CASE STUDY: Checkerboard Copula Defined by Sums of Random Variables(meanabs_err, meansquare_err, cvar_risk)

Background

An m-dimensional copula where $m \geq 2$, is a continuous, m-increasing, probability distribution function $C: [0,1]^m \rightarrow [0,1]$on the unit m-dimensional hyper-cube with uniform marginal probability distributions.A checkerboardcopula is a distribution with a corresponding density defined almost everywhere by a step functionon 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 inton^midentical small hyper-cubes$I_i$with constant densityon each one.It is supposed that the density on$I_i$is defined by the expression$n^{m-1}h_i$, where $h_i$is an element of hyper-matrix$\boldsymbol{h} = [h_i] \in R^{n^m}$ with $h_i \in [0,1]$.This case study builds a copula for the case when distributions of marginals and distributions of sums of some random variables are available. This approach is used for estimating copula for the losses of an insurance company with different lines of business. Some informationabout losses of business lines is available and the objective is to build a copula of the joint distribution. We present optimization problem statements and results of optimization for several exampleproblems as presented in [1]. We assume that three empirical probability distribution functions$F_W(w), F_X(x), F_Y(y)$of random variables W, X, Y are available (with 1000 observations).We want to find a checkerboard copula on $n \times n \times n$ grid. We consider the following two cases.

CASE 1 (Optimization Problems 1-3). Empirical distribution $F_Z(z)$ for the sum $Z = W + X + Y$ is available with K=16observations,$z_1, \dots, z_{16}$. We suppose that these observations are equally probable and the distribution function$F_Z(z)$ takes K values,$\frac{1}{K}, \frac{2}{K}, \dots, \frac{K}{K}$.

CASE 2 (Optimization Problems 4-6).
Let usdefine three random values:$Z_1 = W + X, Z_2 = W + Y, Z_3 = X + Y$. We suppose that empirical Distributions,$F_{Z_1}(z), F_{Z_2}(z), F_{Z_3}(z)$,are specified by$K$=16 observations $z_1^1, \dots, z_{16}^1$from the distribution$F_{Z_1}(z)$, observations $z_1^2, \dots, z_{16}^2$from the distribution$F_{Z_2}(z)$, and observations $z_1^3, \dots, z_{16}^3$from the distribution$F_{Z_3}(z)$. We suppose that these observations areequally probable andthat every distribution function$F_{Z_i}(z), i = 1,2,3$, takes K values, $\frac{1}{K}, \frac{2}{K}, \dots, \frac{K}{K}$.

We solved problems with parameters m =3,$n = 10$ with *Mean-absolute Error, Mean-squared Error, and CVaR Absolute Error*(with confidence levels $\alpha = 0.9, 0.99$).

References

[1] Kuzmenko, V., Salam, R., Pavlikov, K.,Uryasev, S. Checkerboard copula defined by sums of random variables. In preparation.
[2] Piantadosi, J., Howlett, P., and Borwein, J. (2012). Copulas with maximum entropy. Optimization Letters, 6, 99-125.
[3] Boland,J., Howlett, P.,Piantadosi,J., and R. Zakaria.(2016). Modelling and simulation of volumetric rainfall for acatchment in the Murray-Darling basin, ANZIAM J, 58 2, 119–142.
[4] Borwein J., Howlett, P., Piantadosi, J.(2014). Modelling and simulation of seasonal rainfall using theprinciple of maximum entropy, Entropy, 16,2, 747–769.

Notations

m =number of random values;

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

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

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

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

I_i = $\boldsymbol{i}$-thsmall hyper-cube,$\boldsymbol{i} = (i_1, i_2, i_3)$;

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

$Ind(a \le b) = \begin{cases} 1, & \text{if } a \le b, \\ 0, & \text{otherwise;} \end{cases}$ = Indicator function;

$\gamma_{i_1 i_2 i_3}(z)$ = coefficients of the loss function in Case 1, $i_1, i_2, i_3 = 1, \dots n, z \in [0,1]$;

$$\gamma_{i_1 i_2 i_3}(z) = \begin{cases} n^{-1}, & \text{if} F_W^{-1}\big(a(i_1)+1\big) + F_X^{-1}\big(a(i_2)+1\big) + F_Y^{-1}\big(a(i_3)+1\big) \le z, \\ 0, & \text{if} F_W^{-1}\big(a(i_1)+1\big) + F_X^{-1}\big(a(i_2)+1\big) + F_Y^{-1}\big(a(i_3)+1\big) \ge z \\ \frac{1}{2} n^{-1}, & \text{otherwise;} \end{cases}, \quad i_1, i_2, i_3 = 1, \dots, n;$$

