Spring 2019

A Development of Transfer Entropy in Continuous-Time

Christopher David Edgar

A Development of Transfer Entropy in Continuous-Time

by

Christopher David Edgar

Bachelor of Science
University of South Carolina Aiken 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Mathematics

College of Arts and Sciences

University of South Carolina

2019

Accepted by:

Joshua Cooper, Major Professor

Hong Wang, Committee Member

Zhu Wang, Committee Member

Stephen Dilworth, Committee Member

Yan Tong, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Dedication

To my family.

Acknowledgments

Simply put, I owe everything that I have to my family. I am forever indebted to them for all they have done for me. Even though we had little, they all worked tirelessly to give me the best life that they could. They are the reason I am at this point in my life and they always loved and supported me even at my absolute worst. Selflessly, you gave me the utmost support throughout my pursuit of this goal. I am forever grateful for you all and love you more than anything.

I also want to thank some people who are not immediate family. First of all, to all of the members of Coop's Coop: Hays Whitlatch, Gregory Clark, Erin Hanna and Anton Swifton, I say thank you all so much for your support and friendship and I would have never made it through the challenge of qual's, comp's, and coursework if not for you all. Secondly, I am thankful for all of the friends I have made and thank the brilliant people I have had the privilege of working with over the past five years. A special thanks to all of my committee, Duncan Wright, Josh Grice, Shuai Yuan and especially Harsh Mehta; thanks for putting up with me for five years in that mold infested closet we call an office. I also thank Sharon Gregory, Gary Ricketts, Koffi Fadimba, David Jaspers, Reginald Koo, and Rao Li for inspiring me early in my life to study mathematics.

Last but certainly not least, I thank my advisor Joshua Cooper. You took a chance on me and gave me the opportunity to pursue this goal when you could have easily turned me away due to numerous reasons. Thank you for your patience, your belief in me, and the three and a half years you spent working with me even when I was out of ideas and unable to push forward. I came to grad school with two goals

in mind; I have achieved both of them thanks to you. I cannot thank you enough for all that you have done for me and I hope you are proud of the work we have done together.

Abstract

The quantification of causal relationships between time series data is a fundamental problem in fields including neuroscience, social networking, finance, and machine learning. Amongst the various means of measuring such relationships, information-theoretic approaches are a rapidly developing area in concert with other methods. One such approach is to make use of the notion of transfer entropy (TE). Broadly speaking, TE is an information-theoretic measure of information transfer between two stochastic processes. Schreiber's 2001 definition of TE characterizes information transfer as an informational divergence between conditional probability mass functions. The original definition is native to discrete-time stochastic processes whose comprising random variables have a discrete state space. While this formalism is applicable to a wealth of practical scenarios, there is a wide range of circumstances under which the processes of interest are indexed over an uncountable set (usually an interval). One can generalize Schreiber's definition to handle the case when the random variables comprising the processes have state space $\mathbb{R}$ via the Radon-Nikodym Theorem, as demonstrated by Kaiser and Schreiber in 2002. A rigorous treatment of TE among processes that are either indexed over an uncountable set or do not have $\mathbb{R}$ as the state space of their comprising random variables has been lacking in the literature. A common workaround to this theoretical deficiency is to discretize time to create new stochastic processes and then apply Schreiber's definition to these resulting processes. These time discretization workarounds have been widely used as a means to intuitively capture the notion of information transfer between processes in continuous-time, that is, those which are indexed by an interval. These ap-

proaches, while effective and practicable, do not provide a native definition of TE in continuous-time. We generalize Schreiber's definition to the case when the processes are comprised of random variables with a Polish state space and generalize further to the case when the indexing set is an interval via projective limits. Our main result, Theorem 5, is a rigorous recasting of a claim made by Spinney, Propenko, and Lizier in 2016, which characterizes when continuous-time TE can be obtained as a limit of discrete-time TE.

In many applications, the instantaneous transfer entropy or transfer entropy rate is of particular interest. Using our definitions, we define the transfer entropy rate as the right-hand derivative of the expected pathwise transfer entropy (EPT) defined in Section 2.3. To this end, we use our main results to prove some of its properties, including a rigorous version of a result stated without proof in work by Spinney, Propenko, and Lizier regarding a particularly well-behaved class of stationary processes. We then consider time-homogeneous Markov jump processes and provide an analytic form of the EPT via a Girsanov formula, and finally, using a corollary of our main result, we demonstrate how to apply our main result to a lagged Poisson point process, providing a concrete example of two processes to which our aforementioned results apply.

Table of Contents

List of Figures

Chapter 1

Introduction

Broadly speaking, information theory is the subfield of mathematics which deals with information and the fundamental limits of communication. The field came to be as a result of Claude Shannon's seminal paper *A Mathematical Theory of Communication* [47] which quantified information precisely and established limits on information transmission. A full review of information theory is beyond the scope of this thesis; however, we provide a concise summary of topics that will be encountered throughout this work in Section 1.2, then introduce the object of central focus in this thesis in Section 1.3, then we turn to a discussion on its applications in Section 1.4 and recent work in development of an estimator in Section 1.5. We conclude with a discussion of the organization of this manuscript. It should be noted that the majority of the information-theoretic frameworks used in this note make use of ideas from probability theory; thus, we preempt our discussion of information theory with a primer on probability theory in Section 1.1.

1.1 A review of probability

Definition 1.1.1. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a nonempty set, $\mathcal{F}$ is a nonempty $\sigma-$algebra of subsets of Ω and $\mathbb{P}$ is a *probability measure*, that is, a nonnegative countably additive set function mapping $\mathcal{F}$ into $[0, 1]$ such that $\mathbb{P}(\Omega) = 1$.

Remark 1. Elements of $\mathcal{F}$ are often called *events* and an element of $\mathcal{F}$ with $\mathbb{P}$

measure 0 is called a $\mathbb{P}$-*null set*.

Definition 1.1.2. Suppose Σ is some nonempty set and $\mathcal{X}$ is a $\sigma-$algebra of subsets of Σ. A function $X : \Omega \mapsto \Sigma$ is a *random variable (rv)* if $X : (\Omega, \mathcal{F}) \mapsto (\Sigma, \mathcal{X})$ is measurable, i.e.

$$X^{-1}(A) \in \mathcal{F}, \forall A \in \mathcal{X}.$$

We say that Σ is the *state space* of X. If $X(\Omega)$ is countable, then we say that X is a *discrete random variable*.

The following definition defines modes of convergence that will be of particular use in this thesis.

Definition 1.1.3. Suppose $(X_n)_{n\geq 1}$ and X are random variables.

1. We say $X_n \to X$ a.s. (*almost surely*) if there exists $\Omega' \in \mathcal{F}$ such that $\mathcal{P}(\Omega') = 1$ and
$$X_n(\omega) \to X_\omega, \text{ as } n \to \infty,$$
for all $\omega \in \Omega'$.

2. We say that X_n *converges to* X *in probability*, denoted $X_n \xrightarrow{\mathbb{P}} X$, if for each $\epsilon > 0$,
$$\mathbb{P}\left(\left|X_n - X\right| > \epsilon\right) \to 0, \text{ as } n \to \infty.$$

Definition 1.1.4. Suppose X is a rv mapping $\Omega \mapsto \Sigma$. A *realization* of X is an element x of Σ such that $X(\omega) = x$ for some $\omega \in \Omega$.

Definition 1.1.5. Suppose X is a rv. The *probability distribution* of X is the measure P_X on $(\Sigma, \mathcal{X})$ defined by

$$P_X(A) = \mathbb{P}(\{X \in A\}), \text{ for } A \in \mathcal{X}.$$

Furthermore, the $\sigma-$*algebra generated by* X, denoted $\sigma(X)$, is defined by

$$\sigma(X) = \left\{X^{-1}(A) \,\middle|\, A \in \mathcal{X}\right\}.$$

Remark 2. Note that $\sigma(X)$ is the smallest $\sigma-$ algebra for which X is measurable.

The next definition makes precise the concept of independence.

Definition 1.1.6. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

1. Events $A_i \in \mathcal{F}(1 \leq i \leq n)$ are *independent* if, for all subsets J of $[n]$,

$$\mathbb{P}\left(\cap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}\left(A_i\right)$$

2. Suppose $X_1, X_2, \ldots, X_n$ are random variables mapping into $(\Sigma, \mathcal{X})$. Then $X_1, X_2, \ldots, X_n$ are *independent* if

$$\mathbb{P}\left(\cap_{i=1}^{n} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} \mathbb{P}\left(\{X_i \in B_i\}\right),$$

where $B_i \in \mathcal{X}$. To this end, we define an infinite sequence of random variables, $(X_i)_{i \geq 1}$ to be *independent* if for all $n \geq 1$, $X_1, X_2, \ldots, X_n$ are independent.

Definition 1.1.7. Suppose $(\Sigma, \mathcal{X})$ is a measurable space. If μ and ν are two measures on $\mathcal{X}$, then μ is *absolutely continuous with respect to* ν if

$$\nu(A) = 0 \implies \mu(A) = 0, \forall A \in \mathcal{X}.$$

If ν is absolutely continuous with respect to μ, we denote it as $\mu \ll \nu$.

Definition 1.1.8. The *expected value* (or *expectation* or *mean*) of an integrable random variable X, denoted $\mathbb{E}_{\mathbb{P}}[X]$, is defined by

$$\mathbb{E}_{\mathbb{P}}[X] = \int_{\Omega} X d\mathbb{P}.$$

Definition 1.1.9. Suppose $(\Sigma, \mathcal{X})$ is a measurable space and μ is a measure on $(\Sigma, \mathcal{X})$. Then Σ is $\sigma-$*finite under* μ if Σ is the union of countably many subsets of Σ with finite measure under μ. If Σ is $\sigma-$finite under μ, then we say that the measure space $(\Sigma, \mathcal{X}, \mu)$ is a $\sigma-$*finite measure space.*

Clearly probability measures are $\sigma-$finite as they are finite measures in their own right. The notion of $\sigma-$finite measure spaces is of high regard in this work as it is necessary for the conclusion of the following theorem.

Theorem 1 (Radon-Nikodym Theorem)**.** Suppose $(\Sigma, \mathcal{X}, \mu)$ is a $\sigma-$finite measure space and let ν be a measure on $\mathcal{X}$ such that $\mu \ll \nu$. Then there exists a unique, measurable, nonnegative function $\frac{d\mu}{d\nu} : \Sigma \mapsto [0, \infty)$ up to $\nu-$null sets such that

$$\mu(A) = \int_A \frac{d\mu}{d\nu} d\nu,$$

$\forall A \in \mathcal{X}$. We say that $\frac{d\mu}{d\nu}$ is the *Radon-Nikodym (RN) derivative* of μ with respect to ν.

Definition 1.1.10. Suppose $\alpha \subset \mathbb{R}$ is an arbitrary indexing set. A *stochastic process*, $\{X_t\}_{t\in\alpha}$, is a collection of RV's taking values in a common measurable space $(\Sigma, \mathcal{X})$. If α is a countable set, then $\{X_t\}_{t\in\alpha}$ is a *discrete-time stochastic process.* If α is an interval, then $\{X_t\}_{t\in\alpha}$ is a *continuous-time stochastic process.* The functions $t \mapsto X_t(\omega)$ mapping α into Σ are called the *sample paths* of the process X.

If $(\Sigma, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then it is convention to call $\{X_t\}_{t\in\alpha}$ a *real-valued stochastic process.*

Definition 1.1.11. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and α is some indexing set.

1. A *filtration*, $(\mathcal{F}_t)_{t\in\alpha}$ is a nondecreasing collection of sub-σ-algebras of $\mathcal{F}$. We call the 4-tuple $\left(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in\alpha}\right)$ a *filtered probability space.*

2. A stochastic process $(X_t)_{t\in\alpha}$ is *adapted* to the filtration $(\mathcal{F}_t)_{t\in\alpha}$ if X_t is a $\mathcal{F}_t$-measurable rv for each $t \in \alpha$.

Definition 1.1.12. A family of integrable random variables $(X_t)_{t\in\alpha}$ is *uniformly integrable (UI)* if

$$\lim_{K\to\infty}\left(\sup_{t\in\alpha}\mathbb{E}_{\mathbb{P}}\left[X_t\mathbb{1}_{\{|X_t\geq K|\}}\right]\right)=0.$$

1.2 Entropy, Kullback-Leibler divergence, and mutual information

Definition 1.2.1. Suppose $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space. Define the *information* of an event $A\in\mathcal{F}$, denoted $I\left(A\right)$, by

$$I\left(A\right)=-\log\left(\mathbb{P}\left(A\right)\right)$$

Remark 3. We adhere to the convention that the base of the logarithm in Definition 1.2.1 is 2, thus entropy is measured in *bits*. We adhere to the convention of letting $0\log(0)=0$.

Definition 1.2.2. The *entropy* of a probability measure μ, denoted $H(\mu)$ is given by

$$H(\mu)=-\mathbb{E}_{\mathbb{P}}\left[\log\left(\mu\right)\right].$$

Furthermore, define the entropy of a rv X, denoted $H(X)$, as the entropy of its probability distribution P_X, i.e.

$$H(X)=-\mathbb{E}_{\mathbb{P}}\left[\log\left(P_X\right)\right].$$

If X is a discrete rv with range $\alpha(X)$ and distribution P_X, then

$$H(X)=-\mathbb{E}\left[\log\left(P_X\right)\right]=\sum_{x\in\alpha(X)}P_X(x)\log\left(\frac{1}{P_X(x)}\right).$$

Similarly, If $P_X\ll\mu$, where μ denotes Lebesgue measure, then from Theorem 1, there exists a probability density function $p_X:\Sigma\mapsto[0,\infty)$ such that

$$H(X)=-\mathbb{E}\left[\log\left(P_X\right)\right]=-\int_{\Sigma}p_X(x)\log\left(p_X(x)\right)d\mu(x).$$

Intuitively, the entropy of a distribution is the amount (usually measured in bits) of information gained upon observing an event drawn from the distribution.

Example 1.

Suppose $X \sim Bernoulli\left(\frac{1}{2}\right)$ and $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$, where $\mu, \sigma > 0$. Then

$$H(X) = -\mathbb{E}\left[\log\left(P_X\right)\right] = \sum_{x \in \{0,1\}} P_X(x) \log\left(\frac{1}{P_X(x)}\right) = \log 2$$

and

$$\begin{aligned} H(Y) = -\mathbb{E}\left[\log\left(P_Y\right)\right] &= -\int_{\mathbb{R}} p_Y(y) \log\left(p_Y(y)\right) d\mu(y) \\ &= -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) \\ &= \frac{1}{2} \ln\left(2\pi\sigma^2\right) + \frac{1}{2} \\ &= \frac{1}{2} \ln\left(2\pi e \sigma^2\right). \end{aligned}$$

Definition 1.2.3. Suppose P and M are measures on a space $(\Sigma, \mathcal{X})$ with $P \ll M$. The *Kullback-Leibler divergence*, or simply, the *KL-divergence*, of M from P, denoted $KL\left(P||M\right)$, is defined by

$$KL\left(P||M\right) = \mathbb{E}_P\left[\log\left(\frac{dP}{dM}\right)\right].$$

Remark 4. Intuitively, the KL - divergence of one measure to another is the informational distance between the two measures. A source coding interpretation of KL-divergence is as follows: $KL\left(P||M\right)$ is the average number of additional bits needed to encode a sample, assuming it's drawn from distribution M instead of P.

