

Copula Entropy

Theory and Applications

Jian MA, PhD
majian03@gmail.com

Hitachi

Feb 16, 2023

Contents

1 Theory

- Copula Theory
- Theory of Copula Entropy
- Estimating Copula Entropy

2 Applications

- Association Discovery
- Structure Learning
- Variable Selection
- Causal Discovery
- Time Lag Identification
- Multivariate Normality Test

3 Summary

Copula Theory

Definition (Copula)

^a Given N random variables $\mathbf{X} = (X_1, \dots, X_N) \in \mathcal{R}^N$. Let $\{u_i = F_i(x_i), i = 1, \dots, N\}$ be the marginal distribution functions of \mathbf{X} . A N -dimensional copula $C : \mathcal{I}^N \rightarrow \mathcal{I}(\mathcal{I} = [0, 1])$ of \mathbf{X} is a function with following properties:

- ① C is grounded and N -increasing
 - ② $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$.

^a Roger B Nelsen. *An introduction to copulas*. Springer, 2007.

- the theory on **representation** of statistical dependence in probability
 - copula function contains all the dependence information between random variables
 - a probability function on unit cubic

Copula Theory

Theorem (Sklar's Theorem)

^a Given a random vector $\mathbf{X} = (X_1, \dots, X_N)$, its CDF $\mathbf{F}(\mathbf{x})$ can be represented as

$$\mathbf{F}(\mathbf{x}) = C(u_1, \dots, u_N), \quad (1)$$

where C is a copula function, $\{u_i\}$ are marginal distribution functions of \mathbf{X} . If $\{F_i\}$ are continuous, then C is unique.

^aM. Sklar. "Fonctions de répartition à n dimensions et leurs marges". In: *Publ. Inst. Statist. Univ. Paris* 8 (1959), pp. 229–231.

- the core of copula theory
 - there exists a copula function for each multivariate probability function

Copula Theory

Corollary

The probabilistic density function (PDF) $p(\mathbf{x})$ of \mathbf{X} can be represented as

$$p(\mathbf{x}) = c(\mathbf{u}) \prod_{i=1}^N p_i(x_i), \quad (2)$$

where $\{p_i, i = 1, \dots, N\}$ are marginal density functions of \mathbf{X} , and c is copula density.

- separating dependence representation with properties of individual variables

Copula Entropy: Theory

Definition (Copula Entropy)

Let \mathbf{X} be random variables with marginals \mathbf{u} and copula density $c(\mathbf{u})$. Copula Entropy of \mathbf{X} is defined as

$$H_c(\mathbf{x}) = - \int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u}. \quad (3)$$

- a special type of Shannon entropy
 - an ideal measure of statistical independence
 - distribution-free

Copula Entropy: Theory

Theorem

Mutual Information of \mathbf{X} is equivalent to its negative copula entropy.

$$I(\mathbf{x}) = -H_c(\mathbf{x}). \quad (4)$$

Corollary

$$H(\mathbf{x}) = \sum_i H_i(x_i) + H_c(\mathbf{x}). \quad (5)$$

- the bridge between copula theory and information theory¹

¹ Jian Ma and Zengqi Sun. "Mutual information is copula entropy". In: *Tsinghua Science & Technology* 16.1 (2011). See also arXiv preprint arXiv:0808.0845 (2008), pp. 51–54.

Copula Entropy: Theory

- Axiomatic Properties of Copula Entropy
 - multivariate
 - symmetric
 - non-negative, 0 iff independence
 - invariant to monotonic transformation
 - equivalent to correlation coefficient in Gaussian cases
- An ideal measure compared with others

Table: Comparison with other independence measures.

	Copula Entropy	Distance Correlation	HSIC
Definition	copula based	generalised corr	corr in RKHS
Multivariate	Yes	distance multivariance	dHSIC
Invariance	monotonic trans	No	No
Gaussianity	equivalent to cc	unclear	unclear
Computation	low	high	high

Copula Entropy: Estimation

- Non-Parametric Estimation Method²

- ① estimating empirical copula density with rank statistics
- ② estimating copula entropy with kNN entropy estimation method

- Advantages

- distribution-free, non-parametric
- tuning-free, insensitive to parameters
- good convergence
- easy to implement
- low computation burden

² Jian Ma and Zengqi Sun. "Mutual information is copula entropy". In: *Tsinghua Science & Technology* 16.1 (2011). See also arXiv preprint arXiv:0808.0845 (2008), pp. 51–54.

Copula Entropy: Application I

Association Discovery³

³ Jian Ma. "Discovering Association with Copula Entropy". In: *arXiv preprint arXiv:1907.12268* (2019).

