Copula Entropy

Theory and Applications

MA Jian
majian.2003@tsinghua.org.cn

Contents

Lecture 5: Theoretical applications (3)

Copula Entropy: Application 7

Multivariate Normality Test[1]

[1] Jian Ma. "Multivariate Normality Test with Copula Entropy". In: *arXiv preprint arXiv:2206.05956* (2022).

Copula Entropy: Multivariate Normality Test

- Problem
  - To test the hypothesis that the distribution of data is normal distribution

$$H_0 : \mathbf{X} \sim N(\mu, \Sigma); H_1 : \mathbf{X} \nsim N(\mu, \Sigma) \tag{1}$$

- Significance
  - Normal distribution is the most important distribution in probability theory due to central limit theorem
  - Normality is a common assumption of many statistical tools
  - Testing normality is widely needed in real applications
- Related Methods
  - characteristics function based
  - moments based
  - skewness and kurtosis
  - energy distance based
  - entropy based
  - Wasserstein distance based

Copula Entropy: Multivariate Normality Test

- The proposed statistic

$$T_{mvnt} = H_c(\mathbf{x}) - H_c(\mathbf{x}_n), \tag{2}$$

where $\mathbf{x}_n$ is the Gaussian random vector with the same correlation matrix as $\mathbf{x}$.

  - defined as the difference of copula entropies
  - $T_{mvnt} = 0$ if normal distributions

- The estimator
  - the first term in (2) can be estimated with the non-parametric CE estimator;
  - the second term in (2) can be estimated easily by first estimating the correlation matrix Σ_x of $\mathbf{x}$ and then calculating the result according to (3).

$$H_c(\mathbf{x}_n) = \frac{1}{2}\log|\Sigma_x|. \tag{3}$$

Copula Entropy: Multivariate Normality Test

Simulation Experiments

- Data
  - bivariate normal copula with normal and exponential marginals
  - bivariate Gumbel copula with normal marginals
- Compared methods
  - Mardia's
  - Royston's
  - Henze and Zirkler's
  - Doornik and Hansen's, and
  - the energy distance based test by Rizzo and Székely

Copula Entropy: Multivariate Normality Test

Simulation Results

See the results in **Lecture 6: Evaluation**

Copula Entropy: Application 8

Copula Hypothesis Testing[2]

[2] Jian Ma. "Testing Copula Hypothesis with Copula Entropy". In: *arXiv preprint arXiv:2510.22722* (2025).

Copula Entropy: Copula Hypothesis Testing

- Problem
  - To test which copula hypothesis is accepted for sample data

$$H_0 : c_{\mathbf{x}}(\mathbf{u}) = c(\mathbf{u}); H_1 : c_{\mathbf{x}}(\mathbf{u}) \neq c(\mathbf{u}) \tag{4}$$

- Significance
  - Testing copula hypothesis is a fundamental problem in applications of copula theory
  - Many copula hypothesis test exist for specific types of copula function
  - A general method for any copula function is needed
- Related Methods
  - Gaussian copula hypothesis testing
  - Archimedeanity test
  - Copula Information Criteria
  - Goodness-of-Fit tests for copulas

Copula Entropy: Copula Hypothesis Testing

- The proposed statistic

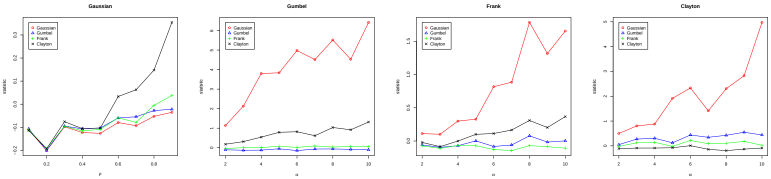
$$T_c(\mathbf{X}_T|c) = H_c(\mathbf{X}_T|c) - H_c(\mathbf{X}_T|c_{\mathbf{x}}) \tag{5}$$

  - defined as the difference of true copula entropy and the copula entropy of hypothesis c
  - $T_c(\mathbf{X}_T|c) = 0$ if $c_{\mathbf{x}} = c$
- The estimator
  - The second term in (5) can be estimated directly from data with the nonparametric CE estimator;
  - The first term can be estimated parametrically in the 3 steps:
    1. estimate empirical copula density $\widehat{\mathbf{u}}$ from $\mathbf{X}_T$;
    2. estimate the parameters α of copula density c with $\widehat{\mathbf{u}}$;
    3. calculate the CE of the copula hypothesis with the following equation:

$$H_c(\mathbf{X}_T|c) = -E(\log c(\widehat{\mathbf{u}}; \alpha)). \tag{6}$$

Copula Entropy: Copula Hypothesis Testing

Simulation



(a) Gaussian (b) Gumbel (c) Frank (d) Clayton

Figure: Simulation results on Gaussian Copula and Archimedean Copulas.