KL-divergence does not satisfy all of the axioms of a metric; thus, it is labeled as a divergence. Specifically, it is not symmetric: In general, the KL-divergence from M to P is not necessarily the same as that of P to M. Typically, P represents a *ground truth* distribution and M represents some approximation of P. For example, in supervised machine learning, P may represent an empirical distribution of observed data and M may represent a distribution imposed on the data via a model, in which case one typically opts to minimize $KL\left(P||M\right)$ with respect to model parameters.

As a consequence of Jensen's inequality,

$$KL\left(P||M\right) \geq 0 \text{ with } KL\left(P||M\right) = 0 \iff P = M,$$

and is explicitly assymetric, that is, $KL\left(P||M\right) \neq KL\left(M||P\right)$ in general. Note that if Σ in Definition 1.2.3 is a discrete space, then

$$KL\left(P||M\right) = \sum_{x\in\Sigma} P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

and if $P \ll \mu$ and $M \ll \mu$, where μ denotes Lebesgue measure on $\mathbb{R}$, then

$$KL\left(P||M\right) = \int_{\Sigma} \frac{dP}{d\mu}\log\left(\frac{\frac{dP}{d\mu}}{\frac{dM}{d\mu}}\right) dP.$$

It should also be noted that $KL\left(P||M\right) = \infty$ in the case that P is not absolutely continuous with respect to M.

Example 2. Suppose $0 < p, q < 1$. Then

$$KL\left(Bernoulli\left(p\right)\middle|\middle|Bernoulli\left(q\right)\right) = p\log\left(\frac{p}{q}\right) + (1-p)\log\left(\frac{1-p}{1-q}\right).$$

The following definition makes rigorous the concept of *shared* information between random variables.

Definition 1.2.4. Suppose $(\Sigma_X, \mathcal{X})$ and $(\Sigma_Y, \mathcal{Y})$ are measurable spaces and that X and Y are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with state spaces Σ_X and Σ_Y, respectively. The *mutual information* between X and Y, denoted $I\left(X, Y\right)$, is defined by

$$I\left(X, Y\right) = KL\left(P_{XY}||P_X \times P_Y\right),$$

where P_X and P_Y are the marginal distributions of X and Y, respectively and $P_X \times P_Y$ is the product measure on $(\Sigma_X \times \Sigma_Y, \mathcal{X} \bigotimes \mathcal{Y})$

Remark 5. Put simply, mutual information between two variables is the average amount of information gained in observing the state of one variable, given full information of the other. Unlike KL-divergence, mutual information is symmetric, as is evident by the definition. Note that if X and Y are independent RV's, then $P_{XY} = P_X \times P_Y$, thus Definition 1.2.3 implies that

$$I\left(X, Y\right) \geq 0,$$

with equality if and only if X and Y are independent and that $H(X) = I(X, X)$. Also, if Σ_X and Σ_Y are discrete, then

$$I(X,Y) = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} P_{X,Y}(x,y) \log \left(\frac{P_{X,Y}(x,y)}{P_X(x) P_Y(y)} \right).$$

1.3 Transfer entropy

The notion of quantifying the transfer of information from one stochastic process to another is a fundamental question in many fields. Mutual information is void of directionality; thus, it is insufficient to quantify information flow. Some efforts to address this limitation include the use of time-lagged mutual information, in which mutual information is calculated between a time-lagged version of the source process and the destination process. This approach remedies the directional deficiency of mutual information; however, it is still deficient as it incorporates the measurement of statically shared information between the processes or a common history imposed by an exogenous force as pointed out in [45]. In this vein, Schreiber proposed a new measure of information transfer which overcomes some limitations of former approaches. Broadly speaking, Schreiber quantified information transfer as a KL-divergence amongst conditional probabilities. The resulting quantity is called transfer entropy and is defined as the following:

Definition 1.3.1. Suppose X and Y are discrete time stochastic processes composed of discrete RV's each with state space $(\Sigma, \mathcal{X})$ and $k, l \geq 1$ are integers. The *transfer entropy from Y to X at n with history window lengths k and l*, denoted $\mathbb{T}^{(k,l)}_{Y \to X}(n)$, is given by

$$\mathbb{T}^{(k,l)}_{Y \to X}(n) =$$

$$\sum_{\substack{x_n \in X_n(\Omega), \\ x^{n-1}_{n-k-1} \in x^{n-1}_{n-k-1}(\Omega), \\ y^{n-1}_{n-l-1} \in Y^{n-1}_{n-l-1}(\Omega)}} P_1\left(x_n, \left(x^{n-1}_{n-k-1}\right), \left(y^{n-1}_{n-l-1}\right)\right) \log \frac{P_2\left(x_n \mid \left(x^{n-1}_{n-k-1}\right), \left(y^{n-1}_{n-l-1}\right)\right)}{P_3\left(x_n \mid \left(x^{n-1}_{n-k-1}\right)\right)}, \quad (1.1)$$

where $x_n \in \Sigma$, $\left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right) \in \Sigma^{k+1}$,

$$P_1\left(x_n, \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right) = \mathbb{P}_{X_n, \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)}\left(x_n, \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right),$$

$$P_2\left(x_n, \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right) = \mathbb{P}_{X_n|(X_{n-k-1}^{n-1}), (Y_{n-l-1}^{n-1})}\left[x_n \middle| \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right],$$

and

$$P_3\left(x_n, \left(x_{n-k-1}^{n-1}\right)\right) = \mathbb{P}_{X_n|\left(X_{n-k-1}^{n-1}\right)}\left[x_n \middle| \left(x_{n-k-1}^{n-1}\right)\right].$$

Y is called the *source process* and X is called the *destination process*.

Although Definition 1.3.1 may seem complicated, the underlying idea is relatively straightforward. $\mathbb{T}_{Y \to X}^{(k,l)}(n)$ is simply a measure of the average reduction of uncertainty in the present value of X (i.e. X_n) due to knowledge of the past of Y $\left(\text{i.e. } \left(Y_{n-l-1}^{n-1}\right)\right)$ given knowledge of the past of X $\left(\text{i.e. } \left(X_{n-k-1}^{n-1}\right)\right)$. Even more simply, TE measures the average amount of information the past of the source process provides about the present of the destination process that the past of the destination process does not already provide. Figure 1.1 illustrates the scheme underlying TE.

Note that this formalism overcomes the limitations of using mutual information to quantify information transfer in former approaches as it takes into account *only* dependencies due to the source process. Transfer entropy can also be characterized as an instance of KL-divergence, as (1.3.1) can be written as a KL-divergence of the conditional probability measure $\mathbb{P}_{X_n|\left(X_{n-k-1}^{n-1}\right)}\left[x_n \middle| \left(x_{n-k-1}^{n-1}\right)\right]$ from $\mathbb{P}_{X_n|(X_{n-k-1}^{n-1}), (Y_{n-l-1}^{n-1})}\left[x_n \middle| \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right]$, that is, the deviation from the assumption that X_n is independent of (Y_{n-l-1}^{n-1}) given (X_{n-k-1}^{n-1}). More succinctly,

$$\mathbb{T}_{Y \to X}^{(k,l)}(n) = KL\left(P_2\left(x_n, (x_{n-k-1}^{n-1}), (y_{n-l-1}^{n-1})\right) \middle\| P_3\left(x_n, (x_{n-k-1}^{n-1})\right)\right),$$

implying that transfer entropy indeed quantifies information flow as it is explicitly asymmetric, that is $\mathbb{T}_{Y \to X}^{(k,l)}(n) \neq \mathbb{T}_{X \to Y}^{(k,l)}(n)$ in general, and is nonnegative for all processes and choices of history window lengths.

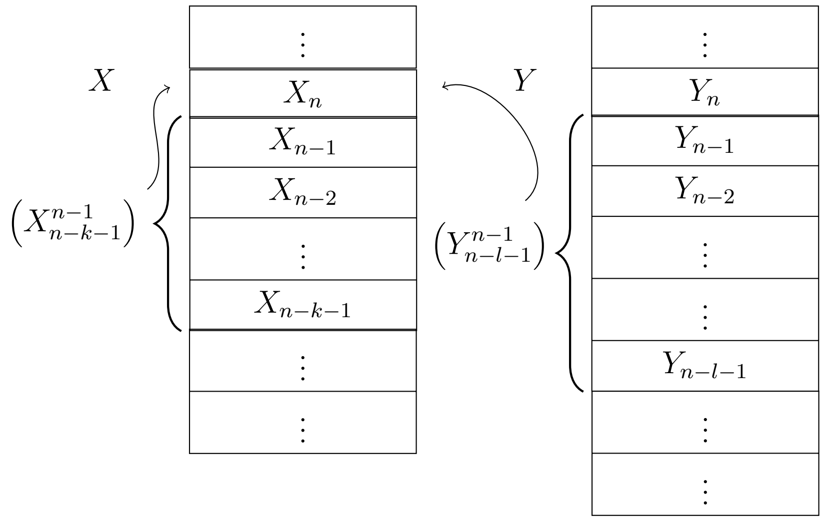

Figure 1.1: A schematic of transfer entropy from process Y to process X.

Observation 1. If indeed X_n is independent of (y_{n-l-1}^{n-1}) given (X_{n-k-1}^{n-1}), then

$$\mathbb{T}_{Y \to X}^{(k,l)}(n) = 0.$$

Definition 1.3.1 can be modified to handle the case when the RV's composing the processes in question are not discrete as demonstrated in [26]. In this case, the expression in (1.1) becomes an integral of a function of probability density functions as opposed to a sum of a function of probability mass functions.

Remark 6. Information transfer has also been measured by Granger causality [17], a predictive statistical tool which makes use of vector autoregression for prediction[1]. Broadly speaking, Y is Granger causal to X whenever X is better predicted using a model that uses both X's and Y's history, than from one that includes exclusively X's history.

1.4 Applications

TE improves upon inadequacies of former approaches of measuring information transfer and has properties that make its application more suitable for some scenarios over

[1]TE and Granger Causality have been shown to be equivalent up to a factor of 2 in Gaussian Processes as shown in [6]

other frameworks in this realm. The definition makes no assumptions about neither the destination nor source processes, making it an appealing tool for scenarios in which quantifying information transfer is of high regard. There is an abundance of applications of TE in the literature, and we provide a short survey of its uses in a wide-ranging spectrum of domains.

Biological research often requires analyses of data that can be riddled with complicated dependencies and noise. TE has been used in a wide variety of topics in this field. In [10], TE is used to detect and quantify dynamics in animal groups. Among other applications in this field, TE has been used to detect leadership among groups of bats through analysis of trajectory paths [38], infer gene regulatory networks [50], and has even been used in epidemiology [5] and cardiology [34, 35, 42]. Among biological subfields, neuroscience has a vast number of applications of TE in the literature, so much so that there exists literature solely devoted to surveying its application in the field [52]. TE has been used on functional MRI, electroencephalography and magnetoencephalography datasets to provide meaningful insights in neural connectivity [24, 36], localization of information storage [54], encoding relationships among neurons [7, 49], and time scale effects on the frequency content of visual stimuli [8]. There exists an open-source toolbox supporting both the functionality for management of datasets common to this field, including those mentioned previously, and the application of TE to such datasets in an efficient manner [33].

In addition to biology, information transfer is of particular relevance in finance as well. TE has a wide-range of applications in this field. For example, in [44], TE is used on stock data from various companies chosen via the S&P 1200 global index with sufficient liquidity to measure intra-sector influence amongst companies. It is used in [29] to analyze indices of stock markets in numerous countries to measure geographic influence of stock indices. Similarly, Sensoy [46] and Sandoval [44] apply TE to analyze exchange rates of stock using various stock data. In [12], the effect of credit risk on market risk is analyzed via TE using iTraxxx and VIX data [1].

Recently, TE has been used in social media analysis, to measure various aspects of influence in social media. In [51], TE is used on a Twitter dataset to measure influences between pairs of users in a small user network on the basis of *tweet* content. The authors found that high TE is a significant predictor of *mentions* on the platform. Saike He et al.[20] use TE to quantify peer influence in online social networks in which part of user activity is internally generated. TE has also been used in machine learning to improve performance of recurrent neural networks in [21] and [37].

1.5 Continuous-time TE, binning and estimators

Schreiber's definition of TE is one of broad utility. However, it suffers from a major theoretical deficiency: specifically, it is only applicable to data with a discretized time basis. In the literature, the most popular approach used as a means of compensation of this limitation is time-discretization [22, 31, 32]. This approach initially discretizes the continuous-time processes under consideration and then uses Schreiber's definition on the resulting processes. This approach, however, has in some cases been shown to suffer from erroneous convergence results as demonstrated in [3] and appears ill-equipped to manage key mechanisms responsible for information transfer as it cannot detect interactions below the resolution of the imposed discretization [48]. However, a large portion of applications using this method to estimate TE on continuous-time data have obtained satisfactory results, e.g., in [15, 22, 51]. This suggests that discrete treatments of continuous-time stochastic processes is worthwhile in some contexts to estimate TE.

Entropy estimation is a fundamental problem in information theory and applied statistics. It has been shown in [40] that there is no unbiased estimator of entropy,

[1] iTraxx and VIX are types of indices of stock.

and most plug-in estimators suffer from underestimation issues as demonstrated in [19]. One widely used estimator of entropy is due to Kozachencko and Leonenko [28], in which an approximate nearest neighbors search is utilized to estimate entropy. In recent developments, efforts have been made to modify the KL estimator to estimate TE and mutual information [57].

1.6 Organization of this thesis

In Chapter 2, we generalize Schreiber's definition of TE to the case when the state space of the destination and source process is an arbitrary Polish space that is meaningful even when probability densities (or probability mass functions) are either intractable or nonexistent (this can occur in lieu of absolute continuity between probability distributions and Lebesgue measure.). Motivated by the Radon-Nikodym Theorem and regular conditional probability measures, we define TE in such a context as an expected KL-divergence between conditional measures. We then generalize further and address TE in a continuous-time setting, that is, when the indexing set is an interval as opposed to a countable set. We develop measures over an appropriate measurable space for our continuous-time framework, and justify their existence via a seminal result due to [43] regarding projective limits of projective systems of probability spaces. Furthermore, in Section 2.3, we use these measures to define the pathwise transfer entropy (PT) and define expected pathwise transfer entropy (EPT); the latter is the continuous-time version of TE measuring information transfer over an interval. After comparing our definitions with those presented in the current literature and defining a type of necessary consistency, we prove our main result, Theorem 5. Theorem 5 establishes necessary and sufficient conditions for the attainability of our continuous-time definition of TE as a limit of discrete time TE; that is, when a discrete treatment of continuous-time processes recovers our continuous-time definition. We conclude the section with some consequences of our main result.

In Chapter 3, we define the transfer entropy rate (or instantaneous transfer entropy) as a right hand derivative of the EPT and prove some of its properties, especially those relevant to the case when the destination and source processes possess a stationarity property. We conclude the section with sufficient conditions for continuity of PT and EPT.

In Chapter 4, we consider time-homogeneous Markov jump processes (THMJP). We define conditional transition and escape rates as a limit of conditional measures and provide an expression of the EPT via a Girsanov formula in terms of these rates. In this vein, we consider the case when the source process is a thinned version of the destination process and provide an expression for the EPT in this context. We conclude with an application of Corollary 5.2, which permits the use of our main result to a time-lagged Poisson point process.

Chapter 2

Continuous-time transfer entropy

2.1 Discrete-time TE generalization

In this section, we present a definition of TE between processes whose comprising random variables have an arbitrary Polish state space. This definition is a generalization of Definition 2.4 and indeed recovers the original definition under suitable conditions as demonstrated in Example 3.