Copula Entropy: Association Discovery

- Problem
 - To discover association relationship between random variables from data
- History
 - An old and fundamental problem since statistics birth
- Related Methods
 - Pearson Correlation Coefficient
 - Regression

Copula Entropy: Association Discovery

- Traditional association measures
 - Pearson Correlation Coefficient

$$r_{XY} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\delta_X \delta_Y} \quad (6)$$

- Spearman's ρ and Kendall's τ

$$\rho_{XY} = 12 \int_u \int_v C(u, v) dudv - 3 \quad (7)$$

$$\tau_{XY} = 4 \int_u \int_v C(u, v) dC(u, v) - 1 \quad (8)$$

- Why Copula Entropy?

Table: Theoretical comparison between CE and CC.

	CC	CE
linearity	linear	nonlinear
Order	2	≥ 2
Assumption	Gaussian	None
variate	bivariate	multivariate

Copula Entropy: Association Discovery

Experiments on the NHANES data

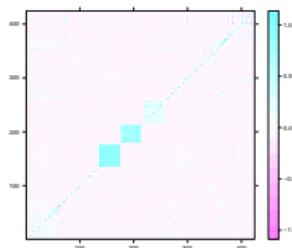
- Objectives of NHANES
 - to monitor trends and emerging issues of population health
 - to investigate its relationship with risk factors, nutritions and environmental exposures, etc.
- NHANES (2013-2014)
 - 14,332 persons from 30 different survey locations were selected;
 - Of those selected, 10,175 interviewed and 9,813 examined;
 - 5 groups of data: demographics, dietary, examination, laboratory, and questionnaire.
- Experimental data

The laboratory data, which includes 423 variables from blood, urine, oral rinse and vaginal/Penile swabs.
- Missing values

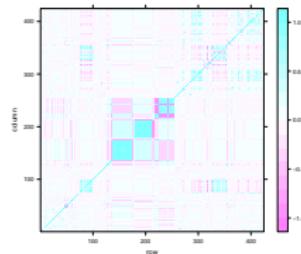
The missing values were filled with the mean of their corresponding variables.

Copula Entropy: Association Discovery

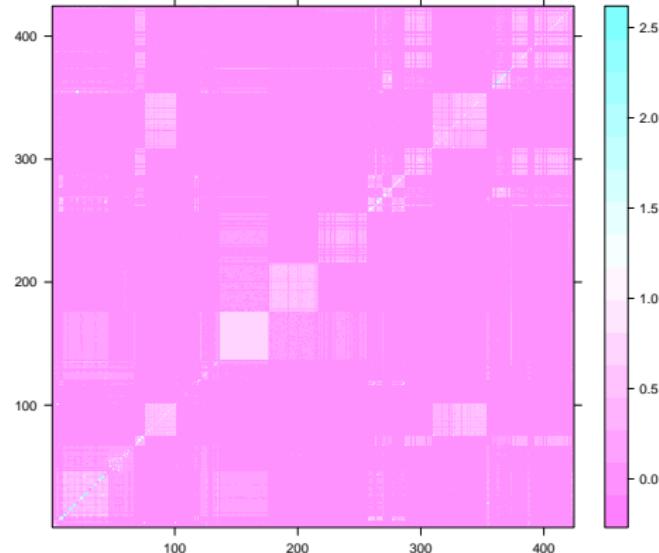
- Results - Correlation matrices



Pearson's r



Spearman's ρ



Copula Entropy

Copula Entropy: Association Discovery

- Results - Variable groups with meanings

Table: Variable groups with biomedical meanings discovered with CE.

Group	Index	Variables
1	288-302	Polycyclic Aromatic Hydrocarbones (PAH) - Urine
	68-75	Copper, Selenium & Zinc - Serum
	395-420	Urine Metals
2	358-373	Blood Lead, cadmium, total Mercury, Selenium, and Manganese
	269-276	Blood mercury: inorganic, ethyl and methyl
3	277-287	Oral Glucose Tolerance Test
	258-262	Insulin
	7-9	Cholesterol-LDL, Triglyceride&Apolipoprotein(ApoB), WTSAF2YR-Fasting Subsample 2 Year MEC Weight, LBXAPB-Apolipoprotein (B) (mg/dL), LBDAPBSI-Apolipoprotein (B) (g/L)
4	10-46	Standard Biochemistry Profile
	137-176	Human Papillomavirus (HPV) - Oral Rinse
5	76-101	Personal Care and Consumer Product chemicals and Metabolites
	327-353	Phthalates and Plasticizers Metabolites - Urines

Copula Entropy: Application II

Structure Learning⁴

⁴ Jian Ma and Zengqi Sun. "Dependence structure estimation via copula". In: *arXiv preprint arXiv:0804.4451* (2008).