See the results in **Lecture 6: Evaluation**

Copula Entropy: Application 9

Two-Sample Test[3]

[3] Jian Ma. "System Identification with Copula Entropy". In: *arXiv preprint arXiv:2304.12922* (2023).

Copula Entropy: Two-Sample Test

- Problem
  - To test the hypothesis that two samples are from a same distribution

$$H_0 : P_0 = P_1; H_1 : P_0 \neq P_1 \tag{7}$$

- Significance
  - a basic hypothesis testing problem
  - Symmetry test and change point detection can be formulated as two-sample test problem
  - has many real applications in many areas, such as politics, medicine, etc
- Related Methods
  - T-test or F-test
  - Kernel-based two-sample test
  - Kolmogorov-Smirnov test
  - Mutual information based test

Copula Entropy: Two-Sample Test

- The proposed statistic

$$T_{tst} = H_c(\mathbf{X}, Y_0) - H_c(\mathbf{X}, Y_1), \tag{8}$$

where $\mathbf{X} = (X_1, X_2)$ is for two samples $X_1 = \{X_{11}, \cdots, X_{1m}\}$ and $X_2 = \{X_{21}, \cdots, X_{2n}\}$, and $Y_0 = (1_1, \cdots, 1_{m+n})$ and $Y_1 = (0_1, \cdots, 0_m, 1_1, \cdots, 1_n)$ are the labels for the null and the alternative hypothesis respectively.

  - non-parametric multivariate two-sample test
  - defined as the difference between the copula entropies of the null and the alternative hypothesis;
  - T_{tst} is small if H_0 is true.

- The estimator
  - estimating the two terms in (8);
  - calculating the estimated statistic.

Copula Entropy: Two-Sample Test

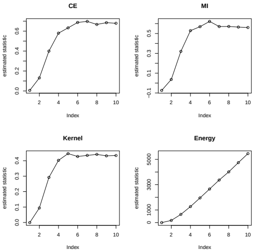
Simulation Experiments

- Data
  - bivariate normal distribution with different means
  - bivariate normal distribution with different variances
  - bivariate Gaussian copula with normal and exponential marginals
- Compared methods
  - Kernel-based test
  - Energy distance-based test
  - Mutual information-based test

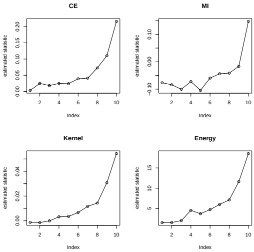
Copula Entropy: Two-Sample Test

Simulation Results

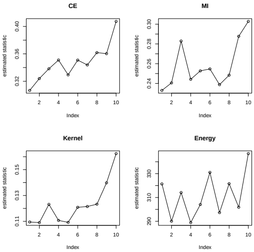
- Bivariate normal distribution with different means



- Bivariate normal distribution with different variances



- Bivariate normal copula with different variances



See the results in **Lecture 6: Evaluation**

Copula Entropy: Application 10

Change Point Detection[4]

[4] Jian Ma. "Root Cause Analysis on Energy Efficiency with Transfer Entropy Flow". In: *arXiv preprint arXiv:2401.05664* (2024).

Copula Entropy: Change Point Detection

- Problem
  - To detect single or multiple disrupt change in time series data
- Significance
  - a basic problem in time series analysis
  - can be solved with two-sample test problem
  - has aboard real applications in many areas, such as geoscience, biology, manufacturing, etc.
- Existing Methods
  - CUSUM
  - Kernel-based method
  - two-sample test based
  - multiple detection with binary segmentation

Copula Entropy: Change Point Detection

- Idea

$$H_0 : P_b(\mathbf{x}) = P_a(\mathbf{x}); H_1 : P_b(\mathbf{x}) \neq P_a(\mathbf{x}) \tag{9}$$

