Copula Entropy

Theory and Applications

MA Jian
majian.2003@tsinghua.org.cn

Contents

Lecture 1: Theory of Copula Entropy (1)

Probability Theory

Given random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$, their **joint probability distribution function**[1] is defined as

$$F(\mathbf{x}) = P(X_1 < x_1, \ldots, X_n < x_n), \tag{1}$$

and the corresponding **marginal function** is defined as

$$F_i(x_i) = \int P(\mathbf{x}) d\mathbf{x} \setminus dx_i. \tag{2}$$

[1] Andrey Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer Berlin, 1933.

Probability Theory

Given random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ with joint distribution function $P(\mathbf{x})$ and marginal functions $P_i(x_i), i = 1, \ldots, n$, **conditional distribution function** is defined as

$$P(\mathbf{x}|x_i) = \frac{P(\mathbf{x})}{P_i(x_i)}. \tag{3}$$

Probability Theory

Given random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ with joint distribution function $P(\mathbf{x})$ and marginal functions $P_i(x_i), i = 1, \ldots, n$, we say $\mathbf{X}$ are **mutually independent** iff

$$P(\mathbf{x}) = \prod_i P_i(x_i). \tag{4}$$

Copula Theory

Definition (Copula)

[a] Given n random variables $\mathbf{X} = (X_1, \ldots, X_n)$. Let $\mathbf{u}$ be margins $u_i = P_i(x_i), i = 1, \ldots, n$ of $\mathbf{X}$, then n dimensional copula $C : I^n \to I, I = [0,1]$ of $\mathbf{X}$ should satisfy the following properties:

1. C is non-decreasing on any subset of I^n;
2. $0 \leq C(\mathbf{u}) \leq 1$;
3. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$.

[a] Roger B Nelsen. *An Introduction to Copulas.* Springer New York, NY, 2006.

- 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 any n dimensional random variables* $\mathbf{X}$ *with joint distribution* $P(\mathbf{x})$ *and margins* $P_i(x_i)$. *Then there exists a copula* $C(\mathbf{u})$ *such that*

$$P(\mathbf{x}) = C(P_1(x_1), \ldots, P_n(x_n)). \tag{5}$$

If P_i *are continuous, then* C *is unique. Conversely, if* C *is a copula and* P_i *are distribution functions, then the function* P *defined by* (5) *is a joint distribution with margins* P_i.

[a] Abe Sklar. "Fonctions de repartition a n dimensions et leurs marges". In: *Publ. Inst. Statist. Univ. Paris* 8 (1959), pp. 229–231.

- the core of copula theory

Copula Theory

Corollary

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

$$p(\mathbf{x}) = c(\mathbf{u}) \prod_{i=1}^{n} p_i(x_i), \tag{6}$$

where $\{p_i, i = 1, \ldots, n\}$ are marginal density functions of $\mathbf{X}$, and $c(\mathbf{u}) = \frac{\partial C(\mathbf{u})}{\partial u_1 \cdots \partial u_n}$ is copula density.

- separating dependence representation from properties of individual variables
- $c(\mathbf{u}) = 1$ iff independent

Copula Theory

Copula functions

- Elliptical Copulas: Gaussian, Student, Cauchy
- Archimedean Copulas: Clayton, Frank, Gumbel, Joe, etc.
- Vine Copulas
- Empirical Copulas
- ...

Inference on Copula functions

- Construction
- Estimation (MLE)
- Hypothesis testing (GoF,Symmetry,Archimedeanity,Gaussianity)
- Sampling

Copula Theory

Copula based dependence measures[2]

- Kendall's τ
- Spearman's ρ
- Blomqvist's β
- Gini's γ

Limitations

- bivariate
- linear

[2] Friedrich Schmid et al. "Copula-based measures of multivariate association". In: *Copula Theory and Its Application*. 2010, pp. 209–236.

Information Theory

Definition (Shannon Entropy)

[a] Given random variables $X \in R^n$ and their pdf $p(\mathbf{x})$, Shannon entropy is defined as

$$H(\mathbf{x}) = -\int_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}. \tag{7}$$

[a] Thomas M Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2005.

Information Theory

Khinchin's Axioms of Entropy[3]

- **Continuity**: $H(\mathbf{x})$ is continuous;
- **Maximality**: $H(\mathbf{x})$ is maximized for uniform distribution;
- **Expansibility**: $H(\mathbf{x}, 0) = H(\mathbf{x})$;
- **Additivity**: $H(X, Y) = H(X) + H(Y|X)$.

[3] A.I. Khinchin. *Mathematical Foundations of Information Theory*. Dover, 1957.

Information Theory

Definition (Mutual Information)

[a] Given a pair of random variables (X, Y) and their pdf $p(x, y)$ and margins $p(x), p(y)$, mutual information is defined as

$$I(x; y) = \int_x \int_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dxdy. \tag{8}$$

[a] Thomas M Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2005.