Copula Entropy: Structure Learning

- Problem
 - To learn statistical structure among random variables from data
- Graph Representation
 - A probability density is represented with a directed or undirected graph, of which each node represents a random variable, and each edge represents a (conditional) dependence relation between two random variables
- Related Methods
 - Chow-Liu Algorithm

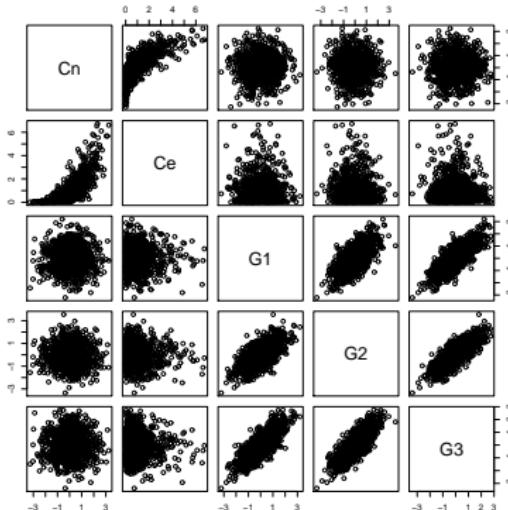
Copula Entropy: Structure Learning

- Our Algorithm
 - ① computing dependence matrix \mathbf{W}_x of data x with CE estimation
 - ② constructing dependence structure T from \mathbf{W}_x with MST algorithm
- Advantages
 - distribution-free, non-parametric
 - tuning-free, insensitive to parameters
 - easy to implement
 - low computation burden

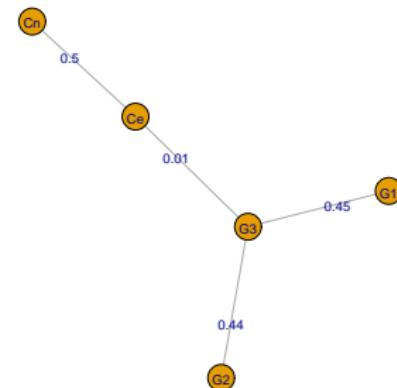
Copula Entropy: Structure Learning

- Simulated Experiment

5 random variables: the first three are Gaussian and the others two are governed by Gaussian copula with margins as normal distribution and exponential distribution respectively



Simulated data



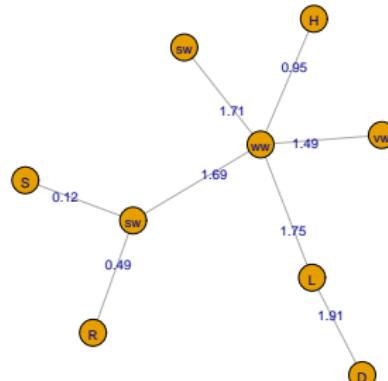
Learned graph

Copula Entropy: Structure Learning

Experiment on real data

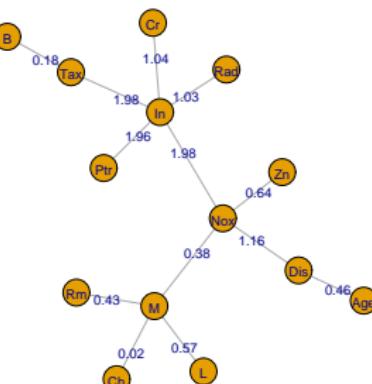
- Abalone data

Predicting the age of abalone from physical measurements



- Boston housing data

Concerns housing values in suburbs of Boston



Copula Entropy: Application III

Variable Selection⁵

⁵ Jian Ma. "Variable Selection with Copula Entropy". In: *Chinese Journal of Applied Probability and Statistics* 37.4 (2021), pp. 405–420.

Copula Entropy: Variable Selection

- Problem
 - To select a 'right' subset of variables from the whole group for building classification or regression models with good predictability and interpretability
- History
 - An old and basic problem in statistics and machine learning
- Related Problems
 - Feature Selection
 - Model Selection

Copula Entropy: Variable Selection

Existing methods - Likelihood with penalty

- Information Criteria
with penalty on the number of parameters in the models

$$\text{AIC} = -2L + 2p \quad (9)$$

$$\text{BIC} = -2L + p \log N \quad (10)$$

- Penalized GLMs
with penalty on the nonzero coefficients in the GLMs

- LASSO
- Ridge Regression
- Elastic Net

$$\min_{\beta} \{L(\beta; y, \mathbf{X}) + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2\} \quad (11)$$

- Adaptive LASSO

$$\min_{\beta} \{L(\beta; y, \mathbf{X}) + \lambda \sum_{j=1}^p w_j |\beta_j|\} \quad (12)$$

Copula Entropy: Variable Selection

Existing methods - Statistical independence measures

- Distance Correlation

$$\text{dCor}(X, Y) = \frac{\nu^2(X, Y)}{\sqrt{\nu^2(X)\nu^2(Y)}}, \quad (13)$$

where $\nu^2(X, Y)$ be distance covariance.