- The proposed method
  - Single change point detection
    solved with copula entropy based two-sample test

$$t_c = \arg\max_t T_{tst}(\mathbf{X}_b, \mathbf{X}_a; t) \tag{10}$$

  - Multiple change point detection
    Single change point detection with binary segmentation
- Merits
  - nonparametric method based on CE estimator
  - applicable to univariate or multivariate cases
  - distribution-free, universally applicable
  - almost tuning-free due to the scale-free test statistic

Copula Entropy: Change Point Detection

Simulation Experiments

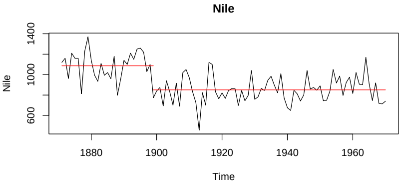
- Univariate and multivariate multiple change points
  - changes in mean
  - changes in mean and variance
  - changes in variance
- Compared methods
  - The methods in the R package **changepoint**
  - Kernel based method in the R package **ecp**

See the results in **Lecture 6: Evaluation**

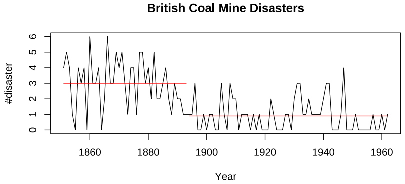
Copula Entropy: Change Point Detection

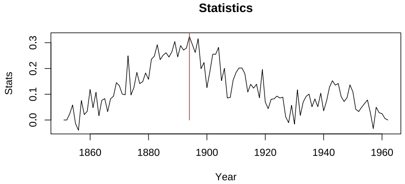
Real data experiment

- The Nile data



- British coal mine disasters data





Copula Entropy: Application 11

Symmetry Test[5]

[5] Jian Ma. "Testing Symmetry with Copula Entropy based Two-Sample Test". In: *ChinaXiv preprint ChinaXiv:202505.00167* (2025).

Copula Entropy: Symmetry Test

- Problem
  - To test symmetry of distribution of samples

$$H_0 : p(u-x) = p(u+x); H_1 : p(u-x) \neq p(u+x) \tag{11}$$

- Significance
  - a fundamental principle of physics
  - an assumption of many statistical models and methods
  - an essential theme in hypothesis testing
- Related work
  - Symmetry tests with copulas
  - Testing symmetry of copulas
  - Entropy based symmetry tests

Copula Entropy: Symmetry Test

- The proposed statistic

$$T_{sym}(X) = T_{tst}(\tilde{X}, -\tilde{X}). \tag{12}$$

  - defined as the statistic of two-sample test on sample and its symmetric transformation
  - $T_{sym}(X) = 0$ if distribution is symmetric

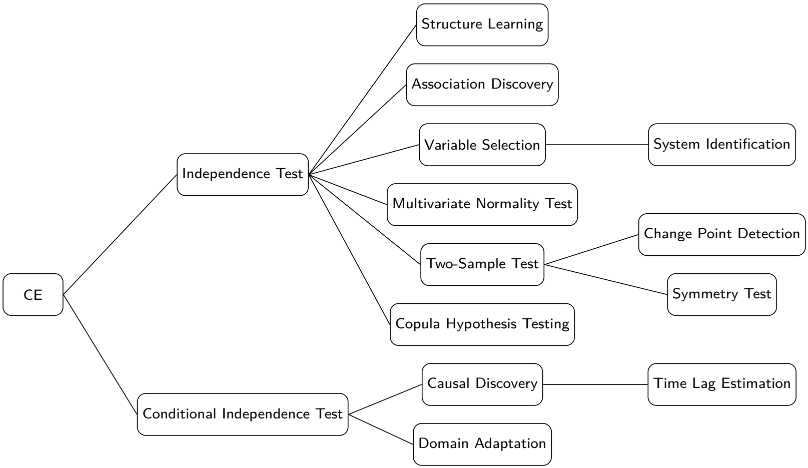
- The estimator
  Then the proposed method composed of two simple steps:
  1. deriving $\tilde{X}$ from X by $\tilde{X} = X - \tilde{u}$;
  2. estimating T_{sym} by performing two-sample test on $\tilde{X}$ according to (12).

- More symmetry tests
  can be proposed similarly with other symmetry transformations, such as exchangeability test by permutation.

Copula Entropy: Symmetry Test

Simulation Experiments
See the results in **Lecture 6: Evaluation**

CE based Framework of Statistical Methodologies



Monograph

Jian Ma. "Copula Entropy: Theory and Applications". In: *arXiv preprint arXiv:2025.18168* (2025)

Thanks!

Q&A