CASE STUDY: Checkerboard Copula Defined by Spearman Rho Coefficients (entropyr)

Background

An m-dimensional copula where $m \geq 2$, is a continuous, m -increasing, probability distribution function $\mathrm{C}: [0,1]^{\mathrm{m}} \rightarrow [0,1]$ on the unit m-dimensional hyper-cube with uniform marginal probability distributions. A checkerboard copula is a distribution with a corresponding density $c: [0,1]^m \rightarrow [0, \infty)$ defined almost everywhere by a step function on an m -uniform subdivision of the hyper-cube. I.e., the checkerboard copula is a distribution on the unit hypercube $[0,1]^m$ defined by subdividing the hyper-cube into n^m identical small hyper-cubes $\mathrm{I_i}$ with constant density on each one. It is supposed that the density on $I_{\boldsymbol{i}}$, where $\boldsymbol{i} = (i_1, i_2, \dots, i_m)$ is defined by the expression $n^{m-1}h_i$, where $\mathrm{h_i}$ is an element of hyper-matrix $\boldsymbol{h} = [h_{\boldsymbol{i}}] \in \boldsymbol{R}^{\boldsymbol{n}^{\boldsymbol{m}}}$ with $h_{\boldsymbol{i}} \in [0,1]$. This case study demonstrates how to build a join distributions for correlated losses with known Spearman's rank correlation coefficients (which are called also grade correlation coefficients). If $\mathrm{x_r}$ are observations from a real valued random variable $X_r$with cumulative distribution function F_r , then the grade of $\mathrm{x_r}$ is given by $u_r = F_r(x_r)$. The grade u_r can be regarded as an observation of the uniform random variable $U_r = F_r(x_r)$ on [0, 1] with mean 1/2 and variance 1/12. The grade correlation coefficient for the continuous random variables X_r and X_s where $r < s$ is defined as the correlation for the grade random variables $U_r = F_r(x_r)$ and $U_s = F_s(x_s)$ by the formula

$$\rho_{rs} = \frac{E[(U_r - 1/2)(U_s - 1/2)]}{E[(U_r - 1/2)^2]^{1/2} E[(U_s - 1/2)^2]^{1/2}} = 12(E[U_r U_s] - 1/4).$$

The Spearman rho correlation coefficient for the checkerboard copula is given by

$$\rho_{rs} = 12\left(\frac{1}{n^3} \sum_{\boldsymbol{i} \in \{1,2,\dots,n\}^m} h_i \, (i_r - 1/2)(i_s - 1/2) - 1/4\right).$$

We present an optimization problem statement and results of optimization for the following example problem. We wish to select a multiply stochastic hyper-matrix $\boldsymbol{h} = [h_{\boldsymbol{i}}]$ to match known grade correlation coefficients ρ_{rs} in such a way that the entropy is maximized. We solved the problem with m =5 and $n = 4, 6, 8, 10$.

References

[1] Kuzmenko, V., Salam, R., Pavlikov, K., Uryasev, S. Checkerboard copula defined by sums of random variables. In preparation.
[2] Nelsen, R. B.: An Introduction to Copulas. Springer Lecture Notes in Stat., New York (1999).
[3] Piantadosi, J., Howlett, P., and Borwein, J. (2012). Copulas with maximum entropy Optimization Letters, 6, 99-125.
[4] Boland, J., Howlett, P., Piantadosi, J., and R. Zakaria. (2016). Modelling and simulation of volumetric rainfall for a catchment in the Murray-Darling basin, ANZIAM J, 58 2, 119–142.
[4] Borwein J., Howlett, Piantadosi, J. (2014). Modelling and simulation of seasonal rainfall using the principle of maximum entropy, Entropy, 16, 2, 747–769.

Notations

m = 5 number of random values;

n = number of sub-intervals in the partition of the interval [0,1];

$0 = a(1) < a(2) < \cdots < a(n+1) = 1$ is the partition of the interval [0,1],

where, $a(j) = \frac{j-1}{n}, j = 1, \dots. n+1$;

n^m = number of identical small hyper-cubes in the unit hyper-cube;

$I_{\boldsymbol{i}}$ = $\boldsymbol{i}$-th small hyper-cube, $\boldsymbol{i} = (i_1, i_2, i_3, i_4, i_5)$**;**

$\boldsymbol{h} = [h_{\boldsymbol{i}}] = [h_{i_1 i_2 i_3 i_4 i_5}]$ = hyper-matrix with $h_{i_1 i_2 i_3 i_4 i_5} \in [0,1]$;

$$\pi_r^c \boldsymbol{x} = \begin{cases} (x_2, \dots, x_m) & \text{if } r = 1 \\ (x_1, \dots x_{r-1}, x_{r+1}, \dots, x_m) & \text{if } r = 2, \dots, m-1 \\ (x_1, \dots, x_{m-1}) & \text{if } r = m \end{cases}$$ = complementary projection of $R^m \to R^{m-1}$;

$J(\boldsymbol{h}) = (-1)\left[\frac{1}{n}\sum_{\boldsymbol{i}\in\{1,2,\dots,n\}^5} h_{\boldsymbol{i}} \log_e h_{\boldsymbol{i}} + (m-1)\log_e\right]$ = entropy of $\boldsymbol{h}$.

Maximization of $J(\boldsymbol{h})$ is equivalent to maximization of $(-1)\sum_{\boldsymbol{i}\in\{1,2,\dots,n\}^5} h_{\boldsymbol{i}} \log_e h_{\boldsymbol{i}}$.

Optimization Problem

find hyper-matrix $\boldsymbol{h}$ *maximizing*

$$max_{\boldsymbol{h}}\,(-1) \sum_{\boldsymbol{i}\in\{1,2,\dots,n\}^5} h_{\boldsymbol{i}} \log_e h_{\boldsymbol{i}}$$

subject to constraints

$$12\left(\frac{1}{n^3} \sum_{\boldsymbol{i}\in\{1,2,\dots,n\}^5} h_i\,(i_r - 1/2)(i_s - 1/2) - 1/4\right) = \rho_{rs}, \qquad 1 \le r < s \le 5$$

$$\sum_{\boldsymbol{\pi}_r^c \boldsymbol{i}\in\{1,2,\dots,n\}^4} h_{\boldsymbol{i}} = 1, \qquad i_r \in \{1,2,\dots,n\}, \qquad r = 1,\dots,5$$

$$h_{\boldsymbol{i}} \ge 0, \qquad \boldsymbol{i} \in \{1,2,\dots,n\}^5$$