# สมบัติไม่แปรเปลี่ยนของตัววัดการขึ้นต่อกัน

นายอธิวัฒน์ กิจวนิชย์ภาสุ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558

INVARIANCE PROPERTIES OF DEPENDENCE MEASURES

Mr. Atiwat Kitvanitphasu

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Mathematics
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2015

| | |
|---|---|
| Thesis Title | INVARIANCE PROPERTIES OF DEPENDENCE MEASURES |
| By | Mr. Atiwat Kitvanitphasu |
| Field of Study | Mathematics |
| Thesis Advisor | Associate Professor Songkiat Sumetkijakan, Ph.D. |

---

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Dean of the Faculty of Science

(Associate Professor Polkit Sangvanich, Ph.D.)

THESIS COMMITTEE

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Chairman

(Associate Professor Wicharn Lewkeeratiyutkul, Ph.D.)

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Thesis Advisor

(Associate Professor Songkiat Sumetkijakan, Ph.D.)

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . Examiner

(Assistant Professor Nattakarn Chaidee, Ph.D.)

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . External Examiner

(Assistant Professor Pattrawut Chansangiam, Ph.D.)

อธิวัฒน์ กิจวนิชย์ภาสุ : สมบัติไม่แปรเปลี่ยนของตัววัดการขึ้นต่อกัน. (INVARIANCE PROPERTIES OF DEPENDENCE MEASURES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ.ดร.ทรงเกียรติ สุเมธกิจการ, 81 หน้า.

ลีให้การวางนัยทั่วไปของตัววัดการขึ้นต่อกันชนิดไม่สมมาตรและมีพื้นจากคอปูลาซึ่งรวมถึงมาตรวัดการขึ้นต่อกันของทรัตช์นิกด้วย เงื่อนไขพอเพียงที่ชัดเจนซึ่งทำให้มาตรวัดการขึ้นต่อกันของลีกลายเป็นมาตรวัดการขึ้นต่อกันอย่างถูกต้องได้ถูกให้ไว้และถูกพิสูจน์อย่างเคร่งครัด โดยข้อสนับสนุนของสมบัติของมาตรวัดการขึ้นต่อกันชนิดไม่สมมาตรของลี เราจึงทำการสมมาตรตัววัดไม่สมมาตรของลีพร้อมทั้งศึกษาสมบัติของมัน โดยเฉพาะอย่างยิ่ง เรายังวิเคราะห์สมบัติสำคัญของตัววัดการขึ้นต่อกัน ได้แก่ ความแจ่มชัด, ความสามารถในการตรวจจับความเป็นอิสระต่อกันหรือการขึ้นต่อกันด้วยค่าสุดขีด $0, 1$ ตามลำดับ และการไม่แปรเปลี่ยนภายใต้ประเภทของการแปลง โดยกล่าวอย่างเฉพาะเจาะจง เราพบผ่านตัวอย่างหลายตัวอย่างว่าตัววัดการขึ้นต่อกันที่สามารถตรวจวัดการขึ้นต่อกันที่มีขนาดใหญ่กว่าจะมีแนวโน้มที่ไม่แปรเปลี่ยนภายใต้การแปลงจำนวนมากตามไปด้วย สัมประสิทธิ์ข้อมูลสูงสุดเชิงความน่าจะเป็น (MIC) ได้ถูกพิสูจน์ว่าเป็นตัววัดการขึ้นต่อกัน ท้ายที่สุด เราแสดงว่าไม่มีตัววัดการขึ้นต่อกันที่ทั้งไม่แปรเปลี่ยนต่อการแปลงทางเดียวโดยแท้และสามารถวัดการขึ้นต่อกันอย่างสมบูรณ์ได้

| | | | |
|---|---|---|---|
| ภาควิชา | คณิตศาสตร์และวิทยาการคอมพิวเตอร์ | ลายมือชื่อนิสิต | ............................ |
| สาขาวิชา | คณิตศาสตร์ | ลายมือชื่อ อ.ที่ปรึกษาหลัก | .................. |
| ปีการศึกษา | 2558 | | |

## # 5672130723 : MAJOR MATHEMATICS
KEYWORDS : MEASURE OF DEPENDENCE/ SCHWEIZER-WOLFF/RÉNYI'S POSTULATES

ATIWAT KITVANITPHASU : INVARIANCE PROPERTIES OF DEPENDENCE MEASURES. ADVISOR: ASSOC. PROF. SONGKIAT SUMETKIJAKAN, Ph.D., 81 pp.

Li gave a generalization of non-symmetric copula-based dependence measure, such as the Trutschnig's measure of dependence. A precise sufficient condition which makes Li's generalization a non-symmetric measure of dependence is given and proved rigorously. Supported by its non-symmetric dependence measure properties, we symmetrize the Li's non-symmetric measure of dependence and investigate its properties. Specifically, we analyze the key properties of dependence measures including well-defined property, abilities to detect independence and dependence at the two extreme values $0, 1$ respectively, and invariance under the certain types of transformations. In particular, we find, via several examples, that a dependence measure possessing an ability to detect a larger class of dependences tends to be invariant under an accordingly large class of transformations. The probabilistic version of maximal information coefficient (MIC) is also proved to be a dependence measure. Lastly, we show that there does not exist a dependence measure which is both invariant under strictly monotonic transformations and able to catch complete dependence.

| | | |
|---|---|---|
| Department: | Mathematics and Computer Science | Student's Signature ...................... |
| Field of Study: | Mathematics | Advisor's Signature ...................... |
| Academic Year: | 2015 | |

# ACKNOWLEDGEMENTS

First of all, I would like to express my largest gratitude to my thesis advisor, Associate Professor Dr.Songkiat Sumetkijakan. He always give me invaluable advice, patience to my lack of skill and discipline, viewpoint to tackle the problem in both mathematics and real life, and infinitely many things that make me to this point. Also, I would like to acknowledge Associate Professor Dr.Wicharn Lewkeeratiyutkul, the chairman, Assistant Professor Dr.Nattakarn Chaidee and Assistant Professor Dr.Pattrawut Chansangiam, the committee members, for enlightening me with insightful suggestions. And I thank Professor Hui Li for sending his valuable manuscript via Research Gate.

Moreover, my sincere thank is for the Withayabandhit scholarship. By this scholarship, I could pass the first year of my study without any financial pressure.

Apart from the thesis, I would like to deeply thank all of my instructors and faculty officers, in particular Professor Dr.Patanee Udomkavanich, Associate Professor Dr.Wicharn Lewkeeratiyutkul, and Dr.Pongdate Montagantirud, to teach me what mathematics really is. My friends in the program, especially the copula study group, Mr.Natee Pitiwan, and Mr.Atiratch Laoharenoo, also make my life as a mathematic student full of fruitful mathematical discussion and happiness.

Last, but not least, finishing this degree is impossible without families of me and my girlfriend. Their continuous encouragement and understanding enhance me to strong in both good and bad days. Any potential contribution of this work is dedicated to all of people I mention while errors and flaws are solely responsible by me.

# CONTENTS

# CHAPTER I
# INTRODUCTION

In exploring the nature, the first and foremost component needed to identify is the relation between objects of interest. This fact is inevitably universal, from natural sciences to social sciences and even epistemology. Associations must be illustrated or proved before discussing their rationales, thus systematic ways to clarify associations are crucial indeed.

Formally, probability and statistic theory provides rigorous tools to attack this task: consider objects under study as random variables or random samples. In this framework, uncertainty in reality is not an obstacle for detecting the dependence between two random objects in both population and sample levels. Measuring dependence probabilistically and statistically plays an eminent role in the uncertainty-embedded fields such as economics, finance, engineering, geology, meteorology, etc.

There have been many explicitly defined dependence measures proposed. For examples, see [4, 5, 9, 10, 12, 13, 18, 19]. Ideally, a function of two random variables is considered a dependence measure if it fulfills a certain set of desirable properties:

- Well-Defined Property: A dependence measure must be well-defined in the sense that the same pair of random variables whose distribution functions are continuous must yield identical measured values of dependence.
- Normalization Property: The possible values of dependence measure should form closed interval. Conventionally, the range of dependence measure is $[0, 1]$.
- Independence Detectability Property: Its zero value should be able to iden-

tify independence between random variables.

- Dependence Detectability Property: The maximum value of a dependence measure should be able to identify maximum dependence between random variables. For different types of dependence measures, maximum dependence may refer to complete dependence, mutual complete dependence, or monotonic dependence. Given random variables $X$ and $Y$, $X$ and $Y$ are said to be *completely dependent (CD)* if either there is a Borel measurable function $f$ such that $Y = f(X)$ with probability one or there exists a Borel measurable function $g$ making $X = g(Y)$ almost surely. They are called *mutually completely dependent (MCD)* if $Y$ is a Borel invertible measurable transformation of $X$. Finally, they are said to be *monotonically dependent* if $Y$ is a strictly monotonic transformation of $X$.

- Invariant Property: Dependence between random variables should not vary when transforming one or both random variables by functions of a certain type. If a dependence measure $R$ satisfies $R(f(X), g(Y)) = R(X, Y)$ for any strictly monotonic functions $f, g$ and random variables $X$ and $Y$, then $R$ is called *monotonic invariant.* If this property holds for all injective functions, then $R$ is said to be *injective invariant.*

Furthermore, is interesting in association between random variables, then a measure of this type of dependence should be symmetric. That is; the order of random variables when computing dependence measure must not affect their associations. This type of dependence measures is called by us a *symmetric dependence measure*. Otherwise, it is called a non-symmetric dependence measure. The latter is also useful despite the lack of symmetry because under some circumstances, one-sided dependence is more suitable.

However, there have been proposed dependence measures whose properties were not investigated thoroughly as dependence measures. Firstly, the Trutschnig's $\zeta_1$ was introduced in [20]. This dependence measure is derived from the topology induced by derivative-based metric on the space of bivariate copulas, defined as joint distribution functions of a pair of random variables uniformly distributed on $[0, 1]$. By its construction, $\zeta_1$ is not symmetric. And it is proved to be able to detect the dependence and independence while the other properties were not studied. Secondly, the more generalized non-symmetric dependence measure of Li, denoted as $\tau$, is proposed in [10]. It can be considered as the generalized version of $\zeta_1$. Although its ability to detect complete dependence is proved, its lack of concrete definition of what Li called *distance-like function* makes it unclear to claim this as a dependence measure. We will also propose a way to *symmetrize* the non-symmetric measures of dependence.

In [14], Reshef et al. introduced maximal information coefficient (MIC) as a dependence measure. In the first version, it is proposed as a statistic. As the statistics of random samples, it has many attractive properties since its zero value can detect the independence between random pair of data samples and its value converges to one if the pair of random samples are drawn from the distribution of complete dependent random variables when sample size goes to infinity. Moreover, statistical MIC is invariant under the transformations of that data samples with strictly increasing maps. Later, the probabilistic MIC appeared in [15]. Nonetheless, this paper focused on proving that the statistical MIC is a consistent estimator of the probabilistic MIC and providing algorithm to compute it efficiently. Its properties as a probabilistic measure of dependence has not been explored yet.

Lastly, from observation, dependence measures seem to have trade-off between the stronger types of invariance and the stronger types of ability to detect dependences. That is; the larger class under which it is invariant, the wider class of dependence having maximum value of dependence measure. It is then natural to ask whether there is a dependence measure which is both invariant in the strong mode (such as bijective transformation) and able to catch the dependence in the stronger mode (such as monotonic dependence). It may be possible by the existences of concrete dependence measure having that pair of properties or it has contradiction in an axiomatic level.

Here are the objectives of this thesis:

(i) Investigate the dependence measure properties of $\zeta_1$;

(ii) Give additional assumptions to the *distance-like* function in the definition of $\tau$, show rigorous proof of its properties, and provide its symmetrization and its corresponding properties;

(iii) Explore the properties of the probabilistic MIC as a dependence measure;

(iv) Find possible interrelationships between each type of invariance and dependence detectability properties axiomatically.

Throughout this thesis, we attempt to fill in the tables summarizing the dependence measures satisfying pairs of invariance and dependence detectability properties. For non-symmetric dependence measures, here is the summary table.

| | MD detect. | CD detect. |
|---|---|---|
| Monotone inv. | | |
| Injective inv. | Impossible | $\tau$ (including $\zeta_1, r$) |

Table 1.1: The summary table for non-symmetric dependence measures.

For convenience, we recall the definitions of $\zeta_1$, $r$, and $\tau$. Here, $A$ is the copula of random variables $X, Y$ having continuous marginal distributions. Moreover, denote the closed interval $[0,1]$ as $\mathbb{I}$.

- $\zeta_1$ from Trutschnig (see [20]) is defined as

$$\zeta_1(Y \mid X) := 3D_1(A, \Pi) = 3\int_{\mathbb{I}}\int_{\mathbb{I}} |\partial_1 A(x,y) - \partial_1 \Pi(x,y)| dxdy$$

  where $\partial_1$ denotes the partial derivative operator with respect to the first place and $\Pi$ is defined by $\Pi(x,y) = xy$ for $x, y \in \mathbb{I}$.

- $r$ from Dette, Siburg, and Stoimenov (see [5]) is defined as

$$r(Y \mid X) := 6 \int_{\mathbb{I}} \int_{\mathbb{I}} (\partial_1 A(x,y) - \partial_1 \Pi(x,y))^2 dx dy.$$

- $\tau$ from Li (see [10]) is defined as

$$\tau(Y \mid X) := \frac{\tau(A)}{\tau(M)}$$

where, for any copula $C$,

$$\tau(C) := \int_{\mathbb{I}} \int_{\mathbb{I}} \varphi(\partial_1 C(x,y) - \partial_1 \Pi(x,y)) dx dy,$$

$\varphi \colon [-1,1] \to [0,\infty]$ continuous, convex, and satisfies $\varphi(-x) = \varphi(x)$ for all $x \in [-1,1]$, and $\varphi(x) = 0$ if and only if $x = 0$. Recall that $M(x,y) := \min\{x,y\}$.

And the following is the table for symmetric dependence measures:

| | MD detect. | MCD detect. | CD detect. |
|---|---|---|---|
| Monotone inv. | $\sigma$ | $\tilde{\tau}$, $\omega$, $\nu_\alpha$ | |
| Injective inv. | Impossible | | $S$, $\omega_*$, $\nu_*$ |

Table 1.2: The summary table for symmetric dependence measures.

Let us recall the definitions of dependence measures appearing in the table.

- Maximal correlation ($S$) from Rényi (see [13]) is defined by

$$S(X,Y) = \sup_{f,g} R(f(X), g(Y))$$

where the supremum is taken over the set of all Borel measurable functions and $R$ is the Pearson's correlation of two random variables.

- $\sigma$ from Schweizer and Wolff (see [18]) is defined by

$$\sigma(X,Y) := 12 \int_{\mathbb{I}^2} |A - \Pi| d\lambda_2.$$

- $\omega$ from Siburg and Stoimenov (see [19]) is defined by

$$\omega(X,Y) := \sqrt{3}\|A - \Pi\|_S := \sqrt{3}\int_{\mathbb{I}}\int_{\mathbb{I}} \|\nabla(A(x,y) - xy)\|^2 dxdy$$

  where $\|\cdot\|$ denotes the usual two-dimensional Euclidean norm and $\nabla$ means the Laplacian.

- $\omega_*$ from Ruankong, Santiwipanont, and Sumetkijakan (see [16]) is defined by

$$\omega_*(X,Y) := \left(3 \sup_{U,Y\in\mathfrak{I}(\mathcal{C})} \|U * A * V\| - 2\right)^{1/2}$$

  where $*$ is the product of copulas defined by

$$A * B(x,y) = \int_{\mathbb{I}} \partial_2 A(x,t)\partial_1 B(t,y)dt$$

  and the invertibility of copula $C$ means $C^T * C = C * C^T = M$ where $C^T(x,y) := C(y,x)$. The class of invertible copulas is written as $\mathfrak{I}(\mathcal{C})$.

- $\nu_*, \nu_\alpha$ from Kamnitui, Santiwipanont, and Sumetkijakan (see [8]) is defined by

$$\nu_*(X,Y) := \sqrt{6|[A]|_* - 3} = \sqrt{6 \sup_{U,Y\in\mathfrak{I}(\mathcal{C})} |[U * A * V]| - 3}$$

  and, for $\alpha \in (0,1)$,

$$\nu_\alpha(X,Y) := \alpha\nu(X,Y) + (1-\alpha)\nu_*(X,Y)$$

  where

$$|[C]| := \int_{\mathbb{I}^2} \left(C^T * C + C * C^T\right) d\lambda_2$$

  and $\nu(X,Y) := \sqrt{6|[A]| - 3}$.

- $\tilde{\tau}$ is defined by

$$\tilde{\tau}(X,Y) := \frac{1}{2}(\tau(Y \mid X) + \tau(X \mid Y)).$$

The blank cells in the tables mean there have not been any dependence measures satisfying those pairs of properties and the word *impossible* clearly denotes that there are no dependence measures satisfying that pair of properties. For the case of MIC, its properties are proved and an enlightening example illustrating the possibility to possess stronger invariant properties will be given.

The structure of this thesis is given as follow: this first chapter is an introduction. It provides the motivations and leading objectives. The second chapter is preliminaries. It contains a collection of prerequisites for proving the main results later. After one has all the tools for attacking the problems, the main results are proved and presented separately in three chapters. Chapter three is devoted to the main results on the dependence measure properties of $\zeta_1$ and $\tau$, and a lemma on incompatibility between an invariant property and a maximum dependence detectability. In the last chapter, we have three tasks. The first section is the symmetrization of Li's $\tau$ into a symmetric one while the section on the properties of MIC follows. The end of this chapter will be a concluding remark on further research agenda for symmetric dependence measures along with its version of incompatibility lemma.

# CHAPTER II
# PRELIMINARIES

Throughout this thesis, we denote $\mathbb{N}, \mathbb{Q}$, and $\mathbb{R}$ as the set of all natural, rational, and real numbers respectively. The other frequently seen notation is $\mathbb{I}$, a symbol for the closed unit interval $[0, 1]$.

## 2.1 Measure Theory

Firstly, the most fundamental object is the so-called $\sigma$-algebra.

**Definition 2.1.** Let $\mathcal{F}$ be a collection of subsets of a nonempty set $X$, then $\mathcal{F}$ is said to be a *$\sigma$-algebra of $X$* if the following are satisfied:

(i) $\phi \in \mathcal{F}$;

(ii) $F \in \mathcal{F}$ implies $F^c \in \mathcal{F}$;

(iii) $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F}$ implies $\cup_{n\in\mathbb{N}} F_n \in \mathcal{F}$.

The pair $(X, \mathcal{F})$ will be called a *measurable space* and elements of $\sigma$-algebra are called *measurable sets*.

From definition, $\{\phi, X\}$ and the power set of $X$ are $\sigma$-algebras, which can be thought of as minimum and maximum $\sigma$-algebras. However, under many circumstances, we need some in-between size $\sigma$-algebra containing some desired family of subsets. To achieve this point, the following lemma helps.

**Lemma 2.2.** *The intersection of a nonempty collection of $\sigma$-algebras is also a $\sigma$-algebra.*

**Corollary 2.3.** *Given a collection $\mathcal{E}$ of subsets of set $X$. Then, there exists the smallest $\sigma$-algebra containing $\mathcal{E}$. This $\sigma$-algebra is called the $\sigma-$algebra generated by $\mathcal{E}$ and denoted as $\sigma(\mathcal{E})$.*

The most frequently seen $\sigma$-algebra of this type is, given topological space $(X, \tau)$, $\sigma(\tau)$ the *Borel $\sigma$-algebra of $X$* and we also write $\mathcal{B}(X)$ customarily. In this work, we deal with $\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{I}), \mathcal{B}(\mathbb{R}^2)$, and $\mathcal{B}(\mathbb{I}^2)$ mainly. The elements in the Borel $\sigma$-algebra are called *Borel sets*. Moreover, since the Borel $\sigma$-algebra is generated by all open sets, one can prove that almost all of the subsets of real line, says singleton, closed, open, clopen interval are Borel sets.

Since we have a measurable space, the set function defined to measure the measurable set can be proposed.

**Definition 2.4.** For any measurable space $(X, \mathcal{F})$, a function $\mu \colon \mathcal{F} \to [0, \infty)$ is called a *measure* if

(i) $\mu(\phi) = 0$;

(ii) for disjoint collection $\{F_n\}_{n \in \mathbb{N}}$ of $\mathcal{F}$,

$$\mu\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mu(F_n).$$

The triplet $(X, \mathcal{F}, \mu)$ is called a *measure space*.

Intuitively, length of interval on the real line seems to be measure. But it requires more work to generalize it to a measure.

**Definition 2.5.** Let $\mathcal{E} = \{(a, b), \varnothing : a, b \in \mathbb{R}\}$. Define $\lambda^*$, called *Lebesgue outer measure* on $\mathbb{R}$, by

$$\lambda^*(E) = \inf\left\{\sum_{n=1}^{\infty} l(E_n) : E_n \in \mathcal{E}, E \subseteq \cup_{n=1}^{\infty} E_n\right\}$$

where $l(a, b) = b - a$ for all $a, b \in \mathbb{R}$ such that $a < b$.

This means that, given a subset $E$ of real line, Lebesgue outer measure is the sum of the length of bases of usual topology on real line generating smallest open set containing $E$. The following is the definition of measurability in the sense of Lebesgue.

**Definition 2.6.** Let $\lambda^*$ be the Lebesgue outer measure on $\mathbb{R}$ and $B \subseteq \mathbb{R}$. A set $B$ is called *Lebesgue measurable* if

$$\lambda^*(E) = \lambda^*(E \cap B) + \lambda^*(E \cap B^c)$$

for all $E \subseteq \mathbb{R}$.

And the following defines what Lebesgue measure is.

**Theorem 2.7.** *Let $\lambda^*$ be the Lebesgue outer measure on $\mathbb{R}$ and $\mathcal{E}$ be a collection of all Lebesgue measurable subsets of $\mathbb{R}$. Then the following properties hold:*

*(i) $\mathcal{E}$ is a $\sigma$-algebra on $\mathbb{R}$;*

*(ii) The restriction of $\lambda^*$ to $\mathcal{E}$ is a measure on $\mathcal{E}$, called Lebesgue measure on $\mathbb{R}$ and denoted as $\lambda$.*

*Moreover, any Borel set is Lebesgue measurable.*

From Lebesgue measure, it has related concept called a Lebesgue-Stieltjes measure on $\mathbb{R}$. The definition is given as follow:

**Definition 2.8.** Let $F\colon \mathbb{R} \to \mathbb{R}$ be a bounded increasing and right-continuous function and $\mathcal{E}$ be a family of all Lebesgue measurable subsets, then $\mu_F((a,b)) := F(b) - F(a)$ extends to a finite measure on $\mathbb{R}$. This measure $\mu_F$ is called *Lebesgue-Stieltjes measure*

In fact, one can construct measure space from relatively weak structure of collection of subsets and, then, extend it to the rich structure likes $\sigma$-algebra later.

**Definition 2.9.** Let $X$ be a set and $\mathcal{P} \subseteq \mathscr{P}(X)$ where $\mathscr{P}(\cdot)$ denotes the power set. Then $\mathcal{P}$ is called a *semi-ring* if the following conditions hold:

(i) If $A, B \in \mathcal{P}$, then $A \cap B \in \mathcal{P}$.

(ii) If $A, B \in \mathcal{P}$, then $A \setminus B = \cup_{n=1}^{k} C_n$ where $C_n \in \mathcal{P}$ for some $k \in \mathbb{N}$

**Definition 2.10.** Let $X$ be a set and $\mathcal{P} \subseteq \mathscr{P}(X)$. Then a function $\mu : \mathcal{P} \to [0, \infty]$ is called a *pre-measure* if the following conditions hold:

(i) If $\varnothing \in \mathcal{P}$, then $\mu(\varnothing) = 0$.

(ii) If $A_1, A_2, \ldots, A_k$ are elements in $\mathcal{P}$ such that $A_i \cap A_j = \varnothing$ for all $i \neq j$ and $\cup_{n=1}^{k} A_n \in \mathcal{P}$, then $\mu(\cup_{n=1}^{k} A_k) = \sum_{n=1}^{k} \mu(A_n)$.

(iii) If $E \subseteq \cup_{n=1}^{\infty} A_n$ where $E, A_n \in \mathcal{P}$ for all $n \in \mathbb{N}$, then $\mu(E) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

**Remark 2.11.** From Definition 2.10, the pre-measure $\mu$ is called *$\sigma$-finite* if there is a sequence $\{A_k\}_{k=1}^{\infty}$ of elements in $\mathcal{P}$ such that $\cup_{k=1}^{\infty} A_k = X$ and $\mu(A_k) < \infty$ for all $k \in \mathbb{N}$. Furthermore, the measure space having $\sigma-$finite measure is called a *$\sigma$-finite measure space*.