- Hilbert-Schmidt Independence Criterion (HSIC)

$$\text{dHSIC}(P(\mathbf{X})) = ||\Pi(P(X_1) \otimes \dots \otimes P(X_d)) - \Pi(P(\mathbf{X}))||, \quad (14)$$

where Π be the mean embedding function associated with kernel functions.

Copula Entropy: Variable Selection

- CE based method

To select variables based on ranks of their negative CE values with target

- Advantages

- model-free, non-parametric
- tuning-free, insensitive to parameters
- interpretable with physical meanings
- supported by rigorous math
- science instead of art, compared with existing methods
- easy to implement, low computation burden

Copula Entropy: Variable Selection

Experiments on the UCI heart disease data⁶

- Overview of the data

The data set contains 4 databases (899 samples) concerning heart disease diagnosis. All attributes are numeric-valued. The data was collected from the four following locations:

- Cleveland clinic foundation;
- Hungarian Institute of Cardiology, Budapest;
- V.A. medical center, long beach, CA;
- University hospital, Zurich, Switzerland.

- Attributes

The data has 76 attributes (#58 'num' for diagnosis). Of them, 13 attributes are recommended by professionals as clinical relevant.

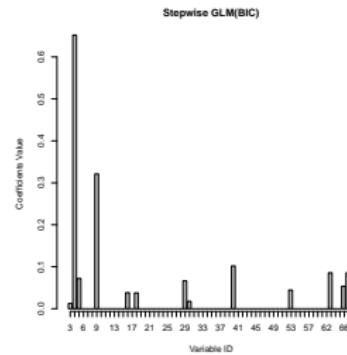
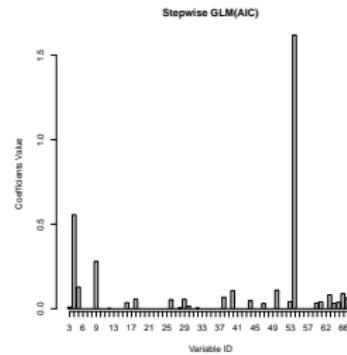
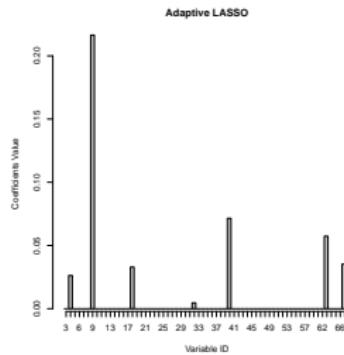
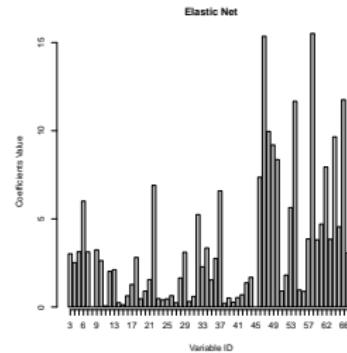
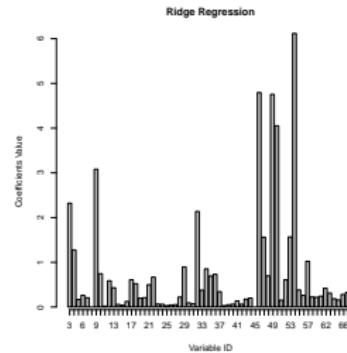
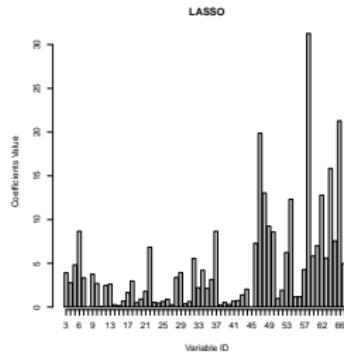
Table: Recommended attributes.

ID	3	4	9	10	12	16	19
Name	age	sex	cp	trestbps	chol	fbps	restecg
ID	32	38	40	41	44	51	58
Name	thalach	exang	oldpeak	slope	ca	thal	num

⁶Arthur Asuncion and David Newman. *UCI machine learning repository*. 2007.

