Simulation and Goodness-of-fit Test of Copulas

Yiran Chen, Major Prof: Giray Ökten

Applied and Computational Mathematics, Department of Mathematics, Florida State University

Introduction

A **copula** is a multivariate distribution with dependent variables. It separates the dependence structure from the marginals. Copulas have been used widely in quantitative finance and insurance, such as option pricing and risk management.

Simulation plays a fundamental role in applications. There are two types of simulation methods for copulas, which can be compared through goodness-of-fit tests.

Simulation Methods

Sampling procedures for a d-dimensional copula C can be viewed as a transformation

$$\phi_C : [0,1]^k \to [0,1]^d \qquad \text{i.e.} \qquad \phi_C(\boldsymbol{U}) = \boldsymbol{V}$$

for some $k \geq d$.

1. **Stochastic Representation** writes a random vector in terms of other random vectors we know how to simulate efficiently. For example, if a random vector $\boldsymbol{X} \sim t_d(\mu, P, \nu)$, then it has the following representation

$$\boldsymbol{X} \stackrel{d}{=} \mu + P^{1/2} \frac{\sqrt{\nu}}{\sqrt{X_\nu^2}} \boldsymbol{Z}.$$

2. **Conditional Distribution Method** is given by conditional distribution transformation: for $\boldsymbol{U} \sim \mathrm{U}[0,1]^d$,

$$\begin{cases} V_1 = U_1 \\ V_2 = C^{-1}(U_2|V_1) \\ \dots \\ V_d = C^{-1}(U_d|V_1,\dots,V_{d-1}) \end{cases}$$

Then $\boldsymbol{V} = (V_1, \dots, V_d)$ is a sample from copula C.

Algorithm

A General Simulation Algorithm Using the CDM

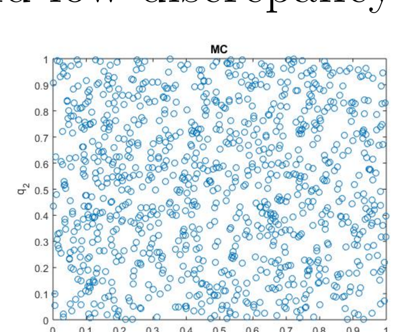
i. *Simulate u_1, q_2, q_3, …, q_d independently from $\mathbf{U}(0,1)$.*

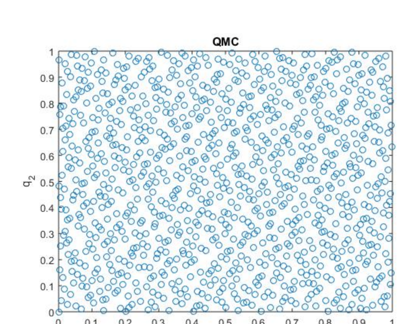
ii. *Set $u_k = C_k^{-1}(q|u_1,\dots,u_{k-1})$, for $k = 2, 3, \dots, d$.*

Then vector $(u_1, \dots, u_d)$ is the desired sample from C.

Monte Carlo vs Quasi-Monte Carlo

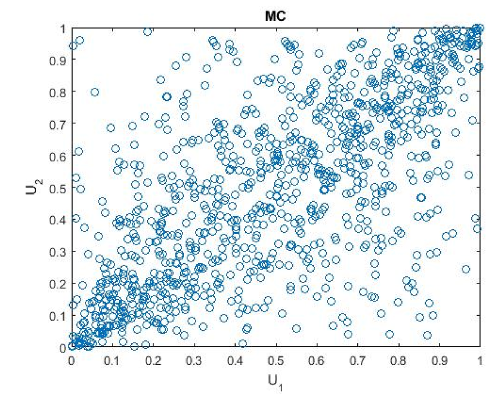
1000 realizations of two-dimensional pseudo-random number sequence (left) and low-discrepancy sequence (right):

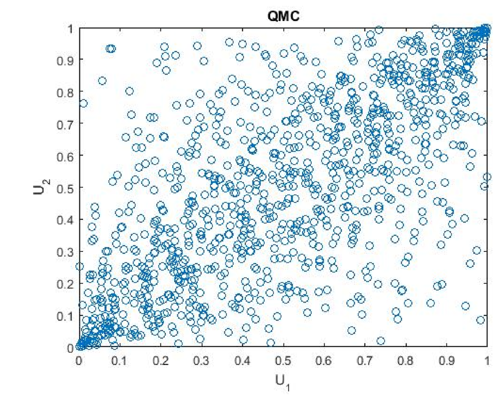




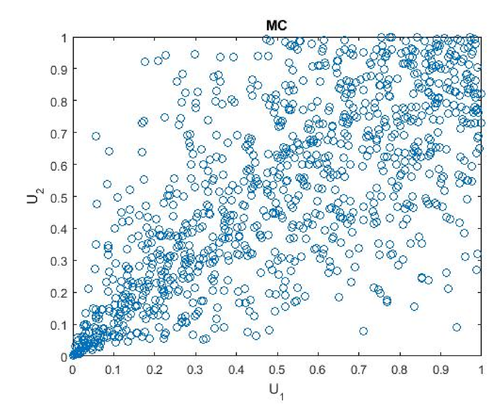
Examples for Bivariate Copulas

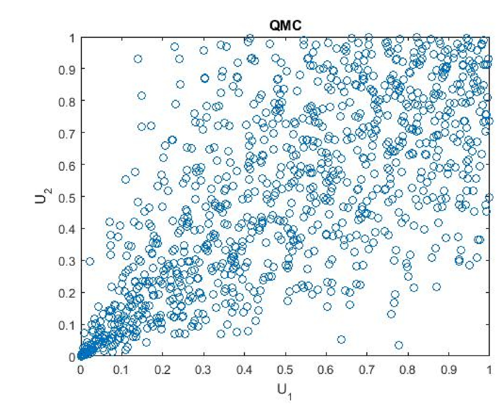
Case 1: 1000 realizations of a t copula with three degree of freedom and correlation parameter $\rho = 1/\sqrt{2}$, generated by a pseudo-random generator (left) and by a quasi-random number generator (right).