Suppose $X := \{X_n\}_{n\geq 1}$ and $Y := \{Y_n\}_{n\geq 1}$ are stochastic processes adapted to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 1}, \mathbb{P})$ such that for each $n \geq 1$, X_n and Y_n are random variables taking values in a Polish state space Σ, that is, a completely metrizable, separable space and let $\mathcal{X}$ be a σ-algebra of subsets of Σ. Let $\mathbb{P}_n$ denote the probability distribution of the random variable X_n (Sometimes we will mean by this a conditional probability distribution.)

For integers $k, l, n \geq 1$, let

$$\left(X_{n-k-1}^{n-1}\right) = (X_{n-k-1}, X_{n-k}, ..., X_{n-1})$$

and

$$\left(Y_{n-l-1}^{n-1}\right) = (Y_{n-l-1}, Y_{n-l}, ..., Y_{n-1})\,.$$

Since Σ is Polish, for each $k, l, n \geq 1$, there exist functions, often called regular conditional probability measures[1], $\mathbb{P}_n^{(k,l)}\left[X_n \middle| \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)\right]$ and $\mathbb{P}_n^{(k)}\left[X_n \middle| \left(X_{n-k-1}^{n-1}\right)\right]$ mapping $\mathcal{F}_n \times \Omega$ into $[0, 1]$ with the following properties:

[1]The existence of regular conditional probability measures is guaranteed on Polish spaces (see Theorem 6.16 of [39])

1. For each $\omega \in \Omega$,

$$\mathbb{P}_n^{(k)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right) \right] (\cdot, \omega) \tag{2.1}$$

and

$$\mathbb{P}_n^{(k,l)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right), \left(Y_{n-l-1}^{n-1} \right) \right] (\cdot, \omega) \tag{2.2}$$

are measures on $(\Sigma, \mathcal{X})$

2. $\forall A \in \mathcal{F}_n$, the mappings

$$\omega \mapsto \mathbb{P}_n^{(k,l)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right), \left(Y_{n-l-1}^{n-1} \right) \right] (A, \omega)$$

and

$$\omega \mapsto \mathbb{P}_n^{(k,l)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right) \right] (A, \omega)$$

are $\mathcal{F}_n$-measurable random variables.

3. $\forall \omega \in \Omega, A \in \mathcal{F}_n$, we have both

$$\mathbb{P}_n^{(k,l)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right), \left(Y_{n-l-1}^{n-1} \right) \right] (A, \omega) =$$
$$\mathbb{P}_n^{(k,l)} \left[\{ X_n \in A \} \,\middle|\, \left\{ B \in \sigma \left(\left(X_{n-k-1}^{n-1} \right), \left(Y_{n-l-1}^{n-1} \right) \right) \,\middle|\, \omega \in B \right\} \right].$$

and

$$\mathbb{P}_n^{(k)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right) \right] (A, \omega) =$$
$$\mathbb{P}_n^{(k)} \left[\{ X_n \in A \} \,\middle|\, \left\{ B \in \sigma \left(\left(X_{n-k-1}^{n-1} \right) \right) \middle| \, \omega \in B \right\} \right].$$

To this end, the conditional probabilities

$$\mathbb{P}_n^{(k,l)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right), \left(Y_{n-l-1}^{n-1} \right) \right] (\cdot, \omega)$$

and

$$\mathbb{P}_n^{(k)} \left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1} \right) \right] (\cdot, \omega)$$

are only defined if both $\left\{ B \in \sigma \left(\left(X_{n-k-1}^{n-1} \right) \right) \middle| \, \omega \in B \right\}$ and $\left\{ B \in \sigma \left(\left(X_{n-k-1}^{n-1} \right), \left(Y_{n-l-1}^{n-1} \right) \right) \,\middle|\, \omega \in B \right\}$ are not $\mathbb{P}$-null sets. We will assume this throughout this work whenever dealing with conditional probabilities.

Notation 1. For sake of convenience, we let

$$\mathbb{P}_n^{(k)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right)\right](A,\omega) = \mathbb{P}_n^{(k)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right)\right](\omega)(A)$$

and

$$\mathbb{P}_n^{(k,l)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)\right](A,\omega) = \mathbb{P}_n^{(k,l)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)\right](\omega)(A),$$

whenever $n, k, l \geq 1, \omega \in \Omega$, and $A \in \mathcal{F}_n$.

The following definition generalizes Schreiber's definition of TE for discrete-time processes.

Definition 2.1.1. Suppose $n, k, l \geq 1$ are integers. Suppose further that Σ is a Polish space and that

$$\mathbb{P}_n^{(k)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)\right](\omega) \ll \mathbb{P}_n^{(k)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right)\right](\omega), \tag{2.3}$$

for each $\omega \in \Omega$. The *transfer entropy from Y to X at n with history window lengths k and l*, denoted $\mathbb{T}_{Y\to X}^{(k,l)}(n)$, is defined by

$$\begin{aligned}\mathbb{T}_{Y\to X}^{(k,l)}(n) &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}_n^{(k,l)}\left[X_n | \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)\right]}\left[\log \frac{d\mathbb{P}_n^{(k,l)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)\right]}{d\mathbb{P}_n^{(k)}\left[X_n \,\middle|\, \left(X_{n-k-1}^{n-1}\right)\right]}\right]\right] \\ &= \int_{\Omega}\left(KL\left(\mathbb{P}_n^{(k,l)}[X_n \,\middle|\, (X_{n-k-1}^{n-1}), (Y_{n-l-1}^{n-1})](\cdot) \,\middle|\middle|\, \mathbb{P}_n^{(k)}[X_n \,\middle|\, (X_{n-k-1}^{n-1})](\cdot)\right)\right) d\mathbb{P}.\end{aligned} \tag{2.4}$$

As in Definition 1.1, we call X the *destination process* and Y the *source process.*

Observation 2. Due to [56], we have for each $n \geq 1$ the following:

1. For fixed $k, l \geq 1$, $\mathbb{T}_{Y\to X}^{(k,l)}$ is a measurable function from $\mathbb{N}$ into the extended nonnegative real line.

2. $KL\left(\mathbb{P}_n^{(k,l)}[X_n \,\middle|\, (X_{n-k-1}^{n-1}), (Y_{n-l-1}^{n-1})](\omega) \,\middle|\middle|\, \mathbb{P}_n^{(k)}[X_n \,\middle|\, (X_{n-k-1}^{n-1})](\omega)\right) \geq 0$, for each $\omega \in \Omega$.

3. $\frac{d\mathbb{P}_n^{(k,l)}[X_n\,|\,(X_{n-k-1}^{n-1}),(Y_{n-l-1}^{n-1})](\cdot)}{d\mathbb{P}_n^{(k)}[X_n\,|\,(X_{n-k-1}^{n-1})](\cdot)}(\cdot)$ is $\mathcal{F}\times\mathcal{X}$-measurable as X is adapted to $\mathcal{F}$.

4. $KL\left(\mathbb{P}_n^{(k,l)}[X_n\Big|(X_{n-k-1}^{n-1}),(Y_{n-l-1}^{n-1})](\omega)\ \Big|\Big|\ \mathbb{P}_n^{(k)}[X_n\Big|(X_{n-k-1}^{n-1})](\omega)\right)$ is $\mathcal{F}-$ measurable for each $\omega\in\Omega$.

Example 3. Suppose X and Y are discrete processes, that is, for each $n\geq 1$, both $X_n(\Omega)$ and $Y_n(\Omega)$ are countable. Then

$$\mathbb{T}_{Y\to X}^{(k,l)}(n)=$$

$$\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}_n^{(k,l)}[X_n\,|\,(X_{n-k-1}^{n-1}),(Y_{n-l-1}^{n-1})]}\left[\log\frac{d\mathbb{P}_n^{(k,l)}\left[X_n\Big|\left(X_{n-k-1}^{n-1}\right),\left(Y_{n-l-1}^{n-1}\right)\right]}{d\mathbb{P}_n^{(k)}\left[X_n\Big|\left(X_{n-k-1}^{n-1}\right)\right]}\right]\right]$$

$$=\sum_{\substack{x_{n-k-1}^{n-1}\in X_{n-k-1}^{n-1}(\Omega)\\ y_{n-l-1}^{n-1}\in Y_{n-l-1}^{n-1}(\Omega)}}\mathbb{P}_{X_{n-k-1}^{n-1},Y_{n-l-1}^{n-1}}\left(x_{n-k-1}^{n-1},y_{n-l-1}^{n-1}\right)\times$$

$$\sum_{x_n\in X_n(\Omega)}\mathbb{P}_{X_n|\left(X_{n-k-1}^{n-1}\right),\left(Y_{n-l-1}^{n-1}\right)}\left[x_n\Big|\left(x_{n-k-1}^{n-1}\right),\left(y_{n-l-1}^{n-1}\right)\right]\times$$

$$\log\frac{\mathbb{P}_{X_n|\left(X_{n-k-1}^{n-1}\right),\left(Y_{n-l-1}^{n-1}\right)}\left[x_n\Big|(x_{n-k-1}^{n-1}),\left(y_{n-l-1}^{n-1}\right)\right]}{\mathbb{P}_{X_n|\left(X_{n-k-1}^{n-1}\right)}\left[x_n\Big|\left(x_{n-k-1}^{n-1}\right)\right]}$$

$$=\sum_{\substack{x_n\in X_n(\Omega),\\ x_{n-k-1}^{n-1}\in x_{n-k-1}^{n-1}(\Omega),\\ y_{n-l-1}^{n-1}\in Y_{n-l-1}^{n-1}(\Omega)}}\mathbb{P}_{X_n,X_{n-k-1}^{n-1},Y_{n-l-1}^{n-1}}\left(x_n,x_{n-k-1}^{n-1},y_{n-l-1}^{n-1}\right)\times$$

$$\log\frac{\mathbb{P}_{X_n|\left(X_{n-k-1}^{n-1}\right),\left(Y_{n-l-1}^{n-1}\right)}\left[x_n\Big|\left(x_{n-k-1}^{n-1}\right),\left(y_{n-l-1}^{n-1}\right)\right]}{\mathbb{P}_{X_n|\left(X_{n-k-1}^{n-1}\right)}\left[x_n\Big|\left(x_{n-k-1}^{n-1}\right)\right]},$$

where the RN-derivatives have become quotients of *probability mass functions* since the processes are discrete. The above demonstrates that Schreiber's initial definition of transfer entropy is indeed a special case of our more general definition of TE. Furthermore, if $(\Sigma,\mathcal{X})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$ and the joint probability measure $\mathbb{P}_{X_n,\left(X_{n-k-1}^{n-1}\right),\left(Y_{n-l-1}^{n-1}\right)}$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^{(1+k+l)}$, then there exist RN-derivatives (*probability densities*)

$$p_{X_n,\left(X_{n-k-1}^{n-1}\right),\left(Y_{n-l-1}^{n-1}\right)},p_{X_n|\left(X_{n-k-1}^{n-1}\right),\left(Y_{n-l-1}^{n-1}\right)}\text{ and }p_{X_n|\left(X_{n-k-1}^{n-1}\right)},\tag{2.5}$$

that act as the probability mass functions in Definition 1.1. In regards to our definition in this setting, $\mathbb{R}$ is indeed Polish, thus assuming (2.3) we can apply our definition and expanding the expression in (2.4) yields

$$\begin{aligned}
&\mathbb{T}^{(k,l)}_{Y\to X}(n) \\
&= \int_{\mathbb{R}^{(1+k+l)}} p_{X_n,\left(X^{n-1}_{n-k-1}\right),\left(Y^{n-1}_{n-l-1}\right)}\left(x_n,\left(x^{n-1}_{n-k-1}\right),\left(y^{n-1}_{n-l-1}\right)\right)\times \\
&\qquad \log\left(\frac{p_{X_n\mid\left(X^{n-1}_{n-k-1}\right),\left(Y^{n-1}_{n-l-1}\right)}\left(x_n\mid\left(x^{n-1}_{n-k-1}\right),\left(y^{n-1}_{n-l-1}\right)\right)}{p_{X_n\mid\left(X^{n-1}_{n-k-1}\right)}\left(x_n\mid\left(x^{n-1}_{n-k-1}\right)\right)}\right)d\mu_{(1+k+l)},
\end{aligned}$$

where $\mu_{(1+k+l)}$ denotes Lebesgue measure on $\mathbb{R}^{(1+k+l)}$. This expression is exactly that for TE in this case as presented in [26], thus our definition recovers the correct expression for TE in the case that $(\Sigma,\mathcal{X})=(\mathbb{R},\mathcal{B}(\mathbb{R}))$ as well.

Note that the definition for $\mathbb{T}^{(k,l)}_{Y\to X}(n)$ in Definition 1.1 is different than Definition 1 of [48]. According to [48],

$$\begin{aligned}
\mathbb{T}^{(k,l),SPL}_{Y\to X}(n) := \mathbb{T}^{(k,l)}_{Y\to X}(n) &= \mathbb{E}_{\mathbb{P}}\left[\log\frac{d\mathbb{P}^{(k,l)}_n[X_n\mid(X^{n-1}_{n-k-1}),(Y^{n-1}_{n-l-1})]}{d\mathbb{P}^{(k)}_n[X_n\mid(X^{n-1}_{n-k-1})]}(\omega)\right] \\
&= \int_\Omega \log\frac{d\mathbb{P}^{(k,l)}_n[X_n\mid(X^{n-1}_{n-k-1}),(Y^{n-1}_{n-l-1})]}{d\mathbb{P}^{(k)}_n[X_n\mid(X^{n-1}_{n-k-1})]}(\omega)d\mathbb{P}(\omega).
\end{aligned} \tag{2.6}$$

This is an ambiguous expression as presented. Note that the RN-derivative,

$$\frac{d\mathbb{P}^{(k,l)}_n[X_n\mid(X^{n-1}_{n-k-1}),(Y^{n-1}_{n-l-1})]}{d\mathbb{P}^{(k)}_n[X_n\mid(X^{n-1}_{n-k-1})]},$$

is not a function of Ω, but is rather by definition a function mapping Σ into $\mathbb{R}_{\geq 0}$, thus the treatment of ω in (2.6) is inconsistent with the definition of the RN-derivative. Furthermore, the conditional measures

$$\mathbb{P}^{(k,l)}_n[X_n\mid(X^{n-1}_{n-k-1}),(Y^{n-1}_{n-l-1})]$$

and

$$\mathbb{P}^{(k)}_n[X_n\mid(X^{n-1}_{n-k-1})]$$

are random measures, thus $\omega \in \Omega$ should be fixed *a priori* before treating them as measures on $(\Sigma, \mathcal{X})$. One could, in principle, interpret the RHS of (2.6) as

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}}\Big[\log \frac{d\mathbb{P}_n^{(k,l)}[X_n \mid (X_{n-k-1}^{n-1}), (Y_{n-l-1}^{n-1})]}{d\mathbb{P}_n^{(k)}[X_n \mid (X_{n-k-1}^{n-1})]}(\omega)\Big] = \\ &\int_\Omega \log \frac{d\mathbb{P}_n^{(k,l)}[X_n \mid (X_{n-k-1}^{n-1}), (Y_{n-l-1}^{n-1})](\omega)}{d\mathbb{P}_n^{(k)}[X_n \mid (X_{n-k-1}^{n-1})](\omega)}(X_n(\omega))d\mathbb{P}. \end{aligned} \tag{2.7}$$