$L(\boldsymbol{h}, j) = \frac{j}{16} - \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \gamma_{i_1 i_2 i_3}(z_j)\, h_{i_1 i_2 i_3}$= loss function in Case 1 for j-th scenario, $j = 1, \dots, 16$;

$\gamma_{i_1 i_2}(z), \gamma_{i_1 i_3}(z), \gamma_{i_2 i_3}(z)$ = coefficients of loss functions in Case 2, $i_1, i_2, i_3 = 1, \dots n, z \in [0,1]$;

$$\gamma_{i_1 i_2}(z) = \begin{cases} n^{-1}, & \text{if} F_W^{-1}\big(a(i_1)+1\big) + F_X^{-1}\big(a(i_2)+1\big) \le z, \\ 0, & \text{if} F_W^{-1}\big(a(i_1)+1\big) + F_X^{-1}\big(a(i_2)+1\big) \ge z \\ \frac{1}{2} n^{-1}, & \text{otherwise;} \end{cases} \qquad i_1, i_2 = 1, \dots, n;$$

$$\gamma_{i_1 i_3}(z) = \begin{cases} n^{-1}, & \text{if} F_W^{-1}\big(a(i_1)+1\big) + F_Y^{-1}\big(a(i_3)+1\big) \le z, \\ 0, & \text{if} F_W^{-1}\big(a(i_1)+1\big) + F_Y^{-1}\big(a(i_3)+1\big) \ge z \\ \frac{1}{2} n^{-1}, & \text{otherwise;} \end{cases} \qquad i_1 i_3 = 1, \dots, n;$$

$$\gamma_{i_2 i_3}(z) = \begin{cases} n^{-1}, & \text{if} F_X^{-1}\big(a(i_2)+1\big) + F_Y^{-1}\big(a(i_3)+1\big) \le z, \\ 0, & \text{if} F_X^{-1}\big(a(i_2)+1\big) + F_Y^{-1}\big(a(i_3)+1\big) \ge z \\ \frac{1}{2} n^{-1}, & \text{otherwise;} \end{cases} \qquad i_2, i_3 = 1, \dots, n;$$

$$\left.\begin{aligned} L_{i_1 i_2}(\boldsymbol{h}, j) &= \frac{j}{16} - \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \gamma_{i_1 i_2}(z_j^1) \sum_{i_3=1}^{n} h_{i_1 i_2 i_3} \\ L_{i_1 i_3}(\boldsymbol{h}, j) &= \frac{j}{16} - \sum_{i_1=1}^{n} \sum_{i_3=1}^{n} \gamma_{i_1 i_3}(z_j^2) \sum_{i_3=1}^{n} h_{i_1 i_2 i_3} \\ L_{i_2 i_3}(\boldsymbol{h}, j) &= \frac{j}{16} - \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \gamma_{i_2 i_3}(z_j^3) \sum_{i_3=1}^{n} h_{i_1 i_2 i_3} \end{aligned}\right\}$$
= loss functions in Case 2, i_1, i_2, i_3=1,...,n, j=1,2,...16;

$err(\cdot)$ = one of the following error functions: Mean SquaredError denoted by$\varepsilon_{MSE}(\cdot)$,Mean Absolute Error denoted by$\varepsilon_{MAE}(\cdot)$, CVaR Absolute Error with confidence level 0.9 and 0.99 d enoted by $\varepsilon_{CVaR_{0.9}}(|\cdot|)$ and $\varepsilon_{CVaR_{0.99}}(|\cdot|)$;

$\varphi(\cdot) = \frac{1}{3}\left\{err\left(L_{i_1 i_2}(\cdot)\right) + err\left(L_{i_1 i_3}(\cdot)\right) + err\left(L_{i_2 i_3}(\cdot)\right)\right\}$ = weighted average of the error functions = objective function in Case 2.