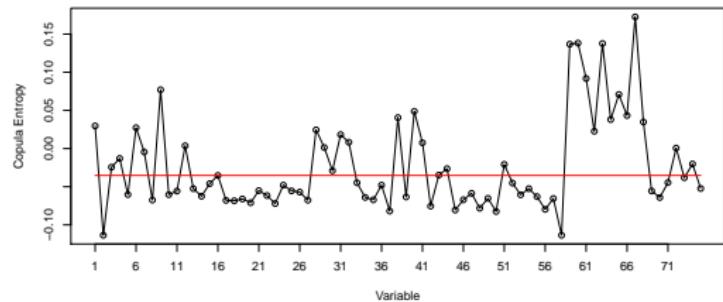
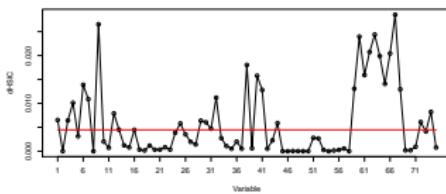
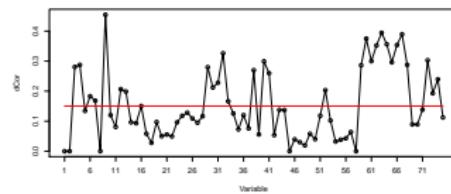
Copula Entropy: Variable Selection

- Results - Coefficients of penalized likelihood based models



Copula Entropy: Variable Selection

- Results - with statistical dependence measures (dCor, dHSIC, CE)



Copula Entropy: Variable Selection

- Results - Prediction accuracy

the selected variables present the best prediction accuracy.

Model	Accuracy(%)
SVM(Recommended variables)	84.20
SVM(CE)	84.76
SVM(dCor)	82.76
SVM(dHSIC)	84.54
Stepwise GLM(AIC)	51.8
Stepwise GLM(BIC)	49.1
LASSO	79.2
Ridge Regression	63.0
Elastic Net	75.9
Adaptive LASSO	35.7

Copula Entropy: Variable Selection

- Results - Selected variables

Copula Entropy selects more 'right' variables than the other methods do.

Method	Selected Variables' ID	✓
Recommended variables	3,4,9,10,12,16,19,32,38,40,41,44,51	13
CE	3,4,6,7,9,12,16,28-32,38,40,41,44,51,59-68	11
dHSIC	3,4,6,7,9,12,13,16,25,29-32,38,40,41,44,59-68	10
dCor	3,4,6,7,9,12,13,16,28-33,38,40,41,52,59-68	9
Stepwise GLM(AIC)	3,4,5,9,12,16,18,20,26,29,30,32,40,44,47,50,53,54,60,61,63,65-67	8
Stepwise GLM(BIC)	3,4,5,9,16,18,29,30,40,53,63,66,67	5
Adaptive LASSO	4,6,9,18,32,40,63,67	4
LASSO		
Ridge Regression	all except 8,45	-
Elastic Net		

Copula Entropy: Application IV

Causal Discovery⁷

⁷ Jian Ma. "Estimating Transfer Entropy via Copula Entropy". In: *arXiv preprint arXiv:1910.04375* (2019).

Copula Entropy: Causal Discovery

- Problem
 - To infer causality from time series data by *estimating Transfer Entropy*
- History & Significance
 - Causality is one of the oldest topics in philosophy.
 - Causal discovery is a central problem of all sciences.
- Correlation vs Causality
 - Correlation does not mean causation.
 - Correlation is only helpful for prediction while causality means intervention and control.

Copula Entropy: Causal Discovery

- Causality measures

- Wiener's Principle

Cause should improve the prediction of effect.

- Granger Causality

improvement measured by the variance of prediction error

$$\delta^2(Y_{t+1}|Y_t, X_t) < \delta^2(Y_{t+1}|Y_t) \quad (15)$$

- Transfer Entropy

improvement on the uncertainty of prediction measured by Shannon entropy

$$TE = \sum p(Y_{t+1}, Y^t, X_t) \log \frac{p(Y_{t+1}|Y^t, X_t)}{p(Y_{t+1}|Y^t)} \quad (16)$$

$$= H(Y_{t+1}|Y^t) - H(Y_{t+1}|Y^t, X_t) \quad (17)$$

$$= I(Y_{t+1}, X_t | Y^t) \quad (18)$$

- Issue on TE

difficult to estimate, some think impossible without model assumptions

Copula Entropy: Causal Discovery

- TE via CE

Proposition

Transfer Entropy can be represented with only Copula Entropy.

$$T_{x \rightarrow y} = -H_c(Y_{t+1}, Y^t, X_t) + H_c(Y_{t+1}, Y^t) + H_c(Y^t, X_t) - H_c(Y^t) \quad (19)$$

- Non-parametric Estimator of TE
 - ① estimating three or four CE terms in (19);
 - ② calculating TE for these estimated CEs.
- inheriting all the merits of non-parametric CE estimation