- multivariate mutual information remains no consensus

Information Theory

Theorem

[a] *Mutual information equals the entropy of one variable reducing its entropy conditionally on the other variable.*

$$I(x;y) = H(y) - H(y|x). \tag{9}$$

[a] Thomas M Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2005.

Information Theory

Theorem

[a] *Mutual information equals the difference between marginal entropies and joint entropy.*

$$I(x;y) = H(x) + H(y) - H(x,y). \tag{10}$$

[a] Thomas M Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2005.

Information Theory

Theorem

[a] *Mutual information is nonnegative with 0 iff independent.*

$$I(x; y) \geq 0. \tag{11}$$

[a] Thomas M Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2005.

Information Theory

Definition (Conditional Mutual Information)

[a] Given random variables (X, Y, Z), conditional mutual information for (X, Y) conditionally on Z is defined as

$$I(x; y|z) = \int_x \int_y \int_z p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} dxdydz. \quad (12)$$

[a] Thomas M Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2005.

Theory of Copula Entropy

Definition (Copula Entropy)

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

$$H_c(\mathbf{x}) = -\int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u}. \tag{13}$$

[a] 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.

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

Theory of Copula Entropy

Theorem

[a] *Mutual Information of* $\mathbf{X}$ *is equivalent to negative copula entropy.*

$$I(\mathbf{x}) = -H_c(\mathbf{x}). \tag{14}$$

[a] 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.

- the bridge between copula theory and information theory

Theory of Copula Entropy

Proof.

$$I(\mathbf{x}) = \int_{\mathbf{x}} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\prod_i p(x_i)} d\mathbf{x} \tag{15}$$

$$= \int_{\mathbf{x}} c(\mathbf{u}) \prod_i p(x_i) \log c(\mathbf{u}) d\mathbf{x} \tag{16}$$

$$= \int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u} \tag{17}$$

$$= -H_c(\mathbf{x}) \tag{18}$$

□

Theory of Copula Entropy

Corollary

[a] *Joint entropy of random variables equals the sum of marginal entropies and copula entropy.*

$$H(\mathbf{x}) = \sum_i H_i(x_i) + H_c(\mathbf{x}). \tag{19}$$

[a] 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.

- the bridge between copula theory and information theory

Theory of Copula Entropy

Theorem

[a] *Given random variables* (X, Y, Z)*, their conditional mutual information can be represented as follows:*

$$I(x; y|z) = H_c(x, z) + H_c(y, z) - H_c(x, y, z). \tag{20}$$

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

- only determined by copula functions, irrelevant to margins
- rebuilding the foundation of information theory with copula entropy

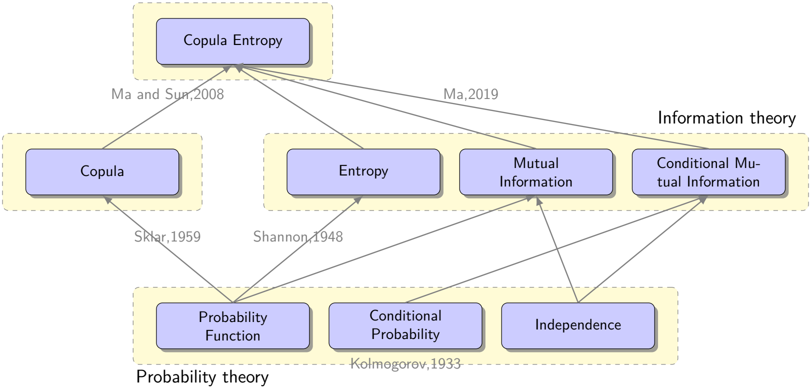
Theory of Copula Entropy

Proof.

$$I(x;y|z) = \int_x \int_y \int_z p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)} dxdydz \tag{21}$$
$$= \int_x \int_y \int_z p(x,y,z) \log \frac{p(x,y,z)p(z)}{p(x,z)p(y,z)} dxdydx \tag{22}$$
$$= I(x,y,z) - I(x,z) - I(y,z) \tag{23}$$
$$= H_c(x,z) + H_c(y,z) - H_c(x,y,z). \tag{24}$$

$\square$

Relationship between theories



Properties of Copula Entropy

Multivariate by nature

Property (Multivariate)

For any random variables $\mathbf{X} \in R^n, n > 1$, *there exists* $H_c(\mathbf{x})$.

Properties of Copula Entropy

Properties of entropy

Property (Continuity)

Given random variables $\mathbf{X}$, *if* $\mathbf{x_n} \to \mathbf{x}$, *then*

$$\lim_{\mathbf{x_n} \to \mathbf{x}} H_c(\mathbf{x_n}) = H_c(\mathbf{x}). \tag{25}$$

Property (Symmetry)

Given random variables X *and* Y, *then*

$$H_c(x, y) = H_c(y, x). \tag{26}$$

Property (Additivity)

Given independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$, *then*

$$H_c(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} H_c(\mathbf{x}_i). \tag{27}$$

Properties of Copula Entropy

Nonpositivity

Property (Nonpositivity)

Given random variables $\mathbf{X}$, *then*

$$H_c(\mathbf{x}) \leq 0, \tag{28}$$

with equality if and only if $\mathbf{X}$ *are independent.*

Proof.

