Potential Statistical Evidence for Copula Model Selection Based on Rényi Distance

Morteza Mohammadi*[1], Majid Hashempour[2], Mahdi Emadi[3]

[1]Department of Statistics, School of Sciences, University of Zabol, Zabol, Iran

[2]Department of Statistics, University of Hormozgan, Bandar Abbas, Iran

[3]Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

---

Abstract: In recent years, the study of copula has made significant progress, and the need for a simple and reliable method to select the appropriate copula model is essential. This paper proposes the use of copula-based Rényi distance as a criterion capable of evaluating potential statistical evidence in an experiment for model selection. In addition, we present two nonparametric estimation methods, empirical and Bernstein, to estimate this distance. Finally, we evaluate the accuracy of these methods through a simulation study.

Keywords: Copula model, Potential statistical evidence, Rényi distance.
Mathematics Subject Classification (2020): 62N05, 62F10, 62E10.

---

*Corresponding author: `mo.mohammadi@um.ac.ir`

1 Introduction

Copulas offer a flexible way to model multivariate dependencies, but choosing the right copula family is crucial for accurate statistical analysis. Parametric models can be restrictive, and nonparametric models lack precision, highlighting the need for a reliable and straightforward approach to copula model selection.

Evidence inference is a modern approach that relies solely on data for inference, avoiding subjective influences such as prior information. Royall (1997) introduced key principles of evidential inference, emphasizing the likelihood function as central to this method. Unlike traditional frequentist or Bayesian methods, Royall's paradigm focuses explicitly on evidence, criticizing the Neyman-Pearson and Fisher approaches for their lack of this concept. In essence, potential statistical evidence refers to data indicating that a hypothesis or model merits further exploration. This likelihood-based approach provides a coherent way to interpret data and guide research inquiries.

The Rényi divergence of order (or alpha-divergence) between two probability density functions f and g is defined as:

$$R_\alpha(f\|g) = \frac{1}{\alpha - 1} \log\left(\int f(x)^\alpha g(x)^{1-\alpha}\, dx\right)$$

where α controls the sensitivity of the divergence. As α approaches 1, the Rényi divergence converges to the Kullback-Leibler divergence. Rényi divergence has applications in hypothesis testing, coding theory, and assessing model fit. Researchers use it to compare distributions, quantify information loss, and select appropriate models. Remember that the Rényi divergence provides a flexible way to measure the difference between distributions, and it complements other statistical tools.

The objective of our paper is to establish a novel connection between copula model selection and symmetric Rényi distance that allow for the measure the potential statistical evidence. For ease of exposition, we focus exclusively on the bivariate continuous case. The second section of the paper will provide an overview of copulas. The relationship between the Rényi distance and potential statistical evidence will be investigated in the third section, and the copula-based Rényi distance will be introduced in the fourth section. In Section 5, nonparametric estimators for the copula-based Rényi distance will be presented, and finally, the accuracy of these estimators will be evaluated by a simulation study.

2 Copulas

Copulas serve as powerful tools for modeling dependence structures in multivariate data. Rather than focusing solely on individual marginal distributions, copulas capture the underlying dependence among variables. Copulas allow us to separate the joint distribution into univariate margins and a copula function. Consider a random vector (X, Y). Suppose its marginals are continuous, meaning their cumulative distribution functions (CDFs) are continuous functions. By applying the probability integral transform to each component, we transform (X, Y) into a new random vector with marginals that are uniformly distributed on the interval $[0, 1]$. Sklar's theorem establishes a connection between the joint distribution function and its copula as:

$$F(x, y) = C(F_X(x), F_Y(y)), \qquad (x, y) \in R^2, \tag{1}$$

where F is the joint CDF of (X, Y), F_X and F_Y are the marginal distributions of X and Y, and C is the copula function that describes the dependence structure.

Copulas provide a flexible framework for modeling various dependence patterns. By combining copulas with appropriate marginal distributions (e.g., Gaussian, t-distribution), we can construct families of bivariate distributions tailored to specific applications. These distributions are essential for risk assessment, portfolio optimization, and reliability analysis. In summary, copulas offer a rich area for exploration and practical use. In Table 1, three commonly used copulas including Gaussian, Clayton and Gumbel copulas and their characteristics are presented.