This expression has meaning; however, we can not recover Schreiber's definition [45]. The fundamental issue is that the RN-derivative in the integrand is only dependent on a single ω, which as written, is evaluated at the same realization of X_n that corresponds to the sample point which generates the conditional distributions that define the RN-derivative itself. Thus when one takes the integral over Ω, one does not capture the entire conditional distribution of X_n given the events in $\sigma\left(\left(X_{n-k-1}^{n-1}\right)\right)$ and $\sigma\left(\left(Y_{n-l-1}^{n-1}\right)\right)$ that contain the sample point. For example, if X and Y are discrete random processes, we get

$$\begin{aligned} &\int_\Omega \log \frac{d\mathbb{P}_n^{(k,l)}[X_n \mid (X_{n-k-1}^{n-1}), (Y_{n-l-1}^{n-1})](\omega)}{d\mathbb{P}_n^{(k)}[X_n \mid (X_{n-k-1}^{n-1})](\omega)}(X_n(\omega))d\mathbb{P} \\ &= \int_\Omega \log \frac{\mathbb{P}_n^{(k,l)}\left[X_n = X_n(\omega) \mid \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)\right](\omega)}{\mathbb{P}_n^{(k)}\left[X_n = X_n(\omega) \mid \left(X_{n-k-1}^{n-1}\right)\right](\omega)} d\mathbb{P} \\ &= \sum_{\substack{x_{n-k-1}^{n-1} \in X_{n-k-1}^{n-1}(\Omega) \\ y_{n-l-1}^{n-1} \in Y_{n-l-1}^{n-1}(\Omega)}} \mathbb{P}_{X_{n-k-1}^{n-1}, Y_{n-l-1}^{n-1}}(x_{n-k-1}^{n-1}, y_{n-l-1}^{n-1}) \times \\ &\qquad \sum_{x_n \in X_n(\Omega)} \log \frac{\mathbb{P}_{X_n \mid \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)}\left[x_n \mid \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right]}{\mathbb{P}_{X_n \mid \left(X_{n-k-1}^{n-1}\right)}\left[x_n \mid \left(x_{n-k-1}^{n-1}\right)\right]} \\ &\neq \sum_{\substack{x_{n-k-1}^{n-1} \in X_{n-k-1}^{n-1}(\Omega) \\ y_{n-l-1}^{n-1} \in Y_{n-l-1}^{n-1}(\Omega)}} \mathbb{P}_{X_{n-k-1}^{n-1}, Y_{n-l-1}^{n-1}}\left(x_{n-k-1}^{n-1}, y_{n-l-1}^{n-1}\right) \times \\ &\qquad \sum_{x_n \in X_n(\Omega)} \mathbb{P}_{X_n \mid \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)}\left[x_n \mid \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right] \times \\ &\qquad \log \frac{\mathbb{P}_{X_n \mid \left(X_{n-k-1}^{n-1}\right), \left(Y_{n-l-1}^{n-1}\right)}\left[x_n \middle| \left(x_{n-k-1}^{n-1}\right), \left(y_{n-l-1}^{n-1}\right)\right]}{\mathbb{P}_{X_n \mid \left(X_{n-k-1}^{n-1}\right)}\left[x_n \middle| \left(x_{n-k-1}^{n-1}\right)\right]} \end{aligned}$$

$$= \mathbb{T}^{(k,l)}_{Y \to X}(n).$$

As demonstrated in Example 3, Definition 1.1 accurately recovers the definition presented in [45] by taking two integrals: one which is a KL-divergence among conditional measures over the *present* (i.e. X_n) of the destination process given a specific set of events in the *past* of X and Y, and another which integrates this KL-divergence over all possible configurations of the *past* via integration over Ω. We note that the notion of using two expectations to properly represent conditional versions of information-theoretic measures has been done in previous literature (See Section 3 of [2], (14) of [4] and (3) of [11].).

2.2 Projective limits and construction of path measures

We now turn our attention to the main purpose of this thesis, namely, the construction of TE in continuous-time. We restrict our attention to the case when the uncountable indexing set is an interval. Let $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ be an interval whose elements we will sometimes refer to as *times*. Analogous to the setup for discrete-time TE, we suppose $X := \{X_t\}_{t\in\mathbb{T}}$ and $Y := \{Y_t\}_{t\in\mathbb{T}}$ are stochastic processes adapted to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\in\mathbb{T}}, \mathbb{P})$ such that for each $t \in \mathbb{T}$, X_t and Y_t are random variables taking values in the measurable state space $(\Sigma, \mathcal{X})$ where Σ is a Polish space and $\mathcal{X}$ is a $\sigma-$algebra of subsets of Σ. In this section we begin our construction of continuous-time TE by introducing conditional measures on the space of sample paths of X. These measures will act as the continuous-time analogues of the random conditional probabilities

$$\mathbb{P}^{(k,l)}_{n} \left[X_n \middle| \left(X^{n-1}_{n-k-1} \right), \left(Y^{n-1}_{n-l-1} \right) \right]$$

and

$$\mathbb{P}^{(k)}_{n} \left[X_n \middle| \left(X^{n-1}_{n-k-1} \right) \right]$$

in Definition 2.1.1. The following seminal result in [43] will be paramount in the formulation of these measures.

Theorem 2. Let $\mathbb{A}$ be any index set and D the set of all its finite subsets directed by inclusion. Let $(\Sigma_t, \mathcal{X}_t)_{t\in\mathbb{A}}$ be a family of measurable spaces where Σ_t is a topological space and $\mathcal{X}_t$ is a σ-field containing all the compact subsets of Σ_t. Suppose, for $\alpha \in D$, $\Sigma_\alpha = \times_{t\in\alpha}\Sigma_t, \mathcal{X}_\alpha = \bigotimes_{t\in\alpha}\mathcal{X}_t$, and $\mathbb{P}_\alpha : \mathcal{X}_\alpha \mapsto [0,1]$ so that $(\Sigma_\alpha, \mathcal{X}_\alpha, \mathbb{P}_\alpha)$ is a probability space. If for each $\alpha \in D$, $\mathbb{P}_\alpha$ is inner regular relative to the compact subsets of $\mathcal{X}_\alpha$, i.e., for any $A \in \mathcal{X}_\alpha$, $\mathbb{P}_\alpha = \sup\left\{\mathbb{P}_\alpha(C) : \ C \text{ is a compact subset of } A\right\}$, and $\pi_{\alpha\beta} : \Sigma_\beta \mapsto \Sigma_\alpha$ $(\beta \geq \alpha)$, $\pi_\alpha = \pi_{\alpha\mathbb{A}} : \times_{t\in\mathbb{A}}\Sigma_t \mapsto \Sigma_\alpha$ for $\alpha, \beta \in D$ are coordinate projections, then there exists a unique probability measure $\mathbb{P}_\mathbb{A}$ on the space $(\times_{t\in\mathbb{A}}\Sigma_t, \bigotimes_{t\in\mathbb{A}}\mathcal{X}_t)$ such that $\forall\alpha \in D$,

$$\mathbb{P}_\alpha = \mathbb{P}_\mathbb{A} \circ \pi_\alpha^{-1}, \tag{2.8}$$

if and only if $\left\{(\Sigma_\alpha, \mathcal{X}_\alpha, \mathbb{P}_\alpha, \pi_{\alpha\beta})_{\beta\geq\alpha} : \alpha, \beta \in D\right\}$ is a *projective system* with respect to mappings $\{\pi_{\alpha\beta}\}$, that is,

(1) $\pi_{\alpha\beta}^{-1}\left(\mathcal{X}_\alpha\right) \subset \mathcal{X}_\beta$ so that $\pi_{\alpha\beta}$ is $\left(\mathcal{X}_\beta, \mathcal{X}_\alpha\right)-$measurable.

(2) for any $\alpha \leq \beta \leq \lambda$, $\pi_{\alpha\beta} \circ \pi_{\beta\lambda} = \pi_{\alpha,\lambda}$, $\pi_{\alpha\alpha} = id_\alpha$ and

(3) $\mathbb{P}_\alpha = \mathbb{P}_\beta\pi_{\alpha\beta}^{-1}$, whenever $\alpha \leq \beta$.

Due to Corollary 15.27 of [1], the same result holds without the inner regularity of $\mathbb{P}_{\{\cdot\}}$ whenever Σ_t is a Polish space for each $t \in \mathbb{A}$. Furthermore, the same result holds if D is the set of countably finite subsets of $\mathbb{A}$ (Corollary 4.9.16 of [13]).

We will work in the case $\mathbb{A} = [t_0, T) \subset \mathbb{T}$, where $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ is a closed and bounded interval. As shown in the proof of Theorem 1 (See [43].), the projective limit $\sigma-$algebra, $\bigotimes_{t\in\mathbb{A}}\mathcal{X}_t$, is generated by $\bigcup\limits_{\alpha\in D}\pi_\alpha^{-1}\left(\mathcal{X}_\alpha\right)$, that is,

$$\bigotimes_{t\in\mathbb{T}}\mathcal{X}_t = \sigma\left(\bigcup_{\alpha\in D}\pi_\alpha^{-1}\left(\mathcal{X}_\alpha\right)\right).$$

If $\alpha, \beta \in D$ with $\alpha < \beta$, then due to (1) of Theorem 1,

$$\pi_\alpha^{-1}(\mathcal{X}_\alpha) = (\pi_{\alpha\beta} \circ \pi_\beta)^{-1} (\mathcal{X}_\alpha) \subset \pi_\beta^{-1}(\mathcal{X}_\beta). \tag{2.9}$$

Consequently, $(\pi_\alpha^{-1}(\mathcal{X}_\alpha))_{\alpha \in D}$ is a filtration ordered by set inclusion which generates $\bigotimes_{t \in \mathbb{A}} \mathcal{X}_t$ and from (2.8) we have

$$\mathbb{P}_{\mathbb{A}} \mid_{\pi_\alpha^{-1}(\mathcal{X}_\alpha)} = \mathbb{P}_\alpha \circ \pi_\alpha. \tag{2.10}$$

In our case, we assume that $\Sigma_t = \Sigma$ and $\mathcal{X}_t = \mathcal{X}$, $\forall t \in \mathbb{T}$.

Now let $s, r > 0$ be such that $(t_0 - \max(s, r), T) \subset \mathbb{T}$. The numbers s and r are in place to act as the continuous analogues of the positive integers k and l in Definition 2.1.1. Observe that they need not be integers as is the case with k and l in Definition 1.1.

For each $\Delta t > 0$, define the *comb set* $D_{\Delta t} \subset \mathbb{T}$ by

$$\begin{aligned} &D_{\Delta t} = \\ &\left\{ \left\lfloor \frac{t_0}{\Delta t} \right\rfloor \Delta t - \left(\left\lfloor \frac{W}{\Delta t} \right\rfloor - 1 \right) \Delta t, \ldots, \left\lfloor \frac{t_0}{\Delta t} \right\rfloor \Delta t, \left\lfloor \frac{t_0}{\Delta t} \right\rfloor \Delta t + \Delta t, \ldots \right. \\ &\qquad \left. \ldots \left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - 2\Delta t, \left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - \Delta t, \left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t \right\}, \end{aligned}$$

where $W = \max(s, r)$. Given $\Delta t > 0$, we can use the comb set $D_{\Delta t}$ to construct two probability measures on the measurable space

$$\left(\times_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \Sigma =: \Sigma^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor}, \bigotimes_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathcal{X} =: \bigotimes^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor} \mathcal{X} \right).$$

Specifically, for $\Delta t > 0$, let $A_m^{\Delta t, X} = \{X_m \in B_m\}$, $A_m^{\Delta t, Y} = \{Y_m \in B_m\}$, $X_{m,k}^{\Delta t} = \sigma\left(\left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (m+k+1)\Delta t}^{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (m+1)\Delta t}\right)\right)$, and $Y_{m,k,l}^{\Delta t} = \sigma\left(\left(Y_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+l+1)\Delta t}^{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}\right)\right)$, for $m = 0, 1, \cdots, \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$.

Then

$$\begin{aligned} &\prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \left(A_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}^{\Delta t, X} \middle| \alpha_X^{i, \Delta t} \right) \\ &= \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \left(\mathbb{P}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \middle| X_{i,k}^{\Delta t} \right) \right) (\omega) \left(B_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \right), \end{aligned} \tag{2.11}$$

for some $\omega \in \Omega$,

where $k = \left\lfloor \frac{s}{\Delta t} \right\rfloor$ and $\alpha_X^{i,\Delta t} = \bigcap_{j=\left\lfloor \frac{T}{\Delta t} \right\rfloor - (i + \left\lfloor \frac{s}{\Delta t} \right\rfloor + 1)}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - (i+1)} A_{j\Delta t}^{\Delta t, X}$, and that

$$\prod_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \left(A_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t}^{\Delta t, X} \middle| \left(\alpha_X^{i,\Delta t}\right) \bigcap \left(\alpha_Y^{i,\Delta t}\right) \right)$$
$$= \prod_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} \left(\mathbb{P}_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \left(X_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \middle| X_{i,k}^{\Delta t}, Y_{i,k,l}^{\Delta t} \right) \right)(\omega) \left(B_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \right), \tag{2.12}$$

for some $\omega \in \Omega$,

where $l = \left\lfloor \frac{r}{\Delta t} \right\rfloor$ and $\alpha_Y^{i,\Delta t} = \bigcap_{j=\left\lfloor \frac{T}{\Delta t} \right\rfloor - (i + \left\lfloor \frac{r}{\Delta t} \right\rfloor + 1)}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - (i+1)} A_{j\Delta t}^{\Delta t, Y}$.

Given $\omega \in \Omega, \Delta t > 0$, define the measures $\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(k)}$ and $\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(k,l)}$ on the space $(\Sigma, \mathcal{X})$ for each $i = 0, 1, \cdots, \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$, by

$$\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(k)} \left(B_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \right) = \left(\mathbb{P}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \left(X_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \middle| X_{i,k}^{\Delta t} \right) \right)(\omega) \left(B_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \right) \tag{2.13}$$

and

$$\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(k,l)} \left(B_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \right)$$
$$= \left(\mathbb{P}_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \left(X_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \middle| X_{i,k}^{\Delta t}, Y_{i,k,l}^{\Delta t} \right) \right)(\omega) \left(B_{\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t} \right). \tag{2.14}$$

Notation 2. For $\Delta t', \Delta t > 0$, we write $\Delta t' \mid \Delta t$ whenever there exists a positive integer m such that $\Delta t = m\Delta t'$.

Suppose $k = \left\lfloor \frac{s}{\Delta t} \right\rfloor$ and $l = \left\lfloor \frac{r}{\Delta t} \right\rfloor$. If for each $\omega \in \Omega$, the systems

$$\left\{ \left(\Sigma^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor}, \bigotimes^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor} \mathcal{X}, \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(k)}, \pi_{D_{\Delta t} D_{\Delta t'}} \right)_{\substack{0 < \Delta t' < \Delta t \\ \Delta t' | \Delta t}} \middle| \ \Delta t > 0 \right\}$$

and

$$\left\{ \left(\Sigma^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor}, \bigotimes^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor} \mathcal{X}, \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(k,l)}, \pi_{D_{\Delta t} D_{\Delta t'}} \right)_{\substack{0 < \Delta t' < \Delta t \\ \Delta t' | \Delta t}} \middle| \ \Delta t > 0 \right\}$$

are projective systems with respect to coordinate projections $\left\{\pi_{D_{\Delta t} D_{\Delta t'}}\right\}$, then as a consequence of Theorem 2, there exist unique probability measures

$$\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}](\omega)$$

and

$$\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}](\omega)$$

on the measurable space $\left(\times_{t\in[t_0,T)}\Sigma, \bigotimes_{t\in[t_0,T)} \mathcal{X}\right)$ such that

$$\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}](\omega)\Big|_{\mathcal{F}_{\Delta t}^{[t_0,T)}} = \left(\prod_{i=0}^{\lfloor \frac{T}{\Delta t}\rfloor - \lfloor \frac{t_0}{\Delta t}\rfloor - 1} \mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(k)}\right) \circ \pi_{D_{\Delta t}} \tag{2.15}$$

and

$$\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}](\omega)\Big|_{\mathcal{F}_{\Delta t}^{[t_0,T)}} = \left(\prod_{i=0}^{\lfloor \frac{T}{\Delta t}\rfloor - \lfloor \frac{t_0}{\Delta t}\rfloor - 1} \mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(k,l)}\right) \circ \pi_{D_{\Delta t}}, \tag{2.16}$$

where $\mathcal{F}_{\Delta t}^{[t_0,T)} = \pi_{D_{\Delta t}}^{-1}(\mathcal{X}_{D_{\Delta t}})$.