Since $I(\mathbf{x}) \geq 0$, then $H_c(\mathbf{x}) \leq 0$.
Iff $\mathbf{x}$ are independent, $I(\mathbf{x})$ and therefore $H_c(\mathbf{x})$ is 0. $\square$

Properties of Copula Entropy

Invariance to monotonic transformation

Property (Invariance to monotonic transformation)

Given random variables (X, Y), and monotonic increasing functions f and g defined on them, then

$$H_c(x, y) = H_c(f(x), g(y)). \tag{29}$$

Proof.

Since $c(\mathbf{u})$ is invariant to monotonic increasing transformation f, g:

$$c(x, y) = c(f(x), g(y)), \tag{30}$$

then $H_c(\mathbf{x})$ that is defined with only $c(\mathbf{x})$ is as well. $\square$

Properties of Copula Entropy

Marginal Free

Property (Marginal free)

$H_c(\mathbf{x})$ of random variables $\mathbf{X} \in R^n$ is determined by copula function and irrelevant to marginal functions.

Properties of Copula Entropy

All orders

Property (All orders)

$H_c(\mathbf{x})$ measures information of all orders dependence between random variables.

Properties of Copula Entropy

Equivalence to correlation coefficient under Gaussianity

Property (Equivalence to correlation coefficient matrix under Gaussianity)

Given random variables $\mathbf{X} \sim N(\mu, \Sigma)$ *governed by Gaussian distribution with correlation coefficient matrix* Σ_ρ, *then*

$$H_c(\mathbf{x}) = \frac{1}{2} \log |\Sigma_\rho|. \tag{31}$$

- second order case for Property 8 (all orders)

Properties of Copula Entropy

Equivalence to correlation coefficient under Gaussianity

Proof.

Given n random variables $\mathbf{X} \sim N(\mu, \Sigma)$, then

$$
\begin{aligned}
H_c(\mathbf{x}) &= H(\mathbf{x}) - \sum_{i=1}^{n} H(x_i) && (32)\\
&= \frac{1}{2}\log(2\pi e)^n|\Sigma| - \sum_{i=1}^{n}\frac{1}{2}\log 2\pi e\delta_i^2 && (33)\\
&= \frac{1}{2}\log|\Delta\Sigma\Delta| && (34)\\
&= \frac{1}{2}\log|\Sigma_\rho|. && (35)
\end{aligned}
$$

δ_i^2 and Δ denote variance of individual variables and matrix with diagonal elements as inverse standard variances δ_i^{-1}. □

Properties of Copula Entropy

Rényi's system for axiomatic properties of dependence measure[4]

1. δ is defined on a pair of random variables (X, Y);
2. $\delta_{X,Y} = \delta_{Y,X}$;
3. $0 \leq \delta_{X,Y} \leq 1$;
4. if X, Y are independent, then $\delta_{X,Y} = 0$;
5. if X, Y has monotonic functional relationship, then $\delta_{X,Y} = 1$;
6. if f, g are monotonic functions on X, Y, then $\delta_{X,Y} = \delta_{f(X),g(Y)}$;
7. if (X, Y) are governed by Gaussian distribution, then $\delta_{X,Y} = |\rho_{X,Y}|$, where $\rho_{X,Y}$ is correlation coefficient.

[4] Alfréd Rényi. "On measures of dependence". In: *Acta Mathematica Academiae Scientiarum Hungarica* 10 (1959), pp. 441–451.

Properties of Copula Entropy

Axiomatic properties of copula entropy

1. H_c is defined on any random vectors $\mathbf{X}$;
2. H_c is continuous;
3. $H_c(X, Y) = H_c(Y, X)$;
4. H_c is additive;
5. $H_c(\mathbf{X}) \leq 0$;
6. if $\mathbf{X}$ are mutual independent, then $H_c(\mathbf{X}) = 0$;
7. if f, g are monotonic increasing functions on X, Y, then $H_c(X, Y) = H_c(f(X), g(Y))$;
8. if $\mathbf{X}$ are governed by Gaussian distribution, then $H_c(\mathbf{X}) = \frac{1}{2}\log|\Sigma_\rho|$, where Σ_ρ is correlation coefficient matrix.

Properties of Copula Entropy

Table: Comparison with other independence measures.

	Copula Entropy	Distance Correlation	HSIC
Definition	copula based	generalized corr	corr in RKHS
Multivariate	by nature	distance multivariance	dHSIC
Invariance	monotonic trans	No	No
Gaussanity	equivalent to cc	unclear	unclear

Monograph

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

Thanks!

Q&A