On copulas constructed with Bernoulli and Coxian-2 distributions

Christopher Blier-Wong

Joint work with Hélène Cossette, Sébastien Legros and Etienne Marceau
Laboratoire ACT&RISK
École d'actuariat
Université Laval, Québec, Canada

christopher.blier-wong.1@ulaval.ca

15th International Conference of the ERCIM WG on Computational and Methodological Statistics
December 17 2022

Overview

Dependence and copulas

- We consider a random vector $\boldsymbol{X}$
- Described by a multivariate cdf

$$F_{\boldsymbol{X}}(x_1, \ldots, x_d) = \Pr(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

- Two components of $F_{\boldsymbol{X}}$:
  1. Marginal cdfs

$$F_1(x_1) = \Pr(X_1 \leq x_1), \ldots, F_n(x_n) = \Pr(X_d \leq x_d)$$

  2. Dependence structure that binds the marginal cdfs

$$F_{\boldsymbol{X}}(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

- Independence copula: $C(\boldsymbol{u}) = u_1 \times \cdots \times u_d$
- Copulas important in applications: let us study the effect of dependence

FGM copulas

- Farlie-Gumbel-Morgenstern (FGM) copula
- Implicit copula from FGM distributions [Eyraud, 1936], [Morgenstern, 1956], [Farlie, 1960], [Gumbel, 1960]
- Expression:

$$C(\boldsymbol{u}) = \prod_{m=1}^{d} u_m \left(1 + \sum_{k=2}^{d} \sum_{1 \leq j_1 < \cdots < j_k \leq d} \theta_{j_1 \ldots j_k} \overline{u}_{j_1} \overline{u}_{j_2} \ldots \overline{u}_{j_k}\right) \quad \boldsymbol{u} \in [0,1]^d,$$

- where $\overline{u}_j = 1 - u_j$, $j \in \{1, \ldots, d\}$
- Number of parameters: $d^\star = 2^d - d - 1$
- Bivariate:

$$C(u_1, u_2) = u_1 u_2 (1 + \theta_{12} \overline{u}_1 \overline{u}_2)$$

- Trivariate:

$$C(\boldsymbol{u}) = u_1 u_2 u_3 \left(1 + \theta_{12} \overline{u}_1 \overline{u}_2 + \theta_{13} \overline{u}_1 \overline{u}_3 + \theta_{23} \overline{u}_2 \overline{u}_3 + \theta_{123} \overline{u}_1 \overline{u}_2 \overline{u}_3\right)$$

FGM copulas

Advantages of FGM copulas

- Admits multiple dependence shapes ($2^d - d - 1$ parameters)
- Quadratic marginals: easy to integrate
- Admits closed-form expressions for many risk measures (association coefficients) [Nelsen, 2006, Genest and Favre, 2007]
- Admits closed-form expressions of many risk measures for aggregate risks and random sums
  [Bargès et al., 2009, Bargès et al., 2011, Cossette et al., 2013, Cossette et al., 2019]

FGM copula

Disadvantage of FGM copulas

- Admits weak dependence (adequate for insurance)
- **Symmetric dependence structures**
- Dependence parameters are not easy to interpret (for now)
- High-dimensional sampling and estimation: tedious (for now)
- Random vectors are difficult to compare (for now)

A d-variate FGM copula exists if $\boldsymbol{\theta} \in \mathcal{T}_d$, where

$$\mathcal{T}_d = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{d^\star} : 1 + \sum_{k=2}^{d} \sum_{1 \leq j_1 < \cdots < j_k \leq d} \theta_{j_1 \ldots j_k} \varepsilon_{j_1} \varepsilon_{j_2} \ldots \varepsilon_{j_k} \geq 0 \right\},$$

for $\{\varepsilon_{j_1}, \varepsilon_{j_2}, \ldots, \varepsilon_{j_k}\} \in \{-1, 1\}^d$ [Cambanis, 1977]