Notation 3. We let $\Omega_X^{[t_0,T)}$ denote the set of sample paths of the process X.

2.3 Pathwise transfer entropy and expected pathwise transfer entropy

The purpose of this section is to use the measures

$$\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}](\cdot)$$

and

$$\mathbb{P}_{X\,|\,X,\{Y\}}^{(s,r)}\left[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}\right](\cdot)$$

to define transfer entropy over an interval of the form $[t_0, T)$ with history window lengths $r, s > 0$. Unlike Definition 2.4, we give the logarithm of the RN-derivative its own name as we will later prove various properties about it alone.

Definition 2.3.1. Suppose $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ is a closed and bounded interval and that $[t_0, T) \subset \mathbb{T}$. For $\omega \in \Omega, x_{t_0}^T \in \Omega_X^{[t_0,T)}$, and $r, s > 0$ such that $(t_0 - \max(s, r), T) \subset \mathbb{T}$, define the *pathwise transfer entropy from Y to X on $[t_0, T)$ at $x_{t_0}^T$ with history window lengths r and s*, denoted $\mathcal{PT}_{Y \to X}^{(s,r)} |_{t_0}^T (\omega, x_{t_0}^T)$, by

$$\mathcal{PT}_{Y \to X}^{(s,r)} |_{t_0}^T (\omega, x_{t_0}^T) = \log \frac{d\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}](\omega)}{d\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}](\omega)} \left(x_{t_0}^T\right), \tag{2.17}$$

whenever $\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}](\omega)$ and $\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}](\omega)$ exist with

$$\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}](\omega) \ll \mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}](\omega).$$

Observation 3. For each $\omega \in \Omega$, $\mathcal{PT}_{Y \to X}^{(s,r)} |_{t_0}^T (\omega, \cdot)$ maps $\Omega_X^{[t_0,T)}$ into the extended nonnegative real line $\mathbb{R}_{\geq 0} \cup \{\infty\}$ and $\mathcal{PT}_{Y \to X}^{(s,r)} |_{t_0}^T (\omega, \cdot)$ is unique $\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}](\omega)$ a.s. due to Theorem 1.

The following is our definition of transfer entropy over an interval of the form $[t_0, T)$[2].

Definition 2.3.2. Suppose $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ is a closed and bounded interval and $[t_0, T) \subset \mathbb{T}$. For $r, s > 0$ such that $(t_0 - \max(s, r), T) \subset \mathbb{T}$, the *expected pathwise transfer entropy from Y to X on $[t_0, T)$ with history window lengths r and s*, denoted $\mathcal{EPT}_{Y \to X}^{(s,r)} |_{t_0}^T$, is defined by

$$\mathcal{EPT}_{Y \to X}^{(s,r)} |_{t_0}^T = \begin{cases} \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}_{X|X,\{Y\}}^{(s,r)}}\left[\log \frac{d\mathbb{P}_{X|X,\{Y\}}^{(s,r)}\left[X_{t_0}^T \middle| X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}\right]}{d\mathbb{P}_X^{(s)}\left[X_{t_0}^T \middle| X_{t_0-s}^{t_0}\right]}\right]\right] & , \mathbb{P}_{X|X,\{Y\}}^{(s,r)}\left[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \left\{Y_{t_0-r}^T\right\}\right](\omega) \\ & \ll \mathbb{P}_X^{(s)}\left[X_{t_0}^T \mid X_{t_0-s}^{t_0}\right](\omega), \\ & \forall \omega \in \Omega \\ \infty & , \text{otherwise} \end{cases} \tag{2.18}$$

[2] One could, in principle, construct a similar definition in the case that the interval was of the form $[t_0, T]$, via following the procedure outlined in Section 2.2 with comb sets of the form $\widetilde{D_{\Delta t}} := \left\{T, T - \Delta t, T - 2\Delta t, \ldots, T - \left\lfloor \frac{\max(s,r)}{\Delta t} \right\rfloor \Delta t\right\}$ rather than $D_{\Delta t}$.

whenever the path measures $\mathbb{P}^{(s,r)}_{X|X,\{Y\}}[X^T_{t_0} \mid X^T_{t_0-s}, \{Y^T_{t_0-r}\}](\omega)$ and $\mathbb{P}^{(s)}_X[X^T_{t_0} \mid X^T_{t_0-s}](\omega)$ exist for each $\omega \in \Omega$.

For clarity, we emphasize that the expectation in (2.18) is understood as the integral

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}^{(s,r)}_{X|X,\{Y\}}} \log \frac{d\mathbb{P}^{(s,r)}_{X|X,\{Y\}}\left[X^T_{t_0}\middle|X^{t_0}_{t_0-s}, \{Y^T_{t_0-r}\}\right]}{d\mathbb{P}^{(s)}_X\left[X^T_{t_0}\middle|X^{t_0}_{t_0-s}\right]}\right] \\ &= \int_\Omega KL(\omega)\, d\mathbb{P}(\omega), \end{aligned} \tag{2.19}$$

where

$$\begin{aligned} &KL(\omega) = \\ &\int_{\Omega^{[t_0,T)}_X} \log\left[\frac{d\mathbb{P}^{(s,r)}_{X|X,\{Y\}}[X^T_{t_0} \mid X^{t_0}_{t_o-s}, \{Y^T_{t_0-r}\}](\omega)}{d\mathbb{P}^{(s)}_X[X^T_{t_0} \mid X^{t_0}_{t_o-s}](\omega)}\left(x^T_{t_0}\right)\right] \\ &\qquad d\mathbb{P}^{(s,r)}_{X|X,\{Y\}}[X^T_{t_0} \mid X^{t_0}_{t_o-s}, \{Y^T_{t_0-r}\}](\omega), \end{aligned}$$

and note that this is consistent with the expression in (2.4) for discrete-time TE. Furthermore, note that (2.19) is an expectation of the KL-divergence among conditional measures induced by the dynamics of the processes over the space of paths of X and the EPT is always real-valued.

2.4 Counterexamples to claims in the literature

Question 1. When, in principle, can continuous-time TE be obtained as a limit of discrete-time TE?

The current section and Section 2.5 are devoted to answering this question. This notion is captured through the following claim made in [48] stated without proof listed as Remark 1.

Claim 1. We recover an approximation to the quantities in this formalism given a discretization of a continuous-time process by recognizing, due to the linearity of the expectation operator,

$$T_{Y\to X}^{(s,r)}\Big|_{t_0}^{T} = \lim_{\Delta t\to 0} \sum_{i=0}^{\frac{T}{\Delta t}-\frac{t_0}{\Delta t}-1} \mathbb{T}_{Y\to X}^{(k,l),\Delta t,SPL}\left(\frac{T}{\Delta t}\Delta t - i\Delta t\right), \tag{2.20}$$

$$\left(\text{where } T_{Y\to X}^{(s,r)}\Big|_{t_0}^{T} = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}]}{d\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}]}\right]\right),$$

where this limit exists, such that the relevant path measures are convergent in such a procedure, and where Δt defines the discretization scheme.

The claim presents a very interesting notion of the relationship between continuous-time TE and discrete-time; however, the statement of the claim in its own right is void of the level of rigor expected in mathematical exposition. Additional elaboration and precision is needed to make (2.20) realizable. The intention of the material herein is to remedy these theoretical deficiencies of the claim as well as to recast it in a manner which is mathematically defensible. We first demonstrate the necessity of these efforts by giving a counterexample to the claim as it stands.

Example 4. For the purposes of only this counterexample, we will redefine $\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(s,r)}$ and $\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(s)}$ in (2.13) and (2.14) as

$$\begin{aligned}
&\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(s,r)}(A) \\
&= \mathbb{P}_{\frac{T}{\Delta t}\Delta t - i\Delta t}\Big(X_{\frac{T}{\Delta t}\Delta t - i\Delta t} \in A \mid X_{\frac{T}{\Delta t}\Delta t-(i+k+1)\Delta t}^{\frac{T}{\Delta t}\Delta t-(i+1)\Delta t}, Y_{\frac{T}{\Delta t}\Delta t-(i+l+1)\Delta t}^{\frac{T}{\Delta t}\Delta t-(i+1)\Delta t}\Big)(\omega)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(s)}(A) \\
&= \mathbb{P}_{\frac{T}{\Delta t}\Delta t - i\Delta t}\Big(X_{\frac{T}{\Delta t}\Delta t - i\Delta t} \in A \mid X_{\frac{T}{\Delta t}\Delta t-(i+k+1)\Delta t}^{\frac{T}{\Delta t}\Delta t-(i+1)\Delta t}\Big)(\omega).
\end{aligned}$$

where $k = \lfloor \frac{s}{\Delta t} \rfloor, l = \lfloor \frac{r}{\Delta t} \rfloor$.

Let $t \in \mathbb{Q}, \mathbb{T} = [t_0, T),\ s, r > 0$ be such that $[t_0 - \max(s,r), T) \subset \mathbb{T}$. Let $\{\Delta t_j^{\mathbb{Q}}\}_{j\geq 1}$

and $\{\Delta t_j^{\mathbb{I}}\}_{j\geq 1}$ be sequences in $\mathbb{Q}$ and $\mathbb{I}$, respectively, with both converging to 0 as $j \to \infty$. Suppose $Z \sim \text{Bern}\left(\frac{1}{2}\right)$ and for each $t \in \mathbb{T}$, let

$$X_t = \chi_{\mathbb{Q}}(t)Z,$$

$$Y_t = Z$$

and let

$$X = (X_t)_{t\in\mathbb{T}},$$

$$Y = (Y_t)_{t\in\mathbb{T}}.$$

Let $\omega \in \Omega$ satisfy $Z(\omega) = 1$. Note that

$$\pi_t(X_{t_0}^T(\omega)) = 1, \forall t \in \mathbb{Q}$$

and

$$\pi_t(X_{t_0}^T(\omega)) = 0, \forall t \in \mathbb{I}.$$

Also observe that

$$\mathbb{P}_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-i\Delta t_j^{\mathbb{Q}}}\left(X_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-i\Delta t_j^{\mathbb{Q}}} = 1 \mid X_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-(i+k+1)\Delta t_j^{\mathbb{Q}}}^{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-(i+1)\Delta t_j^{\mathbb{Q}}}, Y_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-(i+l+1)\Delta t_j^{\mathbb{Q}}}^{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t-(i+1)\Delta t_j^{\mathbb{Q}}}\right)(\omega) = 1,$$

and

$$\mathbb{P}_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-i\Delta t_j^{\mathbb{Q}}}\Big(X_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-i\Delta t_j^{\mathbb{Q}}} = 1 \mid X_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-(i+k+1)\Delta t_j^{\mathbb{Q}}}^{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-(i+1)\Delta t_j^{\mathbb{Q}}}\Big)(\omega) = 1$$

for all $i = 0, 1, \ldots, \frac{T}{\Delta t_j^{\mathbb{Q}}} - \frac{t_0}{\Delta t_j^{\mathbb{Q}}} - 2, \frac{T}{\Delta t_j^{\mathbb{Q}}} - \frac{t_0}{\Delta t_j^{\mathbb{Q}}} - 1$

$$\mathbb{P}_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}}\Big(X_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}} = 0 \mid X_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+k+1)\Delta t_j^{\mathbb{I}}}^{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+1)\Delta t_j^{\mathbb{I}}}, Y_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+l+1)\Delta t_j^{\mathbb{I}}}^{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+1)\Delta t_j^{\mathbb{I}}}\Big)(\omega) = 1,$$

and

$$\mathbb{P}_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}}\Big(X_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}} = 0 \mid X_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+k+1)\Delta t_j^{\mathbb{I}}}^{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+1)\Delta t_j^{\mathbb{I}}}\Big)(\omega) = 1$$

for all $i = 1, 2, \ldots, \frac{T}{\Delta t_j^{\mathbb{I}}} - \frac{t_0}{\Delta t_j^{\mathbb{I}}} - 2, \frac{T}{\Delta t_j^{\mathbb{I}}} - \frac{t_0}{\Delta t_j^{\mathbb{I}}} - 1$ and

$$\mathbb{P}_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}}\Big(X_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}} = 1 \mid X_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+k+1)\Delta t_j^{\mathbb{I}}}^{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+1)\Delta t_j^{\mathbb{I}}}, Y_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+l+1)\Delta t_j^{\mathbb{I}}}^{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+1)\Delta t_j^{\mathbb{I}}}\Big)(\omega) = 1,$$

and

$$\mathbb{P}_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}}\Big(X_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-i\Delta t_j^{\mathbb{I}}}=1 \mid X^{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+1)\Delta t_j^{\mathbb{I}}}_{\frac{T}{\Delta t_j^{\mathbb{I}}}\Delta t_j^{\mathbb{I}}-(i+k+1)\Delta t_j^{\mathbb{I}}}\Big)(\omega)=\frac{1}{2}$$

when $i=0$. Now

$$\begin{aligned}
&\lim_{j\to\infty}\prod_{i=0}^{\frac{T}{\Delta t_j^{\mathbb{Q}}}-\frac{t_0}{\Delta t_j^{\mathbb{Q}}}-1}\frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{Q}}}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{Q}}}}\Big(\pi_{\frac{T}{\Delta t_j^{\mathbb{Q}}}\Delta t_j^{\mathbb{Q}}-i\Delta t_j^{\mathbb{Q}}}(X_{t_0}^T(\omega))\Big)\\
&=\lim_{j\to\infty}\prod_{i=0}^{\frac{T}{\Delta t_j^{\mathbb{Q}}}-\frac{t_0}{\Delta t_j^{\mathbb{Q}}}-1}\frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{Q}}}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{Q}}}}\Big(1\Big)\\
&=\lim_{j\to\infty}\prod_{i=0}^{\frac{T}{\Delta t_j^{\mathbb{Q}}}-\frac{t_0}{\Delta t_j^{\mathbb{Q}}}-1}\frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{Q}}}(\{1\})}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{Q}}}(\{1\})}\\
&=1\\
&\neq 2\\
&=\lim_{j\to\infty}\Big(1^{\frac{t-t_0}{\Delta t_j^{\mathbb{I}}}-1}\Big)\left(\frac{1}{\frac{1}{2}}\right)\\
&=\lim_{j\to\infty}\left(\prod_{i=0}^{\frac{T}{\Delta t_j^{\mathbb{I}}}-\frac{t_0}{\Delta t_j^{\mathbb{I}}}-1}\frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{I}}}(\{0\})}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{I}}}(\{0\})}\right)\left(\frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},0,\Delta t_j^{\mathbb{I}}}(\{1\})}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},0,\Delta t_j^{\mathbb{I}}}(\{1\})}\right)\\
&=\lim_{j\to\infty}\left(\prod_{i=1}^{\frac{T}{\Delta t_j^{\mathbb{I}}}-\frac{t_0}{\Delta t_j^{\mathbb{I}}}-1}\frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{I}}}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{I}}}}\Big(\pi_{t-i\Delta t_j^{\mathbb{I}}}(X_{t_0}^T(\omega))\Big)\right)\left(\frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},0,\Delta t_j^{\mathbb{I}}}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},0,\Delta t_j^{\mathbb{I}}}}\Big(\pi_t(X_{t_0}^T(\omega))\Big)\right)\\
&=\lim_{j\to\infty}\prod_{i=0}^{\frac{T}{\Delta t_j^{\mathbb{I}}}-\frac{t_0}{\Delta t_j^{\mathbb{I}}}-1}\frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{I}}}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{I}}}}\Big(\pi_{t-i\Delta t_j^{\mathbb{I}}}(X_{t_0}^T(\omega))\Big),
\end{aligned}$$

thus the limit in (2.20) does not exist.

