Risk Management - Problems 2
(due date: April 13, 2025)

Jakub Gizbert-Studnicki
Maciej Trzetrzelewski
Maciej A. Nowak

Mark Kac Complex Systems Research Center
Jagiellonian University
Kraków, Poland
Summer semester (2024/25)
Jagiellonian University WFAIS.IF-Y491.0

Multivariate distributions

1. Consider multivariate random variables X_i with a covariance matrix elements $C_{ij}^{(X)} \equiv Cov(X_i, X_j) \equiv \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$. Show that random variables Y_α being linear combinations of X_i, such that $Y_\alpha = \sum_i v_{\alpha i} X_i$, where $v_{\alpha i}$ are constant weights, have the covariance matrix:

$$C_{\alpha\beta}^{(Y)} \equiv Cov(Y_\alpha, Y_\beta) = \sum_{i,j} v_{\alpha i} C_{ij} v_{\beta j}$$

   In the matrix form: $C_{\alpha\beta}^{(Y)} = \boldsymbol{v}_\alpha^T \boldsymbol{C}^{(X)} \boldsymbol{v}_\beta$, if $\boldsymbol{v}_\alpha$, $\boldsymbol{v}_\beta$ are column vectors.

2. Show that the (Gaussian) $d-$dimensional random vector $\boldsymbol{Y}_d \equiv \boldsymbol{X}_d/\sqrt{d}$, where $\boldsymbol{X}_d$ is drawn from a multivariate Gaussian distribution with zero mean vector ($m_i = 0$) and unit covariance matrix ($C_{ij} = \delta_{ij}$) converges to the (unit) $d-$dimensional sphere, i.e. $||\boldsymbol{Y}_d||^2 \equiv \boldsymbol{Y}_d^T \boldsymbol{Y}_d \rightarrow 1$ for $d \rightarrow \infty$.

   (a) Numerically: generate a (large, e.g. $N = 10^4$) sample of $\boldsymbol{Y}_d$ for $d = 1, 2^2, 3^2, ..., 10^2$ and compute the mean and variance of $||\boldsymbol{Y}_d||^2$ and draw them as a function of d. (HINT: as for $C_{ij} = \delta_{ij}$ the elements of the random vector $\boldsymbol{X}_d$ are independent, one can simply generate each of them separately from a one-dim Gaussian)

   (b) Analytically: one can use the Central Limit Theorem (HINT: the sum of squares of d iid standard Gaussian variables $X_i \sim N(0, 1)$: $Q = \sum_{i=1}^d X_i^2$ has the Chi-Squared distribution with d degrees of freedom.)

   Note that this is also the case for any (normalized, i.e. shifted and rescaled: $\frac{\boldsymbol{X}_i - \boldsymbol{m}_i}{C_{ii}^{1/2}}$) correlated Gaussian variables, as introducing correlation is just a rotation of the sphere (see Principal Components).

3. (Contour) plot the joint PDF of the 2-dimensional probability distribution defined by the Gumbel Copula:

$$\text{CDF: } C(u, v) = \exp\left[-\left(\left(\ln\frac{1}{u}\right)^\theta + \left(\ln\frac{1}{v}\right)^\theta\right)^{1/\theta}\right]$$

with standard Gaussian marginals $N(0,1)$ for the (coupling) parameter $\theta = 1, 1.5, 2$.

Note: for $\theta > 1$ this is not a multivariate Gaussian but it has Gaussian marginal distributions !

4. The file *dat.txt* contains sample bivariate data (each line is a 2-dim random vector with components separated by a space), generated using the Gaussian Copula with some (unknown) marginals.

   (a) Compute numerically and plot empirical CDFs of the marginal distributions.

   NOTE: Compute empirical CDFs in the whole range of x_i but plot them in range $x_i \in [-2, 2]$, $(i = 1, 2)$.

   HINT: empirical CDF for the sample with N elements:

$$P_{\leq}(x) = \frac{\text{number of elements in the sample} \leq x}{N} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{X_i \leq x}$$

   (b) Estimate the (only) parameter of the 2-dim Gaussian Copula, i.e. the correlation coefficient ρ.

   HINT: Using empirical marginal CDFs convert the sample data into the 2-dim uniform distribution and then (using inverse Gaussian CDF) into the 2-dim Gaussian and compute the covariance ρ, see (the opposite) example in the Lecture notes, p. 47.