The following definition is the general version of outer measure.

**Definition 2.12.** Let $X$ be a set. Then a function $\mu^* : \mathscr{P}(X) \to [0, \infty]$ is called an *outer measure* if the following conditions hold:

(i) $\mu^*(\varnothing) = 0$.

(ii) If $A_1, A_2 \in \mathscr{P}(X)$ such that $A_1 \subseteq A_2$, then $\mu^*(A_1) \leq \mu^*(A_2)$.

(iii) If $A_1, A_2, \ldots,$ are elements in $\mathscr{P}(X)$ for all $n \in \mathbb{N}$, then $\mu(\cup_{n=1}^{\infty} A_k) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

And, since outer measure is defined, measurability with respect to outer measure is also determined.

**Definition 2.13.** Let $\mu^*$ be an outer measure on $X$ and $B \subseteq X$. A set $B$ is called *$\mu^*$-measurable* if

$$\mu^*(E) = \mu^*(E \cap B) + \mu^*(E \cap B^c)$$

for all $E \subseteq X$.

**Proposition 2.14.** *Let $\mu^*$ be an outer measure on $X$ and $\mathcal{M}$ be the collection of all $\mu^*$-measurable subsets of $X$. Then the following properties hold:*

*(i) $\mathcal{M}$ is a $\sigma$-algebra on $X$.*

*(ii) The restriction of $\mu^*$ to $\mathcal{M}$ is a complete measure.*
*That is; for any set $A$ such that $\mu^*(A) = 0$, if $B \subseteq A$, then $B \in \mathcal{M}$.*

**Theorem 2.15.** *Let $X$ be a set, $\mathcal{P} \subseteq \mathscr{P}(X)$ such that $\mathcal{P} \neq \varnothing$ and $\mu : \mathcal{P} \to [0, \infty]$ be a function. Then a function $\mu^* : \mathscr{P}(X) \to [0, \infty]$ defined by*

$$\mu^*(E) = \begin{cases} 0 & \text{if } E = \varnothing \\ \inf\{\sum_{k=1}^{\infty} \mu(E_k) : E \subseteq \cup_{k=1}^{\infty} E_k, E_k \in \mathcal{P}\} & \text{if } E \neq \varnothing \end{cases}$$

*is an outer measure, called outer measure induced by $\mu$.*

From Theorem 2.15, the restriction of $\mu^*$ to $\mathcal{M}$, the collection of all $\mu^*$-measurable subsets, is called the Carathéodory measure induced by $\mu$, denoted by $\bar{\mu}$. And the following consequence is one of standard measure theoretic machinery to extend pre-measure on semiring to be a measure on a set of measurable sets with respect to outer measure.

**Theorem 2.16.** *(Carathéodory-Hahn extension theorem) Let $X$ be a set. If $\mu$ is a pre-measure on a semi-ring $\mathcal{P} \subseteq \mathscr{P}(X)$, then the Carathéodory measure $\bar{\mu}$ induced by $\mu$ is an extension of $\mu$. Moreover, if $\mu$ is $\sigma$-finite, then $\bar{\mu}$ is the unique measure extending $\mu$.*

This extension is powerful tool to construct many measures since many collection of subsets in consideration is semiring, for example the collection of interval or rectangle in $\mathbb{R}, \mathbb{R}^2$ respectively. In fact, given two $\sigma-$finite measure spaces, one can construct a product measure.

**Theorem 2.17.** *Let $(X, \mathcal{F}, \mu), (Y, \mathcal{G}, \nu)$ be two $\sigma-$finite measure spaces and define $\mathcal{F} \otimes \mathcal{G}$ to be the $\sigma-$algebra generated by the set of Cartesian product of elements of*

*$\mathcal{F}, \mathcal{G}$ respectively, which is called measurable rectangles. Then, there is the unique measure $\mu \times \nu$ on $\mathcal{F} \otimes \mathcal{G}$ such that, for $F \in \mathcal{F}$ and $G \in \mathcal{G}$,*

$$\mu \times \nu(F \times G) = \mu(F)\nu(G).$$

*This measure is called a product measure $\mu \times \nu$.*

In this thesis, the product measure we used often is the product measure of two Lebesgue measures, and it will be denoted as $\lambda_2$.

For the finite measure space eqiupped with topology, the *support of measure* $\mu$, denoted by $\mathrm{supp}(\mu)$, is defined to be the smallest closed measurable subset of full measure. Note that any measurable set which does not contain any part of support will have measure zero.

Next, since we finish introducing the notion of measure, one can define what measuable function is.

**Definition 2.18.** Let $(X, \mathcal{F})$ and $(Y, \mathcal{F}')$ be measurable spaces. A function $f : X \to Y$ is called $(\mathcal{F}, \mathcal{F}')$-measurable if $f^{-1}(E) \in \mathcal{F}$ for all $E \in \mathcal{F}'$. If $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, then a $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{F}))$-measurable function is said to be *Borel measurable.*

Because Borel set comes from open sets, continuous function is Borel measurable. Moreover, monotonic function is also Borel measurable. Basic operation of several Borel measurable function, such as addition, scalar and pointwise multiplication, quotient with non-zero function, taking absolute value and square root can give a new measurable function from the old. And, given a sequence of measurable functions, its corresponding limit supremum and infemum are measurable, thus limit is measurable if exist. Lastly, composition preserves measurability of two measurable function.

The notion of almost everywhere properties are defined below.

**Definition 2.19.** Let $(X, \mathcal{F}, \mu)$ be a measure space. The property $P(x)$ holds *$\mu$-almost everywhere*, denoted by $P(x)$ $\mu$-a.e., if there is a measurable set $N \subseteq X$ such that $\mu(N) = 0$ and $P(x)$ holds for each $x \in X \setminus N$.

An important class of measurable functions is simple function. This type of functions will be used to construct integration later.

**Definition 2.20.** Let $(X, \mathcal{F})$ be a measurable space. Define, for any measurable set $A$, the *characteristic function of* $A$ by the function $\chi_A \colon X \to \{0,1\}$ such that $\chi_A(x) := 1$ for $x \in A$ and $\chi_A(x) = 0$ otherwise. A measurable function $f : X \to \mathbb{R}$ is called a *simple function* if it has finite range. Equivalently, a function $f : X \to \mathbb{R}$ is a simple function if and only if there exist real numbers $a_1, a_2, \ldots, a_n$ and pairwise disjoint measurable sets $A_1, A_2, \ldots, A_n$ such that $X = \cup_{i=1}^n A_i$ and $f = \sum_{i=1}^n a_i \chi_{A_i}$.

One may ask the question whether Borel measurable injective function has Borel measurable inverse. The answer is yes, though this fact is not trivial and require advanced notions like analytic set and polish space, which is out-of-scope for this study. This is the theorem called *Lusin-Suslin theorem*

**Theorem 2.21.** *(Lusin-Suslin) Any Borel measurable one-to-one map has Borel measurable inverse.*

The actual usage of simple function is to approximate the non-negative measurable function.

**Proposition 2.22.** *Let $f : X \to \mathbb{R}$ be nonnegative measurable function, then there exists an increasing sequence $\{f_n\}_{n=1}^\infty$ of nonnegative simple functions such that $f_n \to f$.*

For the simple function, the integral is defined as follow.

**Definition 2.23.** Let $(X, \mathcal{F}, \mu)$ be a measure space and $s$ a nonnegative simple function, which can written as $s = \sum_{i=1}^n a_i \chi_{A_i}$ where $a_1, a_2, \ldots, a_n$ are nonnegative real values and $A_1, A_2, \ldots, A_n$ such that $X = \cup_{i=1}^n A_i$ are pairwise disjoint measurable sets. The integral of $s$ with respect to the measure $\mu$ is defined by

$$\int_X s d\mu = \sum_{i=1}^n a_i \mu(A_i).$$

And the properties on integral of simple function are given here.

**Proposition 2.24.** *Let $s, t$ be nonnegative simple funcitons on $X$ and $c \geq 0$. Then*

*(i)* $\int_X csd\mu = c \int_X sd\mu$,

*(ii)* $\int_X (s+t)d\mu = \int_X sd\mu + \int_X td\mu$,

*(iii) If* $s \leq t$, *then* $\int_X sd\mu \leq \int_X td\mu$.

And this is the definition of integral in general.

**Definition 2.25.** Let $f : X \to [0, \infty]$ be a measurable function. The *integral of $f$ with respect to the measure $\mu$* is defined by

$$\int_X f d\mu = \sup\{\int_X sd\mu : s \text{ is simple and } 0 \leq s \leq f\}.$$

If $E \in \mathcal{F}$, then the integral of a measurable function $f\chi_E$ with respect to the measure $\mu$ is defined as

$$\int_E f d\mu = \int_X f\chi_E d\mu.$$

**Proposition 2.26.** *Let $f, g : X \to [0, \infty]$ be measurable, and $A, B \in \mathcal{F}$. Then*

*(i) If* $f \leq g$, *then* $\int_X f d\mu \leq \int_X gd\mu$,

*(ii) If* $A \subseteq B$, *then* $\int_A f d\mu \leq \int_B f d\mu$,

*(iii) If* $c > 0$, *then* $\int_X cf d\mu = c \int_X f d\mu$,

*(iv) If* $\mu(A) = 0$, *then* $\int_A f d\mu = 0$.

**Proposition 2.27.** *Let $f : X \to [0, \infty]$ be measurable, and $A \in \mathcal{F}$. Then*

*(i) If* $\int_A f d\mu = 0$, *then* $f(x) = 0$ *a.e. on* $A$,

*(ii) If* $\int_A f d\mu < \infty$, *then* $f(x) < \infty$ *a.e. on* $A$.

The existence of approximate simple function makes defining integral be well-defined. The following well-known theorem yields many properties.

**Theorem 2.28.** *[The Monotone Convergence Theorem] Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions on $X$. If*

*(i)* $f_1(x) \leq f_2(x) \leq \cdots \leq \infty$ *for every $x \in X$, and*

*(ii)* $\lim_{n\to\infty} f_n(x) = f(x)$ *for each $x \in X$,*

*then* $\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$.

Monotone convergence theorem, along with properties of integral of simple functions, induces many results.

**Proposition 2.29.** *Let $f, g : X \to [0, \infty]$ be measurable functions. Then*

$$\int_X (f+g)d\mu = \int_X f d\mu + \int_X g d\mu.$$

**Corollary 2.30.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions on $X$. Then*

$$\int_X \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_X f_n d\mu.$$

**Corollary 2.31.** *Let $f$ be a nonnegative measurable function on $X$, and $\{E_n\}_{n\in\mathbb{N}}$ a sequence of disjoint measurable sets. Then*

$$\int_{\cup_{n=1}^{\infty} E_n} f d\mu = \sum_{n=1}^{\infty} \int_{E_n} f d\mu.$$

In fact, integral can be defined for real-valued measurable function, even in case of complex-valued measurable function.

**Definition 2.32.** Let $f$ be a real valued measurable function on $X$. Let $f^+ := \max\{f, 0\}$ and $f^- := \min\{-f, 0\}$. Then the integral of $f$ with respect to $\mu$ is defined by

$$\int_X f d\mu = \int_X f^+ d\mu + \int_X f^- d\mu$$

if at least one of the integral on the right side is finite.

Note that for any real-valued measurable function $f$, one can write $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

The following formula gives a useful tool to convert the integral with respect to this measure into other integral with respect to other measure.

**Theorem 2.33.** *(Change of Variables) Let $(X, \mathcal{F}_1, \mu)$ be a measure space, $(Y, \mathcal{F}_2)$ be another measurable space and $g : X \to Y$ be a measurable function. Define a measure $\nu$ on $Y$ by $\nu = \mu(g^{-1}(B))$ for all measurable sets $B \subseteq Y$. If $f : Y \to \bar{\mathbb{R}}$ is measurable, then $\int_X f \circ g d\mu = \int_Y f d\nu$.*

In the case of multidimensional Lebesgue measure space, the change of variable formula will have more concrete form. For this thesis, the version of the invertible linear transformation suffices.

**Theorem 2.34.** *Let $n \in \mathbb{N}$, $f \colon \mathbb{R}^n \to \mathbb{R}^n$ be a Borel measurable function with respect to $n-$fold Lebesgue measure and $T \colon \Omega \to \mathbb{R}^n$ be a continuously differentiable injective map on open subset $\Omega \subseteq \mathbb{R}^n$ such that $D_x T$, the Jacobian matrix of $T$ evaluated at $x \in \mathbb{R}^n$, is invertible for all $x \in \mathbb{R}^n$. Then, $f \circ T$ is also Borel measurable. Moreover, if $f$ is nonnegative or integrable,*

$$\int_\Omega f \circ T(x) d\lambda_n(x) = \int_{T(\Omega)} f(x) |\det D_x T|^{-1} d\lambda_n(x).$$

And, one of the main integral inequality involving convex function is introduced by Jensen.

**Theorem 2.35.** *Let $f \colon X \to \mathbb{R}$ be a bounded measurable function on finite measure space and $\varphi$ a convex function from $\mathbb{R}$ to $\mathbb{R}$. Then, one has*

$$\varphi\left(\int_X f d\mu\right) \leq \int_X (\varphi \circ f) d\mu.$$

*The strict equality holds if and only if $f$ is almost surely constant function with respect to $\mu$.*

From integration, one can define the following.

**Definition 2.36.** Define the set, for $1 \leq p \leq \infty$,

$$\mathcal{L}^p(X) := \left\{ f \colon X \to \mathbb{R} \mid \int_X |f|^p d\mu < \infty \right\}$$

and $\|f\|_p := \left(\int_X |f|^p d\mu\right)^{\frac{1}{p}}$. Then the space $\mathcal{L}^p(X)$ is a vector space. Moreover, the subspace $\mathcal{L}^p(X)/\sim$ where $\sim$ denotes the almost everywhere equality relation normed vector space. This one is called $L^p$ *space*. And the set

$$L^\infty(X) := \{f\colon X \to \mathbb{R} \mid \exists M \geq 0, |f| \leq M \text{ a.e.}\}$$

along with

$$\|f\|_\infty := \inf\{M \geq 0 \mid |f| \leq M \text{ a.e.}\}$$

is also normed vector space called an $L^\infty$ *space*.

When considering convergence of measurable functions, there are two modes of convergences in this work, and one in term of random variables which is defined later.

**Definition 2.37.** A sequence of measurable functions $\{f_n\}_{n\in\mathbb{N}}$ is said to *converge almost everywhere to measurable function* $f$ if there is a measurable set of full measure such that $f_n \to f$ on that set.

**Definition 2.38.** Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions from a measure space $(X, \mathcal{F}, \mu)$ to $\mathbb{R}^2$ and $f\colon (X, \mathcal{F}, \mu) \to \mathbb{R}^2$ be another measurable function. Then, one says that $f_n \to f$ *almost uniformly* if for any $\epsilon > 0$, there is $\eta_\epsilon \in \mathcal{F}$ such that $\mu(\eta_\epsilon) < \epsilon$ and $f_n \to f$ uniformly on $X \setminus \eta_\epsilon$. This definition also works for the case of real-valued functions.

In the case of a finite measure space, the following theorem is useful to see the relationship between these two type of convergences. This theorem is due to Egoroff.

**Theorem 2.39.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions from finite measure space $(X, \mathcal{F}, \mu)$ to $\mathbb{R}$ and $f\colon (X, \mathcal{F}, \mu) \to \mathbb{R}$ be another measurable function. Then, if $f_n \to f$ almost everywhere, then $f_n \to f$ almost uniformly.*

And, lastly, in the case of measurable function on closed unit interval equipped with Borel $\sigma-$algebra and Lebesgue measure, the following type of function has a prominent role in ergodic theory and this study.

**Definition 2.40.** The Borel measurable function $\varphi\colon \mathbb{I} \to \mathbb{I}$ is said to be *measure preserving* if for any $B \in \mathcal{B}(\mathbb{I})$, $\lambda(\varphi^{-1}(B)) = \lambda(B)$.

## 2.2 Probability and Copula Theory

Any finite measure space whose measure of total space is one will be called a *probability space* while that measure will be called a *probability measure*, always write as $(\Omega, \mathcal{F}, P)$, and measurable set in this case is called an *event*. Real-valued measurable function on probablity space is called a *random variable* and the integral of a random variable with respect to the probability measure will be called an *expectation*, i.e. given random variable $X$, define $EX := \int_\Omega X dP$. *Variance of* $X$ is the expectation of the random variable $(X - EX)^2$. Note that variance exists if $|X|^2$ has finite integral, that is $X$ has the finite second moment. The set of elements $\omega \in \Omega$ which yields $X(\omega) \neq 0$ is called *support of* $X$ while support of probability measure is the complement of union of all open sets having the probability measure of zero. Note that support of the probability measure must be one. In several case, support can be smaller than the full set $\Omega$. A vector-valued function whose components are random variable on a common probability space is called a *random vector*. Characteristic function will be called *indicator function* in this context and is written by $1_B$ for measurable set $B$.

Above paragraph is just a collection of probability notions that are borrowed from measure theory. And all contents following this will be the pure probability theory.

**Theorem 2.41.** *Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$, and $P_X(A) := P(X^{-1}(A))$ for all $A \in \mathcal{B}(\mathbb{R})$. Then $P_X$ is a probability measure. This new measure is called a distribution or push-forward measure of $X$.*

**Definition 2.42.** Let $X$ be a random variable on $\Omega$, and

$$F_X(x) := P(\{\omega : X(\omega) \leq x\})$$

for each $x \in \mathbb{R}$. Then $F_X$ is said to be the *cumulative distribution function (cdf)* of $X$. In the case of random vector whose component is $X, Y$, the *joint cumulative of* $X, Y$ is the function $F_{X,Y} : \mathbb{R}^2 \to \mathbb{I}$ defined by, for $x, y \in \mathbb{R}^2$,

$$F_{X,Y}(x, y) := P(X \leq x, Y \leq y)$$

And the cdf of each component is called a *marginal cdf.*

One of the random variables seen usually is uniform random variable. In case of finite set, a uniform random variable on this set is the random variable whose support is this set and has equal probabilty of occurring each outcome is one over size of that set. On the other hand, *uniform random variable on compact interval* $[a, b]$ is the random variable which takes value in $[a, b]$ and has cdf $F(x) := x/(b-a)$ for $x \in [a, b]$, $F(x) := 0$ if $x < a$ and $F(x) := 1$ if $x > b$. It is denoted by $X \sim \text{Uniform}([a, b])$.

Univariate cdf is monotonic, has limit as 0 when argument tends to minus infinity and 1 if argument goes to infinity, and is right-continuous.

The random variable is said to be *continuous* if its cdf is a continuous function and, on the contrary, is *discrete* if its cdf is just an increasing function with at most countably infinite point.

By change of variable formula, one has two equivalent definition of expectation of random variable $X$.

$$EX := \int_\Omega X dP = \int_\mathbb{R} x dF_X(x).$$

One of the topic distinguishing probability from being merely pure measure theory is independence.

**Definition 2.43.** Let $\{B_\alpha \colon \alpha \in \Lambda\}$ be a family of events in probability space $(\Omega, \mathcal{F}, P)$. $\{B_\alpha \colon \alpha \in \Lambda\}$ is said to be *(mutually) independent* if for any finite subfamily $\{\alpha_1, \ldots, \alpha_k\} \subseteq \Lambda$,

$$P\left(\bigcap_{j=1}^{k} B_{\alpha_j}\right) = \prod_{j=1}^{k} P\left(B_{\alpha_j}\right).$$

The class $\{\mathcal{E}_\alpha : \alpha \in \Lambda\}$ is called *independent* if for any choice of event $E_\alpha \in \mathcal{E}_\alpha$, $\{E_\alpha \colon \alpha \in \Lambda\}$ is independent. Last but not least, the collection of random variables $\{X_\alpha \colon \alpha \in \Lambda\}$ is *independent* if the family of $\sigma$-algebras generated by random variable $X_\alpha$ for $\alpha \in \Lambda$ is independent where $\sigma$-algebra generated by random variable $X$, says $\sigma(X)$, is defined by

$$\sigma(X) := X^{-1}(\mathcal{B}(\mathbb{R})) := \{X^{-1}(B) \colon B \in \mathcal{B}(\mathbb{R})\} \tag{2.1}$$

Equivalently, random variables $\{X_\alpha \colon \alpha \in \Lambda\}$ is *independent* if for any finite subcollection $\{\alpha_1, \dots, \alpha_k\} \subseteq \Lambda$ and $B_{\alpha_j} \in \mathcal{B}(\mathbb{R})$ for $j \in \{1, \dots, k\}$,

$$P(X_{\alpha_j} \in B_{\alpha_j}, j \in \{1, \dots, k\}) = \prod_{j=1}^{k} P(X_{\alpha_j} \in B_{\alpha_j}) \tag{2.2}$$

In fact, especially in this work, the case will be restricted to the two random variable case. And, by some measure theoretic technique, it suffices to check independence between two random variables $X, Y$, by just showing that whether joint cdf can be reduced to just multiplication of two marginal cdf.

From above characterization of independence of set of random variables, one can see that independence in random variables holds if and only if joint distribution of subcollection of random variables equals product of marginal distributions of each components for any finite subfamily of all of random variables. This viewpoint is handy for computing task.

On the other hand, given two random variables $X$ and $Y$, complete dependence can be defined in the following way:

**Definition 2.44.** Given some probability space, random variable $Y$ is said to be *completely dependent* on a random variable $X$ if there is a Borel measurable function $f$ such that $P(Y = f(X)) = 1$. In case that $Y$ is completely dependent on $X$ and vice versa, $X$ and $Y$ are *mutually complete dependent.* In case that the function $f$ is strictly monotonic, $Y$ is said to be *monotonic dependent on $X$*.

The other feature of probability theory differentiate itself away from general measure theory is conditioning. Conditioning is mainly the augmentation of expectation and probability with some sort of given class of events, sub $\sigma$-algebra $\mathcal{A}$ of $\mathcal{F}$ in particular.

**Definition 2.45.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{A} \subseteq \mathcal{F}$ a sub $\sigma$-algebra, and $X$ a random variable on above space. Then, the *conditional expectation of $X$ given $\mathcal{A}$* is a random variable $Y$ which is $\mathcal{A}$-measurable and, for $A \in \mathcal{A}$,

$$\int_A Y(\omega)P(d\omega) = \int_A X(\omega)P(d\omega) \tag{2.3}$$

One can see that, by Radon-Nikodym theorem[1], conditional expectation exists and unique a.e. by, given a random variable $X$ and a sub $\sigma$-algebra $\mathcal{A}$, defining set function $\nu$ on $\mathcal{A}$ as the right hand side of identity (2.3). Obviously, $\nu$ is absolutely continuous with respect to $P$ and, thus, the corresponding Radon-Nikodym derivative is indeed conditional expectation and it is unique a.e. as we desire. Up to this point, we denote the conditional expectation of $X$ given $\mathcal{A}$ as $E(X \mid \mathcal{A})$ or $E^{\mathcal{A}}X$. The property (2.3) is called the *averaging property* of conditional expectation.

One of the special case is when the information for conditioning $X$ is other random variable $Y$. Then, to deal with this situation, define the conditional expectation of $X$ given $Y$ as $E(X \mid Y) := E(X \mid \sigma(Y))$.

Next, *conditional probability given* $\mathcal{A}$ can be defined as $P(B \mid \mathcal{A}) := E(1_B \mid \mathcal{A})$, which is naturally similar to unconditional distribution as well.

The problem arises when one wants to find conditional probability of event $B$ given a point $Y = y$ since event $\{Y = y\}$ may have zero measure. To make this possible, use the fact that there exists the measurable function $\phi$ such that $E(X \mid Y) = \phi \circ Y$ $P$-a.e. *The conditional expectation of* $X$ *given* $Y = y$ or a *version of conditional of* $X$ *given* $Y$.

Moreover, the following concept is needed for developing further results.

**Definition 2.46.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A mapping $\kappa\colon \mathbb{R}\times\mathcal{B}(\mathbb{R}) \to \mathbb{I}$ is called as *Markov kernel* if the following are satisfied:

(i) For all $B \in \mathcal{B}(\mathbb{R})$, $x \mapsto \kappa(x, B)$ is a measurable function; and

(ii) For all $x \in \mathbb{R}$, $B \mapsto \kappa(x, B)$ is a probability measure.

From the notion of Markov kernel, it induces an important concept which is used frequently in the studies on theoretical stochastic processes and copula theory called regular conditional distribution.