Case 2: 1000 realizations of a Clayton copula with $\theta = 2$, generated by a pseudo-random number generator (left) and by a quasi-random number generator (right).





Goodness-of-fit Test

Let C be the underlying d-variate copula. Suppose one wants to test the goodness-of-fit (GoF) hypothesis that C belongs to a copula family C_θ with parameter $\theta \in \Theta$, where Θ is the parameter space:

$$\mathcal{H}_0 : C \in \mathcal{C} = \{C_\theta; \theta \in \Theta\} \qquad \textbf{vs.} \qquad \mathcal{H}_1 : C \notin \mathcal{C} = \{C_\theta; \theta \in \Theta\}.$$

1. **Rank data:** Suppose we have N samples of the d-variate vector $\boldsymbol{X}_j = (X_{j,1}, \dots, X_{j,d})$, for $j = 1, \dots, N$. The rank data $\boldsymbol{Z}_j$ are given by

$$\boldsymbol{Z}_j = (Z_{j,1}, \dots, Z_{j,d}) = \left(\frac{R_{j,1}}{N+1}, \dots, \frac{R_{j,d}}{N+1}\right), \quad j = 1, \dots, N;$$

where $R_{j,i}$ is the rank of $X_{j,i}$ amongst $(X_{1,i}, \dots, X_{N,i})$.

2. **Conditional Probability Integral Transform (CPIT):** Let $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ denote a random vector with marginal distributions $F_i(z_i)$ and conditional distribution $F_{i|1\dots i-1}(Z_i \leq z_i | Z_1 = z_1, \dots, Z_{i-1} = z_{i-1})$ for $i = 1, \dots, d$. **CPIT** of $\boldsymbol{Z}$ is given by

$$V_1 = \mathbb{P}(Z_1 \leq z_1) = F_1(z_1),$$
$$V_2 = \mathbb{P}(Z_2 \leq z_2 | Z_1 = z_1) = F_{2|1}(z_2|z_1),$$
$$\vdots$$
$$V_d = \mathbb{P}(Z_d \leq z_d | Z_1 = z_1, \dots, Z_{d-1} = z_{d-1}) = F_{d|1\dots d-1}(z_d|z_1, \dots, z_{d-1}).$$

Then $\boldsymbol{V} = (V_1, \dots, V_d)$ has components that are i.i.d. $\mathbf{U}(0,1)$ distributed.

CPIT Approach

One GoF test is designed based on the CPIT:

$\{\mathbf{X}_j\}_{j=1}^N$: N samples of $\mathbf{X}_j = (X_{j,1}, \dots, X_{j,d})$ from C_θ

$\{\mathbf{Z}_j\}_{j=1}^N$: N samples from the rank data

Apply CPIT on each $\mathbf{Z}_j$

$\{\mathbf{V}_j\}_{j=1}^N$: N i.i.d. samples from $\mathbf{U}[0,1]^d$

Dimension reduction by transformation function $T(x) = \Phi^{-1}(x)^2$:
$W_{B_j} = \sum_{i=1}^d T(V_{j,i}) = \sum_{i=1}^d \Phi^{-1}(V_{j,i})^2$

$\{W_{B_j}\}_{j=1}^N$: N univariate samples from χ_d^2 distribution

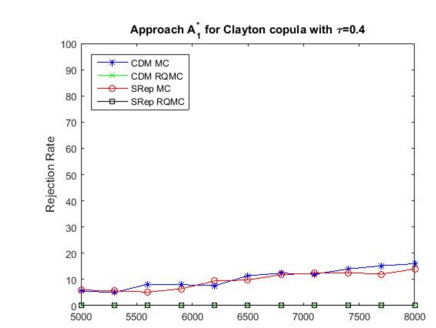
Apply Anderson-Darling test on $\{W_{B_j}\}_{j=1}^N$

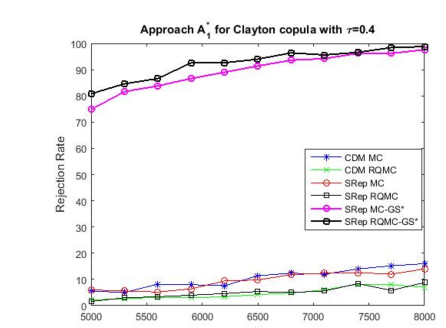
$\widehat{\mathcal{T}}$: AD Statistic with known distribution

Numerical Results

The null hypothesis is C_θ is a t copula.

1. Use samples from Clayton copula generated by different simulation methods.
2. Rejection rates are defined as the number of rejections of null hypothesis divided by the total number of repetitioins. Here, repeat each test 500 times to obtain the rejection rate.
3. For RQMC, Halton sequence with a random start (left) and linear scrambled Faure sequences (right) are used.





Conclusions/Current Research

1. Better GoF tests: we designed a new test based on the collision test.
2. Better simulation methods: we showed GS^* improves rejection rates dramatically.

References

[1] Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance, 15(7-8), 675-701.*

[2] Lindskog, F., RiskLab, E. T. H. (2000). Modelling dependence with copulas and applications to risk management. *Doctoral dissertation, Master Thesis, ETH Zürich.*

[3] Nguyen, N., Ökten, G. (2016). The acceptance-rejection method for low-discrepancy sequences. *Monte Carlo Methods and Applications, 22(2), 133-148.*