Copula Entropy: Causal Discovery

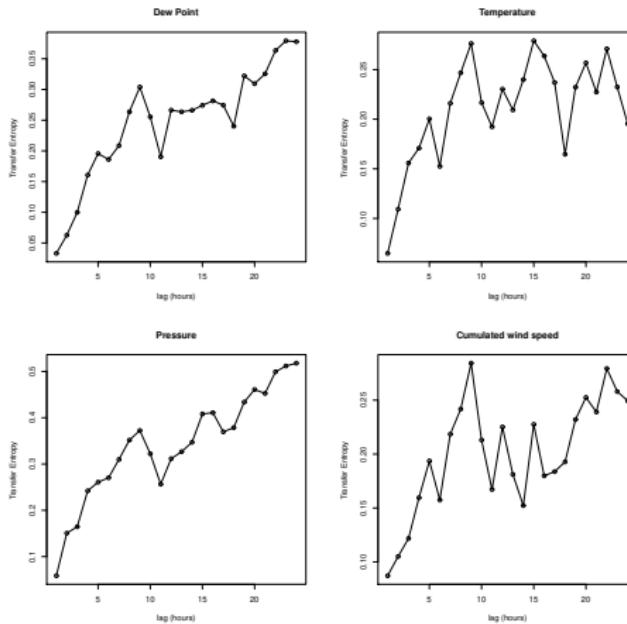
Experiments on the UCI Beijing PM2.5 data⁸

- Overview of the data
 - Time
 - hourly data from 2010-01-01 to 2014-12-31, which results in 43824 samples with missing values.
 - Observations
 - PM2.5 data of US Embassy in Beijing
 - Meteorological data from Beijing Capital International Airport
 - Meteorological factors
 - dew point, temperature, pressure, cumulated wind speed, combined wind direction, cumulated hours of snow, cumulated hours of rain.
- Experimental data
 - the first four factors used in the experiments;
 - 1000 samples without missing values (2010-04-02~2010-05-14).

⁸Arthur Asuncion and David Newman. *UCI machine learning repository*. 2007.

Copula Entropy: Causal Discovery

Results: Effects of meteorological factors on PM2.5



- Two phrases

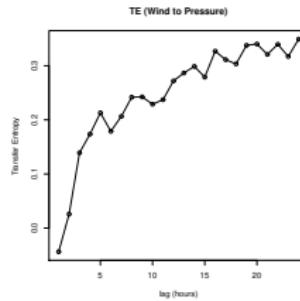
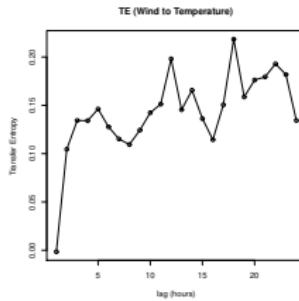
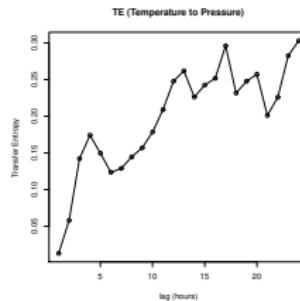
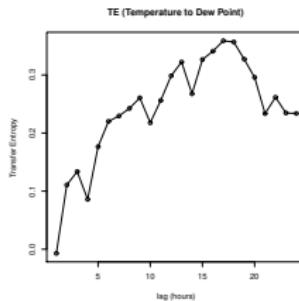
- Sharp increase phrase: the first 9 hours time lag, and peak at about 9 hours lag;
- Flat increase phrase: TE of Dew point and pressure increase with relatively flat rate while TE of temp. and cumulated wind speed does increase any more.

- Interpretation

- The effects do not show immediately and are cumulating processes.

Copula Entropy: Causal Discovery

Results - Effects between meteorological factors



- Temp. to Dew Point & Pressure
- Wind to Temp. & Pressure
 - Wind changes temperature in 3 hours later and
 - Wind changes pressure in 5 hours later.

Copula Entropy: Application V

Time Lag Identification⁹

⁹ Jian Ma. "Identifying Time Lag in Dynamical Systems with Copula Entropy based Transfer Entropy". In: *arXiv preprint arXiv:2301.06037* (2023).

Copula Entropy: Time Lag Identification

- Problem
 - To identify time lag in dynamical systems with copula entropy based transfer entropy
- Significance
 - Time lag is ubiquitous in physical, social, and biological systems.
 - Identifying time lag is of fundamental importance in applications of dynamical systems.
- Related Methods
 - Auto-correlation
 - Time-delayed mutual information

Copula Entropy: Time Lag Identification

- Our method

- ➊ estimating transfer entropies on time lag horizon from data with the CE-based estimator
- ➋ identifying the time lag associated with the maximum TE value