[1] see any measure theory textbook, such as [17]

**Definition 2.47.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{A} \subseteq \mathcal{F}$ a sub $\sigma$-algebra, and $X$ a random variable on above space. Then (a version of) *regular conditional distribution of* $X$ *given* $\mathcal{A}$ is a Markov kernel $\kappa_{X,\mathcal{A}}$ from $\Omega \times \mathcal{F}$ to $\mathbb{I}$ such that, for any $F \in \mathcal{F}$,

$$\kappa_{X,\mathcal{A}}(\omega, F) = P(X \in F \mid \mathcal{A})(\omega) \tag{2.4}$$

And it can be proved that given some prescribed random variable and sub $\sigma$-algebra, there exists corresponding regular conditional distribution uniquely a.e., see Kallenberg (2002) and Klenke (2013). In particular, the case of interest in the context of this study is when the sub $\sigma$-algebra is generated by other random variable, says we have $Y$ and $\sigma(X)$. Under this circumstance, we have $\kappa_{Y,X} := \kappa_{Y,\sigma(X)}$. Note that for the case mentioned recently, one has, for all $B \in \mathcal{B}(\mathbb{R})$,

$$\kappa_{Y,X}(X(\omega), B) = E(1_B \circ Y \mid X)(\omega) \text{ a.e.} \tag{2.5}$$

Regular conditional distribution is not only useful as the tool for characterize stochastic process, but also handy for computing conditional expectation of a measurable function of random variables $X, Y$ given value of $X$.

And, since one has (regular) conditional distribution function, one can define relative term of independence as follow.

**Definition 2.48.** Let $X, Y, Z$ be random variables. then one says $X, Z$ are *conditionally independent* given $Y$ if, for any Borel sets $B_1, B_2$,

$$P(X \in B_1, Z \in B_2 \mid Y) = P(X \in B_1 \mid Y)P(Z \in B_2 \mid Y) \text{ a.e.}$$

There are many modes of convergence for sequences of random variables or vectors. But one relevant and weakest mode will be introduced here.

**Definition 2.49.** Let $\{\vec{X}_n\}$ be a sequence of random vectors and $\vec{X}$ be other random vector. Then, $\vec{X}_n$ is said to *converge in distribution to* $\vec{X}$ if joint cdf of $\vec{X}_n$ converges to joint cdf of $\vec{X}$ for every continuity point.

Before going further, one introduces, for any subset of Euclidean space $S$, $\mathcal{P}(S)$ as the set of all random vectors or random variables whose supports are $S$.

For the space of real-valued random variables $\mathcal{P}(\mathbb{R})$, the function $R\colon \mathcal{P}(\mathbb{R})^2 \to \mathbb{R}$ is said to be *dependence measure in the sense of Rényi* if the following are fulfilled:

(i) $R$ is well-defined;

(ii) $R$ is *normalized*, i.e. $0 \le R(X,Y) \le 1$ for any $X, Y \in \mathcal{P}(\mathbb{R})$;

(iii) For $X, Y \in \mathcal{P}(\mathbb{R})$, $R(X,Y) = 0$ if and only if $X, Y$ are independent;

(iv) $R$ is *complete dependence detectable*, i.e. for all $X, Y \in \mathcal{P}(\mathbb{R})$, $R(X,Y) = 1$ if either $Y$ is CD on $X$ or vice versa;

(v) $R$ is *bijective invariant*, i.e. for any Borel injections $f, g$, $R(f(X), g(Y)) = R(X,Y)$ for any $X, Y \in \mathcal{P}(\mathbb{R})$;

(vi) If (X,Y) are jointly normal random vector with correlation coefficient $\rho$, then $R(X,Y) = |\rho|$.

Then, Schweizer and Wolff criticized the Rényi's postulation to be too strong and, in that day, there is only maximal correlation which can satisfy all of Rényi's axiom (see [18]). Moreover, maximal correlation takes unit value too often since it can be one whenever one of two random variables is just the Borel measurable function of another. Thus, they propose the new postulation.

For the space of real-valued random variables $\mathcal{P}(\mathbb{R})$, the function $R\colon \mathcal{P}(\mathbb{R})^2 \to \mathbb{R}$ is said to be *dependence measure in the sense of Schweizer-Wolff* if the following are fulfilled:

(i) $R$ is well-defined;

(ii) $R$ is *normalized*, i.e. $0 \le R(X,Y) \le 1$ for any $X, Y \in \mathcal{P}(\mathbb{R})$;

(iii) For $X, Y \in \mathcal{P}(\mathbb{R})$, $R(X,Y) = 0$ if and only if $X, Y$ are independent;

(iv) $R$ is *monotonic dependence detectable*, i.e. for all $X, Y \in \mathcal{P}(\mathbb{R})$, $R(X,Y) = 1$ if and only if $Y$ is MD on $X$;

(v) $R$ is *monotonic invariant*, i.e. for any strictly monotonic mappings $f, g$, $R(f(X), g(Y)) = R(X, Y)$ for any $X, Y \in \mathcal{P}(\mathbb{R})$;

(vi) If $(X, Y)$ is jointly normal random vector with correlation coefficient $\rho$, then $R(X, Y) = \varphi(|\rho|)$ for some strictly increasing function $\varphi$;

(vii) If $\{(X_n, Y_n)\}_{n\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^2)$ such that $(X_n, Y_n) \to (X, Y)$ in distribution to some random vector $(X, Y) \in \mathcal{P}(\mathbb{R}^2)$, then $\lim_{n\to\infty} R(X_n, Y_n) = R(X, Y)$.

After the study of Schweizer and Wolff, Siburg and Stoimenov invented the dependence measure called as $\omega$ which satisfying maximal dependence detectability in the middle of two extreme dependence structures, says monotonic and complete dependence (see [19]).

For the space of real-valued random variables $\mathcal{P}(\mathbb{R})$, the function $R\colon \mathcal{P}(\mathbb{R})^2 \to \mathbb{R}$ is said to be *dependence measure in the sense of Siburg-Stoimenov* if the following is fulfilled:

(i) $R$ is well-defined;

(ii) $R$ is *normalized*, i.e. $0 \leq R(X, Y) \leq 1$ for any $X, Y \in \mathcal{P}(\mathbb{R})$;

(iii) For $X, Y \in \mathcal{P}(\mathbb{R})$, $R(X, Y) = 0$ if and only if $X, Y$ are independent;

(iv) $R$ is *mutually dependence detectable*, i.e. for all $X, Y \in \mathcal{P}(\mathbb{R})$, $R(X, Y) = 1$ if and only if $Y$ and $X$ are MCD;

(v) $R$ is *monotonic invariant*, i.e. for any strictly monotonic mappings $f, g$, $R(f(X), g(Y)) = R(X, Y)$ for any $X, Y \in \mathcal{P}(\mathbb{R})$;

(vi) If $\{(X_n, Y_n)\}_{n\in\mathbb{N}} \subseteq \mathcal{P}(\mathbb{R}^2)$ such that $(X_n, Y_n) \to (X, Y)$ in distribution to some random vector $(X, Y) \in \mathcal{P}(\mathbb{R}^2)$, then $\lim_{n\to\infty} R(X_n, Y_n) = R(X, Y)$.

Under some circumstances, one may consider the real-valued function on the Cartesian product of the sets of random variables which have almost the same properties with dependence measures in above senses but symmetry. This will be called a *non-symmetric dependence measure* and one writes $R(Y \mid X)$ instead of $R(X, Y)$ for any random variables $X, Y$. *Invariance property* for a non-symmetric dependence measure is defined as $R(Y \mid f(X)) = R(Y \mid X)$ for any $f$ in some kinds of functions such as monotonic or injective transformations. And, *complete/ monotonic dependent detectability* is the property that $R(Y|X) = 1$ if and only if $Y = f(X)$ for $f$ Borel measurable/monotonic function.

Since the dependence measure in any sense need invariance, one probabilistic object which is reasonable for using as ingredient for dependence measure is copula.

**Definition 2.50.** *A (bivariate) copula* $C$ is the function $C\colon \mathbb{I}^2 \to \mathbb{I}$ satisfying the following conditions:

(i) (Boundary condition) $\forall x, y \in \mathbb{I}^2$, $C(x, 0) = 0 = C(0, y)$, $C(x, 1) = x$, and $C(1, y) = y$; and

(ii) (2-increasing property) for any rectangle $R := [x_1, x_2] \times [y_1, y_2]$ where $x_1, x_2, y_1, y_2 \in \mathbb{I}$,

$$V_C(R) := C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1) \geq 0. \tag{2.6}$$

The set of all copulas will be denoted by $\mathcal{C}$. And, by standard measure theoretic technique, one can extend $V_C$ to a unique measure $\mu_C$ on $\mathbb{I}^2$ such that $\mu_C(R) = V_C(R)$ for any measurable rectangle $R \in \mathcal{B}(\mathbb{I}^2)$. Moreover, from boundary condition, $\mu_C$ becomes *doubly stochastic*. That is; for any $B \in \mathcal{B}(\mathbb{I})$,

$$\mu_C(B \times \mathbb{I}) = \lambda(B) = \mu_C(\mathbb{I} \times B).$$

Since copula can induce two-fold measure, the notion of support can be considered. The support of copula means the support of measure induced by that copula.

One of the essential properties of copula, inheriting from 2-increasing property, is the non-decreasing in each place. This implies that $\partial_i C$ exists $\lambda$-a.e. on $\mathbb{I}$ and $Ran(\partial_i C) \in \mathbb{I}$ for any $i \in \{1, 2\}$. Moreover, for $C \in \mathcal{C}$, one can define *transposition of* $C$ by $C^T(u, v) := C(v, u)$ for $u, v \in \mathbb{I}$. Obviously, transposition is still copula.

From above properties, one has the following technique for viewing integral of partial derivative of copula as the integral with respect to Lebesgue-Stieltjes measure induced by partial derivative, see [8].

**Proposition 2.51.** *Let $f$ be a bounded measurable function on $\mathbb{I}$ and $C$ be a copula. Then, one has*

$$\int_{\mathbb{I}} f(t)\partial_1 C(x, dt) = \frac{d}{dx}\int_{\mathbb{I}} f(t)\partial_2 C(x,t)dt$$

*for almost all $x \in \mathbb{I}$ with respect to $\lambda$.*

Moreover, to integrate 2-place function on $\mathbb{I}^2$ with respect to $\mu_C$ where $C$ is any copula, the following theorem is a so-called disintegration theorem.

**Proposition 2.52.** *Let $C$ be a copula and $g\colon \mathbb{I}^2 \to \mathbb{R}$ an integrable function with respect to $\mu_C$, then*

$$\int_{\mathbb{I}^2} g(x,y)d\mu_C(x,y) = \int_{\mathbb{I}}\int_{\mathbb{I}} g(x,y)\partial_1 C(x,dy)dx = \int_{\mathbb{I}}\int_{\mathbb{I}} g(x,y)\partial_2 C(dx,y)dy.$$

The following theorem is milestone to make copula become important.

**Theorem 2.53.** *(Sklar) For a pair of random variables $(X,Y)$ on common probability space $(\Omega, \mathcal{F}, P)$ with joint distribution $F_{X,Y}$ and marginal distributions $F_X, F_Y$ of $X, Y$ respectively. Then, there is a copula $C$ satisfying*

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)) \tag{2.7}$$

*for all $x, y \in \bar{\mathbb{R}}$. The copula corresponding to the joint distribution is unique up to $Ran(F_X) \times Ran(F_Y)$. Conversely, given a copula $C$ and distribution functions $F_X, F_Y$ of $X, Y$ respectively, then the 2-place function defined by (2.7) is a joint distribution function of $X, Y$ with the marginals $F_X, F_Y$.*

Sklar's theorem is powerful especially in the case of considering the random variables with continuous distribution functions. Thus, we will assume that every random variables in this work have continuous distribution functions to obtain that any pair of random variables has the unique copula to fulfill the condition in Sklar's theorem. Under this circumstance, denote as $(X, Y) \sim A$ where $A$ is the unique copula satisfying assertion (2.7), and we will call that $A$ *is connecting copula of* $(X, Y)$. In other word, we will say that the copula connecting $(X, Y)$ is denoted by $A_{X,Y}$. Furthermore, Sklar's theorem also give equivalence definition of copula to be the joint distribution function of two uniform random variables on $\mathbb{I}$.

The upcoming property can be considered as *scale invariant property* for copula.

**Proposition 2.54.** *Given random variables $X, Y$ with connecting copula $A_{X,Y}$ and $f, g$ strictly monotonic functions. Then, the following holds true:*

*(i) if both $f, g$ are strictly increasing, then $A_{f(X),g(Y)} = A_{X,Y}$;*

*(ii) if $f$ is strictly increasing and $g$ is strictly decreasing, then $A_{f(X),g(Y)}(u, v) = u - A_{X,Y}(u, 1 - v)$ for all $u, v \in \mathbb{I}$;*

*(iii) if $f$ is strictly decreasing and $g$ is strictly increasing, then $A_{f(X),g(Y)}(u, v) = v - A_{X,Y}(1 - u, v)$ for all $u, v \in \mathbb{I}$;*

*(iv) if both $f, g$ are strictly decreasing, then $A_{f(X),g(Y)}(u, v) = u + v - 1 + A_{X,Y}(1 - u, 1 - v)$ for all $u, v \in \mathbb{I}$.*

These are important examples of copulas.

**Example 2.55.**
- (Counter-monotonic copula) $W(x, y) := \max\{x + y - 1, 0\}$;
- (Independence copula) $\Pi(x, y) := xy$;
- (Co-monotonic copula) $M(x, y) := \min\{x, y\}$.

The reasons to give these examples are, firstly, $W, M$ are pointwise lower, upper bound of copula space. These are called *Fréchet-Hoeffding bounds.* And, additionally, all of them have probabilistic meanings, which are indicated by the following proposition.

**Proposition 2.56.** *For random variables $X, Y$ with continuous distributions, one has* *(i)* $(X, Y) \sim W$ *if and only if there is strictly decreasing function* $f$ *such that* $Y = f(X)$ *with probability 1;*

*(ii)* $(X, Y) \sim \Pi$ *if and only if* $X, Y$ *are independent;*

*(iii)* $(X, Y) \sim M$ *if and only if there is strictly increasing function* $f$ *such that* $Y = f(X)$ *with probability 1.*

For the space of all copulas, one can give the algebraic structure by the following binary operation.

**Definition 2.57.** On the space of all copulas, says $\mathcal{C}$, define the operation $*$ on $\mathcal{C}$ by, for $A, B \in \mathcal{C}$ and $(u, v) \in \mathbb{I}^2$,

$$A * B(u, v) := \int_{\mathbb{I}} \partial_2 A(u, t) \partial_1 B(t, v) dt. \tag{2.8}$$

This operation is named as *(Markov) *-product.*

Darsow, Nguyen and Olsen (see [4]) proved that its is binary operation and $(\mathcal{C}, *)$ is monoid with identity element $M$ and, additionally, null element $\Pi$. Moreover, one can also show straightforwardly that, for $(u, v) \in \mathbb{I}^2$ and $A \in \mathcal{C}$,

(i) $W * A(u, v) = v - A(1 - u, v)$;

(ii) $A * W(u, v) = u - A(u, 1 - v)$;

(iii) $W * A * W(u, v) = u + v - 1 + A(1 - u, 1 - v)$.

Hence, for random variables $X, Y$ with continuous distribution functions whose connecting copula $A_{X,Y} \in \mathcal{C}$ and strictly monotonic mappings $f, g$,

(i) if both $f, g$ are strictly increasing, then $A_{f(X),g(Y)} = A_{X,Y}$;

(ii) if $f$ is strictly increasing and $g$ is strictly decreasing, then $A_{f(X),g(Y)} = A_{X,Y} * W$;

(iii) if $f$ is strictly decreasing and $g$ is strictly increasing, then $A_{f(X),g(Y)} = W * A_{X,Y}$;

(iv) if both $f, g$ are strictly decreasing, then $A_{f(X),g(Y)} = W * A_{X,Y} * W$.

$*$-product can be used to characterize the copula connecting two random variables, if one has information about other random variable which forms the conditional independence.

**Theorem 2.58.** *Let $X, Y, Z$ be random variables such that $X, Z$ are conditionally independent given $Y$ and $C_{X,Y}, C_{Y,Z}, C_{X,Z}$ are corresponding connecting copula, then $C_{X,Z} = C_{X,Y} * C_{Y,Z}$.*

And Ruankong, Santiwipanont, and Sumetkijakan (2013), see [16], proved this lemma.

**Lemma 2.59.** *Let $X, Y$ be random variables and $g\colon \mathbb{R} \to \mathbb{R}$ a Borel measurable function, then $g(X), Y$ are conditionally independent given $X$.*

Combining this result with theorem above yields this result. In the case of Borel injection, the another result will be consequence from Lusin-Suslin theorem.

**Lemma 2.60.** *Let $X, Y$ be random variables and $g\colon \mathbb{R} \to \mathbb{R}$ a Borel measurable function, then $C_{g(X),Y} = C_{g(X),X} * C_{X,Y}$. Moreover if $g$ is Borel injection, then $C_{X,Y} = C_{X,g(X)} * C_{g(X),Y}$.*

In view of algebra, it is natural to ask about invertibility of elements in copula space. One will define as follow.

**Definition 2.61.** Let $C \in \mathcal{C}$. Then, $C$ is said to be *left (right) invertible* if $C^T * C = M$ ($C * C^T = M$). If it is both left and right invertible, it is called *invertible.*

The following proposition gives the characterization of complete dependence between random variables by partial derivatives, and invertibility of the corresponding connecting copula.

**Proposition 2.62.** *Let $X, Y$ be random variables with associating connecting copula $C \in \mathcal{C}$, then the following are equivalent:*

- *$Y$ is completely dependent on $X$;*
- *$C$ is left invertible;*
- *$\partial_1 C \in \{0,1\}$ a.e.*

*Similarly, the following are equivalent:*

- *$X$ is completely dependent on $Y$;*
- *$C$ is right invertible;*
- *$\partial_2 C \in \{0,1\}$ a.e.*

Surprisingly, copula can be connected with the special type of linear maps on $L^\infty$, which is defined below.

**Definition 2.63.** The linear map $T : L^\infty(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda) \to L^\infty(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda)$ is said to be a *Markov operator* if it is positive, has a constant function on $(\mathbb{I}, \mathcal{B}(\mathbb{I}))$ as fixed point, and preserves integral in the sense that, for $f \in L^\infty(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda)$, $\int_{\mathbb{I}}(Tf)(t)dt = \int_{\mathbb{I}} f(t)dt$.

Durante (see [6]) showed that any Markov operator $T : L^\infty(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda) \to L^\infty(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda)$ has the unique extension to $T : L^p(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda) \to L^p(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda)$ for $p \in [1, \infty)$. Thereby, it is legitimate to use the terminology of Markov operator on any type of $L^p$ space, especially $L^1$ as in [20]. And, for convenience, the space of all Markov operators will be denoted by $\mathcal{M}$.

The connection between copula and markov operator is established by Olsen, Darsow, and Nguyen (see [11]), which is stated here.

**Theorem 2.64.** *Algebraic isomorphism between $(\mathcal{C}, *)$ and $(\mathcal{M}, \circ)$ where $\circ$ denotes composition of functions is constituted by the following two mappings:*

*(i)* $\Phi\colon \mathcal{C} \to \mathcal{M}$ *defined by*

$$\Phi(A)(f)(x) := (T_A f)(x) := \frac{d}{dx}\int_{\mathbb{I}} \partial_2 A(x,t) f(t)dt;$$

*(ii)* $\Psi\colon \mathcal{M} \to \mathcal{C}$ *defined by*

$$\Psi(T)(x,y) := A_T(x,y) := \int_{[0,x]} (T1_{[0,y]})(t)dt$$

*for any* $A \in \mathcal{C}$, $f \in L^1(\mathbb{I}, \mathcal{B}(\mathbb{I}), \lambda)$, *and* $x, y \in \mathbb{I}$.

The connection between copula and Markov kernel is shown in [6] in the following statement. Given random variable $X, Y$ with connecting copula $A \in \mathcal{C}$ and $\kappa_A$ a regular conditional distribution of $Y$ given $X$, then

$$\mu_A(B_1 \times B_2) = \int_{B_1} \kappa_A(x, B_2)dx.$$

In particular, $\lambda(B) = \int_{\mathbb{I}} \kappa_A(x, B)dx$. Conversely, any Markov kernel $\kappa$ satisfying above assertion will be automatically a regular conditional distribution of $Y$ given $X$. Accordingly, one may refer to a *Markov kernel of copula* $A$ for instead of regular conditional distribution of underlying random variables directly.

Given a copula space $\mathcal{C}$, define $D_1\colon \mathcal{C}^2 \to \mathbb{R}$ by, for $A, B \in \mathcal{C}$

$$D_1(A,B) := \int_{\mathbb{I}}\int_{\mathbb{I}} |\kappa_A(x,[0,y]) - \kappa_B(x,[0,y])|dydx. \tag{2.9}$$

Trutschnig (see [20]) proved that $(\mathcal{C}, D_1)$ is a complete separable metric space and, for any random variable $X, Y$ with connecting copula $A$, $Y$ is complete dependent to $X$ if and only if $D_1(A, \Pi) = \frac{1}{3}$. These fact will lead us to the construction of Trutschnig's $\zeta_1$, and it symmetric version.

For more convenience, this lemma help us to think of regular conditonal distribution, analytically, in term of derivative and Markov operator.

**Lemma 2.65.** *For any* $A \in \mathcal{C}$ *and* $x, y \in \mathbb{I}$ *,*

$$\kappa_A(x, [0,y]) = \partial_1 A(x,y) = T_A 1_{[0,y]}(x).$$

*Proof.* By the property of Markov kernels of copulas, one has

$$\int_{[0,x]} \kappa_A(t,[0,y])d\lambda(t) = \mu_A([0,x]\times[0,y]) = A(x,y).$$

Since Markov kernel is bounded and integrable, Lebesgue differentiation theorem can be applied here to obtain that

$$\partial_1 A(x,y) = \frac{d}{dx}\int_{[0,x]} \kappa_A(t,[0,y])d\lambda(t) = \kappa_A(x,[0,y]).$$

Furthermore, by isomorphism between copula and Markov operator spaces, one has $A(x,y) = \int_{[0,x]} T_A 1_{[0,y]}(t)d\lambda(t)$. By the same argument to the above assertion, one gets

$$\kappa_A(x,[0,y]) = \partial_1 A(x,Y) = \frac{d}{dx}\int_{[0,x]} T_A 1_{[0,y]}(t)d\lambda(t) = T_A 1_{[0,y]}(x).$$

This finishes the proof. $\square$

By above lemma, one can rewrite $D_1(A,B)$, for $A,B \in \mathcal{C}$ as follow.

$$D_1(A,B) = \int_{\mathbb{I}}\int_{\mathbb{I}} |\partial_1 A(x,y) - \partial_1 B(x,y)|\, dxdy. \tag{2.10}$$

And, in case of modifying $\zeta_1$ into symmetric version, the following are useful.

**Lemma 2.66.** *Let $X,Y$ be random variables such that $(X,Y) \sim A \in \mathcal{C}$, then $(Y,X) \sim A^T$ where $\cdot^T$ is transposition.*

*Proof.* Let $X,Y$ be two random variables with $X \sim F_X$, $Y \sim F_Y$ and $(X,Y)$ has the connecting copula $A$. Then, by Sklar's theorem, one has, for $x,y \in \mathbb{R}$,

$$\begin{aligned} P(Y \le y, X \le x) &= P(X \le x, Y \le y) \\ &= A(F_X(x), F_Y(y)) \\ &= A^T(F_Y(y), F_X(x)). \end{aligned} \tag{2.11}$$

Thanks to Sklar's theorem, again, corresponding copula connecting $(Y,X)$ is unique, thus we are done. $\square$

In fact, from measure preserving functions $f,g\colon \mathbb{I} \to \mathbb{I}$ and uniform random variable $X$ on $\mathbb{I}$, the corresponding connect copula of the random variables $f(X), g(X)$, denoted by $C_{f,g}$ can be found by

$$\begin{aligned} C_{f,g}(x,y) &= P(f(X) \le x, g(X) \le y) \\ &= P(X \in f^{-1}([0,x]) \cap g^{-1}([0,y])) \\ &= \lambda(f^{-1}([0,x]) \cap g^{-1}([0,y])). \end{aligned}$$

Lastly, the following definition of shuffle of Min is given.