In a revised version of [48], the expression in (2.20) is changed to

$$T_{Y\to X}^{(s,r)}\Big|_{t_0}^{T}=\lim_{\Delta t\downarrow 0}\left[\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}_{Y\to X}^{(k,l),\Delta t}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-i\Delta t\right)\right].\tag{2.21}$$

This version still inherits the theoretical issues apparent in Claim 1, thus additional elaboration is still needed. For completeness, we give an example demonstrating that it is also not true in general.

Example 5. For each $t \in \mathbb{Q}^+$, let $\{\epsilon_j^t\}_{j\geq 1}$ be a sequence of irrational numbers converging to 0 as $j \to \infty$ which are linearly independent over $\mathbb{Q}$. For each $j \geq 1$, define

$$\lfloor t \rfloor_{\epsilon_j^t} = \left\lfloor \frac{t}{\epsilon_j^t} \right\rfloor \epsilon_j^t$$

and let

$$U = \bigcup_{t\in\mathbb{Q}} \{\lfloor t \rfloor_{\epsilon_j^t} \mid t \geq t_0, j \geq 1\} \bigcup \mathbb{Q}.$$

Let $t \in \mathbb{Q}^+, \mathbb{T} = [t_0, T)$, $s, r > 0$ be such that $[t_0 - \max(s, r), T) \subset \mathbb{T}$. Let $\{\Delta t_j^{\mathbb{Q}}\}_{j\geq 1}$ be a sequence in $\mathbb{Q}$ converging to 0 as $j \to \infty$. Suppose

$$Z \sim \text{Bern}\left(\frac{1}{2}\right)$$

and for each $t \in \mathbb{T}$, let

$$X_t = \chi_U(t)Z,$$

$$Y_t = Z$$

and let $X = (X_t)_{t\in\mathbb{T}}, Y = (Y_t)_{t\in\mathbb{T}}$. Let $\omega \in \Omega$ satisfy $Z(\omega) = 1$. Note that

$$\pi_t(X_{t_0}^t(\omega)) = 1, \forall t \in \mathbb{Q},$$

$$\pi_t(X_{t_0}^t(\omega)) = 1, \forall t \in U,$$

and

$$\pi_t(X_{t_0}^t(\omega)) = 0, \forall t \in \mathbb{I} \setminus U.$$

Also observe that $\forall j \geq 1$,

$$\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{Q}}}(\{1\}) = 1$$

and

$$\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{Q}}}(\{1\}) = 1$$

for all $i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - 2, \left\lfloor \frac{T}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - 1$

$$\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\varepsilon_j^t}(\{0\}) = 1$$

and since $\overleftarrow{Y}_{\epsilon_j^t} \notin \mathbb{Q}$ from linear independence.

$$\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\varepsilon_j^t}(\{0\}) = 1,$$

for all $i = 1, 2, \ldots \left\lfloor \frac{T}{\varepsilon_j^t} \right\rfloor - \left\lfloor \frac{t_0}{\varepsilon_j^t} \right\rfloor - 2, \left\lfloor \frac{T}{\varepsilon_j^t} \right\rfloor - \left\lfloor \frac{t_0}{\varepsilon_j^t} \right\rfloor - 1$. Furthermore,

$$\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\varepsilon_j^t}(\{1\}) = 1$$

and since $\lfloor t \rfloor_{\epsilon_j^t} - i\epsilon_j^t \notin U, \forall i > 0$

$$\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\varepsilon_j^t}(\{1\}) = \frac{1}{2},$$

whenever $i = 0$. Now

$$\begin{aligned}
&\lim_{j\to\infty} \prod_{i=0}^{\left\lfloor \frac{T}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - 1} \frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{Q}}}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{Q}}}} \left(\pi_{\left\lfloor \frac{T}{\Delta t_j^{\mathbb{Q}}} \right\rfloor \Delta t_j^{\mathbb{Q}} - i\Delta t_j^{\mathbb{Q}}} (X_{t_0}^t(\omega)) \right) \\
&= \lim_{j\to\infty} \prod_{i=0}^{\left\lfloor \frac{T}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - 1} \frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{Q}}}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{Q}}}} (1) \\
&= \lim_{j\to\infty} \prod_{i=0}^{\left\lfloor \frac{T}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t_j^{\mathbb{Q}}} \right\rfloor - 1} \frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t_j^{\mathbb{Q}}}(\{1\})}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t_j^{\mathbb{Q}}}(\{1\})} \\
&= 1 \\
&\neq 2 \\
&= \lim_{j\to\infty} \left(1^{\left\lfloor \frac{t}{\epsilon_j^t} \right\rfloor - \left\lfloor \frac{t_0}{\epsilon_j^t} \right\rfloor - 1} \right) \left(\frac{1}{\frac{1}{2}} \right)
\end{aligned}$$

$$
= \lim_{j\to\infty} \left(\prod_{i=1}^{\left\lfloor \frac{t}{\epsilon_j^t} \right\rfloor - \left\lfloor \frac{t_0}{\epsilon_j^t} \right\rfloor - 1} \frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\varepsilon_j^t}(\{0\})}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\varepsilon_j^t}(\{0\})} \right) \left(\frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},0,\varepsilon_j^t}(\{1\})}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},0,\varepsilon_j^t}(\{1\})} \right)
$$

$$
= \lim_{j\to\infty} \prod_{i=0}^{\left\lfloor \frac{t}{\epsilon_j^t} \right\rfloor - \left\lfloor \frac{t_0}{\epsilon_j^t} \right\rfloor - 1} \frac{d\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\epsilon_j^t}}{d\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\epsilon_j^t}} \left(\pi_{\left\lfloor \frac{t}{\epsilon_j^t} \right\rfloor \epsilon_j^t - i\epsilon_j^t} (X_{t_0}^t(\omega)) \right).
$$

2.5 The attainability of continuous-time TE as a limit of discrete-time TE

As demonstrated in Example 4 and Example 5, the limit in (2.20) is not true in general. This motivates the pursuit of conditions under which the limit in (2.20) is indeed valid. The purpose of this section is to present such conditions for our definitions of continuous-time and discrete-time TE and present our main theorem, which should be regarded as a recasting of the revised version of claim 1 presented in (2.21). We first prove two analysis lemmas that will be used in the proof of our main theorem; then we define a type of consistency between processes that makes the expressions in the main result meaningful; then we provide the main result and conclude with some of its consequences.

Lemma 3. Suppose $N \geq 1$ and $\{\mu_i\}_{i\geq 1}$ and $\{\nu_i\}_{i\geq 1}$ are finite measures on the measurable space $(\mathcal{X}, \Sigma)$ with $\mu_i \ll \nu_i$ for $i = 1, ..., N$. Let $\mu = \prod_{i=1}^N \mu_i$ and $\nu = \prod_{i=1}^N \nu_i$ be product measures on the space $\left(\mathcal{X}^N, \otimes^N \Sigma\right)$. Then $\mu \ll \nu$ and

$$
\prod_{i=1}^N \frac{d\mu_i}{d\nu_i} \left(\pi_i(x_1, x_2, ..., x_N)\right) = \frac{d\mu}{d\nu} \left(x_1, x_2, ..., x_N\right), \;\; \nu \text{ - a.e.},
$$

where $x_i \in \mathcal{X}$, for $i \in [N]$.

Proof. Clearly $\mu \ll \nu$ since $\forall A \in \otimes^N \Sigma$,

$$\begin{aligned}
&\nu(A) = 0 \\
&\implies \exists j \leq N, \text{ such that } \nu_j(\pi_j(A)) = 0 \\
&\implies \mu_j\left(\pi_j(A)\right) = 0 \\
&\implies \mu(A) = 0.
\end{aligned}$$

Fix $E \in \otimes^N \Sigma$ and for $i = 1, 2, \ldots, N$, let

$$E_{x_1,x_2,\ldots,x_i} = \{(x_{i+1}, x_{i+2}, ..., x_N \in \mathcal{X}^{N-i}) \mid (x_1, x_2, ..., x_i, x_{i+1}, ..., x_N) \in E\},$$

where $x_i \in \mathcal{X}, \forall i \in [N]$. Then from the Radon-Nikodym chain rule,

$$\begin{aligned}
\mu(A) &= \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \left(\chi_{E_{x_1,x_2,\ldots,x_{N-1}}}(x_N)\right) d\mu_N\left(x_N\right) ...d\mu_1(x_1) \\
&= \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \left(\left(\chi_{E_{x_1,x_2,\ldots,x_{N-1}}}(x_N)\right) \frac{d\mu_N}{d\nu_N}\left(x_N\right) d\nu_N(x_N)\right) d\mu_{N-1}(x_{N-1})...d\mu_1(x_1) \\
&= \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \left(\chi_{E_{x_1,x_2,\ldots,x_{N-1}}}(x_N) \frac{d\mu_N}{d\nu_N}\left(x_N\right) \frac{d\mu_{N-1}}{d\nu_{N-1}}\left(x_{N-1}\right) ... \frac{d\mu_1}{d\nu_1}(x_1)\right) \prod_{i=1}^{N} d\nu_i(x_i) \\
&= \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \left(\chi_{E_{x_1,x_2,\ldots,x_{N-2}}}\left(x_N, x_{N-1}\right)\right) \prod_{i=1}^{N} \frac{d\mu_i}{d\nu_i}(x_i) \prod_{i=1}^{N} d\nu_i(x_i) \\
&\vdots \\
&\int_{\mathcal{X}^N} \chi_E\left(x_N, x_{N-1}, \ldots, x_2, x_1\right) \prod_{i=1}^{N} \frac{d\mu_i}{d\nu_i}(x_i) \prod_{i=1}^{N} d\nu_i(x_i) \\
&= \int_E \prod_{i=1}^{N} \frac{d\mu_i}{d\nu_i}(x_i) \prod_{i=1}^{N} d\nu_i(x_i). \\
&= \int_E \prod_{i=1}^{N} \frac{d\mu_i}{d\nu_i}(x_i) d\nu\left(x_1, \ldots, x_N\right).
\end{aligned}$$

By the uniqueness of the RN-derivative,

$$\begin{aligned}
\frac{d\mu}{d\nu}\left(x_1, x_2, ..., x_N\right) &= \prod_{i=1}^{N} \frac{d\mu_i}{d\nu_i}\left(x_i\right) \\
&= \prod_{i=1}^{N} \frac{d\mu_i}{d\nu_i}\left(\pi_i(x_1, x_2, ..., x_N)\right) \quad \nu \text{ - a.e.}
\end{aligned}$$

$\square$

The following lemma establishes convergence of KL-divergences in a manner which will be useful for the proof of our main result.

Lemma 4. Suppose $(\Omega, \mathcal{F})$ is a measurable space. Suppose further, $(\mathcal{F}_{\Delta t})_{\Delta t>0}$ is a sequence of decreasing sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F} = \bigcap\limits_{\Delta t>0} \mathcal{F}_{\Delta t}$ and that P and M are probability measures on $(\Omega, \mathcal{F})$ with $P \ll M$. Let $P_{\Delta t} = P\mid_{\mathcal{F}_{\Delta t}}$ and $M_{\Delta t} = M\mid_{\mathcal{F}_{\Delta t}}$ for each $\Delta t > 0$. If $\mathbb{E}_P\left[\log \frac{dP}{dM}\right] < \infty$, then

$$\mathbb{E}_{P_{\Delta t}}\left[\log \frac{dP_{\Delta t}}{dM_{\Delta t}}\right] \to \mathbb{E}_P\left[\log \frac{dP}{dM}\right], \text{ as } \Delta t \downarrow 0. \tag{2.22}$$

Proof. Since probability measures are $\sigma-$finite, the RN-derivatives in (2.22) exist. Suppose $\Delta t > 0$. Observe that for all $A \in \mathcal{F}_{\Delta t}$, we have that

$$\mathbb{E}_M\left[\chi_A \frac{dP_{\Delta t}}{dM_{\Delta t}}\right] = \mathbb{E}_M\left[\chi_A \frac{dP}{dM}\right],$$

implying that

$$\mathbb{E}_M\left[\frac{dP}{dM}\middle|\mathcal{F}_{\Delta t}\right] = \frac{dP_{\Delta t}}{dM_{\Delta t}} \ (M\text{-a.s.}) \tag{2.23}$$

from the definition of conditional expectation. Define $\zeta_{\Delta t} = \frac{dP_{\Delta t}}{dM_{\Delta t}}$ for each $\Delta t > 0$. From (2.23), we get that $\{\zeta_{\Delta t}\}_{\Delta t>0}$ is a uniformly integrable backward martingale since $\zeta_{\Delta t}$ is clearly $M-$integrable for any $\Delta t > 0$ by Theorem 1 and if $\Delta t' > \Delta t$, then $\mathcal{F}_{\Delta t'} \subset \mathcal{F}_{\Delta t}$, thus

$$\begin{aligned}
\mathbb{E}_M\left[\zeta_{\Delta t} \,|\, \mathcal{F}_{\Delta t'}\right] &= \mathbb{E}_M\left[\mathbb{E}_M\left[\frac{dP}{dM}\middle|\mathcal{F}_{\Delta t}\right]\middle|\mathcal{F}_{\Delta t'}\right] \\
&= \mathbb{E}_M\left[\frac{dP}{dM}\middle|\mathcal{F}_{\Delta t'}\right] \\
&= \zeta_{\Delta t'},
\end{aligned}$$

due to the tower property of conditional expectation.

We claim that

$$\zeta_{\Delta t} \to \frac{dP}{dM}, \text{ as } \Delta t \downarrow 0, \ M \text{ -a.s.}. \tag{2.24}$$

To see this, note first that the limit exists a.s and in L_1 due to Theorem 6.1 of [14], that is, there exists some nonnegative $\zeta \in L_1(\Omega, \mathcal{F}, M)$ such that

$$\mathbb{E}_M\left[\left|\zeta_{\Delta t} - \zeta\right|\right] \to 0, \text{ as } \Delta t \downarrow 0.$$

Fix $\Delta t > 0$ and suppose $A \in \mathcal{F}_{\Delta t}$. Then for all $0 < \Delta t' < \Delta t$, we have that $A \in \mathcal{F}_{\Delta t'}$ since $(\mathcal{F}_{\Delta t})_{\Delta t > 0}$ is a decreasing collection of $\sigma-$algebras. As a consequence of the Radon-Nikodym Theorem,

$$P(A) = \mathbb{E}_M \left[\chi_A \zeta_{\Delta t'}\right],$$

implying that $\mathbb{E}_M \left[\chi_A \zeta_{\Delta t'}\right]$ is constant for $0 < \Delta t' < \Delta t$, consequently

$$P(A) = \mathbb{E}_M \left[\chi_A \zeta_{\Delta t'}\right] = \mathbb{E}_M \left[\chi_A \zeta\right].$$

Furthermore, since $\mathcal{F} = \bigcap_{\Delta t > 0} \mathcal{F}_{\Delta t}$ we must have that

$$P(A) = \mathbb{E}_M \left[\chi_A \zeta\right]$$

for all $A \in \mathcal{F}$, proving (2.24).