> *This family, and others like it, are not useful for statistical modeling. However, because calculations of many things can done in closed form, it is useful for demonstrating dependence concepts. [Joe, 2015]*

Stochastic representation

Theorem 1 (Stochastic representation of FGM copulas)

The FGM copula has an equivalent stochastic expression

$$C(\boldsymbol{u}) = E_{\boldsymbol{I}}\left[\prod_{m=1}^{d} u_m \left\{1 + (-1)^{I_m}\overline{u}_m\right\}\right],$$

for $\boldsymbol{u} \in [0,1]^d$, *where* $\boldsymbol{I}$ *is a random vector of symmetric Bernoulli rvs. The parameters of the copula are proportional to the central mixed moments:*

$$\theta_{j_1 \ldots j_k} = (-2)^k E_{\boldsymbol{I}}\left\{\prod_{n=1}^{k}\left(I_{j_n} - \frac{1}{2}\right)\right\},$$

for $k \in \{2, \ldots, d\}$ *and* $1 \leq j_1 < \cdots < j_k \leq d$.

[Blier-Wong et al., 2022] (CSDA)

- $\boldsymbol{U}_0, \boldsymbol{U}_1$: two vectors of independent standard uniform rvs.
- $\boldsymbol{I}$: random vector of multivariate symmetric Bernoulli rvs.
- Let $\boldsymbol{U}$ admits the representation

$$\boldsymbol{U} \stackrel{\mathcal{D}}{=} \sqrt{\boldsymbol{U}_0}\boldsymbol{U}_1^{\boldsymbol{I}} \tag{1}$$

- Then, $F_{\boldsymbol{U}}$ = FGM copula.

Objective of this talk

In this talk, we ask

- How can we change the symmetric hypothesis for $\boldsymbol{I}$ to build a stochastic representation as in (1)?
- What are the properties of this new family of copulas?

Coxian-2 distributions

Lemma 2

Let $I \sim Bern(p)$ and U, U_0, U_1 be independent and uniform rvs. Then,

$$U \overset{\mathcal{D}}{=} U_0^{1-p} U_1^I$$

Proof sketch:

- Let $X \sim Y_0 \sim Exp(\beta_1)$ and $Y_1 \sim Exp(\beta_2)$
- $X \overset{\mathcal{D}}{=} Y_0 + IY_1$
- Then X is Coxian-2 distributed with LST

$$\mathcal{L}_X(t) = (1-p)\frac{\beta_1}{t+\beta_1} + p\frac{\beta_1}{t+\beta_1}\frac{\beta_2}{t+\beta_2} = \frac{1}{1+t}$$

- Solve for β_1, β_2, notice that β_2 is linear in β_1, so fix $\beta_2 = 1$.

GFGM copulas

We construct GFGM copulas with the stochastic representation

Definition 3

- Fix some vector $\boldsymbol{p} = (p_1, \ldots, p_d)$ with $0 < p_j < 1$ for $j \in \{1, \ldots, d\}$.
- Let $\boldsymbol{I}$ be a random vector with $I_j \sim Bern(p_j)$
- Let $\boldsymbol{U}_0$ and $\boldsymbol{U}_1$ be independent random vectors of uniform rvs
- Set

$$\boldsymbol{U} \stackrel{\mathcal{D}}{=} \boldsymbol{U}_0^{\boldsymbol{1}-\boldsymbol{p}} \boldsymbol{U}_1^{\boldsymbol{I}}$$

- Let $C^{GFGM}(\boldsymbol{u}) = \Pr(\boldsymbol{U} \leq \boldsymbol{u})$

Denote the class $\mathcal{C}_{GFGM}$ as the class of copulas for $\boldsymbol{p} \in (0,1)^d$.