**Definition 2.67.** The copula $C$ is a *shuffle of Min* if there exist $k \in \mathbb{N}$, finite collections $\{s_i\}_{i=0}^k, \{t_j\}_{j=0}^k \subseteq \mathbb{I}$ and $\sigma \in S^k$ such that $\{s_i\}_{i=0}^k, \{t_j\}_{i=0}^k$ form partitions of $\mathbb{I}$ and, for $1 \leq i \leq k$, each of $[s_{i-1}, s_i] \times [t_{\sigma(i)-1}, t_{\sigma(i)}]$ is square containing mass of $s_i - s_{i-1}$ of $C$ uniformly on either one of diagonals.

Shuffle of Mins are invertible. This fact will be used later.

## 2.3 Basic Information Theory

In this part, the basic notions of information theory used in this work will be treated here.

**Definition 2.68.** Given discrete random variables $X, Y$ whose supports are the sets $\mathcal{X}, \mathcal{Y}$ respectively, one can define the *(Shannon's) entropy* as

$$H(X) := \sum_{x \in \mathcal{X}} \log \left( \frac{1}{P(X = x)} \right) P(X = x). \tag{2.12}$$

**Definition 2.69.** Let $X, Y$ be discrete random variables or vectors with common support $\mathcal{X}$ and probability mass function $p, q$ respectively. The *Kullback-Leiber divergence*, or shortly *divergence*, of $p$ and $q$ as

$$D(X \parallel Y) := D(p \parallel q) := \sum_{x \in \mathcal{X}} p(x) \log \left( \frac{p(x)}{q(x)} \right). \tag{2.13}$$

Note that, intuitively, divergence is the average logarithmic discrepency between two distributions when using one as the reference for computing expectation. And one of its special caseis one of the most important concept in information theory is mutual information. The definition is given below.

**Definition 2.70.** Let $X, Y$ be discrete random variables whose supports are the sets $\mathcal{X}, \mathcal{Y}$ respectively. The *mutual information* as

$$I(X, Y) := \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \log \left( \frac{P(X = x, Y = y)}{P(X = x)P(Y = y)} \right) P(X = x, Y = y). \tag{2.14}$$

Moreoever, we call $\frac{I(X,Y)}{\log \min\{|\mathcal{X}|, |\mathcal{Y}|\}}$ the *normalized score of* $X, Y$.

Again, one can view mutual information as $I(X,Y) = D(p \parallel q)$ where $p$ is the probability mass function of $(X,Y)$ and $q$ is the probability mass function of $(X,Y')$ where $X,Y'$ are independent and $Y'$ is identically distributed to $Y$.

Note that by convexity of logarithm function and Jensen's inequality, both entropy and divergence are non-negative. Moreover, the mutual information has a bound

$$I(X,Y) \leq \log \min\{H(X), H(Y)\}. \tag{2.15}$$

For the comprehensive treatment on information theory, see [2].

However, all of above deinitions are defined for discrete random variables and random vectors. For the case of continuous random variables and vectors, as assumed throughout this thesis, one may consider cutting the range of random vectors into several pieces and put the probability mass to each piece by using the probability that the values of random vectors lies in that piece. In the case of considering random variables with compact support, the following definition gives us a nice way to partition the range into the collection of subrectangles.

**Definition 2.71.** For the compact interval $[a,b] \subseteq \mathbb{R}$ and natural numbers $k,l$, the *$k$-by-$l$ grid of* $[a,b]^2$ (or, shortenly $k$-$l$ grid) is a partition of $[a,b]^2$ into $k$ columns and $l$ rows respectively. The collection of rows and column of $k$-by-$l$ grid will be denoted as $\{C_i\}_{i=1}^k$ and $\{R_j\}_{j=1}^l$, respectively. The set of all $k$-by-$l$ grids is denoted by $G(k,l)$.

Since one has a grid, any pair of random variables jointly distributed on $\mathbb{I}^2$ can be used to make the new discrete random variables. The method to do this task is given below.

**Definition 2.72.** For random variables $X,Y$ jointly distributed on $\mathbb{I}^2$ and a $k$-by-$l$ grid $G$ with columns $\{C_i\}_{i=1}^k$ and rows $\{R_j\}_{j=1}^l$, $\mathrm{Col}_G(X,Y)$ is the random variable defined by $\mathrm{Col}_G(X,Y)(\omega) := i$ whenever $X(\omega) \in C_i$. And $\mathrm{Row}_G(X,Y)$ is the random variable defined by $\mathrm{Row}_G(X,Y)(\omega) := j$ if $Y(\omega) \in R_j$. Denote random vector

$$(X,Y)_G := (\mathrm{Col}_G(X,Y), \mathrm{Row}_G(X,Y)).$$

This is called as *discretization of* $(X, Y)$ *by grid* $G$.

From this definition, one can define maximal information coefficient.

**Definition 2.73.** For variables $X, Y$ jointly distributed on $\mathbb{I}^2$, $k, l \in \mathbb{N}$, and a $k$-by-$l$ grid $G$, define $I_G(X, Y) := I(\mathrm{Col}_G(X, Y), \mathrm{Row}_G(X, Y))$. And, furthermore, given random vector $(X, Y)$ on $\mathbb{I}^2$ and $k, l \in \mathbb{N}$, define the *k-by-l characteristic entry of (X, Y)* as

$$M_{k,l}(X, Y) := \sup_{G \in G(k,l)} \frac{I_G(X, Y)}{\log \min\{k, l\}}.$$

**Definition 2.74.** For random variables $X, Y$ jointly distributed on $\mathbb{I}^2$, define *mutual information of* $X, Y$, by

$$\mathrm{MIC}(X, Y) := \sup_{k,l \in \mathbb{N} \setminus \{1\}} M_{k,l}(X, Y) = \sup_{k,l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k,l)} \frac{I_G(X, Y)}{\log \min\{k, l\}}. \tag{2.16}$$

From the definition, one can view MIC as the maximal normalized score of those random variables over every possible choices of grids' size and grid-cutting methods.

# CHAPTER III

# NON-SYMMETRIC DEPENDENCE MEASURES

To ensure the well-defined property of dependence measures discussed in this thesis, all random variables are assumed to have continuous distribution function.

## 3.1 Trutschnig's dependence measure

First of all, recall the definition of Trutchnig's dependence measure $\zeta_1$. For the random variables $X, Y$ with connecting copula $C_{X,Y}$, one defines

$$\zeta_1(Y \mid X) := 3D_1(C_{X,Y}, \Pi) = 3\int_{\mathbb{I}}\int_{\mathbb{I}} |\partial_1 C_{X,Y}(x,y) - \partial_1 \Pi(x,y)| dxdy.$$

The properties of $\zeta_1$ are listed in the following single theorem.

**Theorem 3.1.** *For any random variables $X, Y$, $\zeta_1(Y \mid X)$ satisfies all of the following preoperties:*

*(i)* $\zeta_1(Y \mid X)$ *is well-defined;*

*(ii)* $0 \le \zeta_1(Y \mid X) \le 1$;

*(iii)* $\zeta_1(Y \mid X) = 0$ *if and only if* $X, Y$ *are independent;*

*(iv)* $\zeta_1(Y \mid X) = 1$ *if and only if* $Y$ *is completely dependent on* $X$*;*

*(v) For any Borel injection* $f\colon \mathbb{R} \to \mathbb{R}$, $\zeta_1(Y \mid f(X)) = \zeta_1(Y \mid X)$.

*Proof.* In [20], the $\zeta_1$ is proved to have the first four properties. However, the last property, (v) is not proved. To prove this, the following lemma is needed.

**Lemma 3.2.** *For any copulas $A$ and $B$ , $D_1(A * B, \Pi) \le D_1(B, \Pi)$.*

*Proof.* Let $A, B \in \mathcal{C}$, since $\Pi$ is the null element of $\mathcal{C}$,

$$
\begin{aligned}
D_1(A * B, \Pi) &= D_1(A * B, A * \Pi) \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \left| \frac{\partial}{\partial x} A * B(x,y) - \frac{\partial}{\partial x} A * \Pi(x,y) \right| dxdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \left| \frac{\partial}{\partial x} \int_{\mathbb{I}} \frac{\partial}{\partial t} A(x,t) \left( \frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y) \right) dt \right| dxdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \left| \int_{\mathbb{I}} \left( \frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y) \right) \partial_1 A(x, dt) \right| dxdy \\
&\leq \int_{\mathbb{I}} \int_{\mathbb{I}} \int_{\mathbb{I}} \left| \frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y) \right| \partial_1 A(x, dt) dxdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \frac{d}{dx} \int_{\mathbb{I}} \left| \frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y) \right| \partial_2 A(x,t) dtdxdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \left| \frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y) \right| \partial_2 A(1,t) dtdy \\
&\quad - \int_{\mathbb{I}} \int_{\mathbb{I}} \left| \frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y) \right| \partial_2 A(0,t) dtdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \left| \frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y) \right| dtdy \\
&= D_1(B, \Pi).
\end{aligned}
$$

This chain of inequality, de facto, requires several clarifications. The fourth and sixth lines are applications of Proposition 2.51. The fifth line comes from the Jensen's inequality and recall that absolute value is a convex function. The seventh line is the consequence of fundamental theorem of calculus. This finishes the proof. □

The above lemma can be used to prove the fifth property of $\zeta_1(\cdot \mid \cdot)$ Asssuming $X, Y$ are random variables with connecting copula $C_{X,Y}$ and $f$ is a Borel injection. By above lemma and Lemma 2.60, one has

$$D_1(C_{f(X),Y}, \Pi) = D_1(C_{f(X),X} * C_{X,Y}, \Pi) \leq D_1(C_{X,Y}, \Pi)$$

and

$$D_1(C_{X,Y}, \Pi) = D_1(C_{X,f(X)} * C_{f(X),Y}, \Pi) \leq D_1(C_{f(X),Y}, \Pi).$$

Thereby, one has

$$\zeta_1(Y \mid f(X)) = 3D_1(C_{f(X),Y}, \Pi) = 3D_1(C_{X,Y}, \Pi) = \zeta_1(Y \mid X).$$

□

## 3.2 Li's generalization of copula-based dependence measures

The more general version of Truschnig's $\zeta_1$ and, in fact, Dette et al.'s $r$, is Li's $\tau$, which is defined by, for random variables $X, Y$ with connecting copula $C_{X,Y}$,

$$\tau(Y \mid X) := \frac{\tau(C_{X,Y})}{\tau(M)}$$

where $\varphi\colon [-1,1] \to [0,\infty]$ such that $\varphi$ is continuous, convex, $\varphi(x) = 0$ if and only if $x = 0$, and, for any $C \in \mathcal{C}$,

$$\tau(C) := \int_{\mathbb{I}} \int_{\mathbb{I}} \varphi(\partial_1 C(x,y) - \partial_1 \Pi(x,y)) dxdy.$$

Note that both $\zeta_1$ and $\tau$ are special cases of this class of dependence measure. Before going on, the following technical is the main tool for proving various properties of $\tau$, which is called *data-processing inequality (DPI)* in [10].

**Lemma 3.3.** *For any copulas $A, B \in \mathcal{C}$, $\tau(A * B) \leq \tau(B)$. Equality holds if and only if $\partial_1 B(t,y) - y$ is almost surely constant function with respect to measure induced by Lebesgue-Stieltjes measure induced by $\partial_1 A(x,\cdot)$ almost all $x, y \in \mathbb{I}$.*

*Proof.* Let $A, B \in \mathcal{C}$, then one has

$$\begin{aligned}
\tau(A * B) &= \int_{\mathbb{I}} \int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial x} A * B(x,y) - \frac{\partial}{\partial x} A(x,y)\right) dxdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial x} A * B(x,y) - \frac{\partial}{\partial x} A * \Pi(x,y)\right) dxdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial x} \int_{\mathbb{I}} \frac{\partial}{\partial t} A(x,t) \left(\frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y)\right) dt\right) dxdy \\
&= \int_{\mathbb{I}} \int_{\mathbb{I}} \varphi\left(\int_{\mathbb{I}} \left(\frac{\partial}{\partial t} B(t,y) - \frac{\partial}{\partial t} \Pi(t,y)\right) \partial_1 A(x,dt)\right) dxdy
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_{\mathbb{I}}\int_{\mathbb{I}}\int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial t}B(t,y) - \frac{\partial}{\partial t}\Pi(t,y)\right)\partial_1 A(x,dt)dxdy \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} \frac{d}{dx}\int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial t}B(t,y) - \frac{\partial}{\partial t}\Pi(t,y)\right)\partial_2 A(x,t)dtdxdy \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial t}B(t,y) - \frac{\partial}{\partial t}\Pi(t,y)\right)\partial_2 A(1,t)dtdy \\
&\quad - \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial t}B(t,y) - \frac{\partial}{\partial t}\Pi(t,y)\right)\partial_2 A(0,t)dtdy \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi\left(\frac{\partial}{\partial t}B(t,y) - \frac{\partial}{\partial t}\Pi(t,y)\right)dtdy \\
&= \tau(B).
\end{aligned}
$$

For the main argument for showing this chain of inequalities, it uses the same arguments as Lemma 3.2 except the fifth line which this proof imposes Jensen's inequality in the general setup. Thus, by the equality condition of Jensen's inequality, the strict equality of the fifth line holds if and only if $\frac{\partial}{\partial t}B(t,y) - \frac{\partial}{\partial t}\Pi(t,y)$ is almost surely constant on $t$ with respect to measure induced by $\partial_1 A(x,.)$ for almost all $(x,y) \in \mathbb{I}^2$ with respect to $\lambda_2$. □

For the properties of $\tau$, the following theorem summarized all of them.

**Theorem 3.4.** *For any random variables $X, Y$, $\tau(Y \mid X)$ satisfies all of the following properties:*

*(i)* $\tau(Y \mid X)$ *is well-defined;*

*(ii)* $0 \leq \tau(Y \mid X) \leq 1$;

*(iii)* $\tau(Y \mid X) = 0$ *if and only if* $X, Y$ *are independent;*

*(iv)* $\tau(Y \mid X) = 1$ *if and only if* $Y$ *is completely dependent on* $X$*;*

*(v) For any Borel injection* $f\colon \mathbb{R} \to \mathbb{R}$, $\tau(Y \mid f(X)) = \tau(Y \mid X)$.

*Proof.* (i) It's obviously due to the assumption that random variables in this study possess continuous ditribution functions and, thus, corresponding connecting copula exists and is unique.

(ii) To prove for the lower bound, let $X, Y$ be random variables with connecting copula $C$, then one has

$$\begin{aligned}\tau(C) &= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi(\partial_1 C(x,y) - \partial_1 \Pi(x,y))dxdy \\ &= \int_{\mathbb{I}^2} \varphi(\partial_1 C(x,y) - y)d\lambda_2(x,y) \\ &\geq \varphi\left(\int_{\mathbb{I}}\int_{\mathbb{I}} (\partial_1 C(x,y) - y)dxdy\right) \qquad (3.1) \\ &= \varphi\left(\int_{\mathbb{I}} (C(1,y)C(0,y) - y)dy\right) \\ &= \varphi(0) \\ &= 0 \\ &= \tau(\Pi).\end{aligned}$$

The Fubini's theorem is applied from line one to line three while the third line uses the Jensen's inequality to plug the integral into $\varphi$, thus $\tau(C) \geq 0$ as needed. For an upper bound in (ii), since $C * M = C$, Lemma 3.3 inequality says that $\tau(C) = \tau(C * M) \leq \tau(M)$.

(iii) Firstly, let $\tau(C) = 0$. From the proof of (ii), the inequality (3.1) becomes equality if and only if $\partial_1 C(x,y) - y$ is almost surely constant with respect to $\lambda_2$. That is; there exists $K \in \mathbb{R}$ such that $\partial_1 C(x,y) - y = K$ almost all $(x,y) \in \mathbb{I}^2$ with respect to $\lambda_2$. Then, by the fundamental theorem of calculus,

$$\begin{aligned}K &= \int_{\mathbb{I}}\int_{\mathbb{I}} (\partial_1 C(x,y) - y)dxdy \\ &= \int_{\mathbb{I}}\left(\int_{\mathbb{I}} \partial_1 C(x,y)dx - \int_{\mathbb{I}} ydx\right)dy \\ &= \int_{\mathbb{I}} (C(1,y) - C(0,y) - y)dy \\ &= 0.\end{aligned}$$

Thus, $\partial_1 C(x,y) = y$ almost all $(x,y) \in \mathbb{I}^2$ with respect to $\lambda_2$. Thus, $C = \Pi$ by the fundamental theorem of calculus. The converse is trivial since one has $\tau(\Pi) = 0$.

(iv) Let $X, Y$ be random variables with copula $C$. First of all, assume that $Y$ is completely dependent on $X$, $C$ will be left invertible, thus, $C^T * C = M$. Lemma 3.3 implies $\tau(M) = \tau(C^T * C) \leq \tau(C)$. Combined with the upper bound of $\tau(C)$, one has $\tau(C) = \tau(M)$.

To show the other direction of the proof, assume that $\tau(Y \mid X) = 1$, i.e. $\tau(C) = \tau(M)$. Thereby, one has $\tau(C*M) = \tau(M)$. By the equality condition of Lemma 3.3, one has that for almost all $(x, y) \in \mathbb{I}^2$ with respect to $\lambda_2$, $\partial_1 M(t, y) - y$ is almost surely constant on $t$ with respect to measure induced by $\partial_1 C(x, t)$, i.e. there is some $\kappa_y \in \mathbb{R}$ such that $\partial_1 M(t, y) - y = \kappa_y$ for almost all $t \in \mathbb{I}$ with respect to Lebesgue-Stieltjes measure $\mu_{\partial_1 C(x,\cdot)}$ induced by $\partial_1 C(x, \cdot)$.

One claims that for almost all $x \in \mathbb{I}$ with respect to $\lambda$, the support of Lebesgue-Stieltjes measure $\mu_{\partial_1 C(x,\cdot)}$ induced by $\partial_1 C(x, \cdot)$ is a singleton. To prove this claim, denote $M_x := \operatorname{supp}(\mu_{\partial_1 C(x,\cdot)})$ for almost all $x \in \mathbb{I}$ with respect to $\lambda$. Let $a := \inf M_x, b := \sup M_x$ and suppose that $a < b$. Then $\lambda((a, b)) > 0$. From the assertion before the stating the claim, there is $y \in (a, b)$ such that $\partial_1 M(t, y) - y$ is almost surely constant on $t$ with respect to $\mu_{\partial_1 C(x,\cdot)}$.

Define

$$E_{y,x} := \{t \in M_x \colon \partial_1 M(t, y) - y = \kappa_y\} \subseteq M_x.$$

Note that $\mu_{\partial_1 C(x,\cdot)}(E_{y,x}) = 1$. Then, one has $\mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [a, y)) > 0$ because if $\mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [a, y)) = 0$, then $\mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [y, b]) = 1$ due to $E_{y,x} \subseteq [a, b]$, which implies

$$\begin{aligned}
1 &= \mu_{\partial_1 C(x,\cdot)}(E_{y,x}) \\
&= \mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [a, b]) \\
&= \mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap ([a, y) \cup [y, b])) \\
&= \mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [a, y)) + \mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [y, b]) \\
&= \mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [y, b]).
\end{aligned}$$

From the fact that $E_{y,x} \subseteq M_x$, one obtains $\mu_{\partial_1 C(x,\cdot)}(M_x \cap [y,b]) = 1$ since

$$1 \geq \mu_{\partial_1 C(x,\cdot)}(M_x \cap [y,b]) \geq \mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap [y,b]) = 1.$$

Thus, by the fact that $M_x$ is a subset of $[a,b]$ with has unit $\mu_{\partial_1 C(x,\cdot)}$-measure, $\mu_{\partial_1 C(x,\cdot)}(M_x \cap [a,y)) = 0$ since one has

$$\begin{aligned}
1 &= \mu_{\partial_1 C(x,\cdot)}(M_x) \\
&= \mu_{\partial_1 C(x,\cdot)}(M_x \cap [a,b]) \\
&= \mu_{\partial_1 C(x,\cdot)}(M_x \cap ([a,y) \cup [y,b])) \\
&= \mu_{\partial_1 C(x,\cdot)}(M_x \cap [a,y)) + \mu_{\partial_1 C(x,\cdot)}(M_x \cap [y,b]) \\
&= 1 + \mu_{\partial_1 C(x,\cdot)}(M_x \cap [y,b]).
\end{aligned}$$

This means the support of $\mu_{\partial_1 C(x,\cdot)}$ does not contain $[a,y)$ which gives us $\inf M_x \geq y > a$, a contradiction. By the same argument, one can proof that $\mu_{\partial_1 C(x,\cdot)}(E_{y,x} \cap (y,b]) > 0$. From both $[a,y) \cap E_{y,x}$ and $(y,b] \cap E_{y,x}$ have positive $\mu_{\partial_1 C(x,\cdot)}$-measure, there exist $a' < y < b'$ such that $\{a',b'\} \subseteq E_{y,x}$. This means, since $y$ is fixed and $a', b' \in E_{y,x}$, $\partial_1 M(a',y) = \partial_1 M(b',y)$. However, since $\partial_1 M(t,y) = 1_{[0,y]}(x)$ for $x \in \mathbb{I}$ and $a' < y < b'$, it yields

$$\partial_1 M(a',y) = 1_{[0,y]}(a') = 1 > 0 = 1_{[0,y]}(b') = \partial_1 M(b',y),$$

a contradiction. Thus, the proof of this claim is completed.

From above claim, for almost all $x$, define function $f$ to send $x$ to the single point in $\operatorname{supp}(\mu_{\partial_1 C(x,\cdot)})$. one can see that for any integrable function $g$ on $\mathbb{I}^2$, by Proposition 2.52,

$$\begin{aligned}
\int_{\mathbb{I}^2} g(x,y)\,d\mu_C(x,y) &= \int_{\mathbb{I}} \int_{\mathbb{I}} g(x,y)\partial_1 C(x,dy)dx \\
&= \int_{\mathbb{I}} \int_{\{f(x)\}} g(x,y)\partial_1 C(x,dy)dx \\
&= \int_{\mathbb{I}} g(x,f(x))dx.
\end{aligned}$$

That is, integration $g$ with respect to $\mu_C$ coincides with the integration with respect to $\lambda_2$ along the set $\{(x,f(x)) \colon x \in \mathbb{I}\}$. This means copula $C$ has support on this set, thus $C$ is copula connecting $X \sim \text{Uniform}(\mathbb{I})$ and $f(X)$, then we are done.

(v) This property can be proved in the same way to the property (v) of $\zeta_1$. except use Lemma 3.3 rather than Lemma 3.2.

□

## 3.3 Concluding Remark

From the above section, one can see that $\tau$, by this concrete condition on $\varphi$, is the class of non-symmetric dependence measures satisfying several properties, especially it is injective invariant and complete dependence detectable. Note that, since $\tau$ is the generalization of $\zeta_1$ and $r$, all of non-symmetric dependence measures in this work share the same set of properties.

In an axiomatic way, this small lemma will give which pair of properties we cannot expect on non-symmetric dependence measure to possess.

**Lemma 3.5.** *There is no non-symmetric dependence measure which can be both monotonic dependence detectable and injective invariant.*

*Proof.* Suppose that there exists a non-symmetric dependence measure $R(\cdot \mid \cdot)$ such that it is both monotonic dependence detectable and injective invariant and let $X$ be a uniform random variable on $\mathbb{I}$. Since the identity map $\iota\colon \mathbb{I} \to \mathbb{I}$ is strictly increasing, then $R(X \mid X) = R(\iota(X) \mid X) = 1$. Let $f_\alpha\colon \mathbb{I} \to \mathbb{I}$ be the map such that

$$f_\alpha(x) = \begin{cases} x + (1-\alpha) & \text{if } x \in [0,\alpha) \\ x - \alpha & \text{if } x \in [\alpha, 1]. \end{cases}$$

This map is obviously injection, thus one has

$$R(X \mid f_\alpha(X)) = R(X \mid X) = 1.$$

However, $X$ is not a strictly monotonic function of $f_\alpha(X)$ since $f_\alpha^{-1}$ is not monotonic, though even injection, map, this yields a contradiction. □

Thus, from above impossibility lemma and what we prove in this section give us this summarize table.

| | MD detect. | CD detect. |
|---|---|---|
| Monotone inv. | | |
| Injective inv. | Impossible | $\tau$ (including $\zeta_1, r$) |

Table 3.1: Recalling the summary table for non-symmetric dependence measures.