Since $(0, \infty) \ni x \mapsto x \log x$ is convex and $\forall \Delta t > 0$,

$$\mathbb{E}_P \left[\log \zeta_{\Delta t}\right] = \mathbb{E}_{M_{\Delta t}} \left[\zeta_{\Delta t} \log \zeta_{\Delta t}\right] = \mathbb{E}_M \left[\frac{dP_{\Delta t}}{dM_{\Delta t}} \log \frac{dP_{\Delta t}}{dM_{\Delta t}}\right], \tag{2.25}$$

conditional Jensen's inequality and (2.23) imply that

$$\mathbb{E}_M \left[\frac{dP}{dM} \log \frac{dP}{dM} \middle| \mathcal{F}_{\Delta t}\right] \geq \zeta_{\Delta t} \log \zeta_{\Delta t} \ (M_{\Delta t}\text{-a.s.}). \tag{2.26}$$

Taking expectations with respect to M of both sides of (2.26) we get that $\forall \Delta t > 0$,

$$\begin{aligned}
\mathbb{E}_P \left[\log \frac{dP}{dM}\right] &= \mathbb{E}_M \left[\frac{dP}{dM} \log \frac{dP}{dM}\right] \\
&= \mathbb{E}_M \left[\mathbb{E}_M \left[\frac{dP}{dM} \log \frac{dP}{dM} \middle| \mathcal{F}_{\Delta t}\right]\right] \\
&\geq \mathbb{E}_M \left[\frac{dP_{\Delta t}}{dM_{\Delta t}} \log \frac{dP_{\Delta t}}{dM_{\Delta t}}\right],
\end{aligned}$$

thus

$$\begin{aligned}
\mathbb{E}_P \left[\log \frac{dP}{dM}\right] &\geq \limsup_{\Delta t \downarrow 0} \mathbb{E}_M \left[\frac{dP_{\Delta t}}{dM_{\Delta t}} \log \frac{dP_{\Delta t}}{dM_{\Delta t}}\right] \\
&= \limsup_{\Delta t \downarrow 0} \mathbb{E}_{M_{\Delta t}} \left[\frac{dP_{\Delta t}}{dM_{\Delta t}} \log \frac{dP_{\Delta t}}{dM_{\Delta t}}\right] \\
&= \limsup_{\Delta t \downarrow 0} \mathbb{E}_{P_{\Delta t}} \left[\log \frac{dP_{\Delta t}}{dM_{\Delta t}}\right]
\end{aligned}$$

The Radon-Nikodym Theorem guarantees that $\frac{dP}{dM}$ is nonnegative and that $\frac{dP}{dM}\log\frac{dP}{dM}$ is $\mathcal{F}-$ measurable, thus

$$\begin{aligned}
\liminf_{\Delta t\downarrow 0}\mathbb{E}_{P_{\Delta t}}\left[\log\frac{dP_{\Delta t}}{dM_{\Delta t}}\right] &= \liminf_{\Delta t\downarrow 0}\mathbb{E}_{M}\left[\frac{dP_{\Delta t}}{dM_{\Delta t}}\log\frac{dP_{\Delta t}}{dM_{\Delta t}}\right] \\
&\geq \mathbb{E}_{M}\left[\frac{dP}{dM}\log\frac{dP}{dM}\right] \\
&= \mathbb{E}_{P}\left[\log\frac{dP}{dM}\right]
\end{aligned} \tag{2.27}$$

as a consequence of the continuous-time version of Fatou's Lemma and (2.25). Now

$$\mathbb{E}_{P_{\Delta t}}\left[\log\frac{dP_{\Delta t}}{dM_{\Delta t}}\right]\to\mathbb{E}_{P}\left[\log\frac{dP}{dM}\right],\ \text{ as } \Delta t\downarrow 0.$$

$\square$

Let $\mathcal{F}_X^{[t_0,T)}$ be the sub-$\sigma-$algebra of $\bigotimes\limits_{t\in[t_0,T)}\mathcal{X}$ defined by

$$\mathcal{F}_X^{[t_0,T)} = \bigcap_{\Delta t>0}\mathcal{F}_{\Delta t}^{[t_0,T)} \tag{2.28}$$

and observe that $\left(\mathcal{F}_{\Delta t}^{[t_0,T)}\right)_{\Delta t>0}$ is a decreasing collection of $\sigma-$algebras due to (2.9). Herein, it should be understood that when we write $\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0},\{Y_{t_0-r}^T\}](\cdot)$ or $\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}\}](\cdot)$ we are referring to the restriction of these measures to the $\sigma-$algebra $\mathcal{F}_X^{[t_0,T)}$. Furthermore, recall from (2.15) and (2.16) that for all $A\in\mathcal{F}_{\Delta t}^{[t_0,T)},\omega\in\Omega$,

$$\begin{aligned}
&\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0},\{Y_{t_0-r}^T\}](\omega)\Big|_{\mathcal{F}_{\Delta t}^{[t_0,T)}}(A) = \\
&\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(k,l)}\left(\pi_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-i\Delta t}(A)\right),
\end{aligned} \tag{2.29}$$

and

$$\begin{aligned}
&\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}\}](\omega)\Big|_{\mathcal{F}_{\Delta t}^{[t_0,T)}}(A) = \\
&\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(k)}\left(\pi_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-i\Delta t}(A)\right),
\end{aligned} \tag{2.30}$$

where $k = \left\lfloor \frac{s}{\Delta t} \right\rfloor$ and $l = \left\lfloor \frac{r}{\Delta t} \right\rfloor$ and note that (2.29) and (2.30) will be used in the proof of our main result. From here on in, we will ignore writing the projections in (2.29) and (2.30) to avoid cumbersome notation.

Notation 4. We denote by $P_{\Delta t}^{(\omega)}$ and $M_{\Delta t}^{(\omega)}$ the probability measures

$$\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}]\,(\omega)\Big|_{\mathcal{F}_{\Delta t}^{[t_0,T)}}$$

and

$$\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}\}]\,(\omega)\Big|_{\mathcal{F}_{\Delta t}^{[t_0,T)}},$$

respectively. It should be noted that these are measures on the measurable space

$$\left(\Sigma^{\left\lfloor \frac{T}{\Delta t}\right\rfloor - \left\lfloor \frac{t_0}{\Delta t}\right\rfloor}, \bigotimes^{\left\lfloor \frac{T}{\Delta t}\right\rfloor - \left\lfloor \frac{t_0}{\Delta t}\right\rfloor} \mathcal{X}\right).$$

Notation 5. For any $\Delta t > 0$, let

$$\mathbb{T}_{Y\to X}^{(k,l),\Delta t}\left(\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t - i\Delta t\right) = \mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(k,l)} \middle\| M_{\Delta t}^{(k)}\right)\right],$$

for any $i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t}\right\rfloor - \left\lfloor \frac{t_0}{\Delta t}\right\rfloor - 1$, where

$$P_{\Delta t}^{(k,l)} = \mathbb{P}^{(k,l)}\left[X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t - i\Delta t} \middle| \left(X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+k+1)\Delta t}^{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+1)\Delta t}\right), \left(Y_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+l+1)\Delta t}^{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+1)\Delta t}\right)\right],$$

$$M_{\Delta t}^{(k)} = \mathbb{P}^{(k)}\left[X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t - i\Delta t} \middle| \left(X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+k+1)\Delta t}^{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+1)\Delta t}\right)\right],$$

$$\left(X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+k+1)\Delta t}^{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+1)\Delta t}\right) = \left(X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+1)\Delta t}, X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+2)\Delta t}, \cdots, X_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+k+1)\Delta t}\right)$$

and

$$\left(Y_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+l+1)\Delta t}^{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+1)\Delta t}\right) = \left(Y_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+1)\Delta t}, Y_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+2)\Delta t}, \cdots, Y_{\left\lfloor \frac{T}{\Delta t}\right\rfloor \Delta t-(i+l+1)\Delta t}\right).$$

As a means of succinctly capturing all of the conditions which need hold to use TE in our formalism, we define a type of consistency between two processes dependent on the window lengths r and s and the set $[t_0, T)$. This notion of consistency captures the conditions under which our main result, Theorem 5, is of utility.

Definition 2.5.1. Suppose $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ is a closed and bounded interval, $[t_0, T) \subset \mathbb{T}$, and $s, r > 0$ are such that $(t_0 - \max(s, r), T) \subset \mathbb{T}$. Suppose further that $X := \{X_t\}_{t \in \mathbb{T}}$ and $Y := \{Y_t\}_{t \in \mathbb{T}}$ are stochastic processes adapted to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{T}}, \mathbb{P})$ such that for each $t \in \mathbb{T}, X_t$ and Y_t are random variables taking values in the measurable space $(\Sigma, \mathcal{X})$ where Σ is assumed to be a Polish space and $\mathcal{X}$ is a $\sigma-$algebra of subsets of Σ. Y *is (s,r)-consistent upon* X *on* $[t_0, T)$ iff

1. $\forall \omega \in \Omega$, there exist path measures $\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_0 - s}^{t_0}\}]\,(\omega)$ and $\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^T\}]\,(\omega)$ on the space $\left(\Omega_X^{[t_0,T)}, \mathcal{F}_X^{[t_0,T)}\right)$ for which (2.15) and (2.16) hold.

2. $\exists \delta_1 > 0$ such that $\forall \Delta t \in (0, \delta_1), \omega \in \Omega, i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$,

    (a) $\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(\lfloor \frac{s}{\Delta t} \rfloor, \lfloor \frac{r}{\Delta t} \rfloor)} \ll \mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(\lfloor \frac{s}{\Delta t} \rfloor)}$.

    (b) $\frac{d\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(k,l)}}{d\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(k)}} \in L_1\left(\Sigma, \mathcal{X}, \mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(k,l)}\right)$.

3. $\forall \omega \in \Omega, \mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_o-s}^{t_0}, \{Y_{t_0-r}^T\}]\,(\omega) \ll \mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_o-s}^{t_0}]\,(\omega)$.

We call 1.- 3. *consistency conditions.*

Remark 7. For clarity, we assume that the limit of a function $f : \mathbb{R} \mapsto \mathbb{R}$ at a point $c \in \mathbb{R}$ exists iff $\exists L < \infty$ such that $\forall \epsilon > 0, \exists \delta > 0$ such that $x \in (c - \delta, c + \delta) \implies f(x) \in (L - \epsilon, L + \epsilon)$.

We now proceed to our main theorem which should be regarded as our recasted version of Claim 1 in (2.20). We show that integrability (under $\mathbb{P}$) of the EPT is equivalent to our version of the limit in Claim 1 under $(s, r)-$ consistency and a bounding condition.

Theorem 5. Suppose $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ is a closed and bounded interval, $[t_0, T) \subset \mathbb{T}$, Σ is a Polish space and $s, r > 0$ satisfy $(t_0 - \max(s, r), T) \subset \mathbb{T}$. Suppose further that $X := \{X_t\}_{t \in \mathbb{T}}$ and $Y := \{Y_t\}_{t \in \mathbb{T}}$ are stochastic processes adapted to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathbb{P})$ such that for each $t \in \mathbb{T}, X_t$ and Y_t are random variables taking values in the measurable state space $(\Sigma, \mathcal{X})$ and that Y is (s,r)-consistent upon X on $[t_0, T)$.

If $\exists M, \delta_2 > 0$ such that $\forall \Delta t \in (0, \delta_2)$,

$$KL\left(P_{\Delta t}^{(\cdot)}\middle\|M_{\Delta t}^{(\cdot)}\right) \leq M,\ \mathbb{P}\text{-a.s. }, \tag{2.31}$$

then

$$\mathcal{EPT}_{Y\to X}^{(s,r)}\mid_{t_0}^{T} < \infty$$

iff

$$\lim_{\Delta t\downarrow 0}\left[\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}_{Y\to X}^{(k,l),\Delta t}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t - i\Delta t\right)\right] = \mathcal{EPT}_{Y\to X}^{(s,r)}\mid_{t_0}^{T}, \tag{2.32}$$

where $k = \lfloor\frac{s}{\Delta t}\rfloor$ and $l = \lfloor\frac{r}{\Delta t}\rfloor$.

Proof. ($\Rightarrow$) Suppose $\mathcal{EPT}_{Y\to X}^{(s,r)}\mid_{t_0}^{T} < \infty$, let $\delta = \min\{\delta_1, \delta_2\}$ and for each $\omega \in \Omega$, let

$$P^{(\omega)} = \mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^{T} \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^{T}\}]\,(\omega)\,,$$

and

$$M^{(\omega)} = \mathbb{P}_{X}^{(s,r)}[X_{t_0}^{T} \mid X_{t_0-s}^{t_0}]\,(\omega)\,.$$

If $\Delta t \in (0, \delta)$, then (2.31) implies that $KL\left(P_{\Delta t}^{(\omega)}||M_{\Delta t}^{(\omega)}\right)$ is $\mathbb{P}$-integrable and since Σ is $\sigma-$finite under both $\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)}$ and $\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor)}, \forall\omega \in \Omega$, and
$i = 0, 1, \ldots, \left\lfloor\frac{T}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor - 1$, we have that the measurable space
$\left(\Sigma^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}, \bigotimes^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}\mathcal{X}\right)$ is $\sigma-$finite under both $P_{\Delta t}^{(\omega)}$ and $M_{\Delta t}^{(\omega)}$ for each $\omega \in \Omega$,

thus the RN-derivatives in (2.32) exist. Furthermore, we get from Lemma 3 that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{P^{(\omega)}_{\Delta t}}\left[\log\frac{d\left(\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}\right)}{d\left(\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right)}\right]\right]\\ &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{P^{(\omega)}_{\Delta t}}\left[\log\left(\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\left(\frac{d\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}}{d\mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}}\right)\right)\right]\right]\\ &= \sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{P^{(\omega)}_{\Delta t}}\left[\log\frac{d\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}}{d\mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}}\right]\right].\end{aligned}\tag{2.33}$$

Now for each $\Delta t > 0, i = 0, 1, \cdots, \lfloor\frac{T}{\Delta t}\rfloor - \lfloor\frac{t_0}{\Delta t}\rfloor - 1$ and $\omega \in \Omega$, let

$$F^{\omega}_{i,\Delta t}\left(x_0, x_1, \cdots, x_{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\right) = \log\frac{d\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}}{d\mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}}(x_i)$$

for each $\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)$-tuple

$$\left(x_0, x_1, \cdots, x_{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\right) \in \Sigma^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}.$$

Clearly, $F^{\omega}_{i,\Delta t}$ is $\Sigma^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}-$measurable and furthermore $P^{(\omega)}_{\Delta t}-$integrable due to Jensen's inequality since consistency condition 2(b) implies

$$\begin{aligned}&\int_{\Sigma^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}}\left[F^{\omega}_{i,\Delta t}\right]dP^{(\omega)}_{\Delta t}\\ &\leq \log\left(\int_{\Sigma^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}}\left[\frac{d\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}}{d\mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}}\right]dP^{(\omega)}_{\Delta t}\right) < \infty,\end{aligned}$$

Now we apply Fubini's Theorem and obtain

$$\begin{aligned}&\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{P^{(\cdot)}_{\Delta t}}\left[\log\frac{d\mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}}{d\mathbb{P}^{(\cdot),(k)}_{X|\overleftarrow{X},i,\Delta t}}\right]\right]\\ &= \sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{E}_{\mathbb{P}}\left[\int_{\Sigma^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}}\left[\log\frac{d\mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}}{d\mathbb{P}^{(\cdot),(k)}_{X|\overleftarrow{X},i,\Delta t}}\right]d\left(\prod_{j=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},j,\Delta t}\right)\right]\\ &= \sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{E}_{\mathbb{P}}\left[\int_{\Sigma^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor}}F^{(\cdot)}_{i,\Delta t}\,d\left(\prod_{j=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},j,\Delta t}\right)\right]\end{aligned}$$

$$= \sum_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{E}_{\mathbb{P}} \left[S_{i,\Delta t} \left(\int_\Sigma \log \frac{d\mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}}{d\mathbb{P}^{(\cdot),(k)}_{X|\overleftarrow{X},i,\Delta t}} \, d\mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \right) \right]$$

$$= \sum_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t \right).$$

where $S_{i,\Delta t} = \prod_{\substack{j=0 \\ j \neq i}}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \int_\Sigma 1 d\mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},j,\Delta t}$ for $i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$.