Table 1: Copula Formulas and Parameter Ranges

Copula	Formula	Parameter Range
Gaussian	$C(u_1, u_2; \rho) = \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\Phi^{-1}(u_1) - \rho\Phi^{-1}(u_2)\right]^2\right\}$	$-1 \leq \rho \leq 1$
Clayton	$C(u_1, u_2; \theta) = \max\left\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\right\}$	$\theta > 0$
Gumbel	$C(u_1, u_2; \theta) = \exp\left\{-\left((-\log u_1)^\theta + (-\log u_2)^\theta\right)^{1/\theta}\right\}$	$\theta > 1$

The Gaussian Copula is suitable for elliptical dependence, capturing both positive and negative correlations. It is also tail-independent. The Clayton Copula allows specific non-zero lower tail dependency and is commonly used for modeling credit risk and insurance claims. It is an exchangeable copula. The Gumbel Copula permits a specific level of upper tail dependence, finds applications in extreme value theory and hydrology, and is also exchangeable. For further study regarding copulas and their properties refer to Joe (1997).

3 Potential statistical evidence and Rényi distance

In statistical experiments, we often seek evidence to support hypotheses or model choices. Habibi et al. (2006) introduced a pre-experimental criterion based on Kullback-Leibler distance to assess the potential strength of evidence provided by an experiment. However, we can also investigate potential statistical evidence using Rényi distance. Rényi divergence measures the dissimilarity between two probability distributions based on their empirical copulas or density functions. It provides a unified framework for model selection and goodness-of-fit assessment.

The central tenet of evidential inference is the likelihood function. This approach relies on the likelihood principle and the law of likelihood. The likelihood principle asserts that any statistical inference should be based on the likelihood function. The law of likelihood states that based on the likelihood ratio $R(x) = \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$, the data support for $\theta = \theta_1$ is greater (or less) than for $\theta = \theta_0$ when $R(x) > 1$ (or $R(x) < 1$), and the data provide equal support for both parameter values when $R(x) = 1$.

Consider an experimenter who wants to choose between two experiments with approximately equal costs. The experimenter aims to select the experiment that has greater potential statistical evidence. Let x represent observations from experiment E, where X is a random vector with density $f_\theta(x)$.

Habibi et al. (2006) introduced the following measure for assessing potential statistical evidence in experiment E:

$$S_\psi(E) = E_{\theta_1}\Big[\psi(R_E(X))\Big] + E_{\theta_0}\Big[\psi(\frac{1}{R_E(X)})\Big]. \tag{2}$$

Here, ψ is an increasing function. If $\psi(t) = \log(t)$, then symmetric Kullback-Leibler (K-L) distance is obtained. Alos, If $\psi(t) = 1 - \frac{1}{\sqrt{t}}$, then symmetric Hellinger distance is obtained. These distances provide valuable insights into the statistical evidence and divergence between different distributions. Researchers often choose the appropriate measure based on the context and specific goals of their analysis.

The generalization of Criterion B is presented by Abbasnejad and Arghami (2006) as follows:

$$S_{\psi_1,\psi_2}(E) = \psi_2\Big[E_{\theta_1}\big[\psi_1(R_E(X))\big]\Big] + \psi_2\Big[E_{\theta_0}\big[\psi_1(\frac{1}{R_E(X)})\big]\Big]. \tag{3}$$

where both ψ_1 and ψ_2 are non-decreasing or non-increasing functions. If we take $\psi_1(t) = t^{\alpha-1}$ and $\psi_2(t) = \frac{1}{\alpha-1}\log(t)$, $\alpha > 0$ and $\alpha \neq 1$, then

$$\begin{aligned} S_{\psi_1,\psi_2}(E) &= \frac{1}{\alpha-1}\log\Big[E_{\theta_1}\big[\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}\big]\Big] + \frac{1}{\alpha-1}\log\Big[E_{\theta_0}\big[\frac{f_{\theta_0}(x)}{f_{\theta_1}(x)}\big]\Big] \\ &= R_\alpha(f_1\|f_0) + R_\alpha(f_0\|f_1) \\ &= RD_\alpha(f_0\|f_1) \end{aligned} \tag{4}$$

where R_α and RD_α are, respectively, asymmetric and symmetric Rényi distance of f_1 and f_0.