GFGM copulas

Proposition 1

We have that

$$C^{GFGM}(\boldsymbol{u}) = E\left[\prod_{m=1}^{d}\left(u_m^{(1-p_m)^{-1}} - I_m\left\{\frac{u_m^{(1-p_m)^{-1}} - u_m}{p_m}\right\}\right)\right]$$

Proposition 2

We have that

$$C(\boldsymbol{u}) = \prod_{m=1}^{d} u_m \left(1 + \sum_{k=1}^{d} \sum_{1 \leq j_1 < \cdots < j_k \leq d} \nu_{j_1 \ldots j_k} \left(1 - u_{j_1}^{\frac{p_{j_1}}{1-p_{j_1}}}\right) \ldots \left(1 - u_{j_k}^{\frac{p_{j_k}}{1-p_{j_k}}}\right)\right),$$

where

$$\nu_{j_1 \ldots j_k} = E\left[\prod_{n=1}^{k} \frac{I_{j_n} - p_{j_n}}{p_{j_n}}\right]$$

Bivariate GFGM copula

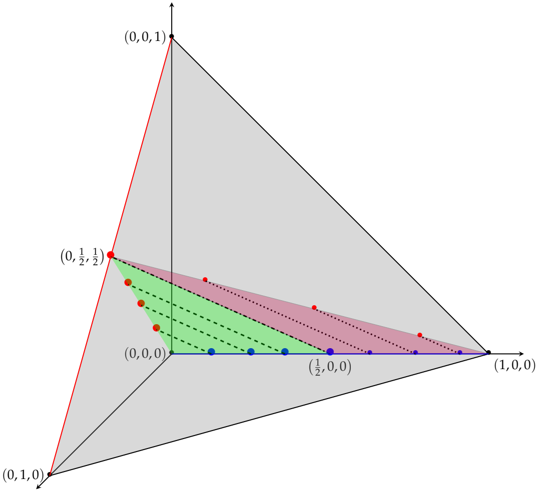
Represent biviariate pmf as linear function

$$\begin{vmatrix} f_{00} & f_{01} \\ f_{10} & f_{11} \end{vmatrix} = \begin{vmatrix} (1-p_1)(1-p_2) + p_1 p_2 \theta & (1-p_1)p_2 - p_1 p_2 \theta \\ p_1(1-p_2) - p_1 p_2 \theta & p_1 p_2 + p_1 p_2 \theta \end{vmatrix}$$

for

$$-\min\left(1, \frac{(1-p_1)(1-p_2)}{p_1 p_2}\right) \leq \theta \leq \min\left(\frac{1-p_1}{p_1}, \frac{1-p_2}{p_2}\right).$$

Bivariate GFGM copula



Proposition 3

The bivariate GFGM copula has the shape

$$C(u,v) = uv\left(1 + \theta\left(1 - u^{\frac{p_1}{1-p_1}}\right)\left(1 - v^{\frac{p_2}{1-p_2}}\right)\right)$$

for

$$-\min\left(1, \frac{(1-p_1)(1-p_2)}{p_1 p_2}\right) \leq \theta \leq \min\left(\frac{1-p_1}{p_1}, \frac{1-p_2}{p_2}\right).$$

Compare with the Huang-Kotz copula [Huang and Kotz, 1999][1].