Copula Entropy: Time Lag Identification

- Simulations

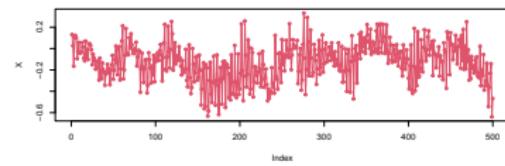
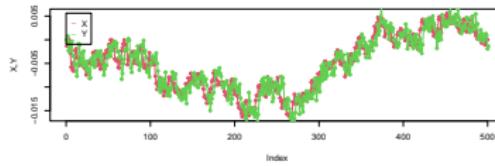
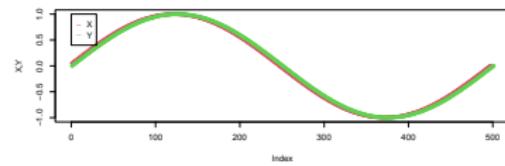
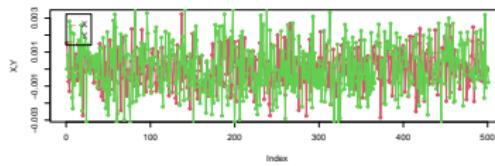
- ① generate trajectories from four simulated dynamical system with respect to different state or output lags
- ② identify the time lag with our method

- Simulated systems

- a system driven by random walk with output lag
- a system driven by sine function with output lag
- Wiener process with output lag
- a first-order linear system with state lag

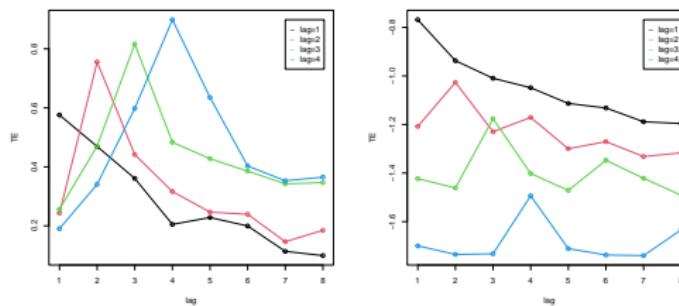
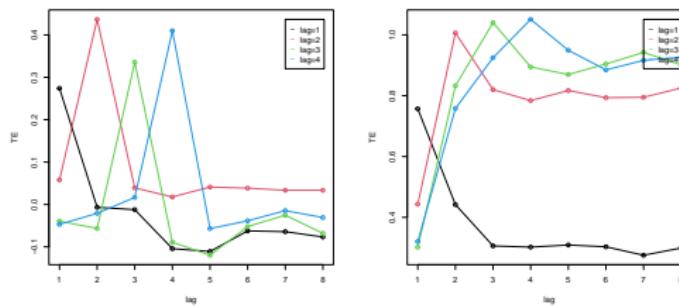
Copula Entropy: Time Lag Identification

- Simulated trajectories



Copula Entropy: Time Lag Identification

- Simulation: Results



Copula Entropy: Time Lag Identification

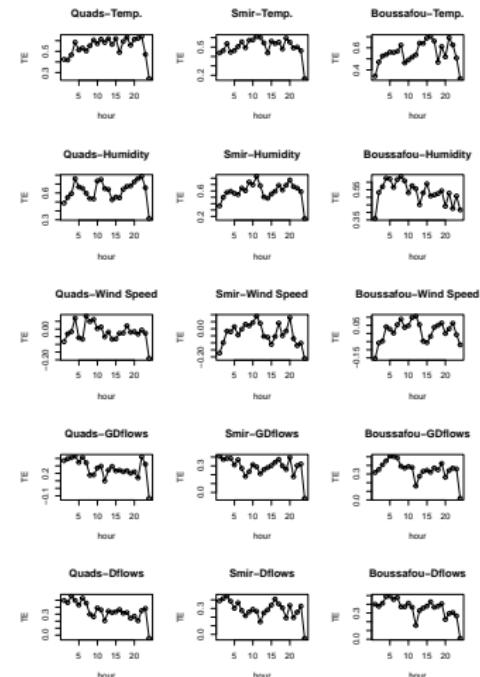
Power consumption of the Tetouan city¹⁰

- Data

- power consumption of 3 networks in 2017
- weather factors, including temperature, humidity, wind speed, general diffuse flows, diffuse flows

- Power consumption forecast

- To identify time lags from weather to power consumption



¹⁰ Arthur Asuncion and David Newman. *UCI machine learning repository*. 2007.

Copula Entropy: Application VI

Multivariate Normality Test¹¹

¹¹ 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
- Significance
 - Normal distribution is the most important distribution in probability theory;
 - 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

$$T_{ce} = H_c(\mathbf{x}) - H_c(\mathbf{x}_n), \quad (20)$$

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

- defined as the difference of copula entropies
 - $T_{ce} = 0$ if normal distributions
- The estimator
 - the first term in (20) can be estimated with the non-parametric CE estimator;
 - the second term in (20) can be estimated easily by first estimating the covariances V_x of \mathbf{x} and then calculating the result according to (21).