The blank spaces mean there is no dependence measure satisfying that pair of properties while the word *impossible* means this pair of properties cannot hold simultaneously. The table tell us that trade-off between stronger type of invariance and stronger type of ability to detect dependence exists. Thus, if one can prove that some particular non-symmetric measure of dependence has the strong invariance property such as injective one, the expectation to hope that it can detect a quite strong type of dependence such as monotonic dependence is certainly fail.

Anyway, since there are still blank space which is neither able to find contradiction nor propose concrete dependence measure to fulfill these set of properties, the room to play with this table is still available. For the other direction, only Dette et al.'s $r$ has its kernel-based estimator and, furthermore, that estimator has nice asymptotic properties. The statistical work dedicated to propose the estimators of the Li's $\tau$, is not happened. Although, the definition of $\tau$ makes introducing its nice estimators become challenging, it worths our consideration since it will cover a large class of estimators of non-symmetric dependence measures.

# CHAPTER IV

# SYMMETRIC DEPENDENCE MEASURES

## 4.1 Symmetric version of Li's dependence measures

Before going on, the addition sufficient condition to construct symmetric version of $\tau$ is needed. $\varphi$ is assumed additionally to be symmetric around zero in the sense that $\varphi(-x) = \varphi(x)$ for all $x \in [-1, 1]$. From this point, now it is the time to propose our modified version of $\tau$. And random variables are assumed to have continuous distribution function as well.

**Definition 4.1.** For random variables $X$ and $Y$, define the *symmetrized* $\tau$, denoted as $\tilde{\tau}$, by

$$\tilde{\tau}(X, Y) := \frac{1}{2}(\tau(Y \mid X) + \tau(X \mid Y)) \tag{4.1}$$

And here is the theorem summarizing properties of $\tilde{\tau}$.

**Theorem 4.2.** *For any random variables $X$ and $Y$, one has*

*(i)* $\tilde{\tau}(X, Y)$ *is well-defined;*

*(ii)* $0 \leq \tilde{\tau}(X, Y) \leq 1$;

*(iii)* $\tilde{\tau}(X, Y) = \tilde{\tau}(Y, X)$;

*(iv)* $\tilde{\tau}(X, Y) = 0$ *if and only if* $X, Y$ *are independent;*

*(v)* $\tilde{\tau}(X, Y) = 1$ *if and only if* $X$ *and* $Y$ *are mutually complete dependent;*

*(vi)* $\tilde{\tau}(f(X), g(Y)) = \tilde{\tau}(X, Y)$ *for any strictly monotonic functions* $f, g$.

*Proof.* Let $X$ and $Y$ be random variables with $(X, Y) \sim C$ for some $C \in \mathcal{C}$.

(i) This property is inherited from the definition of $\tau$.

(ii) Boundedness of $\tilde{\tau}(X,Y)$ into $\mathbb{I}$ is a direct consequence of boundedness of $\tau$.

(iii) Obvious from the definition.

(iv) Assume that $\tilde{\tau}(X,Y) = 0$. Since $\tau(Y \mid X)$ and $\tau(X \mid Y)$ are nonnegative, they must be both zero. So $X$ and $Y$ are independent. The converse is trivial since independence between $X$ and $Y$ implies $\tilde{\tau}(X,Y) = 0$.

(v) Assume unit value of $\tilde{\tau}(X,Y)$ and suppose that $X$ and $Y$ are not mutually completely dependent. That is; without loss of generality, $\tau(Y \mid X) < 1$. Then, since $0 \leq \tau(X \mid Y) \leq 1$, one has and

$$0 \leq \tau(Y \mid X) + \tau(X \mid Y) < 2,$$

thus $\tilde{\tau}(X,Y) < 1$, a contradiction. Thus, maximal value implies the mutual complete dependence between $X$ and $Y$. Conversely, assume that $X, Y$ are mutually completely dependent. Then, $Y$ is completely dependent on $X$ and vice versa. Thus $\tilde{\tau}(X,Y) = 1$.

(vi) For the last property, the following lemma is needed.

**Lemma 4.3.** *Let $\varphi\colon [-1,1] \to [0,\infty)$ be continuous, convex, satisfying $\varphi(-x) = \varphi(x)$ for any $x \in [-1,1]$, and $\varphi(x) = 0$ if and only if $x = 0$, $X$ and $Y$ random variables and $f$ a strictly monotonic function, then*

$$\tau(f(Y) \mid X) = \tau(Y \mid X).$$

*Proof.* Let $X$ and $Y$ be random variables with connecting copula $C_{X,Y} \in \mathcal{C}$ and $f$ a strictly monotonic transformation. By Proposition 2.54, if $f$ is a strictly increasing map, one has $C_{X,f(Y)} = C_{X,Y}$, which implies

$$\tau(f(Y) \mid X) = \tau(Y \mid X).$$

For the case of strictly decreasing $f$, Proposition 2.54 also implies that, for any $x, y \in \mathbb{I}$,

$$C_{X,f(Y)}(x,y) = x - C_{X,Y}(x, 1-y). \tag{4.2}$$

Then, let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $(x, y) \mapsto (x, 1-y)$. This map is continuously differentiable, injective, and its Jacobian $D_{(x,y)}T$ is invertible. Moreover, the absolute value of its determinant is one for every point $x \in \mathbb{R}^2$ and $T(\mathbb{I}^2) = \mathbb{I}^2$. By Theorem 2.34, the setup that $\varphi(-x) = \varphi(x)$ and the equation (4.2), one has

$$\begin{aligned}
\tau(C_{X,f(Y)}) &= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi(\partial_1 C_{X,f(Y)}(x, y) - \partial_1\Pi(x, y))dxdy \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi(\partial_1(x - C_{X,Y}(x, 1-y)) - \partial_1\Pi(x, y))dxdy \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi((1 - \partial_1 C_{X,Y}(x, 1-y)) - y)dxdy \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi(\partial_1\Pi(x, 1-y) - \partial_1 C_{X,Y}(x, 1-y))dxdy \\
&= \int_{\mathbb{I}}\int_{\mathbb{I}} \varphi(\partial_1 C_{X,Y}(x, 1-y) - \partial_1\Pi(x, 1-y))dxdy \\
&= \int_{\mathbb{I}^2} \varphi((\partial_1 C_{X,Y} - \partial_1\Pi) \circ T(x, y))d\lambda_2(x, y) \\
&= \int_{\mathbb{I}^2} \varphi((\partial_1 C_{X,Y} - \partial_1\Pi)(x, y))d\lambda_2(x, y) \\
&= \int_{\mathbb{I}^2} \varphi(\partial_1 C_{X,Y}(x, t) - \partial_1\Pi(x, t))d\lambda_2(x, t) \\
&= \tau(C_{X,Y}).
\end{aligned}$$

□

From the above lemma, we are ready for the proof of invariance property of $\tilde{\tau}$. Let $f, g$ be strictly monotonic functions. Then one can obtain

$$\begin{aligned}
\tilde{\tau}(f(X), g(Y)) &= \frac{1}{2}(\tau(g(Y) \mid f(X)) + \tau(f(X) \mid g(Y))) \\
&= \frac{1}{2}(\tau(g(Y) \mid X) + \tau(f(X) \mid Y)) \\
&= \frac{1}{2}(\tau(Y \mid X) + \tau(X \mid Y)) \\
&= \tilde{\tau}(X, Y).
\end{aligned}$$

The second line comes from the injective invariance of $\tau$, note that strictly monotonic functions are injective, and the third line is the consequence of the above lemma. □

## 4.2 Maximal information coefficient (MIC)

To make MIC be compatible with other dependence measure, all of random variables are also assumed to have continuous distribution function, this assumption is also useful because the proof for ability to detect complete dependence employs copula techniques. Properties of MIC, which is defined by expression (2.16), can be summarized in a single theorem.

**Theorem 4.4.** *For random variables $X$ and $Y$,* MIC *satisfies the following:*

*(i) (Normalization)* $0 \leq \mathrm{MIC}(X,Y) \leq 1$*;*

*(ii) (Symmetry)* $\mathrm{MIC}(X,Y) = \mathrm{MIC}(Y,X)$*;*

*(iii) (Independence Detectability)* $\mathrm{MIC}(X,Y) = 0$ *if and only if $X$ and $Y$ are independent;*

*(iv) (Monotonic Invariance)* $\mathrm{MIC}(\varphi(X),\psi(Y)) = \mathrm{MIC}(X,Y)$ *for any strictly monotonic transformations $\varphi$ and $\psi$;*

*(v) (Complete Dependence Detectability)* $\mathrm{MIC}(X,Y) = 1$ *if either $Y$ is CD on $X$ or $X$ is CD on $Y$.*

*Proof.* (i) Non-negativity comes from the fact that the mutual information is always non-negative. Now, let $X, Y$ be random variables, $k, l$ natural numbers exceeding 1, $G$ a $k$-by-$l$ grid with columns $\{C_i\}_{i=1}^{k}$ and rows $\{R_j\}_{j=1}^{l}$, and consider $H(\mathrm{Col}_G(X,Y))$ and $H(\mathrm{Row}_G(X,Y))$. We will show that $H(\mathrm{Col}_G(X,Y))$ is bounded by $\log k$. Let $p$ be the probability mass function of $\mathrm{Col}_G(X,Y)$ and $q$ be the probability mass function of a discrete uniform random variable on the set $\{1,2,\ldots,k\}$. By the setup, $q(x) = \frac{1}{k}$ for any $x \in \{1, 2, \ldots, k\}$. Thus, by the non-negativity of divergence, one has

$$\begin{aligned}
0 &\leq D(p \parallel q) \\
&= \sum_{x \in \mathcal{X}} p(x) \log \left( \frac{p(x)}{q(x)} \right) \\
&= \sum_{x \in \mathcal{X}} p(x) \log(kp(x)) \\
&= \sum_{x \in \mathcal{X}} p(x) \log k + \sum_{x \in \mathcal{X}} p(x) \log(p(x)) \\
&= \log k - \sum_{x \in \mathcal{X}} p(x)(-\log(p(x))) \\
&= \log k - H(\mathrm{Col}_G(X, Y)).
\end{aligned}$$

This means $H(\mathrm{Col}_G(X, Y)) \leq \log k$. By the same fashion, $H(\mathrm{Row}_G(X, Y))$ is also bounded above by $\log l$.

Finally, from above assertion and the expression (2.15) one has

$$\begin{aligned}
I_G(X, Y) &= I(\mathrm{Col}_G(X, Y), \mathrm{Row}_G(X, Y)) \\
&\leq \min(H(\mathrm{Col}_G(X, Y)), H(\mathrm{Row}_G(X, Y))) \\
&\leq \min(\log k, \log l) \\
&= \log \min\{k, l\}.
\end{aligned}$$

This assertion implies that any $k$-by-$l$ characteristic entry of $(X, Y)$ is always bounded above by 1. As a result, $\mathrm{MIC}(X, Y)$ must not exceed 1.

(ii) First of all, let us observe that, for random variables $X, Y$, $\mathrm{MIC}(X, Y)$ is the supremum of the set

$$\mathcal{I}(X, Y) := \left\{ \frac{I_G(X, Y)}{\log \min\{k, l\}} \mid k, l \in \mathbb{N} \setminus \{1\}, G \in G(k, l) \right\}. \tag{4.3}$$

Note that $I_G(X, Y) = I_{\tilde{G}}(Y, X)$ where $\tilde{G}$ is the transpose partition of $G$, i.e. the column partition of $\tilde{G}$ is from the row partition of $G$ and the row partition of $\tilde{G}$ is from the column partition of $G$. To see this point, let $k, l \in \mathbb{N} \setminus \{1\}$ and $G \in G(k, l)$ with rows and columns $\{C_i\}_{i=1}^k, \{R_j\}_{j=1}^n$. Choose $\tilde{G} \in G(k, l)$ with columns $\tilde{C}_j := R_j$ and rows $\tilde{R}_i := C_i$ for all

$j \in \{1, \ldots, l\}$ and $i = \{1, \ldots, k\}$. One will have $\{X \in C_i\} = \{X \in \tilde{R}_i\}$ and $\{X \in R_j\} = \{X \in \tilde{C}_j\}$ for any such $i, j$, thus the summand of both $I_G(X, Y)$ and $I_{\tilde{G}}(Y, X)$ are identical.

Then, it easily follows that $\mathcal{I}(X, Y) = \mathcal{I}(Y, X)$. Hence, $\mathrm{MIC}(X, Y) = \mathrm{MIC}(Y, X)$.

(iii) Firstly, suppose that $X, Y$ are independent random variables, also denoted by $X \perp Y$. Let $k, l \in \mathbb{N}$ and $G \in G(k, l)$ with columns $\{C_i\}_{i=1}^k$ and rows $\{R_j\}_{j=1}^l$. Since $\{C_i\}_{i=1}^k$ and $\{R_j\}_{j=1}^l$ are partitions of $\mathbb{I}$, all of elements are $\mathcal{B}(\mathbb{I})$-measurable. One yields

$$\begin{aligned} P(\mathrm{Col}_G(X, Y) = i, \mathrm{Row}_G(X, Y) = j) &= P(X \in C_i, Y \in R_j) \\ &= P(X \in C_i) P(Y \in R_j) \\ &= P(\mathrm{Col}_G(X, Y) = i) \\ &\quad \times P(\mathrm{Row}_G(X, Y) = j) \end{aligned}$$

for $(i, j) \in \{1, \ldots, k\} \times \{1, \ldots, l\}$. This implies

$$\log \left( \frac{P(\mathrm{Col}_G(X, Y) = i, \mathrm{Row}_G(X, Y) = j)}{P(\mathrm{Col}_G(X, Y) = i) P(\mathrm{Row}_G(X, Y) = j)} \right) = \log 1 = 0$$

for all $i = 1, \ldots, k$ and $j = 1, \ldots, k$. Thus, one gets

$$\begin{aligned} I_G(X, Y) &= I(\mathrm{Col}_G(X, Y), \mathrm{Row}_G(X, Y)) \\ &= \sum_{i=1}^{k} \sum_{j=1}^{l} \log \left( \frac{P(\mathrm{Col}_G(X, Y) = i, \mathrm{Row}_G(X, Y) = j)}{P(\mathrm{Col}_G(X, Y) = i) P(\mathrm{Row}_G(X, Y) = j)} \right) \\ &\qquad \times P(\mathrm{Col}_G(X, Y) = i, \mathrm{Row}_G(X, Y) = j) \\ &= 0. \end{aligned}$$

One can prove the converse assertion by contrapositive. The following lemma in supplementary material of [14] which is handy for this purpose.

**Lemma 4.5** ([14])**.** *For any $a \in \mathbb{I}$, define $\chi_a$ to be function on $\mathbb{I}$ such that $\chi_a(x) = 0$ if $x \in [0, a)$ and $\chi_a(x) = 1$ otherwise. Then, for any random variables $X, Y$ on $\mathbb{I}$, if $I(\chi_a(X), \chi_b(Y)) = 0$ for all $a, b \in \mathbb{I}$, then $X$ and $Y$ are independent.*

Thus, to prove the contraposition, suppose that $X, Y$ be random variables which are not independent. Then, by above lemma, there is $a, b \in \mathbb{I}$ such that $I(\chi_a(X), \chi_b(Y)) = \delta$ for some $\delta > 0$. Thus, consider a 2-by-2 grid such that it partitions $\mathbb{I}$ on the $x$-axis into 2 pieces by $a$ and, also, partitions $\mathbb{I}$ on the y-axis into 2 pieces by $b$. This implies that

$$\begin{aligned}
I_G(X,Y) &= \sum_{i=1}^{2}\sum_{j=1}^{2} \log\left(\frac{P(\mathrm{Col}_G(X,Y)=i, \mathrm{Row}_G(X,Y)=j)}{P(\mathrm{Col}_G(X,Y)=i)P(\mathrm{Row}_G(X,Y)=j)}\right) \\
&\qquad \times P(\mathrm{Col}_G(X,Y)=i, \mathrm{Row}_G(X,Y)=j) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} \log\left(\frac{P(X \in C_i, Y \in R_j)}{P(X \in C_i)P(Y \in R_j)}\right) P(X \in C_i, Y \in R_j) \\
&= \sum_{m=0}^{1}\sum_{n=0}^{1} \log\left(\frac{P(\chi_a(X)=m, \chi_b(Y)=n)}{P(\chi_a(X)=m)P(\chi_b(Y)=n)}\right) \\
&\qquad \times P(\chi_a(X)=m, \chi_b(Y)=n) \\
&= I(\chi_a(X), \chi_b(Y)) \\
&= \delta > 0.
\end{aligned}$$

The third equality comes from the fact that $\{X \in C_1\} = \{X \in [0,a)\} = \{\chi_a(X) = 0\}$ and $\{X \in C_2\} = \{X \in [a,1]\} = \{\chi_a(X) = 1\}$. Similarly, one also obtains $\{Y \in R_1\} = \{\chi_b(Y) = 0\}$ and $\{Y \in R_2\} = \{\chi_b(Y) = 1\}$. Since the $\mathrm{MIC}(X,Y)$ is the supremum of $I_G(X,Y)$ over all natural numbers $k, l$ exceeding 1 and $G \in G(k,l)$, $\mathrm{MIC}(X,Y)$ is bounded away from zero.

(iv) By the symmetry, it suffices to show that $\mathrm{MIC}(\varphi(X), Y) = \mathrm{MIC}(X,Y)$ for any strictly monotonic mapping $\varphi$ since if we have two strictly monotonic transformations $\varphi, \psi$, then

$$\begin{aligned}
\mathrm{MIC}(\varphi(X), \psi(Y)) = \mathrm{MIC}(X, \psi(Y)) = \mathrm{MIC}(\psi(Y), X) &= \mathrm{MIC}(Y, X) \\
&= \mathrm{MIC}(X, Y).
\end{aligned}$$

Moreover, considering the case of strictly increasing transformations is enough since the other case can be proved similarly. The desired result follows if we can show that $\mathcal{I}(X,Y) = \mathcal{I}(\varphi(X), Y)$. For each $G \in G(k,l)$ with columns

$\{C_i\}_{i=1}^k$ and rows $\{R_j\}_{j=1}^l$, choose $\tilde{G}$ consisting of columns $\{\varphi(C_i)\}_{i=1}^k$ and rows $\{R_j\}_{j=1}^l$. Since $\varphi$ is strictly increasing, one has

$$\{X \in C_i\} = \{\varphi(X) \in \varphi(C_i)\} = \{\varphi(X) \in \tilde{C}_i\},$$

so this yields $I_G(X, Y) = I_{\tilde{G}}(\varphi(X), Y) \in \mathcal{I}(\varphi(X), Y)$.

The other inclusion can be proved similarly using the quasi-inverse of $\varphi$, i.e. the function $\varphi^{[-1]} \colon \mathbb{R} \to \mathbb{R}$ defined by

$$\varphi^{[-1]}(x) := \inf\{y \in \mathbb{R} \colon \varphi(y) \geq x\}.$$

Note that quasi-inverse of strictly increasing function is well-defined and satisfies $\varphi^{[-1]}(\varphi(x)) = x$ for $x \in \mathbb{R}$. Moreover, for any $x, t \in \mathbb{R}$, $\varphi(x) \geq t$ if and only if $x \geq \varphi^{[-1]}$. To see this inclusion, consider $G \in G(k, l)$ with columns $\{C_i\}_{i=1}^k$ and rows $\{R_j\}_{j=1}^l$. Choose $\tilde{G}$ with columns $\{\varphi^{[-1]}(C_i)\}_{i=1}^k$ and the same rows. One can see that, by the properties of quasi-inverse,

$$\{\varphi(X) \in C_i\} = \{X \in \varphi^{[-1]}(C_i)\} = \{X \in \tilde{C}_i\}.$$

Thus, one also has $I_G(\varphi(X), Y) = I_{\tilde{G}}(X, Y)$ and $\mathcal{I}(X, Y) = \mathcal{I}(\varphi(X), Y)$, which implies the equality of $\mathrm{MIC}(X, Y)$ and $\mathrm{MIC}(\varphi(X), Y)$.

(v) For the last result on complete dependence detectability, it requires more demanding analytic tools. Thus, it will be restated and proved in Proposition 4.25. Anyway, this is possible by approximating Borel measurable functions by nice functions which yields the unit MIC along with showing the continuity of MIC. These two main components will finish the proof of this elegant property. $\square$

From now on, all random variables are assumed to be uniformly distributed on $\mathbb{I}$ as every random variables with continuous distribution functions can be transformed into a uniform $[0, 1]$ random variable via a strictly increasing function.

The following type of Borel measurable function defined below is useful later.

**Definition 4.6.** A Borel measurable function $\varphi\colon \mathbb{I} \to \mathbb{I}$ is said to be *measure-preserving* if for any $B \in \mathcal{B}(\mathbb{I})$, $\lambda(\varphi^{-1}(B)) = \lambda(B)$.

In the following lemma, let us state a well-known fact which says that only measure-preserving functions can transform a uniform $[0,1]$ random variable into another uniform $[0,1]$ random variable.

**Lemma 4.7.** *Let $X, Y \sim \mathrm{Uniform}(\mathbb{I})$ and $\varphi\colon \mathbb{I} \to \mathbb{I}$ be a Borel measurable function such that $Y = \varphi(X)$ a.s., then $\varphi$ must be measure-preserving. Moreover, if $X$ is a uniform random variable on $\mathbb{I}$ and $\varphi\colon \mathbb{I} \to \mathbb{I}$ is measure-preserving, then $Y := \varphi(X) \sim \mathrm{Uniform}(\mathbb{I})$.*

*Proof.* Let $X, Y \sim \mathrm{Uniform}(\mathbb{I})$ and $\varphi\colon \mathbb{I} \to \mathbb{I}$ be a Borel measurable function, then, for any $B \in \mathcal{B}(\mathbb{I})$,

$$\lambda(\varphi^{-1}(B)) = P(X \in \varphi^{-1}(B)) = P(\varphi(X) \in B) = P(Y \in B) = \lambda(B).$$

This uses the property that $X, Y \sim \mathrm{Uniform}(\mathbb{I})$, which yields $P(X \in A) = \lambda(A)$ and $P(Y \in B) = \lambda(B)$ for any $A, B \in \mathcal{B}(\mathbb{I})$. On the other side, given $X \sim \mathrm{Uniform}(\mathbb{I})$ and $\varphi\colon \mathbb{I} \to \mathbb{I}$. Define $Y := \varphi(X)$. Then, for any $y \in \mathbb{I}$,

$$P(Y \leq y) = P(\varphi(X) \leq y) = P(X \in \varphi^{-1}([0,y])) = \lambda(\varphi^{-1}([0,y])) = \lambda([0,y]) = y.$$

The third equality comes from the same reason as the first assertion of this lemma and the next one is claimed by measure-preserving property. Since $y \in \mathbb{I}$ is arbitrary, one can say that $Y \sim \mathrm{Uniform}(\mathbb{I})$. $\square$

Measure preserving maps from $\mathbb{I}$ to itself can be approximated by a sequence of some nicer measure-preserving maps. This result is due to [1].

**Theorem 4.8** ([1]). *Let $\varphi\colon \mathbb{I} \to \mathbb{I}$ be a measure-preserving map. Then, there exists a sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ from $\mathbb{I}$ to itself such that all of them are one-to-one piecewise linear measure-preserving functions and $f_n \to \varphi$ almost everywhere*[1].

[1] Throughout this thesis, any one-to-one piecewise linear measure preserving map has *finite* discontinuity points.

The rationale behind using this type of approximation is that one-to-one piecewise linear measure-preserving functions have a useful characterization.