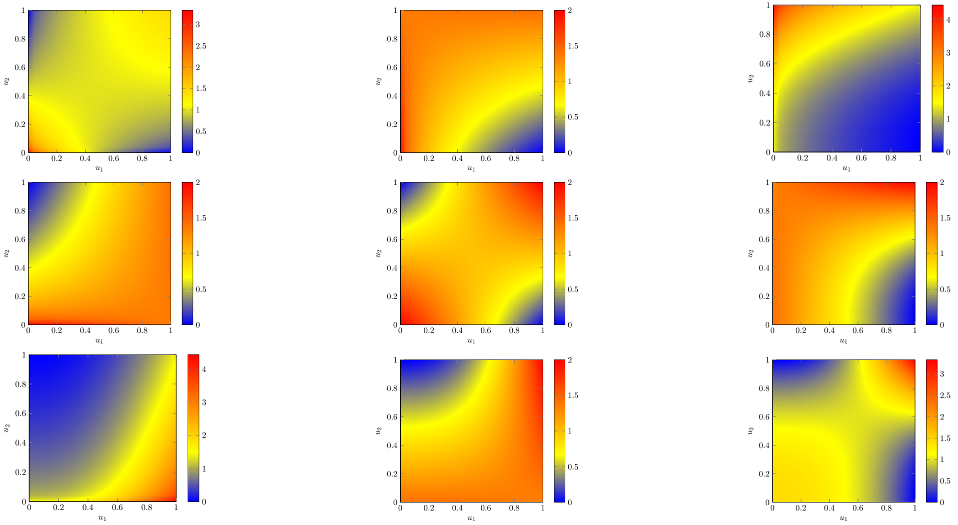
$$C(u,v) = uv\left(1 + a\left(1 - u^{b}\right)\left(1 - v^{b}\right)\right)$$

for $-(\max(1,b))^{-2} \leq a \leq b^{-1}$

[1] Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb. Metrika.

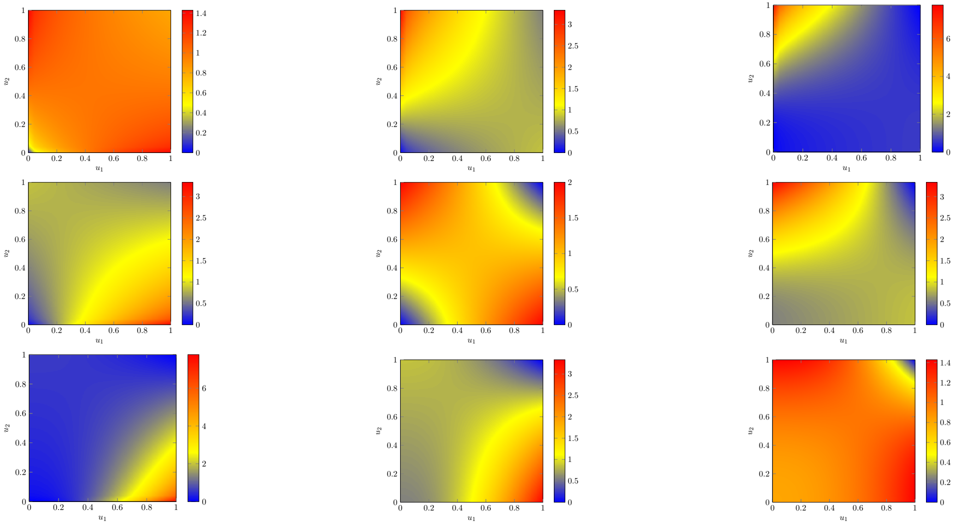
Bivariate GFGM copula

Positive dependence



Bivariate GFGM copula

Negative dependence



Algorithm 0: Stochastic simulation method

Input: Number of simulations n, pmf $f_{\boldsymbol{I}}$

Output: Set of simulations

for $l = 1, \ldots, n$ **do**

  Generate n independent random vectors $\boldsymbol{I}^{(l)}$, $\boldsymbol{U}_0^{(l)}$ and $\boldsymbol{U}_1^{(l)}$;

  for $m = 1, \ldots, d$ **do**

    Compute $U_m^{(l)} = \left(U_{0,m}^{(l)}\right)^{1-p_m} \left(U_{1,m}^{(l)}\right)^{I_m^{(l)}}$;

Return $\boldsymbol{U}^{(l)} = \left(U_1^{(l)}, \ldots, U_d^{(l)}\right), l = 1, \ldots, n.$

Let $\boldsymbol{p} = (p, \ldots, p)$ and $\boldsymbol{I}^+$, a vector of symmetric Bernoulli rvs with

$$\Pr(\boldsymbol{I}^+ = \boldsymbol{0}) = \Pr(\boldsymbol{I}^+ = \boldsymbol{1}) = p.$$

Then $\boldsymbol{I} \preceq_{sm} \boldsymbol{I}^+$ for all $\boldsymbol{I}$.

