Copule d'entropie maximale avec des statistiques d'ordre fixées

Richard Fischer [1] & Cristina Butucea [2] & Jean-François Delmas [3] & Anne Dutfoy [4]

[1] *EDF Recherche & Développement, Management des Risques Industriels, 1 Av. Général de Gaulle, 92141 Clamart; richard.fischer@edf.fr*
[2] *LAMA, Université Paris-Est - Marne-la-Vallée, 5 Bd. Descartes, 77454 Champs-sur-Marne; cristina.butucea@univ-mlv.fr*
[3] *CERMICS, Ecole Nationale des Ponts et Chaussées, 6 & 8 Av. Pascal, 77455 Champs-sur-Marne; delmas@cermics.enpc.fr*
[4] *EDF Recherche & Développement, Management des Risques Industriels, 1 Av. Général de Gaulle, 92141 Clamart; anne.dutfoy@edf.fr*

Résumé. L'objectif de cette communication est de présenter une nouvelle famille de copules absolument continues qui maximisent l'entropie de Shannon lorsque la distribution des statistiques d'ordre associées est imposée. On donne la formule explicite de la densité ainsi que l'entropie qui implique une condition pour que cette entropie soit finie. Cette copule nous permet de construire la loi jointe d'entropie maximale de vecteurs aléatoires dont les lois marginales sont imposées et qui vérifient des relations d'ordre presque sûrement. Avec notre méthode de construction, on peut également donner des familles de copules absolument continues dont la distribution des k plus grandes statistiques d'ordre est fixée. Dans le cas $k = 1$, cela revient à l'étude des copules à diagonale fixée, qui fait l'objet d'un grand intérêt dans la littérature de la théorie des copules.

Mots-clés. statistiques d'ordre, copule, modélisation multivariée, entropie

Abstract. The objective of this communication is to present a new family of absolutely continuous copulas which maximize the Shannon entropy under the constraint that the distribution of the corresponding order statistics is given. We give the explicit formula of its density and also the formula of its entropy, which implies a certain condition on the prescribed distribution of order statistics that ensures the finiteness of this entropy. This copula enables us to give the joint distribution of the maximum entropy random vector with given marginal distributions and with almost surely ordered components. Our approach also provides families of absolutely continuous copulas where only the distribution of the k largest order statistics are fixed. In particular, the case $k = 1$ corresponds to the problem of copulas with given diagonal section, which has been the subject of interest in recent works in the field of copula theory.

Keywords. order statistics, copula, multivariate modelling, entropy

1 Introduction

1.1 Contexte

Dans de nombreux contextes de modélisation des incertitudes en vue de leur propagation à travers des chaînes complexes de simulation numérique, l'ingénieur doit construire un modèle probabiliste prenant à la fois en compte l'information statistique disponible (ou le jugement d'experts), qui conduit le plus souvent à la détermination de lois marginales pour chacune des grandeurs incertaines, mais également les différentes contraintes physiques liant ces grandeurs incertaines, comme par exemple une relation d'ordre entre ces grandeurs. Un exemple consiste en la modélisation d'un système asservi tel que l'évolution d'une grandeur incertaine soit monotone en fonction du paramètre d'asservissement. Dans ce contexte, il n'est pas du tout évident qu'un tel modèle probabiliste existe, et même si c'est le cas, sa construction soulève des difficultés qu'il convient d'étudier spécifiquement. La construction d'un tel modèle est fondamentale si l'objectif est de quantifier un critère de fiabilité probabiliste qui nécessite de définir entièrement la loi de distribution du vecteur des variables aléatoires d'entrée. Les marginales étant imposées, la loi jointe est entièrement définie dès que la copule est fixée. Ce choix peut reposer sur un critère assurant que la copule optimale ne rajoute aucune contrainte. Ce travail étend celui présenté par Lebrun et Dutfoy (2013), dans laquelle les auteurs présentent une caractérisation des lois possibles et proposent une construction admissible mais sous-optimale.

1.2 Description du problème

Supposons que $X_{OS} = (X_{(1)}, \ldots, X_{(d)})$ est un vecteur de statistiques d'ordre de dimension d, c'est à dire:

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(d)} \quad p.s. \tag{1}$$

On note par F_{OS} et f_{OS} sa fonction de répartition et sa densité, respectivement. Les fonctions de répartition de ses marginales sont notées par $F_{(i)}$, $1 \leq i \leq d$. Rappelons la définition de l'entropie de Shannon d'une variable aléatoire Y de dimension d:

$$H(Y) = \begin{cases} -\int_{\mathbb{R}^d} f_Y(x) \log\left(f_Y(x)\right) \, dx, & \text{si } Y \text{ a une densité } f_Y, \\ -\infty & \text{sinon.} \end{cases}$$

On note $H(F_Y) = H(Y)$ où F_Y est la fonction de répartition de Y. On cherche à maximiser l'entropie $H(F_{OS})$ étant donné que les marginales $F_{(i)}$ sont données et que (1) est vérifiée.

1.3 Copules

Les marginales étant imposées, la loi jointe est entièrement définie dès lors que la copule est fixée. La copule d'un vecteur aléatoire permet de caractériser la dépendance entre les différentes composantes d'un vecteur aléatoire sans se préoccuper de ses lois marginales. Une copule est une fonction de répartition définie sur $I^d = [0,1]^d$ dont les marginales sont uniformes sur I. La référence Nelsen (2006) donne un panorama exhaustif sur les copules. Le théorème principal est celui de Sklar qui décompose la fonction de répartition F_Y de Y comme:

$$F_Y(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)),$$

où C est une copule et F_i les fonctions de répartition des marginales Y_i, $1 \leq i \leq d$. Si les F_i sont continues, alors C est unique. L'entropie $H(F_Y)$ se décompose de la façon suivante:

$$H(F_Y) = \sum_{i=1} H(F_i) + H(C).$$

Comme les marginales sont imposées dans notre problème, maximiser $H(F_Y)$ revient à maximiser $H(C)$.