**Proposition 4.9.** *Let $\varphi\colon \mathbb{I} \to \mathbb{I}$ be a Borel measurable function. Then $\varphi$ is 1-1 piecewise linear measure-preserving if and only if there exists a partition $\{x_i\}_{i=0}^k$ of $\mathbb{I}$ satisfying the following conditions:*

*(i)* $0 := x_0 < x_1 < \cdots < x_k = 1$;

*(ii) for any $i \in \{1, \ldots, k\}$, $\varphi$ is linear on $[x_{i-1}, x_i]$ with slope $1$ or $-1$ and the collection $\{\varphi([x_{i-1}, x_i])\}_{i=1}^k$ is also a partition of $\mathbb{I}$.*

*Proof.* Let $\varphi$ be a 1-1 piecewise linear measure-preserving. Then, one has $k \in \mathbb{N}$ and $\{x_i\}_{i=0}^k$ such that $\{x_i\}_{i=0}^k$ forms a partition of $\mathbb{I}$ and fulfills the condition (i), And for each $i \in \{1, \ldots, k\}$, there exists $\{m_i\}_{i=1}^k, \{c_i\}_{i=1}^k \subseteq \mathbb{R}$ such that $\varphi(x) = m_i x + c_i$ for $x \in [x_{i-1}, x_i]$. Suppose that there is $i_0 \in \{1, \ldots, k\}$ such that $m_{i_0} \notin \{-1, 1\}$. Choose $a, b \in \mathbb{I}$ such that $x_{i_0 - 1} < a < b < x_{i_0}$ and consider the set $\tilde{B} := \varphi([a, b])$. By the setup, $\varphi$ is linear function on $[a, b]$ of slope $m_{i_0}$ and intercept term $m_{i_0}$. Thus, one has $\lambda(\tilde{B}) = \lambda(\varphi([a, b])) = |m_{i_0}|(b - a)$. Since $\tilde{B}$ is an subinterval of $\mathbb{I}$, $\tilde{B} \in \mathcal{B}(\mathbb{I})$. Note that, due to the fact that linear function is injective, $\varphi^{-1}(\tilde{B}) = \varphi^{-1}(\varphi([a, b])) = [a, b]$. Once we suppose that $m_{i_0} \notin \{-1, 1\}$, $\lambda(\varphi^{-1}(\tilde{B})) = b - a \neq |m_{i_0}|(b - a) = \lambda(\tilde{B})$, a contradiction to $\varphi$ is measure-preserving. Thus, each piece of the graph of $\varphi$ restricted each slice of partition must have slope of unit magnitude. Moreover, $\{\varphi([x_{i-1}, x_i])\}_{i=1}^k$ is obvious a partition of $\mathbb{I}$ because $\{x_i\}_{i=0}^k$ is partition and $\varphi$ is one-to-one and measure-preserving. The last properties comes from the fact that for each $i = 1, \ldots, k$, $\lambda(\varphi([x_{i-1}, x_i])) = |m_i|(x_i - x_{i-1}) = (x_i - x_{i-1})$ this use the slope properties recently proved. The converse is obvious. $\square$

**Corollary 4.10.** *Let $\varphi\colon \mathbb{I} \to \mathbb{I}$ be a Borel measurable function. Then $\varphi$ is 1-1 piecewise linear measure-preserving if and only if there exists $k \in \mathbb{N}$, partitions $\{s_i\}_{i=0}^k, \{t_i\}_{i=0}^k$ of $\mathbb{I}$, a permutation $\sigma$ of $\{1, \ldots, k\}$, and $\{m_i\}_{i=1}^k$ such that*

*(i)* $0 := s_0 < s_1 < \cdots < s_k = 1$ *and* $0 := t_0 < t_1 < \cdots < t_k = 1$;

*(ii)* $\varphi$ *maps* $[s_{i-1}, s_i]$ *onto* $[t_{\sigma(i)-1}, t_{\sigma(i)}]$ *linearly with slope* $m_i \in \{-1, 1\}$.

**Remark 4.11.** (i) In the characterization from Corollary 4.10, we call the 5-tuple $(k, \{s_i\}_{i=0}^k, \{t_i\}_{i=0}^k, \sigma, \{m_i\}_{i=1}^k)$ a *shuffle representation of* $\varphi$ and we may also write $\varphi \sim (k, \{s_i\}_{i=0}^k, \{t_i\}_{i=0}^k, \sigma, \{m_i\}_{i=1}^k)$.

(ii) If a shuffle representation has $m_i = 1$ for all $i \in \{1, \dots, k\}$, it is called a *straight shuffle*. In this case, one may drop out $\{m_i\}$ from the representation.

(iii) In fact, knowing $\{s_i\}$ and $\sigma$ will lead us to $\{t_i\}$. Sometimes, $\{t_i\}$ may be omitted and closed form of $\varphi$ given the shuffle representation can be derived analytically.

This representation of 1-1 piecewise linear measure-preserving map comes from geometric aspect. Visualize the graph of identity function on $\mathbb{I}$ and partition the $x$-axis into a finite number of intervals which cuts $\mathbb{I}^2$ into vertical stripes. Then rearrange the stripes according to the underlying permutation map allowing a flip about the vertical axis of symmetry.

A special case of one-to-one piecewise linear measure-preserving transformation can give the unit MIC.

**Proposition 4.12.** *Let* $X \sim \text{Uniform}(\mathbb{I})$ *and* $\varphi\colon \mathbb{I} \to \mathbb{I}$ *a one-to-one piecewise linear measure-preserving transformation whose all discontinuity points are rational numbers in* $\mathbb{I}$. *Then* $\text{MIC}(X, \varphi(X)) = 1$.

*Proof.* Let $X \sim \text{Uniform}(\mathbb{I})$ and $\varphi\colon \mathbb{I} \to \mathbb{I}$ a one-to-one piecewise linear measure-preserving transformation whose all discontinuity points are rational numbers in $\mathbb{I}$. By Remark (4.11), $\varphi$ has the shuffle representation $(k, \{s_i\}_{i=0}^k, \{t_i\}_{i=0}^k, \sigma, \{m_i\}_{i=1}^k)$. Then, there are sequences $\{n_i\}_{i=1}^{k-1}$ and $\{d_i\}_{i=1}^{k-1}$ of natural numbers such that $s_i = n_i/d_i$ for all $i = 1, \dots, k-1$. Then, choose $d^* := \prod_{i=1}^{k-1} d_i$. This implies that, for all $i \in \{1, \dots, k-1\}$, $x_i = n_i \prod_{j \neq i} d_j / d^*$. Accordingly, for any $i = 1, \dots, k-1$, $\Delta x_i = (n_i \prod_{j \neq i} d_j - n_{i-1} \prod_{j \neq i-1} d_j)/d^*$. One can notice that $n_i \prod_{j \neq i} d_j - n_{i-1} \prod_{j \neq i-1} d_j$ is a natural number. Thus, each of $[s_{i-1}, s_i]$ can be partitioned again as finite collection of subintervals whose lengths are $1/d^*$.

Then, choose the equipartition grid $G \in G(d^*, d^*)$ of size $1/d^* \times 1/d^*$. Note that this partition divides each square $[s_{i-1}, s_i] \times [t_{\sigma(i)-1}, t_{\sigma(i)}]$ into a collection of small squares of size $1/d^* \times 1/d^*$ and, since $\varphi$ is one-to-one, each piece in $[s_{i-1}, s_i]$ is mapped into a piece in $[\varphi(s_{i-1}), \varphi(s_i)] = [t_{\sigma(i)-1}, t_{\sigma(i)}]$ determined by $\sigma$ only. That is; for any $i \in \{1, \ldots, k\}$ and $j \neq \sigma(i)$, the rectangle $C_i \times R_j$ does not contain the graph of $(X, \varphi(X))$. Moreover, the linear graph in every grid is one of the two diagonal lines. Thereby,

$$
\begin{aligned}
I(X, \varphi(X))_G = & \sum_{i=1}^{d^*} \sum_{j=1}^{d^*} \log \left( \frac{P(\mathrm{Col}_G(X, \varphi(X)) = i, \mathrm{Row}_G(X, \varphi(X)) = j)}{P(\mathrm{Col}_G(X, \varphi(X)) = i) P(\mathrm{Row}_G(X, \varphi(X)) = j)} \right) \\
& \times P(\mathrm{Col}_G(X, \varphi(X)) = i, \mathrm{Row}_G(X, \varphi(X)) = j) \\
= & \sum_{i=1}^{d^*} \sum_{j=1}^{d^*} \log \left( \frac{P(X \in C_i, \varphi(X) \in R_j)}{P(X \in C_i) P(\varphi(X) \in R_j)} \right) P(X \in C_i, \varphi(X) \in R_j) \\
= & \sum_{i=1}^{d^*} \log \left( \frac{P(X \in C_i, \varphi(X) \in R_{\sigma(i)})}{P(X \in C_i) P(\varphi(X) \in R_{\sigma(i)})} \right) P(X \in C_i, \varphi(X) \in R_{\sigma(i)}) \\
= & \sum_{i=1}^{d^*} \log((d^*)^2 P(X \in C_i)) P(X \in C_i) \\
= & \log d^*.
\end{aligned}
$$

The second equality from the last comes from law of total probability, the fact that probability of $X$ lies in $C_i$ and $\varphi(X)$ takes value outside of $R_{\sigma(i)}$ is zero, and $X \sim \mathrm{Uniform}(\mathbb{I})$.

Then, the normalized score is $I(X, \varphi(X))_G / \log d^* = 1$. Since MIC is supremum of normalized score over all $k, l > 1$ and $G \in G(k, l)$, $\mathrm{MIC}(X, \varphi(X))$ must be one. $\square$

In one of the nicest classes of measurable functions, MIC can detect relation of this type with unit value. This result along with the continuity theorem of MIC will be used to prove the detectability of larger classes of dependence later.

Another reason which makes assuming uniform random variables to be uniform on $\mathbb{I}$ is that their joint distribution will be copula. Its measure theoretic properties of the measure induced by a copula play an important role as a tool for bounding probability of uniform random variables on $\mathbb{I}$ taking value on some rectangle with some small uniform bound.

Since the technique used in the proofs is approximation, we introduce some distance between random vectors.

**Definition 4.13.** For random vectors $\vec{X}, \vec{Y} \in \mathcal{P}(\mathbb{I}^2)$, define:

(i) $d_{TV}(\vec{X}, \vec{Y}) := \sup_{B \in \mathcal{B}(\mathbb{I}^2)} \Big( P(\vec{X} \in B) - P(\vec{Y} \in B) \Big)$, and

(ii) $d_{RTV}(\vec{X}, \vec{Y}) := \sup_{R \in \mathcal{R}(\mathbb{I}^2)} \Big( P(\vec{X} \in R) - P(\vec{Y} \in R) \Big)$,

where $\mathcal{R}(\mathbb{I}^2)$ denotes the class of Borel measurable rectangles. $d_{TV}, d_{RTV}$ are called the *total variation* and *rectangular total variation distances*, respectively.

Note that although both of them are quite similar in the constructions, the main difference is that the total variation is the supremum of probabilistic distance over the set of all measurable *sets* on $\mathbb{I}^2$ while the latter are the supremum of the same distance, but over all measurable *rectangles* in $\mathbb{I}^2$. The total variation distance is not the new metric in literature but rectangular version is newly defined here since to prove any properties with copulas, dealing with measurable rectangles is more convenient. Moreover, in the case of discrete random vectors with common support, $d_{TV}$ and $d_{RTV}$ are the same.

The following lemma shows the property of $d_{RTV}$ which is used to prove continuity of MIC.

**Lemma 4.14.** *Let $\vec{X}, \vec{Y}$ be two-dimensional random vectors on $\mathbb{I}^2$, then*

$$d_{RTV}((\vec{X})_G, (\vec{Y})_G) \leq d_{RTV}(\vec{X}, \vec{Y})$$

*for any grid $G \in G(k, l)$ and $k, l \in \mathbb{N} \setminus \{1\}$.*

*Proof.* Let $\vec{X}, \vec{Y}$ be two-dimensional random vectors on $\mathbb{I}^2$ and $G \in G(k, l)$ for natural numbers $k, l$ exceeding 1. Thus, one has

$$\begin{aligned}
&d_{RTV}((\vec{X})_G, (\vec{Y})_G) \\
&= \sup_{B_1, B_2 \in \mathcal{B}(\mathbb{R})} \Big( P((\vec{X})_G \in B_1 \times B_2) - P((\vec{Y})_G \in B_1 \times B_2) \Big) \\
&= \sup_{B_1, B_2 \in \mathcal{B}(\mathbb{R})} \left( P(X_1 \in \bigcup_{i \in B_1} C_i, X_2 \in \bigcup_{j \in B_2} R_j) - P(Y_1 \in \bigcup_{i \in B_1} C_i, Y_2 \in \bigcup_{j \in B_2} R_j) \right) \\
&\leq \sup_{B \in \mathcal{R}(\mathbb{I}^2)} \Big( P(\vec{X} \in B) - P(\vec{Y} \in B) \Big) \\
&= d_{RTV}(\vec{X}, \vec{Y}).
\end{aligned}$$

The inequality comes from the fact that the supremum is taken on a larger set. □

Before going to prove the continuity result of MIC, there are two technical tools which are useful. The first tool is already proposed in [15].

**Proposition 4.15** ([15])**.** *Let $(X, Y)$ and $(\tilde{X}, \tilde{Y})$ be discrete random vectors with common support $\{1, \ldots, k\} \times \{1, \ldots, l\}$. If $0 \leq \delta \leq \frac{1}{4}$, then $|I(X, Y) - I(\tilde{X}, \tilde{Y})|$ is $O\left(\delta \log\left(\frac{1}{\delta} + \delta \log \min\{k, l\}\right)\right)$ whenever $d_{TV}((X, Y), (\tilde{X}, \tilde{Y})) \leq \delta$.*

The second tool is the bound of difference of MIC of two random vectors.

**Lemma 4.16.** *For any random pairs $(X, Y)$ and $(\tilde{X}, \tilde{Y})$,*

$$|\operatorname{MIC}(X, Y) - \operatorname{MIC}(\tilde{X}, \tilde{Y})| \leq \sup_{k, l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k, l)} \left| \frac{I_G(X, Y)}{\log \min\{k, l\}} - \frac{I_G(\tilde{X}, \tilde{Y})}{\log \min\{k, l\}} \right|.$$

*Proof.* Let $(X, Y)$ and $(\tilde{X}, \tilde{Y})$ be two pairs of random variables, then

$$\frac{I_G(X, Y)}{\log \min\{k, l\}} = \left( \frac{I_G(X, Y)}{\log \min\{k, l\}} - \frac{I_G(\tilde{X}, \tilde{Y})}{\log \min\{k, l\}} \right) + \frac{I_G(\tilde{X}, \tilde{Y})}{\log \min\{k, l\}}.$$

By sub-additivity of supremum, one has

$$\begin{aligned}
\sup_{k, l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k, l)} \frac{I_G(X, Y)}{\log \min\{k, l\}} &\leq \sup_{k, l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k, l)} \left( \frac{I_G(X, Y)}{\log \min\{k, l\}} - \frac{I_G(\tilde{X}, \tilde{Y})}{\log \min\{k, l\}} \right) \\
&\quad + \sup_{k, l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k, l)} \frac{I_G(\tilde{X}, \tilde{Y})}{\log \min\{k, l\}} \\
&\leq \sup_{k, l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k, l)} \left| \frac{I_G(X, Y)}{\log \min\{k, l\}} - \frac{I_G(\tilde{X}, \tilde{Y})}{\log \min\{k, l\}} \right| \\
&\quad + \sup_{k, l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k, l)} \frac{I_G(\tilde{X}, \tilde{Y})}{\log \min\{k, l\}}.
\end{aligned}$$

This implies

$$\operatorname{MIC}(X,Y)-\operatorname{MIC}(\tilde{X},\tilde{Y}) \leq \sup_{k,l\in\mathbb{N}\smallsetminus\{1\}} \sup_{G\in G(k,l)} \left|\frac{I_G(X,Y)}{\log\min\{k,l\}}-\frac{I_G(\tilde{X},\tilde{Y})}{\log\min\{k,l\}}\right|.$$

Similarly, one obtains

$$\begin{aligned}\operatorname{MIC}(\tilde{X},\tilde{Y})-\operatorname{MIC}(X,Y) &\leq \sup_{k,l\in\mathbb{N}\smallsetminus\{1\}} \sup_{G\in G(k,l)} \left|\frac{I_G(\tilde{X},\tilde{Y})}{\log\min\{k,l\}}-\frac{I_G(X,Y)}{\log\min\{k,l\}}\right| \\ &= \sup_{k,l\in\mathbb{N}\smallsetminus\{1\}} \sup_{G\in G(k,l)} \left|\frac{I_G(X,Y)}{\log\min\{k,l\}}-\frac{I_G(\tilde{X},\tilde{Y})}{\log\min\{k,l\}}\right|.\end{aligned}$$

So, one obtains

$$\operatorname{MIC}(X,Y)-\operatorname{MIC}(\tilde{X},\tilde{Y}) \geq -\sup_{k,l\in\mathbb{N}\smallsetminus\{1\}} \sup_{G\in G(k,l)} \left|\frac{I_G(X,Y)}{\log\min\{k,l\}}-\frac{I_G(\tilde{X},\tilde{Y})}{\log\min\{k,l\}}\right|,$$

then we are done. ☐

Now we are ready to prove an other version of the continuity theorem of MIC in the original version of [15].

**Proposition 4.17.** *On $\mathcal{P}(\mathbb{I})$ equipped with topology induced by $d_{RTV}$, the map $(X,Y)\mapsto \operatorname{MIC}(X,Y)$ is uniformly continuous.*

*Proof.* Firstly, one will show that the collection of mappings

$$F=\{(X,Y)\mapsto (X,Y)_G\colon k,l\in\mathbb{N}\smallsetminus\{1\}, G\in G(k,l)\}$$

is uniformly equicontinuous with respect to $d_{RTV}$. Let $\epsilon>0$ and $(\tilde{X},\tilde{Y})$ be another random vector. Then, by Lemma 4.14, one yields that $d_{RTV}((X,Y)_G,(\tilde{X},\tilde{Y})_G)$ does not exceed $d_{RTV}((X,Y),(\tilde{X},\tilde{Y}))$. If choosing $\delta:=\epsilon$ and and assume that $d_{RTV}((X,Y),(\tilde{X},\tilde{Y}))\leq\delta$, one has

$$d_{RTV}((X,Y)_G,(\tilde{X},\tilde{Y})_G)\leq d_{RTV}((X,Y),(\tilde{X},\tilde{Y}))\leq\epsilon.$$

Since $d_{RTV}$ and $d_{TV}$ coincide, one can see that the class of mappings $F$ is also uniformly equicontinuous when the domain is equipped with $d_{RTV}$ and range is equipped with $d_{TV}$.

Then, from uniform equicontinuity of $F$ and Proposition 4.15, one has that

$$F' := \{(X,Y) \mapsto \frac{I_G(X,Y)}{\log \min\{k,l\}} : k,l \in \mathbb{N} \setminus \{1\}, G \in G(k,l)\}$$

is also uniformly equicontinuous when the domain is equipped by $d_{RTV}$.

Now we are ready to prove Proposition 4.17. Let $(X,Y)$ be a pair of random variables and $\epsilon > 0$. Then, by uniformly equicontinuity of $F'$, there exists $\delta > 0$ such that for any pair of random variables $(\tilde{X}, \tilde{Y})$ with $d_{RTV}((X,Y),(\tilde{X},\tilde{Y})) < \delta$ and for any $k,l \in \mathbb{N} \setminus \{1\}$ and $G \in G(k,l)$, one has

$$\left| \frac{I_G(X,Y)}{\log \min\{k,l\}} - \frac{I_G(\tilde{X},\tilde{Y})}{\log \min\{k,l\}} \right| < \epsilon. \tag{4.4}$$

Since $k,l \in \mathbb{N} \setminus \{1\}$ and $G \in G(k,l)$ are arbitrary, then one can conclude from inequality (4.4) and Lemma 4.16 that for any random pair $(\tilde{X},\tilde{Y})$ with $d_{RTV}((X,Y),(\tilde{X},\tilde{Y})) < \delta$,

$$|\operatorname{MIC}(X,Y) - \operatorname{MIC}(\tilde{X},\tilde{Y})| \leq \sup_{k,l \in \mathbb{N} \setminus \{1\}} \sup_{G \in G(k,l)} \left| \frac{I_G(X,Y)}{\log \min\{k,l\}} - \frac{I_G(\tilde{X},\tilde{Y})}{\log \min\{k,l\}} \right| < \epsilon.$$

This finishes the proof. □

To use the above continuity theorem, a mode of convergence of sequences of functions is needed.

**Definition 4.18.** Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions from measure space $(\Omega, \mathcal{F}, \mu)$ to $\mathbb{R}^2$ and $f\colon (\Omega, \mathcal{F}, \mu) \to \mathbb{R}^2$ be another measurable function. Then, one says that $f_n \to f$ *almost uniformly* if for any $\epsilon > 0$, there is $N_\epsilon \in \mathcal{F}$ such that $\mu(N_\epsilon) < \epsilon$ and $f_n \to f$ uniformly on $\Omega \setminus N_\epsilon$. This definition also works for the case of real-valued functions.

In our context, almost uniform convergence has prominent role since it implies the convergence of random vectors with respect to $d_{RTV}$.

**Theorem 4.19.** *Let $\{\vec{X}_n\}_{n\in\mathbb{N}}$ be a sequence of two-dimensional random vectors whose components are uniform random variables on $\mathbb{I}$ and $\vec{X}$ is another random vector with* Uniform($\mathbb{I}$) *components such that $\vec{X}_n \to \vec{X}$ almost uniformly. Then, $\vec{X}_n \to \vec{X}$ in $d_{RTV}$ distance.*

*Proof.* Assume that $\vec{X}_n \to \vec{X}$ almost uniformly and let $\epsilon > 0$. Then, there exists $N_\epsilon \in \mathcal{F}$ such that such that $\mu(N_\epsilon) < \epsilon$ and $\vec{X}_n \to \vec{X}$ uniformly on $\Omega \setminus N_\epsilon$. This means there is $N \in \mathbb{N}$ such that $\|\vec{X}_n(\omega) - \vec{X}(\omega)\| < \epsilon$ for any $n \geq N$ and $\omega \in \Omega \setminus N_\epsilon$. Let $B := [a,b] \times [c,d] \subseteq \mathbb{I}^2$. Define, for any $k > 0$ and $\vec{x} \in \mathbb{I}^2$, $B(\vec{x}; k)$ an open ball centered at $\vec{x}$ with radius $k$ and $B_k = \bigcup_{\vec{x} \in B} B(\vec{x}; k)$. One would claim two things.

Firstly, one claims $\{\vec{X} \in B\} \setminus N_\epsilon \subseteq \{\vec{X}_n \in B_\epsilon\} \setminus N_\epsilon$ for $n \geq N$. Let $\omega \in \{\vec{X} \in B\} \setminus N_\epsilon$, i.e. $\vec{X}(\omega) \in B$ and $\omega \in \Omega \setminus N_\epsilon$. Since $\omega \in \Omega \setminus N_\epsilon$, one has $\|\vec{X}_n(\omega) - \vec{X}(\omega)\| < \epsilon$, thus $\vec{X}_n(\omega) \in B_\epsilon$ since $\vec{X}(\omega) \in B$. Moreover, because $\omega \in \Omega \setminus N_\epsilon$, one has $\omega \in \{\vec{X}_n \in B_\epsilon\} \setminus N_\epsilon$.

The second claim proved here is $\{\vec{X}_n \in B_\epsilon\} \setminus N_\epsilon \subseteq \{\vec{X} \in B_{2\epsilon}\}$ for $n \geq N$. Let $\omega \in \{\vec{X}_n \in B_\epsilon\} \setminus N_\epsilon$. This implies $\vec{X}_n(\omega) \in B_\epsilon$ and $\omega \in \Omega \setminus N_\epsilon$. Thus, there is $\vec{x}_0 \in B$ such that $\|\vec{X}_n(\omega) - \vec{x}_0\| < \epsilon$. Furthermore, since $\omega \in \Omega \setminus N_\epsilon$, one has $\|\vec{X}_n(\omega) - \vec{X}(\omega)\| < \epsilon$. By triangle inequality, $\|\vec{X}(\omega) - \vec{x}\| < 2\epsilon$. This fact combined with $\vec{x}_0 \in B$ proves this claim.