4 Copula based Rényi Distance

The Rényi distance between two copula functions, denoted as C_1 and C_2, is given by:

$$\begin{aligned} RD_\alpha(C_1\|C_2) &= R_\alpha(C_1\|C_2) + R_\alpha(C_2\|C_1) \\ &= \frac{1}{\alpha-1}\Big[\log\Big(\int_{[0,1]^2} C_1^\alpha(u,v)C_2^{1-\alpha}(u,v)\ dudv\Big) \\ &\qquad + \log\Big(\int_{[0,1]^2} C_2^\alpha(u,v)C_1^{1-\alpha}(u,v)\ dudv\Big)\Big]. \end{aligned} \tag{5}$$

Here, α is a parameter that influences the sensitivity of the distance measure. As α approaches 1, the Rényi distance converges to the Kullback-Leibler (K-L) divergence. As α approaches infinity, it converges to the Hellinger distance.

A small Rényi distance indicates similarity between the copulas. A large distance implies significant dissimilarity in their dependence structures. Researchers use the Rényi distance for model comparison, goodness-of-fit assessment, and detecting changes in dependence patterns. It provides a flexible alternative to other divergence measures. The Rényi distance between two copula functions, denoted as C_1 and C_2, has several important properties:

1. The Rényi distance is always non-negative: $RD_\alpha(C_1\|C_2) \geq 0$. It measures the divergence or dissimilarity between the two copulas.

2. The distance is symmetric: $RD_\alpha(C_1\|C_2) = RD_\alpha(C_2\|C_1)$. This property ensures that the order of the copulas does not affect the distance.

3. The Rényi distance satisfies the triangle inequality:

$$RD_\alpha(C_1\|C_2) \leq RD_\alpha(C_1\|C_3) + RD_\alpha(C_2\|C_3).$$

In other words, the direct path between C_1 and C_2 is no longer than the sum of the paths through an intermediate copula C_3.

Recently, Mohammadi and Emadi (2023) and Mohammadi et al. (2024) have introduced independence tests based on divergence measures and copulas. Also, Mohammadi et al. (2021) has presented dependence measures based on copula-based symmetric divergence measures.

5 Nonparametric Estimator

To estimate the bopula-based Rényi distance, we propose a nonparametric estimator. This estimator is consistent and computationally efficient. It adapts to the sample size of the data, making it suitable for practical applications. The estimation of Rényi distance for copula model selection can be considered as:

$$\widehat{RD}_\alpha(C_1\|\hat{C}_2) = \frac{1}{\alpha-1}\Big[\log\Big(\int_{[0,1]^2} C_1^\alpha(u,v)\hat{C}_2^{1-\alpha}(u,v)\ dudv\Big) + \log\Big(\int_{[0,1]^2} \hat{C}_2^\alpha(u,v)C_1^{1-\alpha}(u,v)\ dudv\Big)\Big]. \tag{6}$$

Here, C_1 is the competing copula and $\hat{C}_2$ is the nonparametric estimate of the correct copula based on real data.

In this study, we use empirical copula and empirical Bernstein copula methods for the estimation of C_2. The empirical copula is a nonparametric estimator of the copula function based on observed data. The empirical copula is a powerful tool for understanding dependence patterns in real-world data, especially when we don't make strong assumptions about the underlying distributions Let $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ be a random sample of size n from a pair (X, Y). Empirical copula defined as

$$\hat{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} I\{\tilde{U}_i \le u, \tilde{V}_i \le v\}, \tag{7}$$

where $\tilde{U}_i = n\hat{F}_X(x_i)/(n+1)$, $\tilde{V}_i = n\hat{F}_Y(y_i)/(n+1)$ for $i = 1, \cdots, n$, are the pseudo observations and $\hat{F}_X$ and $\hat{F}_Y$ are the empirical cumulative distribution function of the observation X_i and Y_i, respectively.

The empirical copula can be extended nonparametrically as the Bernstein empirical copula by Maldonado et al. (2024) and is formulated as

$$\hat{C}_n^{\mathcal{B}}(u,v;\eta) = \sum_{i=1}^{n}\sum_{j=1}^{n} \hat{C}_n\left(\frac{i}{n},\frac{j}{n}\right) \times \eta(i,j;u,v),$$

where the individual Bernstein weights $\eta(i,j;u,v)$ for the k-th paired value of the u and v vectors are

$$\eta(i,j;u,v) = \binom{n}{i} u^i(1-u)^{n-i} \times \binom{n}{j} v^j(1-v)^{n-j}.$$

The Bernstein empirical copula estimates the dependence structure between components of (X, Y) without assuming any specific parametric form. It adapts the empirical copula using Bernstein polynomials, providing a smoother estimate for arbitrary dimensions.