$$H_c(\mathbf{x}_n) = \frac{1}{2} \log |V_x|. \quad (21)$$

Copula Entropy: Multivariate Normality Test

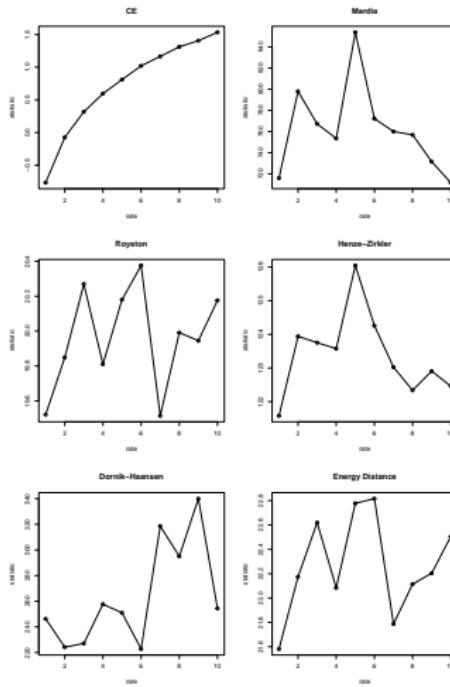
Simulation Experiments

- 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

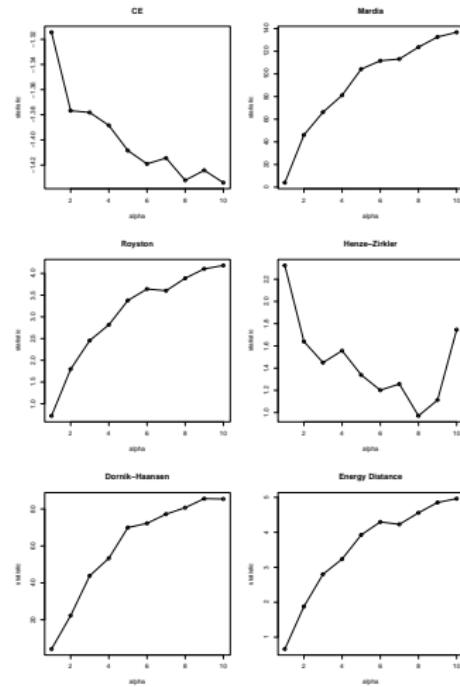
Copula Entropy: Multivariate Normality Test

Simulation Results

- Bivariate normal copula



- Bivariate Gumbel copula



Summary

- The theory of Copula Entropy was developed from copula theory, and a non-parametric method for estimating CE was proposed.
- CE was proposed to test statistical independence and conditional independence (transfer entropy).
- CE was applied to solve six fundamental statistical problems, including association discovery, structure learning, variable selection, causal discovery, time lag identification, and multivariate normality test.

References

- ① Jian Ma and Zengqi Sun. "Mutual information is copula entropy". In: *Tsinghua Science & Technology* 16.1 (2011). See also arXiv preprint arXiv:0808.0845 (2008), pp. 51–54
- ② Jian Ma. "Discovering Association with Copula Entropy". In: *arXiv preprint arXiv:1907.12268* (2019)
- ③ Jian Ma and Zengqi Sun. "Dependence structure estimation via copula". In: *arXiv preprint arXiv:0804.4451* (2008)
- ④ Jian Ma. "Variable Selection with Copula Entropy". In: *Chinese Journal of Applied Probability and Statistics* 37.4 (2021), pp. 405–420
- ⑤ Jian Ma. "Estimating Transfer Entropy via Copula Entropy". In: *arXiv preprint arXiv:1910.04375* (2019)
- ⑥ Jian Ma. "Multivariate Normality Test with Copula Entropy". In: *arXiv preprint arXiv:2206.05956* (2022)
- ⑦ Jian Ma. "Identifying Time Lag in Dynamical Systems with Copula Entropy based Transfer Entropy". In: *arXiv preprint arXiv:2301.06037* (2023)
- ⑧ Jian Ma. "copent: Estimating Copula Entropy and Transfer Entropy in R". In: *arXiv preprint arXiv:2005.14025* (2020)



http://arxiv.org/a/ma_j_3

Softwares

- Official

The **copent**¹² package in R and Python for estimating copula entropy, transfer entropy, and the statistic for multivariate normality test are available on CRAN and PyPI respectively. The source codes are provided on GitHub.



<https://cran.r-project.org/package=copent>



<https://pypi.org/project/copent/>



<https://github.com/majianthu>

- Third-Party

The third-party implementations of the CE estimator include the **Cylcop** package in R, the **MLFinLab** package in Python, the **CopEnt.jl** package and the **CausalityTools.jl** package in Julia, and the **GCMI** package in Matlab and Python.

¹² Jian Ma. "copent: Estimating Copula Entropy and Transfer Entropy in R". In: *arXiv preprint arXiv:2005.14025* (2020).



Enjoy the Power of Copula Entropy!