Now, we choose $\tilde{N} := N$. Then, for all $n \geq \tilde{N}$,

$$
\begin{aligned}
&|P(\vec{X}_n \in B) - P(\vec{X} \in B)| \\
&\leq |P(\vec{X}_n \in B) - P(\vec{X}_n \in B_\epsilon)| + |P(\vec{X}_n \in B_\epsilon) - P(\vec{X} \in B)| \\
&\leq P(\vec{X}_n \in B_\epsilon \setminus B) + \Big| P(\{\vec{X}_n \in B_\epsilon\} \setminus N_\epsilon) - P(\{\vec{X} \in B\} \setminus N_\epsilon) \Big| \\
&\quad + P(\{\vec{X}_n \in B_\epsilon\} \cap N_\epsilon) + P(\{\vec{X} \in B\} \cap N_\epsilon) \\
&\leq P(\vec{X}_n \in B_\epsilon \setminus B) + P((\{\vec{X}_n \in B_\epsilon\} \setminus N_\epsilon) \setminus (\{\vec{X} \in B\} \setminus N_\epsilon)) \\
&\quad + P(\{\vec{X}_n \in B_\epsilon\} \cap N_\epsilon) + P(\{\vec{X} \in B\} \cap N_\epsilon) \\
&< P(\vec{X}_n \in B_\epsilon \setminus B) + P(\{\vec{X} \in B_{2\epsilon}\} \setminus (\{\vec{X} \in B\} \setminus N_\epsilon)) + 2\epsilon \\
&= P(\vec{X}_n \in B_\epsilon \setminus B) + P(\{\vec{X} \in B_{2\epsilon}\} \cap (\{\vec{X} \in B\} \setminus N_\epsilon)^c) + 2\epsilon \\
&= P(\vec{X}_n \in B_\epsilon \setminus B) + P(\{\vec{X} \in B_{2\epsilon}\} \cap (\{\vec{X} \in B\} \cap N_\epsilon^c)^c) + 2\epsilon \\
&= P(\vec{X}_n \in B_\epsilon \setminus B) + P(\{\vec{X} \in B_{2\epsilon}\} \cap (\{\vec{X} \in B\}^c \cup N_\epsilon)) + 2\epsilon \\
&\leq P(\vec{X}_n \in B_\epsilon \setminus B) + P(\vec{X} \in B_{2\epsilon} \setminus B) + P(\{\vec{X} \in B_\epsilon\} \cap N_\epsilon)) + 2\epsilon \\
&< P(\vec{X}_n \in B_\epsilon \setminus B) + P(\vec{X} \in B_{2\epsilon} \setminus B) + 3\epsilon.
\end{aligned}
$$

Then, consider the term $P(\vec{X}_n \in B_\epsilon \setminus B)$. Since the components of $\vec{X}_n$ are Uniform($\mathbb{I}$), one has that its joint probability measure of $\vec{X}_n$ will be induced measure from copula, says $\mu_{C_n}$ for some $C_n \in \mathcal{C}$. Then, one has

$$\begin{aligned}
&P(\vec{X}_n \in B_\epsilon \setminus B) \\
&= \mu_{C_n}(B_\epsilon \setminus B) \\
&\leq \mu_{C_n}(([a-\epsilon, b+\epsilon] \times [c-\epsilon, d+\epsilon]) \setminus ([a,b] \times [c,d])) \\
&\leq \mu_{C_n}(([a-\epsilon, a] \times \mathbb{I}) \cup ([b, b+\epsilon] \times \mathbb{I}) \cup (\mathbb{I} \times [c-\epsilon, c]) \cup (\mathbb{I} \times [d, d+\epsilon])) \\
&\leq \mu_{C_n}([a-\epsilon, a] \times \mathbb{I}) + \mu_{C_n}([b, b+\epsilon] \times \mathbb{I}) \\
&\quad + \mu_{C_n}((\mathbb{I} \times [c-\epsilon, c]) + \mu_{C_n}(\mathbb{I} \times [d, d+\epsilon]) \\
&= 4\epsilon.
\end{aligned}$$

The last line comes from the fact that copula measure is doubly stochastic. By the same fashion, one has $P(\vec{X} \in B_{2\epsilon} \setminus B) \leq 8\epsilon$. Thereby,

$$|P(\vec{X}_n \in B) - P(\vec{X} \in B)| < 15\epsilon.$$

Since the choice of $B$ is arbitrary in $\mathcal{B}(\mathbb{I}^2)$, we are done. $\square$

And, given one-to-one piecewise linear measure-preserving transformation, one finds another one-to-one piecewise linear measure-preserving function whose all discontinuity points are rational numbers and both mappings are *close* enough on the set of almost full measure.

**Proposition 4.20.** *Let $\varphi\colon \mathbb{I} \to \mathbb{I}$ be a one-to-one piecewise linear measure-preserving transformation. Then, for $\epsilon > 0$, there is $\tilde{\varphi}\colon \mathbb{I} \to \mathbb{I}$ a one-to-one piecewise linear measure-preserving function whose all discontinuity points are rational numbers and a set $M_\epsilon \in \mathcal{B}(\mathbb{I})$ such that $\lambda(M_\epsilon) < \epsilon$ and*

$$\sup_{x \in \mathbb{I} \setminus M_\epsilon} |\varphi(x) - \tilde{\varphi}(x)| < \epsilon.$$

*Proof.* Let $\varphi\colon \mathbb{I} \to \mathbb{I}$ be a one-to-one piecewise linear measure-preserving transformation. By Corollary 4.10, one can consider its corresponding shuffle representation $(k, \{s_i\}_{i=0}^k, \{t_i\}_{i=0}^k, \sigma, \{m_i\}_{i=1}^k)$. Without loss of generality, one can consider the case of only straight shuffle representation since dealing with flipped piece is also the same[2]. Then, by definition of shuffle representation, one has, for any $i \in \{1,\dots,k\}$, $t_{\sigma(i)} - t_{\sigma(i)-1} = s_i - s_{i-1}$. This implies the following recurrence relation:

$$t_i - t_{i-1} = s_{\sigma^{-1}(i)} - s_{\sigma^{-1}(i)-1} \text{ for all } i \in \{1,\dots,k\}. \tag{4.5}$$

From Equation (4.5), one has $t_0 = 0$ (by setup) and, for $i = 1,\dots,k$

$$t_i = \sum_{l=2}^{i}(t_l - t_{l-1}) + t_1 = \sum_{l=1}^{i}(t_l - t_{l-1}) = \sum_{l=1}^{i}(s_{\sigma^{-1}(l)} - s_{\sigma^{-1}(l)-1}).$$

By definition of straight shuffle, for each $i \in \{1,\dots,k\}$, the graph of $\varphi$ on $[s_{i-1}, s_i]$ will be the line segment from $(s_{i-1}, t_{\sigma(i)-1})$ to $(s_i, t_{\sigma(i)})$ with unit slope. Thus, one has, for all $i \in \{1,\dots,k\}$ and $x \in [s_{i-1}, s_i]$,

$$\varphi(x) = x + t_{\sigma(i)} - s_i = x + \sum_{l=1}^{\sigma(i)}(s_{\sigma^{-1}(l)} - s_{\sigma^{-1}(l)-1}) - s_i. \tag{4.6}$$

Then, let $0 < \epsilon < \min\limits_{1\le j\le k} \Delta s_j$. By density of $\mathbb{Q} \cap \mathbb{I}$ in $\mathbb{I}$, there are $q_1,\dots,q_{k-1} \in \mathbb{Q} \cap \mathbb{I}$ such that, for any $i \in \{1,\dots,k-1\}$,

$$\lambda([s_{i-1}, s_i]) - \frac{\epsilon}{2k^2} < q_i < \lambda([s_{i-1}, s_i]) \tag{4.7}$$

and define $q_k := 1 - \sum_{j=1}^{k-1} q_j$.

One can claim that $|(s_i - s_{i-1}) - q_i| \le \epsilon/2k$ for any $i \in \{1,\dots,k\}$. For the case of $i \in \{1,\dots,k-1\}$, expression (4.7) yields $-\epsilon/2k^2 < 0 < (s_i - s_{i-1}) - q_i < \epsilon/2k^2$. This implies

$$|(s_i - s_{i-1}) - q_i| < \frac{\epsilon}{2k^2}, \tag{4.8}$$

[2]In case of flipped part, one can get, by the same fashion that $\varphi(x) = x + t_{\sigma(i)} - s_{i-1}$ on that piece if $\varphi \sim (k, \{s_i\}, \{t_i\}, \sigma, m)$. This implies, in case of $\tilde{\varphi} \sim (k, \{\tilde{s}_i\}, \{\tilde{t}_i\}, \sigma, m)$, that for $x \in (\mathbb{I} \smallsetminus M_\epsilon) \cap [s_{i-1}, s_i]$,

$$|\varphi(x) - \tilde{\varphi}(x)| = |(t_{\sigma(i)} - \tilde{t}_{\sigma(i)}) - (s_{i-1} - \tilde{s}_{i-1})|.$$

Thus, the proof in this case can be treated in the same way to the straight shuffle.

so $|(s_i - s_{i-1}) - q_i| < \epsilon/2k$. For $i = k$, one has the following:

$$\begin{aligned}
|(s_k - s_{k-1}) - q_k| &= \left|1 - \sum_{j=1}^{k-1}(s_j - s_{j-1}) - \left(1 - \sum_{j=1}^{k-1} q_j\right)\right| \\
&= \left|\sum_{j=1}^{k-1}((s_j - s_{j-1}) - q_j)\right| \\
&\leq \sum_{j=1}^{k-1} |(s_j - s_{j-1}) - q_j| \\
&< \frac{\epsilon(k-1)}{2k^2} \\
&< \frac{\epsilon}{2k}.
\end{aligned}$$

The second line from the bottom is the consequence of inequality (4.8). Then, define $\{\tilde{s}_i\}_{i=0}^k$ by $\tilde{s}_0 := 0$, $\tilde{s}_k := 1$ and, for $i = 1, \ldots, k-1$, $\tilde{s}_i := \sum_{j=1}^{i} q_j$ and $\tilde{\varphi}\colon \mathbb{I} \to \mathbb{I}$ a one-to-one piecewise linear measure preserving function such that $\tilde{\varphi}$ has shuffle representation $(k, \{\tilde{s}_i\}_{i=0}^k, \sigma)$. The closed form of $\tilde{\varphi}$ is, by the same fashion, for $x \in [\tilde{s}_{i-1}, \tilde{s}_i]$,

$$\tilde{\varphi}(x) = x + \tilde{t}_{\sigma(i)} - \tilde{s}_i = x + \sum_{l=1}^{\sigma(i)} (\tilde{s}_{\sigma^{-1}(l)} - \tilde{s}_{\sigma^{-1}(l)-1}) - \tilde{s}_i. \tag{4.9}$$

Choose $M_\epsilon := \bigcup_{i=0}^k B(s_i, \epsilon/2k) \cap \mathbb{I}$. Obviously, $M_\epsilon \in \mathcal{B}(\mathbb{I})$ and $\lambda(M_\epsilon) < \epsilon$.

And, to finish the proof, one can show that $|\varphi(x) - \tilde{\varphi}(x)| < \epsilon$ for any $x \in (\mathbb{I} \setminus M_\epsilon) \cap [s_{i-1}, s_i]$ and $i = 1, \ldots, k$.

Let $x \in (\mathbb{I} \setminus M_\epsilon) \cap [s_{i-1}, s_i]$. By (4.6) and (4.9), one has

$$\begin{aligned}
|\varphi(x) - \tilde{\varphi}(x)| &= |(x + t_{\sigma(i)} - s_i) - (x + \tilde{t}_{\sigma(i)} - \tilde{s}_i)| \\
&= |(t_{\sigma(i)} - \tilde{t}_{\sigma(i)}) - (s_i - \tilde{s}_i)| \\
&\leq |s_i - \tilde{s}_i| + |t_{\sigma(i)} - \tilde{t}_{\sigma(i)}| \\
&= \left|s_i - \sum_{l=1}^{i} q_l\right| + \left|\sum_{l=1}^{\sigma(i)}(s_{\sigma^{-1}(l)} - s_{\sigma^{-1}(l)-1}) + \sum_{l=1}^{\sigma(i)}(\tilde{s}_{\sigma^{-1}(l)} - \tilde{s}_{\sigma^{-1}(l)-1})\right| \\
&= \left|\sum_{l=1}^{i}((s_l - s_{l-1}) - q_l)\right| \\
&\quad + \left|\sum_{l=1}^{\sigma(i)} \big((s_{\sigma^{-1}(l)} - s_{\sigma^{-1}(l)-1}) - (\tilde{s}_{\sigma^{-1}(l)} - \tilde{s}_{\sigma^{-1}(l)-1})\big)\right|
\end{aligned}$$

$$
\begin{aligned}
&= \left|\sum_{l=1}^{i}((s_l - s_{l-1}) - q_l)\right| \\
&\quad + \left|\sum_{l=1}^{\sigma(i)} \big((s_{\sigma^{-1}(l)} - \tilde{s}_{\sigma^{-1}(l)}) - (s_{\sigma^{-1}(l)-1} - \tilde{s}_{\sigma^{-1}(l)-1})\big)\right| \\
&= \left|\sum_{l=1}^{i}((s_l - s_{l-1}) - q_l)\right| \\
&\quad + \left|\sum_{l=1}^{\sigma(i)} \left(\left(s_{\sigma^{-1}(l)} - \sum_{m=1}^{\sigma^{-1}(l)} q_m\right) - \left(s_{\sigma^{-1}(l)-1} - \sum_{n=1}^{\sigma^{-1}(l)-1} q_n\right)\right)\right| \\
&= \left|\sum_{l=1}^{i}((s_l - s_{l-1}) - q_l)\right| \\
&\quad + \left|\sum_{l=1}^{\sigma(i)} \left(\sum_{m=1}^{\sigma^{-1}(l)} ((s_m - s_{m-1}) - q_m) - \sum_{n=1}^{\sigma^{-1}(l)-1} ((s_n - s_{n-1}) - q_n)\right)\right| \\
&= \left|\sum_{l=1}^{i}((s_l - s_{l-1}) - q_l)\right| + \left|\sum_{l=1}^{\sigma(i)}((s_{\sigma^{-1}(l)} - s_{\sigma^{-1}(l)-1}) - q_{\sigma^{-1}(l)})\right| \\
&\leq \sum_{l=1}^{i} |(s_l - s_{l-1}) - q_l| + \sum_{l=1}^{\sigma(i)} |(s_{\sigma^{-1}(l)} - s_{\sigma^{-1}(l)-1}) - q_{\sigma^{-1}(l)}| \\
&< \frac{\epsilon i}{2k} + \frac{\epsilon \sigma(i)}{2k} \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon.
\end{aligned}
$$

Since $i \in \{1, \ldots, k\}$ and $x \in (\mathbb{I} \setminus M_\epsilon) \cap [s_{i-1}, s_i]$ are arbitrary, we are done for this proof. □

To extend the theorem of [1] to the tool ready to prove complete dependence detectability, the well-known Egoroff's theorem in measure theory is handy. For the proof of this theorem, see real analysis textbooks such as [17].

**Theorem 4.21** ([17])**.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions from finite measure space $(\Omega, \mathcal{F}, \mu)$ to $\mathbb{R}$ and $f\colon (\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ be another measurable function. Then, if $f_n \to f$ almost surely, then $f_n \to f$ almost uniformly.*

Now, one obtains the weaker version of Theorem 4.8.

**Theorem 4.22.** *Let $\varphi\colon \mathbb{I} \to \mathbb{I}$ be measure-preserving. Then there exists a sequence of one-to-one piecewise linear measure-preserving maps $\{f_n\}_{n\in\mathbb{N}}$ whose discontinuity points of $f_n$ are rational numbers for $n \in \mathbb{N}$ such that $f_n \to \varphi$ almost uniformly.*

*Proof.* Assume that $\varphi$ is measure-preserving maps on $\mathbb{I}$. Theorem 4.8 says there is a one-to-one piecewise linear measure-preserving $\{\varphi_n\}_{n\in\mathbb{N}}$ such that $\varphi_n \to \varphi$ almost surely. Egoroff's theorem implies that $\varphi_n \to \varphi$ weak uniformly. Then, let $\epsilon > 0$. By weakly uniform uniform convergence, there is $N_\epsilon \in \mathcal{B}(\mathbb{I})$ such that $\lambda(N_\epsilon) < \epsilon/2$ and $\sup_{x\in\mathbb{I}\setminus N_\epsilon} |\varphi(x) - \varphi_n(x)| < \epsilon/2$. For any $n \in \mathbb{N}$, thanks to Lemma 4.7, one has corresponding one-to-one piecewise linear measure-preserving $f_n$ whose discontinuity points are rational numbers and $M_\epsilon \in \mathcal{B}(\mathbb{I})$ such that $\lambda(M\epsilon) < \epsilon/2$ and $\sup_{x\in\mathbb{I}\setminus M_\epsilon} |f_n(x) - \varphi_n(x)| < \epsilon/2$. Choose $\tilde{M_\epsilon} := N_\epsilon \cup M_\epsilon$. One has $\lambda(\tilde{M_\epsilon}) < \epsilon$ and for any $x \in \mathbb{I} \setminus \tilde{M_\epsilon}$,

$$|\varphi(x) - f_n(x)| \leq |\varphi(x) - \varphi_n(x)| + |\varphi_n(x) - f_n(x)| < \epsilon.$$

This finishes the proof. $\square$

And, the following is one of the key lemma.

**Lemma 4.23.** *Let $X \sim$ Uniform$(\mathbb{I})$ and $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions on $\mathbb{I}$ to $\mathbb{R}$ and $f\colon \mathbb{I} \to \mathbb{R}$ be another measurable function such that $f_n \to f$ almost uniformly. Then $(X, f_n(X)) \to (X, f(X))$ almost uniformly too.*

*Proof.* Assume that $X \sim$ Uniform$(\mathbb{I})$ and $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of measurable functions on $\mathbb{I}$ to $\mathbb{R}$ and $f\colon \mathbb{I} \to \mathbb{R}$ be another measurable function such that $f_n \to f$ almost uniformly. Let $\epsilon > 0$, then there exists $N_\epsilon \in \mathcal{B}(\mathbb{I})$ such that $\lambda(N_\epsilon) < \epsilon$ and $|f_n(x) - f(x)| < \epsilon$ for any $x \in \mathbb{I} \setminus N_\epsilon$. Define $\tilde{B} := \{X \in N_\epsilon\}$. One has $\tilde{B}$ is measurable set since $X$ is random variable and $P(\tilde{B}) = P(X \in N_\epsilon) = \lambda(N_\epsilon) < \epsilon$ since $X \sim$ Uniform$(\mathbb{I})$. Moreover, for any $\omega \in \Omega \setminus \tilde{B}$, $X(\omega) \in \mathbb{I} \setminus N_\epsilon$, so

$$\|(X(\omega), f_n(X(\omega))) - (X(\omega), f(X(\omega)))\| = |f_n(X(\omega)) - f(X(\omega))| < \epsilon.$$

The proof is done here. $\square$

Thus, one has following powerful corollary.

**Corollary 4.24.** *Let* $X, Y \sim \text{Uniform}(\mathbb{I})$ *and* $\varphi \colon \mathbb{I} \to \mathbb{I}$ *be a Borel measurable function such that* $Y = \varphi(X)$. *Then, there exists a sequence of one-to-one piecewise linear measure-preserving maps* $\{f_n\}_{n\in\mathbb{N}}$ *whose discontinuity point of* $f_n$ *are rational numbers for* $n \in \mathbb{N}$ *such that* $(X, f_n(X)) \to (X, f(X))$ *in* $d_{RTV}$.

*Proof.* Let $X, Y \sim \text{Uniform}(\mathbb{I})$ and $\varphi \colon \mathbb{I} \to \mathbb{I}$ be a Borel measurable function such that $Y = \varphi(X)$. Then, by Lemma 4.7, $\varphi$ must be measure-preserving. By Theorem 4.22, there is a sequence of one-to-one piecewise linear measure-preserving functions $\{f_n\}_{n\in\mathbb{N}}$ whose discontinuity point of $f_n$ are rational numbers for $n \in \mathbb{N}$ such that $f_n \to \varphi$ almost uniformly. By the above lemma, $(X, f_n(X)) \to (X, f(X))$ almost uniformly too. Lastly, by Theorem 4.19, $(X, f_n(X)) \to (X, f(X))$ in $d_{RTV}$. □

From this point, we are ready to prove dependence detectability of MIC. For convenience, we recall the statement and prove it here.

**Proposition 4.25.** $\text{MIC}(X, Y) = 1$ *if either* $Y$ *is CD on* $X$ *or* $X$ *is CD on* $Y$.

*Proof.* Firstly, assume $X, Y \sim \text{Uniform}(\mathbb{I})$ such that there is a Borel measurable function $f \colon \mathbb{I} \to \mathbb{I}$ which making $Y = f(X)$ almost surely. By Corollary 4.24, the sequence of one-to-one piecewise linear measure-preserving maps $\{f_n\}_{n\in\mathbb{N}}$ whose discontinuity points of $f_n$ are rational numbers for $n \in \mathbb{N}$ such that $(X, f_n(X)) \to (X, f(X))$ in $d_{RTV}$ exists. Proposition 4.12 claims that for every $n \in \mathbb{N}$, one has $\text{MIC}(X, f_n(X)) = 1$. Finally, because of continuity of MIC with respect to $d_{RTV}$ and $(X, f_n(X)) \to (X, f(X))$ in the mode of $d_{RTV}$, one has $\text{MIC}(X, Y) = \text{MIC}(X, f(X)) = 1$. For the case that $X$ is completely dependent on $Y$, symmetry of MIC helps us to use the result proved here. This task of proving complete dependence detectable is completed now. □

## 4.3 Concluding Remark

From the results in the first two sections, one can conclude that $\tilde{\tau}$ is a class of symmetric dependence measure which is mutually complete dependence detectable and monotonic invariant. MIC is proved to have complete dependence detectability and monotonic invariance. Also, the incompatibilty can be proved in the same way to the nonsymmetric case since the counterexample to impose contradiction is bijection indeed. Thus, in the same fashion on the non-symmetric case, the summary table for symmetric dependence measures is given.

| | MD detect. | MCD detect. | CD detect. |
|---|---|---|---|
| Monotone inv. | $\sigma$ | $\tilde{\tau}$, $\omega$, $\nu_\alpha$ | |
| Injective inv. | Impossible | | $S$, $\omega_*$, $\nu_*$ |

Table 4.1: Recalling the summary table for symmetric dependence measures.

One may observe about the absence of MIC in this table. From the above section, MIC seems to be placed at the upper-right cell of the table since it can was proved to be both monotonic invariant and complete dependence detectable. However, there is a case that invariance and the detectability go beyond what we proved.

Let $X \sim \text{Uniform}(\mathbb{I})$ and, for $\alpha \in (0,1)$, define the map $f_\alpha \colon \mathbb{I} \to \mathbb{I}$ as, for $x \in \mathbb{I}$, $f_\alpha(x) := \frac{x}{\alpha}\chi_{[0,\alpha)}(x) + \frac{x-1}{\alpha-1}\chi_{[\alpha,1]}(x)$. Its graph is illustrated in Figure 4.1. Consider the random pairs $(f_{1/3}(X), f_{2/3}(X))$.

By tedious but straightforward calculation, $\text{MIC}(f_{1/3}(X), f_{2/3}(X)) = 1$. From $\text{MIC}(X,X) = 1$, one of interesting point is that in this case, MIC is invariant ever when transforming random variables by functions which are not one-to-one. For the second important point, note that MIC is one in this case while $f_{2/3}(X)$ is not completely dependent on $f_{1/3}(X)$ and vice versa. For the derivation of these results, see appendix.

Although this is just an example, general properties await further for explorations. Possibility of new results on MIC's invariance and dependence detectability properties still loom large. Since MIC already has a consistent statistical version, the study on finding its estimator is not needed.

# APPENDIX

The appendix is devoted to proving some results on the example in the concluding remark. Since the calculation is tedious but straightforward, showing the proof here may be more appropriate.

Let us recall the setup. Let $X \sim \text{Uniform}(\mathbb{I})$ and the map $f_\alpha \colon \mathbb{I} \to \mathbb{I}$, for $\alpha \in (0,1)$, is defined by

$$f_\alpha(x) = \begin{cases} \dfrac{x}{\alpha} & \text{if } x \in [0, \alpha) \\ \dfrac{x-1}{\alpha-1} & \text{if } x \in [\alpha, 1]. \end{cases}$$

This function is called a tent map, and its graph is shown in Figure 4.1.

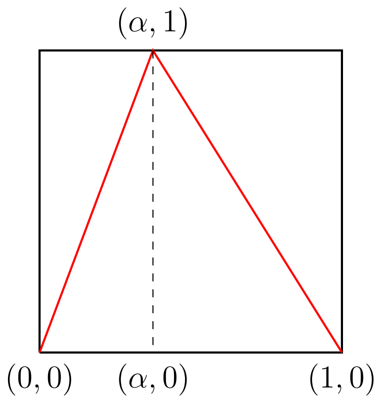


Figure 4.1: The graph of $f_\alpha$ on $\mathbb{I}$.

Obviously, this function is not strictly monotone, yet clearly measure preserving. For completeness, we include a proof here in the following lemma.