6 Simulation Study

We perform extensive simulations to assess the performance of our method. In this study, we investigated the estimation of the Rényi distance for the baseline copula using two methods: empirical estimation and the Bernstein method for non-parametric copula estimation. For our simulations, we utilized Gaussian, Clayton, and Gumbel copulas, considering various parameters based on weak (0.2), moderate (0.5), and strong (0.8) dependency levels as defined by Kendall's tau (τ). Additionally, we analyzed two sample sizes 50 and 200. The data were simulated from the correct copula (C_1) with parameters corresponding to the values of Kendall's tau. Subsequently, the Renyi distance was computed between these simulated data and the competing copula (C_2), for which the parameters were estimated using the maximum likelihood method.

The simulation results in Table 2 for moderate dependency level ($\tau = 0.5$) indicated that the accuracy of the Bernstein method in estimating the Renyi distance for the baseline copula outperformed that of the empirical method. Consequently, selecting copulas based on the Bernstein method yields more precise results. Furthermore, as the sample size increases, the accuracy of the estimates in selecting the appropriate copula also improves.

Table 2: Simulation Results for Renyi Distance Estimation ($\tau = 0.5$)

Correct Copula	Competing Copula	Sample Size 50		Sample Size 200	
		Bernstein	Empirical	Bernstein	Empirical
Gaussian	Gaussian	**0.1152**	0.1708	**0.0769**	0.1320
	Clayton	0.3373	0.3449	0.3063	0.3100
	Gumbel	0.3401	0.3508	0.3063	0.3174
Clayton	Gaussian	0.2264	0.2627	0.1884	0.2305
	Clayton	**0.1213**	0.1748	**0.0900**	0.1396
	Gumbel	0.3429	0.3547	0.3110	0.3233
Gumbel	Gaussian	0.2328	0.2671	0.2001	0.2331
	Clayton	0.2382	0.2720	0.2026	0.2328
	Gumbel	**0.1261**	0.1810	**0.0929**	0.1484

Conclusion

The increasing complexity of statistical modeling has highlighted the importance of copulas in capturing dependencies among random variables. This paper has addressed the critical need for a straightforward and reliable method for selecting the appropriate copula model. By utilizing the copula-based Rényi distance as a criterion for model selection, we have demonstrated its effectiveness in evaluating potential statistical evidence in experimental settings. Through our comparative analysis of two nonparametric estimation methods—empirical and Bernstein—we found that the Bernstein method consistently outperforms the empirical method, particularly as sample sizes increase. The simulation study revealed that the accuracy of the Bernstein method enhances the reliability of copula selection, making it a valuable tool for researchers and practitioners alike. Overall, our findings underscore the utility of the copula-based

Rényi distance in model selection and contribute to the ongoing development of robust statistical methodologies. Future research could explore further refinements to these estimation techniques and their applications across various fields, thereby extending the impact of copula theory in practical scenarios.

References

Abbasnejad, M., and Arghami, N. R. (2006), Potential Statistical Evidence in Experiments and Renyi Information, *Journal of the Iranian Statistical Society*, **5**(1), 39–52.

Habibi, A., Arghami, N. R., and Ahmadi, J. (2006), Statistical evidence in experiments and in record values, *Communications in Statistics-Theory and Methods*, **35**(11), 1971–1983.

Joe H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

Maldonado, H. V. M., Erdely, A., Díaz-Viera, M., and Rios, L. (2024), Fast procedure to compute empirical and Bernstein copulas, *Applied Mathematics and Computation*, **477**, 128827.

Mohammadi, M., and Emadi, M. (2023), Nonparametric tests of independence using copula-based Renyi and Tsallis divergence measures, *Statistics, Optimization and Information Computing*, **11**(4), 949–962.

Mohammadi, M., Emadi, M., and Amini, M. (2021), Bivariate Dependency Analysis using Jeffrey and Hellinger Divergence Measures based on Copula Density Estimation by Improved Probit Transformation, *Journal of Statistical Sciences*, **15**(1), 233–254.

Mohammadi, M., Emadi, M., and Amini, M. (2024), Testing bivariate independence based on -divergence by improved probit transformation method for copula density estimation, *Communications in Statistics-Simulation and Computation*, **53**(1), 499–517.

Royall, R. (1997), *Statistical Evidence: A Likelihood Paradigm* (Vol. 71), CRC Press, New York.