2 Copules avec des statistiques d'ordre fixées

Dans la suite, on va montrer qu'on peut associer à X_{OS} une variable aléatoire échangeable $\tilde{X}$ dont la distribution des statistiques d'ordre est la même que celle de X_{OS} (cf Navarro et Spizzichino (2010)). Si on note par $\tilde{C}$ sa copule, on peut montrer que $H(C)$ est maximale si et seulement si $H(\tilde{C})$ est maximale. La copule $\tilde{C}$ est telle que si U est une variable aléatoire à marginales uniformes et de copule $\tilde{C}$, alors la fonction de répartition de la i-ème plus grande valeur $U_{(i)}$ est donnée par $\delta_{(i)} = F_{(i)}(G^{-1})$, avec $G(t) = \frac{1}{d}\sum_{i=1}^{d} F_{(i)}(t)$. On s'intéresse dans cet article à trouver la copule qui maximise l'entropie étant donné que les $\delta_{(i)}$ sont fixées pour $1 \leq i \leq d$.

3 Résultats

Sous certaines conditions sur les fonctions $\delta_{(i)}$, on montre que la densité $\tilde{c}^*$ de la copule optimale $\tilde{C}^*$ est symmétrique et est de la forme, pour $u \in I^d$:

$$\tilde{c}^*(u) = \prod_{i=1}^{d} a_i(u_{(i)})$$

où $u_{(i)}$ est la i-ème plus grande composante du vecteur u. Ce résultat a été obtenu par l'application du théorème de maximisation d'entropie sous contraintes linéaires de Borwein et al. (1994), comme il a déjà été appliqué à d'autres contraintes linéaires de rang (cf Meeuwissen et Bedford (1997) et Bedford et Lewis (2013)). Les fonctions a_i, $1 \leq i \leq d$ ne dépendent que de $\delta_{(i)}$ et leurs formes explicites seront présentées. Pour l'entropie, on a:

$$H(\tilde{C}^*) = \mathcal{J}(\delta) + \sum_{i=1}^{d} H(\delta_{(i)}) + \log(d!) + (d-1),$$

avec $\mathcal{J}(\delta)$ donné par:

$$\mathcal{J}(\delta) = \sum_{i=2}^{d} \int_I \delta'_{(i)}(t) \log\left(\delta_{(i-1)}(t) - \delta_{(i)}(t)\right) \, dt.$$

L'entropie est donc finie si et seulement si $\mathcal{J}(\delta)$ est finie. On montre également que si seule la distribution de k plus grandes statistiques d'ordres est imposée, alors la densité $\hat{c}^*$ de la copule optimale est de la forme:

$$c_\delta(u) = \prod_{i=1}^{d-k} b(u_{(i)}) \prod_{i=d-k+1}^{d} a_i(u_{(i)})$$

où on peut calculer a_i et b explicitement. Le cas $k = 1$ correspond à imposer la diagonale δ définie par $\delta(t) = C(t, \ldots, t)$. Ce cas a été considéré par Butucea et al. (2013).

À l'aide de $\tilde{c}^*$, on peut établir la densité f^*_{OS} du vecteur optimal X^*_{OS} du problème initial ainsi que son entropie $H(X^*_{OS})$. Cette densité est aussi de la forme produit sur le simplexe $S = \{x \in \mathbb{R}^d; x_1 \leq x_2 \leq \ldots \leq x_d\}$. Les résultats obtenus seront illustrés par des exemples avec plusieurs choix des marginales.

4 Perspectives

Nous envisageons d'appliquer les résultats lors des simulations au sein des études liées à l'activité industrielle d'EDF et puis déterminer si certaines données physiques multivariées issues de retour d'expertise pourraient être modélisé par la copule obtenue. Différentes contraintes pour la variable aléatoire ainsi que différents critères d'optimisation seront également considérés.

Bibliographie

[1] Nelsen, R. B. (2006) *An introduction to copulas, second edition*, Springer, New York

[2] Lebrun, R. et Dutfoy, A. (2013), Copulas for order statistics with prescribed marginal distribution functions *Journal of Multivariate Analysis*, soumis.

[3] Navarro, J. et Spizzichino F. (2010), On the relationships between copulas of order statistics and marginal distributions, *Statist. Probab. Lett.*, 80(5-6), 473–479.

[4] Borwein, J., Lewis, A. et Nussbaum, R. (1994), Entropy minimization, DAD problems, and doubly stochastic kernels, *Journal of Functional Analysis*, 123(2), 264–307.

[5] Meeuwissen, A.M.H. et Bedford, T. (1997), Minimally informative distributions with given rank correlation for use in uncertainty analysis, *Journal of Statistical Computation and Simulation*, 57(1-4), 143–174.
[6] Bedford, T. and Wilson, K.J. (2013), On the construction of minimum information bivariate copula families, *Annals of the Institute of Statistical Mathematics*,1–21
[7] Butucea, C., Delmas, J-F., Dutfoy, A. et Fischer, R. (2013) Maximum entropy copula with given diagonal section, arXiv preprint arXiv:1312.5219