**Lemma 4.26.** *For $\alpha \in (0,1)$, $f_\alpha \colon \mathbb{I} \to \mathbb{I}$ is a measure preserving map.*

*Proof.* Let $\alpha \in (0,1)$ and $[a,b] \subseteq \mathbb{I}$. Then, one yields

$$\begin{aligned}
\lambda(f_\alpha^{-1}([a,b])) &= \lambda(f_\alpha^{-1}([0,b] \setminus [0,a])) \\
&= \lambda(f_\alpha^{-1}([0,b])) - \lambda(f_\alpha^{-1}([0,a)) \\
&= \lambda(\mathbb{I} \setminus (\alpha b, 1-(1-\alpha)b)) - \lambda(\mathbb{I} \setminus [\alpha a, 1-(1-\alpha)a]) \\
&= (1-(1-b+\alpha b-\alpha b)) - (1-(1-a+\alpha a-\alpha a)) \\
&= b-a = \lambda([a,b]).
\end{aligned}$$

Since the set $\{[a,b] \colon a,b \in \mathbb{I} \text{ and } a<b\}$ generates $\mathcal{B}(\mathbb{I})$, we are done. ☐

Note that $f_\alpha(X)$ is still a uniform random variable on $\mathbb{I}$. Thus, for $\alpha, \beta \in (0,1)$, the joint probability distribution of $f_\alpha(X)$ and $f_\beta(X)$ will be the copula $C_{\alpha,\beta}$ where, for any $x, y \in \mathbb{I}$, this copula is defined by

$$\begin{aligned}
C_{\alpha,\beta}(x,y) &= \lambda(f_\alpha^{-1}([0,x]) \cap f_\beta^{-1}([0,y])) \\
&= \lambda\left((\mathbb{I} \setminus (\alpha x, 1-(1-\alpha)x)) \cap (\mathbb{I} \setminus (\beta y, 1-(1-\beta)y))\right) \qquad (4.10) \\
&= \lambda\left(\mathbb{I} \setminus ((\alpha x, 1-(1-\alpha)x) \cup (\beta y, 1-(1-\beta)y))\right).
\end{aligned}$$

To prove the claim, let us first consider the case $\alpha < \beta$. There are following subcases:

- Case 1: $\alpha x \le \beta y$ and $1-(1-\alpha)x > 1-(1-\beta)y$.

  In this case, one has

$$\begin{aligned}
C_{\alpha,\beta}(x,y) &= 1-\lambda((\alpha x, 1-(1-\alpha)x)) \\
&= 1-(1-(1-\alpha)x-\alpha x) \\
&= x.
\end{aligned}$$

- Case 2: $\alpha x \le \beta y$ and $1-(1-\alpha)x \le 1-(1-\beta)y$.

  For this case, if $1-(1-\alpha)x < \beta y$, one has

$$\begin{aligned}
C_{\alpha,\beta}(x,y) &= 1-\lambda((\alpha x, 1-(1-\alpha)x)) - \lambda((\beta b, 1-(1-\beta)y)) \\
&= 1-(1-(1-\alpha)x-\alpha x) - (1-(1-\beta)y-\beta y) \\
&= x+y-1.
\end{aligned}$$

And if $1-(1-\alpha)x \geq \beta y$, one obtains

$$\begin{aligned} C_{\alpha,\beta}(x,y) &= 1-\lambda((\alpha x, 1-(1-\beta)y) \\ &= 1-(1-(1-\beta)y-\alpha x) \\ &= \alpha x + (1-\beta)y. \end{aligned}$$

- Case 3: $\alpha x \geq \beta y$ and $1-(1-\alpha)x > 1-(1-\beta)y$.

  This case cannot happen since from the assumption, one has

$$\alpha(1-\beta)xy \geq (1-\alpha)\beta xy.$$

  However, since we assume $\alpha < \beta$, the consequence is $\alpha(1-\beta) < (1-\alpha)\beta$, which contradicts the recent assertion.

- Case 4: $\alpha x \geq \beta y$ and $1-(1-\alpha)x \leq 1-(1-\beta)y$.

  This case is the same as the first case. One has $C_{\alpha,\beta}(x,y) = y$.

Thus, the closed form of this copula is

$$C_{\alpha,\beta}(x,y) = \begin{cases} x & \text{if } x \leq \frac{\beta}{\alpha}y \text{ and } x < \frac{1-\beta}{1-\alpha}y; \\ y & \text{if } x \geq \frac{\beta}{\alpha}y \text{ and } x \geq \frac{1-\beta}{1-\alpha}y; \\ \alpha x + (1-\beta)y & \text{if } x < \frac{\beta}{\alpha}y, x \geq \frac{1-\beta}{1-\alpha}y \text{ and } x < \frac{1-\beta y}{1-\alpha}; \\ x+y-1 & \text{if } x < \frac{\beta}{\alpha}y, x \geq \frac{1-\beta}{1-\alpha}y \text{ and } x \geq \frac{1-\beta y}{1-\alpha}. \end{cases} \tag{4.11}$$

In our example, we choose $(\alpha,\beta) = (1/3, 2/3)$. Denoting $C := C_{1/3,2/3}$ and rearranging the conditions, one gets more compact closed form as follow:

$$C(x,y) = \begin{cases} x & \text{if } x \leq \frac{1}{2}y; \\ y & \text{if } x > 2y; \\ \alpha x + (1-\beta)y & \text{if } \frac{1}{2}y < x \leq 2y \text{ and } x+y < \frac{3}{2}; \\ x+y-1 & \text{if } \frac{1}{2}y < x \leq 2y \text{ and } x+y \geq \frac{3}{2}. \end{cases} \tag{4.12}$$

The support of this copula is the triangle $\ell_1 \cup \ell_2 \cup \ell_3$ where

$$\begin{aligned} \ell_1 &:= \{(x, 2x)\colon x \in [0, \tfrac{1}{2}]\}; \\ \ell_2 &:= \{(x, \tfrac{3}{2} - x)\colon x \in [\tfrac{1}{2}, 1]\}; \\ \ell_3 &:= \{(x, \tfrac{1}{2}x)\colon x \in \mathbb{I}\}. \end{aligned}$$

The support of $C$ is illustrated in Figure 4.2.

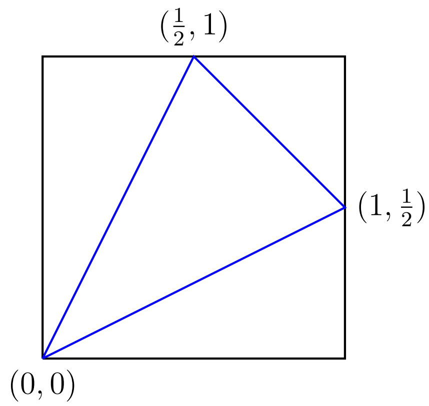


Figure 4.2: The support of copula $C := C_{1/3,2/3}$.

From the support of this copula, it is natural to choose the $2n \times 2n$ equipartition grid of $\mathbb{I}^2$. Thereby, for any block $C_i \times R_j$ of this grid, one has $\mu(C_i \times R_j) = \mu\left(\left[\frac{i-1}{2n}, \frac{i}{2n}\right] \times \left[\frac{j-1}{2n}, \frac{j}{2n}\right]\right)$ where $\mu := \mu_{C_{1/3,2/3}}$. Furthermore, by Sklar's theorem and $f_{1/3}(X), f_{2/3}(X) \sim \text{Uniform}(\mathbb{I})$, one can find joint probability in each block by $P(f_{1/3}(X) \in C_i, f_{2/3}(X) \in R_j) = \mu\left(\left[\frac{i-1}{2n}, \frac{i}{2n}\right] \times \left[\frac{j-1}{2n}, \frac{j}{2n}\right]\right)$ Since we realize support and grid, one can identify blocks containing each line of support and eliminate others when computing $I_G(f_{1/3}(X), f_{2/3}(X))$ from the fact that any measurable set which does not contain any part of support will have zero measure.

**Proposition 4.27.** *The collections*

$$\mathcal{B}_1 := \{C_i \times R_j \colon i \in \{1, \ldots, n\} \text{ and either } j = 2i \text{ or } j = 2i - 1\};$$

$$\mathcal{B}_2 := \{C_i \times R_j \colon i = n+1, \ldots, 2n \text{ and } j = 3n - i + 1\};$$

$$\mathcal{B}_3 := \{C_i \times R_j \colon j \in \{1, \ldots, n\} \text{ and either } i = 2j \text{ or } i = 2j - 1\}$$

*are the collections of blocks those intersect $\ell_1, \ell_2$ and $\ell_3$ respectively.*

*Proof.* To show that $\mathcal{B}_1$ intersect $\ell_1$, it suffices to prove that for any $i = 1, \dots, n$, the rectangle

$$(C_i \times R_{2i-1}) \cup (C_i \times R_{2i}) = \left[\frac{i-1}{2n}, \frac{i}{2n}\right] \times \left[\frac{i-1}{n}, \frac{i}{n}\right]$$

has its lower-left and upper-right vertices, denoted by $\vec{X}_{LL}, \vec{X}_{UR}$ respectively, lying on $\ell_1$. Note that one has $\vec{X}_{LL} = \left(\frac{i-1}{2n}, \frac{i-1}{n}\right)$ and $\vec{X}_{LL} = \left(\frac{i}{2n}, \frac{i}{n}\right)$ for all $i \in \{1, \dots, n\}$. Choose $\theta_1 := 1 - \left(\frac{i-1}{n}\right)$ and $\theta_2 := 1 - \frac{i}{n}$. Since $i = 1, \dots, n$, one can see that $\theta_1 \in (0, 1]$ and $\theta_2 \in [0, 1)$. Moreover, one has

$$\begin{aligned}
\vec{X}_{LL} &= \left(\frac{i-1}{2n}, \frac{i-1}{n}\right) \\
&= \left(\left(1 - \left(1 - \frac{i-1}{n}\right)\right)\frac{1}{2}, 1 - \left(1 - \frac{i-1}{n}\right)\right) \\
&= \left(\theta_1(0) + (1-\theta_1)\left(\frac{1}{2}\right), \theta_1(0) + (1-\theta_1)(1)\right) \\
&= \theta_1(0,0) + (1-\theta_1)\left(\frac{1}{2}, 1\right) \in \ell_1.
\end{aligned}$$

Similarly, one can show that $\vec{X}_{UR} = \theta_2(0,0) + (1-\theta_2)\left(\frac{1}{2}, 1\right) \in \ell_1$. Showing $\mathcal{B}_2$ and $\mathcal{B}_3$ intersect $\ell_2$ and $\ell_3$ respectively can be done in the same fashion. □

Any block $C_i \times R_j$ will be of zero measure with respect to $\mu$ since it does not contain any part of support. Thus, the next step will be finding the measure of blocks in $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$. Thus, with (4.12), finding value of each summand of $I_G(f_{1/3}(X), f_{2/3})$, denoted by $M_{ij}$, is possible.

It is not hard to see that $C_1 \times R_1 \in \mathcal{B}_1 \cup \mathcal{B}_3$. Moreover, one has

$$\mu(C_1 \times R_1) = \mu\left(\left[0, \frac{1}{2n}\right]^2\right) = C\left(\frac{1}{2n}, \frac{1}{2n}\right) = \frac{1}{3} \cdot \frac{1}{2n} + \frac{1}{3} \cdot \frac{1}{2n} = \frac{1}{3n}. \tag{4.13}$$

And one has

$$\begin{aligned}
M_{11} &= \log\left(\frac{\mu(C_1 \times R_1)}{P(f_{1/3}(X) \in C_1)P(f_{2/3}(X) \in R_1)}\right)\mu(C_1 \times R_1) \\
&= \frac{1}{3n}\log\left(\frac{1/3n}{1/4n^2}\right) = \frac{1}{3n}\log\left(\frac{4n}{3}\right).
\end{aligned}$$

The following propositions will give the value of $M_{ij}$ in case of $C_i \times R_j$ is in each $\ell_i$'s.

**Proposition 4.28.** *For any* $(i,j)$*. If* $C_i \times R_j \in \mathcal{B}_1$*, then*

$$M_{ij} = \begin{cases} \frac{1}{3n}\log\left(\frac{4n}{3}\right) & \text{if } (i,j) = (1,1) \\ \frac{1}{6n}\log\left(\frac{2n}{3}\right) & \text{otherwise.} \end{cases}$$

*If* $C_i \times R_j \in \mathcal{B}_2$*, then* $M_{ij} = \frac{1}{3n}\log\left(\frac{4n}{3}\right)$*. And, if* $C_i \times R_j \in \mathcal{B}_3$*, then*

$$M_{ij} = \begin{cases} \frac{1}{3n}\log\left(\frac{4n}{3}\right) & \text{if } (i,j) = (1,1) \\ \frac{1}{6n}\log\left(\frac{2n}{3}\right) & \text{otherwise.} \end{cases}$$

*Proof.* Firstly, one will prove for the case that $C_i \times R_j \in \mathcal{B}_1$. The case $(i,j) = (1,1)$ has been proved already. For the second case, let $i \in \{1, \ldots, n\}$ and either $j = 2i$ or $j = 2i - 1$ where $i \neq j \neq 1$. If $j = 2i - 1$ and considering the rectangle $C_i \times R_{2i-1} = \left[\frac{i-1}{2n}, \frac{i}{2n}\right] \times \left[\frac{2i-2}{2n}, \frac{2i-1}{2n}\right]$, note that its upper- and lower-right vertices lie in the region $\{(x,y) \in \mathbb{I}^2 \colon \frac{1}{2}y \leq x \leq 2y \text{ and } x + y < \frac{3}{2}\}$ and the upper- and lower-left vertices are in $\{(x,y) \in \mathbb{I}^2 \colon x \leq \frac{1}{2}y\}$. Therefore, one has

$$\begin{aligned}
&\mu(C_i \times R_{2i-1}) \\
&= \mu\left(\left[\frac{i-1}{2n}, \frac{i}{2n}\right] \times \left[\frac{2i-2}{2n}, \frac{2i-1}{2n}\right]\right) \\
&= C\left(\frac{i}{2n}, \frac{2i-1}{2n}\right) - C\left(\frac{i}{2n}, \frac{2i-2}{2n}\right) - C\left(\frac{i-1}{2n}, \frac{2i-1}{2n}\right) + C\left(\frac{i-1}{2n}, \frac{2i-2}{2n}\right) \\
&= \left[\frac{1}{3}\left(\frac{i}{2n}\right) + \frac{1}{3}\left(\frac{2i-1}{2n}\right)\right] - \left[\frac{1}{3}\left(\frac{i}{2n}\right) + \frac{1}{3}\left(\frac{2i-2}{2n}\right)\right] - \left(\frac{i-1}{2n}\right) + \left(\frac{i-1}{2n}\right) \\
&= \frac{1}{3}\left(\frac{2i-1-2i+2}{2n}\right) \\
&= \frac{1}{6n}.
\end{aligned}$$

Similarly, by the same argument, $\mu(C_i \times R_{2i}) = \frac{1}{6n}$. Thus, the value of $M_{ij}$ in this case is

$$\begin{aligned}
M_{ij} &= \log\left(\frac{\mu(C_i \times R_j)}{P(f_{1/3}(X) \in C_i)P(f_{2/3}(X) \in R_j)}\right)\mu(C_i \times R_j) \\
&= \frac{1}{6n}\log\left(\frac{1/6n}{1/4n^2}\right) \\
&= \frac{1}{6n}\log\left(\frac{2n}{3}\right).
\end{aligned}$$

To prove this proposition in part of $C_i \times R_j \in \mathcal{B}_2$, let $i = n+1, \ldots, 2n$ and $j = 3n-i+1$ and consider $C_i \times R_{3n-i+1} = \left[\frac{i-1}{2n}, \frac{i}{2n}\right] \times \left[\frac{3n-i}{2n}, \frac{3n-i+1}{2n}\right]$. Note that the upper- and lower-left vertices stay in $\{(x,y) \in \mathbb{I}^2 \colon \frac{1}{2}y \leq x \leq 2y \text{ and } x+y \geq \frac{3}{2}\}$ and the upper- and lower-right vertices lie in $\{(x,y) \in \mathbb{I}^2 \colon \frac{1}{2}y \leq x \leq 2y \text{ and } x+y \leq \frac{3}{2}\}$. Then, one obtains

$$
\begin{aligned}
&\mu(C_i \times R_{3n-i+1}) \\
&= \mu\left(\left[\frac{i-1}{2n}, \frac{i}{2n}\right] \times \left[\frac{3n-i}{2n}, \frac{3n-i+1}{2n}\right]\right) \\
&= C\left(\frac{i}{2n}, \frac{3n-i+1}{2n}\right) - C\left(\frac{i}{2n}, \frac{3n-i}{2n}\right) - C\left(\frac{i-1}{2n}, \frac{3n-i+1}{2n}\right) \\
&\quad + C\left(\frac{i-1}{2n}, \frac{3n-i}{2n}\right) \\
&= \left[\frac{i}{2n} + \frac{3n-i+1}{2n} - 1\right] - \left[\frac{i}{2n} + \frac{3n-i}{2n} - 1\right] - \left[\frac{1}{3}\left(\frac{i-1}{2n}\right) + \frac{1}{3}\left(\frac{3n-i+1}{2n}\right)\right] \\
&\quad + \left[\frac{1}{3}\left(\frac{i-1}{2n}\right) + \frac{1}{3}\left(\frac{3n-i}{2n}\right)\right] \\
&= \frac{1}{2n} - \frac{1}{6n} \\
&= \frac{1}{3n}.
\end{aligned}
$$

Thereby,

$$
\begin{aligned}
M_{ij} &= \log\left(\frac{\mu(C_i \times R_j)}{P(f_{1/3}(X) \in C_i)P(f_{2/3}(X) \in R_j)}\right)\mu(C_i \times R_j) \\
&= \frac{1}{3n}\log\left(\frac{1/3n}{1/4n^2}\right) \\
&= \frac{1}{3n}\log\left(\frac{4n}{3}\right).
\end{aligned}
$$

The last case, when $C_i \times R_j \in \mathcal{B}_3$, the proof is exactly the same to $C_i \times R_j \in \mathcal{B}_1$. now we are done. $\square$

Now we are ready for computing discretized mutual information since we know all of summands involving in this task.

$$
\begin{aligned}
I_G(f_{1/3}(X), f_{2/3}(X)) &= \sum_{i=1}^{2n}\sum_{j=1}^{2n} M_{ij} \\
&= \sum_{i=1}^{n} M_{i,2i} + \sum_{i=2}^{n} M_{i,2i-1} + \sum_{j=1}^{n} M_{2j,j} + \sum_{j=2}^{n} M_{2j-1,j} \\
&\quad + \sum_{i=n+1}^{2n} M_{i,3n-i+1} + \mu_{11} \\
&= 2(2n-1)\frac{1}{6n}\log\left(\frac{2n}{3}\right) + n\frac{1}{3n}\log\left(\frac{4n}{3}\right) + \frac{1}{3n}\log\left(\frac{4n}{3}\right) \\
&= \frac{1}{3}\log\left(\frac{4n}{3}\right) + \left(\frac{2}{3} - \frac{1}{3n}\right)\log\left(\frac{2n}{3}\right) + \frac{1}{3n}\log\left(\frac{4n}{3}\right) \\
&= \frac{1}{3}\log\left(\frac{4n}{3}\right) + \frac{2}{3}\log\left(\frac{2n}{3}\right) - \frac{1}{3n}\left(\log\left(\frac{2n}{3}\right) - \log\left(\frac{4n}{3}\right)\right).
\end{aligned}
$$

Thus, the normalized score takes value of

$$
\begin{aligned}
\frac{I_G(f_{1/3}(X), f_{2/3(X)})}{\log 2n} &= \frac{\frac{1}{3}\log\left(\frac{4n}{3}\right) + \frac{2}{3}\log\left(\frac{2n}{3}\right) - \frac{1}{3n}\left(\log\left(\frac{2n}{3}\right) \; - \log\left(\frac{4n}{3}\right)\right)}{\log 2n} \\
&= \frac{\frac{1}{3}\log\left(\frac{4n}{3}\right) + \frac{2}{3}\log\left(\frac{2n}{3}\right)}{\log 2n} - \frac{1}{3n}\frac{\log\left(\frac{2n}{3}\right) - \log\left(\frac{4n}{3}\right)}{\log 2n}.
\end{aligned}
$$

As $n \to \infty$, the second term goes to 0 and the first term, by L'Hôpital's rule, goes to one since both nominator and denominator converge to infinity and

$$
\lim_{n\to\infty} \frac{\frac{d}{dn}\frac{1}{3}\log\left(\frac{4n}{3}\right) + \frac{2}{3}\log\left(\frac{2n}{3}\right)}{\frac{d}{dn}\log 2n} = \lim_{n\to\infty} \frac{\frac{1}{3}\cdot\frac{3}{4n}\cdot\frac{4}{3} + \frac{2}{3}\cdot\frac{3}{2n}\cdot\frac{2}{3}}{\frac{2}{2n}} = \lim_{n\to\infty}\frac{1}{3} + \frac{2}{3} = 1.
$$

By this type of grid, normalized score becomes 1, so is the MIC. Note that these random variables $f_{2/3}(X)$ is not completely dependent on $f_{1/3}(X)$ since $\partial_1 C(u,v) = \frac{1}{3}$ whenever $(u,v) \in \{(x,y) \in \mathbb{I}^2 \colon \frac{1}{2}y < x \le 2y \text{ and } x + y < \frac{3}{2}\}$. Note that this set has 2-dimensional Lebesgue measure of $\frac{3}{8}$. This implies what we need. Also, since $\partial_2 C = \frac{1}{3}$ on the same set, then $f_{1/3}(X)$ is not completely dependent on $f_{2/3}(X)$.

# REFERENCES


[1] Chou, S.H., Nguyen, T.T.: On Fréchet theorem in the set of measure preserving functions over the unit interval, *Int. J. Math. Math. Sci.*, **13**(2), 373-378 (1990).

[2] Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, (2nd ed.), John Wiley & Sons, Inc. (2006).

[3] Darsow, W.F., Nguyen, B., Olsen, E.T.: Copulas and markov processes, *Illinois J. Math*, **36** (4), 600-642 (1992).

[4] Darsow, W.F., Olsen, E.T.: Norms for copulas, *Int. J. Math. Math. Sci.*, **18** (3), 417-436 (1995).

[5] Dette, H., Siburg, K.F., Stoimenov, P.A.: A copula-based non-parametric measure of regression dependence, *Scand. J. Stat.*, **40**(1), 21-41 (2013).

[6] Duranta, F., Sempi C.: *Principles of Copula Theory*, CRC Press. (2015).

[7] Kallenberg, O.: *Foundations of Modern Probability*, Springer-Verlag. (1997).

[8] Kamnitui, N., Santiwipanont, T., Sumetkijakan, S.: Dependence measuring from conditional variance, *Depend. Model.*, **3**, 98-112 (2015).

[9] Lancaster, H.O.: Correlation and complete dependence of random variables, *Ann. Math. Statist.*, **34**(4), 1315-1321 (1963).

[10] Li, H.: On nonsymmetric nonparametric measures of dependence, *arXiv preprint arXiv:1502.03850*, (2015).

[11] Olsen, E.T., Darsow, W.F., Nguyen, B.: Copulas and Markov operators, *Lecture Notes-Monograph Series*, **28** 244-259, (1996).

[12] Pearson, K., Heron, D.: On theories of association, *Biometrika*, **9**(1/2), 159-315 (1913).

[13] Rényi, A.: On measures of dependence, *Acta. Math. Acad. Sci. Hungar.*, **10**(3), 441-451 (1959).

[14] Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaygh, P.J., Lander, E.S., Mitzenmacher, M., Sabeti, P.C.: Detecting novel associations in large data sets, *Science*, **334**(6062), 1518–1524 (2011).

[15] Reshef, Y.A., Reshef, D.N., Finucane, H.K., Sabeti, P.C., Mitzenmacher, M.: Measuring dependence powerfully and equitably, *arXiv preprint arXiv: 1505.02213*, (2015).

[16] Ruankong, P., Santiwipanont, T., Sumetkijakan, S.: Shuffles of copulas and a new measure of dependence, *J. Math. Anal. Appl.*, **398**(1), 392-402 (2013).

[17] Rudin, W.: *Real and Complex Analysis*, McGraw-Hill International Edition. (1987).

[18] Schweizer, B., Wolff, E.F.: On nonparametric measures of dependence for random variables, *Ann. Statist.*, **9**, 879-885 (1981).

[19] Siburg, K.F., Stoimenov, P.A.: A measure of mutual conplete dependence, *Metrika*, **71**(2), 239-251 (2010).

[20] Trutschnig, W.: On a strong metric on the space of copulas and its induced dependence measure, *J. Math. Anal. Appl.*, **384**(2), 690-705 (2011).

# VITA

| | |
|---|---|
| **Name** | Mr.Atiwat Kitvanitphasu |
| **Date of Birth** | 26 July 1987 |
| **Place of Birth** | Bangkok, Thailand |
| **Education** | M.A. (Economics, English Language Program), |
| | Thammasat University, 2013 |
| | B.A. (Economics), |
| | Thammasat University, 2009 |
| **Scholarship** | Witthayabandhit Scholarship |
| | Bank of Thailand Scholarship |