Verlag, 2006.
20. C. Shannon, *Bell System Technical Journal* **27**, 432–379 (1948).
21. A. Renyi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1961, vol. 1, pp. 547–561.
22. J. Burg, *Maximum entropy spectral analysis*, Ph.D. thesis, Stanford University (1975).
23. J. Havrda, and F. Charvát, *Kybernetika* **3**, 0–3 (1967).
24. C. Tsallis, *Journal of statistical physics* **52**, 479–487 (1988).
25. D. Ellerman, *Synthese* **168**, 119–149 (2009).
26. L. Nivanen, A. Le Mehaute, and Q. Wang, *Reports on Mathematical Physics* **52**, 437–444 (2003).
27. S. Kotz, and J. Van Dorp, *Beyond beta: other continuous families of distributions with bounded support and applications*, World Scientific Pub Co Inc, 2004.