Optimization Problem 1

Find hyper-matrix $\boldsymbol{h}$ *minimizingMean Absolute Error function of loss* $L(\boldsymbol{h}, j)$

$$min_{\boldsymbol{h}}\ meanabs\_err\ L(\boldsymbol{h}, j)$$

subject to the constraints

$$\sum_{i_2=1}^{10} \sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_1 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10} \sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_2 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_2=1}^{10} h_{i_1 i_2 i_3} = 1, i_3 = 1, \dots, 10,$$

$$h_{i_1 i_2 i_3} \geq 0, \qquad i_1, i_{2,} i_3 = 1, \dots, 10.$$

Optimization Problem 2

Find hyper-matrix$\boldsymbol{h}$*minimizing Mean Squared Errorfunction of loss* $L(\boldsymbol{h}, j)$

$$min_{\boldsymbol{h}}\ meansquare\_err\ L(\boldsymbol{h}, j)$$

subject to constraints

$$\sum_{i_2=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_1 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_2 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_2=1}^{10} h_{i_1 i_2 i_3} = 1, i_3 = 1, \dots, 10,$$

$$h_{i_1 i_2 i_3} \geq 0, \qquad i_1, i_{2,} i_3 = 1, \dots, 10.$$

Optimization Problem 3

Find hyper-matrix$\boldsymbol{h}$*minimizingCVaR Absolute Errorfunction of loss* $L(\boldsymbol{h}, j)$*with confidence levels* $\alpha = 0.9$and$\alpha = 0.99$

$$min_{\boldsymbol{h}}\ CVaR_{\alpha}\ |L(\boldsymbol{h}, j)|$$

subject to constraints

$$\sum_{i_2=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_1 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_2 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_2=1}^{10} h_{i_1 i_2 i_3} = 1, i_3 = 1, \dots, 10,$$

$$h_{i_1 i_2 i_3} \geq 0, \qquad i_1, i_{2,} i_3 = 1, \dots, 10.$$

Optimization Problem 4

Find hyper-matrix$\boldsymbol{h}$*minimizing weigted average of the Mean Absolute Error functionsof loss* $L(\boldsymbol{h}, j)$

$$min_{\boldsymbol{h}} \frac{1}{3}\left\{meanabs\_err\left(L_{i_1 i_2}(\boldsymbol{h}, j)\right) + meanabs\_err\left(L_{i_1 i_3}(\boldsymbol{h}, j)\right) + meanabs\_err\left(L_{i_2 i_3}(\boldsymbol{h}, j)\right)\right\}$$

subject to the constraints

$$\sum_{i_2=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_1 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_2 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_2=1}^{10} h_{i_1 i_2 i_3} = 1, i_3 = 1, \dots, 10,$$

$$h_{i_1 i_2 i_3} \geq 0, \qquad i_1, i_{2,} i_3 = 1, \dots, 10.$$

Optimization Problem 5

Find hyper-matrix$\boldsymbol{h}$*minimizing weigted average of the Mean Squared Error functionsof loss* $L(\boldsymbol{h}, j)$

$$min_{\boldsymbol{h}} \frac{1}{3}\Big\{meansquare\_err\left(L_{i_1 i_2}(\boldsymbol{h}, j)\right) + meansquare\_err\left(L_{i_1 i_3}(\boldsymbol{h}, j)\right) + meansquare\_err\left(L_{i_2 i_3}(\boldsymbol{h}, j)\right)\Big\}$$

subject to constraints

$$\sum_{i_2=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_1 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_2 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_2=1}^{10} h_{i_1 i_2 i_3} = 1, i_3 = 1, \dots, 10,$$

$$h_{i_1 i_2 i_3} \geq 0, \qquad i_1, i_{2,} i_3 = 1, \dots, 10.$$

Optimization Problem 6

Find hyper-matrix$\boldsymbol{h}$*minimizing weigted average of the CVaR Absolute Errorfunctionsof loss* $L(\boldsymbol{h}, j)$*with confidence levels* $\alpha = 0.9$ and $\alpha = 0.99$

$$min_{\boldsymbol{h}} \frac{1}{3}\left\{ CVaR_\alpha\left(\left|L_{i_1 i_2}(\boldsymbol{h}, j)\right|\right) + CVaR_\alpha\left(\left|L_{i_1 i_3}(\boldsymbol{h}, j)\right|\right) + CVaR_\alpha\left(\left|L_{i_2 i_3}(\boldsymbol{h}, j)\right|\right)\right\}$$

subject to constraints

$$\sum_{i_2=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_1 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_3=1}^{10} h_{i_1 i_2 i_3} = 1, i_2 = 1, \dots, 10,$$

$$\sum_{i_1=1}^{10}\sum_{i_2=1}^{10} h_{i_1 i_2 i_3} = 1, i_3 = 1, \dots, 10,$$

$$h_{i_1 i_2 i_3} \geq 0, \qquad i_1, i_{2,} i_3 = 1, \dots, 10.$$