Theorem 4 (Identification of maximal dependence under $\preceq_{sm}$)

Let the EDP copula constructed with $\boldsymbol{I}^+$. The expression of the copula is

$$C^{EPD}(\boldsymbol{u}) = (1-p)\prod_{m=1}^{d} u_m^{1/(1-p)} + p\prod_{m=1}^{d}\left(\frac{p-1}{p}u_m^{1/(1-p)} - \frac{u_m}{p}\right)$$

Proposition 4 (Value of ρ^{cL}, ρ^{cU}, ρ^{c} and τ for copulas within $\mathcal{C}^{GFGM}$)

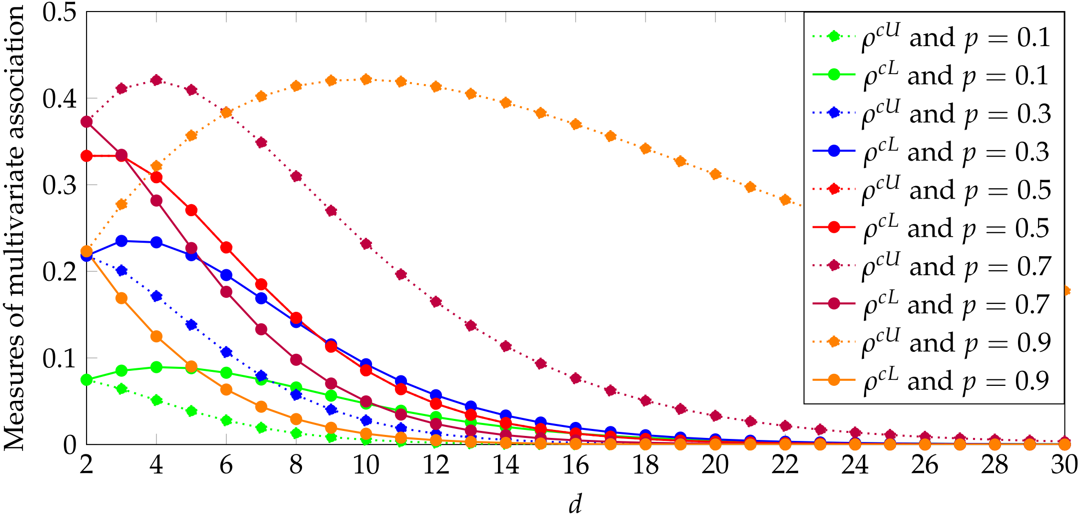
$$\rho^{cL}(\boldsymbol{U}) = \frac{d+1}{2^d - d - 1}\left(\left\{\prod_{m=1}^{d} \frac{1}{2-p_m}\right\} E\left[\prod_{m=1}^{d}(2(1-p_m)+I_m)\right] - 1\right);$$

$$\rho^{cU}(\boldsymbol{U}) = \frac{d+1}{2^d - d - 1}\left(\left\{\prod_{m=1}^{d} \frac{1}{2-p_m}\right\} E\left[\prod_{m=1}^{d}(2-I_m)\right] - 1\right);$$

$$\rho^{c}(\boldsymbol{U}) = \frac{d+1}{2^d - d - 1}\left(\left\{\prod_{m=1}^{d} \frac{1}{2-p_m}\right\} \frac{1}{2}\left(E\left[\prod_{m=1}^{d}(2(1-p_m)+I_m) + \prod_{m=1}^{d}(2-I_m)\right]\right) - 1\right)$$

$$\tau(\boldsymbol{U}) = \frac{1}{2^{d-1}-1}\left(2^d \sum_{\boldsymbol{i}\in\{0,1\}^d} \sum_{\boldsymbol{j}\in\{0,1\}^d} f_{\boldsymbol{I}}(\boldsymbol{i}) f_{\boldsymbol{I}}(\boldsymbol{j}) \prod_{m=1}^{d}\left(\frac{1}{2} - \frac{i_m + j_m}{2p_m} + \frac{j_m(1-p_m)+i_m}{p_m(2-p_m)}\right) - 1\right).$$

Association measures



Conclusion

We reveal a stochastic formulation of GFGM copulas

- Convenient stochastic representation
- Convenient natural representation
- Closed-form expression for many measures of multivariate association, risk measures
- Simulation algorithm

Conclusion

Thanks for your attention!

- christopher.blier-wong.1@ulaval.ca
- Blier-Wong, C. and Cossette, H. and Marceau, E. (2022) Stochastic representation of FGM copulas using multivariate Bernoulli random variables. Computational Statistics & Data Analysis
- Blier-Wong, C. and Cossette, H. and Marceau, E. (2022) Exchangeable FGM copulas
- Blier-Wong, C. and Cossette, H. and Marceau, E. (2022) Risk aggregation with FGM copulas
- Blier-Wong, C. and Cossette, H. and Marceau, E. (2022) Collective risk models with FGM copulas
- Blier-Wong, C. and Cossette, H. and Legros, S. and Marceau, E. (2022) On copulas constructed with Bernoulli and Coxian-2 distributions

References I

Bargès, M., Cossette, H., Loisel, S., and Marceau, É. (2011).
On the Moments of Aggregate Discounted Claims with Dependence Introduced by a Fgm Copula.
ASTIN Bulletin: The Journal of the IAA, 41(1):215–238.

Bargès, M., Cossette, H., and Marceau, É. (2009).
TVaR-based capital allocation with copulas.
Insurance: Mathematics and Economics, 45(3):348–361.

Blier-Wong, C., Cossette, H., and Marceau, E. (2022).
Stochastic representation of FGM copulas using multivariate Bernoulli random variables.
Computational Statistics & Data Analysis, page 28.

References II

Cambanis, S. (1977).
Some properties and generalizations of multivariate Eyraud-Gumbel-Morgenstern distributions.
Journal of Multivariate Analysis, 7(4):551–559.

Cossette, H., Côté, M.-P., Marceau, E., and Moutanabbir, K. (2013).
Multivariate distribution defined with Farlie–Gumbel–Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation.
Insurance: Mathematics and Economics, 52(3):560–572.

Cossette, H., Marceau, E., and Mtalai, I. (2019).
Collective risk models with dependence.
Insurance: Mathematics and Economics, 87:153–168.

References III

Eyraud, H. (1936).
Les principes de la mesure des correlations.
Ann. Univ. Lyon, III. Ser., Sect. A, 1(30-47):111.

Farlie, D. J. (1960).
The performance of some correlation coefficients for a general bivariate distribution.
Biometrika, 47(3/4):307–323.

Genest, C. and Favre, A.-C. (2007).
Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask.
Journal of Hydrologic Engineering, 12(4):347–368.

References IV

Gumbel, E. J. (1960).
Bivariate exponential distributions.
Journal of the American Statistical Association, 55(292):698–707.

Huang, J. S. and Kotz, S. (1999).
Modifications of the Farlie-Gumbel-Morgenstern distributions. A tough hill to climb.
Metrika, 49(2):135–145.

Joe, H. (2015).
Dependence Modeling with Copulas.
Number 134 in Monographs on Statistics and Applied Probability. CRC Press, Boca Raton, FL.

References V

Morgenstern, D. (1956).
Einfache beispiele zweidimensionaler verteilungen.
Mitteilingsblatt fur Mathematische Statistik, 8:234–235.

Nelsen, R. B. (2006).
An Introduction to Copulas.
Springer Series in Statistics. Springer, New York, 2nd ed edition.