Thus

$$\mathbb{E}_{\mathbb{P}} \left[KL \left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t} \right) \right] = \sum_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t \right). \tag{2.34}$$

Since $\mathcal{EPT}^{(s,r)}_{Y \to X} \mid_{t_0}^{T} < \infty$,

$$KL \left(P^{(\cdot)} \middle|\middle| M^{(\cdot)} \right) < \infty$$

for all $\omega \in \Omega \backslash B$ for some $\mathbb{P}$-null set B, which from Lemma 4 implies that

$$\mathbb{E}_{P^{(\cdot)}_{\Delta t}} \left[\log \frac{dP^{(\cdot)}_{\Delta t}}{dM^{(\cdot)}_{\Delta t}} \right] \to \mathbb{E}_{P^{(\cdot)}} \left[\log \frac{dP^{(\cdot)}}{dM^{(\cdot)}} \right], \text{ as } \Delta t \downarrow 0, \ \ \mathbb{P}\text{- a.s.} \tag{2.35}$$

Let

$$g(\omega) = \begin{cases} \lim_{\Delta t \downarrow 0} \left(KL \left(P^{(\omega)}_{\Delta t} || M^{(\omega)}_{\Delta t} \right) \right) & \omega \in \Omega \backslash B \\ 0 & \omega \in B \end{cases}$$

and observe that

$$g \in L_1 (\Omega, \mathcal{F}, \mathbb{P})$$

and

$$\lim_{\Delta t \downarrow 0} KL \left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t} \right) = g \ \ \mathbb{P}\text{-a.s.} \tag{2.36}$$

Since almost sure convergence implies convergence in measure over finite measure spaces, (2.36) implies that

$$KL \left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t} \right) \xrightarrow{\mathbb{P}} g \ \text{ as } \Delta t \downarrow 0. \tag{2.37}$$

Now for each $\epsilon, \Delta t > 0, \omega \in \Omega$, define $h^{\epsilon}_{\Delta t}(\omega)$ by

$$h^{\epsilon}_{\Delta t}(\omega) = \begin{cases} KL \left(P^{(\omega)}_{\Delta t} || M^{(\omega)}_{\Delta t} \right) & \left| KL \left(P^{(\omega)}_{\Delta t} || M^{(\omega)}_{\Delta t} \right) - g(\omega) \right| < \epsilon \\ 0 & \text{otherwise} \end{cases}$$

and note that $h^{\epsilon}_{\Delta t}$ is nonnegative $\forall \epsilon, \Delta t > 0$ due to Gibbs' inequality and converges in probability to g since $\forall \eta > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left\{\left|h_{\Delta t} - g\right| \geq \eta\right\}\right) &= \mathbb{P}\left(\left\{\left|h_{\Delta t} - g\right| \geq \eta\right\} \cap \left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| < \epsilon\right\}\right) \\
&\quad + \mathbb{P}\left(\left\{\left|h_{\Delta t} - g\right| \geq \eta\right\} \cap \left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}\right) \\
&\leq \mathbb{P}\left(\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \eta\right\}\right) \\
&\quad + \mathbb{P}\left(\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}\right) \\
&\to 0, \text{ as } \Delta t \downarrow 0.
\end{aligned}$$

Let $\epsilon > 0$ be arbitrary and observe that

$$\begin{aligned}
\|h^{\epsilon}_{\Delta t} - g\|_{L_1} &= \mathbb{E}_{\mathbb{P}}\left[\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \mathbb{1}_{\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| < \epsilon\right\}}\right] \\
&\quad + \mathbb{E}_{\mathbb{P}}\left[g \mathbb{1}_{\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}}\right] \\
&< \epsilon\, \mathbb{P}\left(\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| < \epsilon\right\}\right) + \mathbb{E}_{\mathbb{P}}\left[g \mathbb{1}_{\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}}\right].
\end{aligned} \tag{2.38}$$

Since $g \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ we have that $\forall \epsilon' > 0, \exists \delta' > 0$ such that

$$\mathbb{P}(A) < \delta' \implies \mathbb{E}_{\mathbb{P}}\left[g \mathbb{1}_A\right] < \epsilon',$$

$\forall A \in \mathcal{F}$. Since $KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) \xrightarrow{\mathbb{P}} g$ as $\Delta t \downarrow 0$, $\exists \delta'' > 0$ such that

$$\mathbb{P}\left(\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}\right) < \delta',$$

$\forall \Delta t \in (0, \delta'')$, implying that

$$\lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[g \mathbb{1}_{\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}}\right] = 0.$$

Now since $\mathbb{P}\left(\left\{\left|KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) - g\right| < \epsilon\right\}\right) \to 1$, as $\Delta t \downarrow 0$, we obtain

$$\lim_{\Delta t \downarrow 0} \|h^{\epsilon}_{\Delta t} - g\|_{L_1} \leq \epsilon$$

from (2.38) and thus

$$\lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \|h^{\epsilon}_{\Delta t} - g\|_{L_1} = 0$$

since $\epsilon > 0$ was arbitrary. In particular,

$$\lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t}\right] = \mathbb{E}_{\mathbb{P}}\left[g\right] = \mathbb{E}_{\mathbb{P}}\left[\lim_{\Delta t \downarrow 0} KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right)\right]. \tag{2.39}$$

We now show that

$$\lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t}\right] = \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right)\right]. \tag{2.40}$$

Note that

$$\lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[\alpha^{\epsilon}_{\Delta t}\right] = 0 \implies \lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t} - KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right)\right] = 0,$$

where

$$\alpha^{\epsilon}_{\Delta t}(\omega) = KL\left(P^{(\omega)}_{\Delta t} || M^{(\omega)}_{\Delta t}\right) \mathbb{1}_{\left\{\left|KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}}(\omega)$$

for $\epsilon, \Delta t > 0, \omega \in \Omega$. Fix $\epsilon > 0$ and note that (2.31) implies

$$0 \leq \mathbb{E}_{\mathbb{P}}\left[\alpha^{\epsilon}_{\Delta t}\right] \leq M\mathbb{P}\left(\left\{\left|KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right) - g\right| \geq \epsilon\right\}\right], \tag{2.41}$$

$\forall \Delta t \in (0, \delta)$. From (2.37), the RHS of (2.41) converges to 0 as $\Delta t \downarrow 0$, thus

$$\lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[\alpha^{\epsilon}_{\Delta t}\right] = 0, \tag{2.42}$$

so

$$\lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t} - KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right)\right] = 0. \tag{2.43}$$

Now from (2.39) and (2.43) we have that $\lim\limits_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right)\right]$ exists since

$$\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right)\right] = \mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t}\right] - \left[\mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t} - KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right)\right]\right],$$

hence

$$
\begin{aligned}
0 &= \lim_{\epsilon\downarrow 0}\lim_{\Delta t\downarrow 0}\mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t} - KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] \\
&= \lim_{\epsilon\downarrow 0}\lim_{\Delta t\downarrow 0}\left(\mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t}\right] - \mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right]\right) \\
&= \lim_{\epsilon\downarrow 0}\lim_{\Delta t\downarrow 0}\left(\mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t}\right]\right) - \lim_{\epsilon\downarrow 0}\lim_{\Delta t\downarrow 0}\left(\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right]\right) \\
&= \lim_{\epsilon\downarrow 0}\lim_{\Delta t\downarrow 0}\left(\mathbb{E}_{\mathbb{P}}\left[h^{\epsilon}_{\Delta t}\right]\right) - \lim_{\Delta t\downarrow 0}\left(\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right]\right)
\end{aligned}
$$

proving (2.40). Now we have

$$
\lim_{\Delta t\downarrow 0}\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] = \mathbb{E}_{\mathbb{P}}\left[\lim_{\Delta t\downarrow 0}KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] \tag{2.44}
$$

from which the result follows as

$$
\begin{aligned}
\mathcal{EPT}^{(s,r)}_{Y\to X}\mid^{T}_{t_0} &= \mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}||M^{(\cdot)}\right)\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[\lim_{\Delta t\downarrow 0}KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] \\
&= \lim_{\Delta t\downarrow 0}\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] \\
&= \lim_{\Delta t\downarrow 0}\left[\sum_{i=0}^{\left\lfloor \frac{T}{\Delta t}\right\rfloor - \left\lfloor \frac{t_0}{\Delta t}\right\rfloor - 1}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t - i\Delta t\right)\right].
\end{aligned}
$$

($\Leftarrow$) Suppose towards a contradiction

$$
\lim_{\Delta t\downarrow 0}\left[\sum_{i=0}^{\left\lfloor \frac{T}{\Delta t}\right\rfloor - \left\lfloor \frac{t_0}{\Delta t}\right\rfloor - 1}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t - i\Delta t\right)\right] = \mathcal{EPT}^{(s,r)}_{Y\to X}\mid^{T}_{t_0} = \infty.
$$

Then

$$
\lim_{\Delta t\downarrow 0}\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] = \infty,
$$

meaning $\exists\delta_3 > 0$ such that $\Delta t\in(0,\delta_3) \implies \mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] > M$. From (2.31),

$$
KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right) \le M \ \mathbb{P}\text{- a.s. } \forall\Delta t\in(0,\delta_2),
$$

hence

$$
M < \mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t}||M^{(\cdot)}_{\Delta t}\right)\right] \le \mathbb{E}_{\mathbb{P}}[M] = M,
$$

$\forall\Delta t\in(0,\min\{\delta_3,\delta_2\})$, a contradiction. $\square$

Due to the following corollary, one can conclude the only if part of Theorem 5 under a slightly weaker version of (2.31).

Corollary 5.1. Let $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ be an interval and $[t_0, T) \subset \mathbb{T}$ and $s, r > 0$ be such that $(t_0 - \max(s, r), T) \subset \mathbb{T}$. Suppose $X := \{X_t\}_{t \in \mathbb{T}}$ and $Y := \{Y_t\}_{t \in \mathbb{T}}$ are stochastic processes adapted to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{T}}, \mathbb{P})$ such that for each $t \in \mathbb{T}, X_t$ and Y_t are random variables taking values in the measurable state space $(\Sigma, \mathcal{X})$ and Y is $(s, r)-$SPL consistent upon X on $[t_0, T)$. If

$$\exists \eta \in L_1 (\Omega, \mathcal{F}, \mathbb{P}), \delta_2 > 0 \text{ such that } \forall \Delta t \in (0, \delta_2), KL \left(P_{\Delta t}^{(\cdot)} \middle\| M_{\Delta t}^{(\cdot)} \right) \leq \eta(\cdot), \ \mathbb{P}\text{-a.s.} \tag{2.45}$$

and

$$\mathcal{EPT}_{Y \to X}^{(s,r)} \mid_{t_0}^{T} < \infty,$$

then

$$\lim_{\Delta t \downarrow 0} \left[\sum_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{T}_{Y \to X}^{(k,l),\Delta t} \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i \Delta t \right) \right] = \mathcal{EPT}_{Y \to X}^{(s,r)} \mid_{t_0}^{T},$$

where $k = \lfloor \frac{s}{\Delta t} \rfloor$ and $l = \lfloor \frac{r}{\Delta t} \rfloor$.

Proof. We need only show that (2.42) in the proof of the forward direction of Theorem 5 is still true assuming (2.45). Since $\eta \in L_1 (\Omega, \mathcal{F}, \mathbb{P})$, this is immediate since for $\epsilon > 0$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[\alpha_{\Delta t}^{\epsilon}(\cdot) \right] &= \mathbb{E}_{\mathbb{P}} \left[KL \left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)} \right) \mathbb{1}_{\left\{ \left| KL \left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)} \right) - g \right| \geq \epsilon \right\}} \right] \\ &\leq \mathbb{E}_{\mathbb{P}} \left[\eta \mathbb{1}_{\left\{ \left| KL \left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)} \right) - g \right| \geq \epsilon \right\}} \right] \to 0, \end{aligned}$$

as $\Delta t \downarrow 0$ due to (2.37). □

The following corollary of Theorem 5 is a key result because it will be used in an application to be explored later in Section 4.2. The conditions in Theorem 5 may be too strong to apply to some common situations. The following weakens these conditions at the cost of the equivalence between the hypotheses and conclusion; however, it is necessary to make a reasonable example work.

Corollary 5.2. Let $\mathbb{T} \subset \mathbb{R}_{\geq 0}$ be an interval and $[t_0, T) \subset \mathbb{T}$ and $s, r > 0$ be such that $(t_0 - \max(s,r), T) \subset \mathbb{T}$. Suppose $X := \{X_t\}_{t\in\mathbb{T}}$ and $Y := \{Y_t\}_{t\in\mathbb{T}}$ are stochastic processes adapted to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathbb{P})$ such that for each $t \in \mathbb{T}, X_t$ and Y_t are random variables taking values in the measurable state space $(\Sigma, \mathcal{X})$ and Y is $(s,r)-$SPL consistent upon X on $[t_0, T)$. If there exists $\gamma > 0$ such that

$$\lim_{\Delta t \downarrow 0} \mathbb{P}\left(B_{\Delta t, \gamma}\right) = 1, \tag{2.46}$$

where

$$B_{\Delta t, \gamma} = \left\{ \omega \in \Omega \;\middle|\; \Delta t' \in (0, \Delta t) \implies KL\left(P^{(\omega)}_{\Delta t'} \middle|\middle| M^{(\omega)}_{\Delta t'}\right) \leq \gamma \right\} \tag{2.47}$$

for $\Delta t, \lambda > 0$, and

$$\mathcal{EPT}^{(s,r)}_{Y\to X} \mid_{t_0}^{T} < \infty,$$

then

$$\lim_{\Delta t \downarrow 0} \left[\sum_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t \right) \right] = \mathcal{EPT}^{(s,r)}_{Y\to X} \mid_{t_0}^{T},$$

where $k = \lfloor \frac{s}{\Delta t} \rfloor$ and $l = \lfloor \frac{r}{\Delta t} \rfloor$.

Proof. As in Corollary 5.1, it suffices to show that (2.42) holds whenever both (2.46) and $\mathcal{EPT}^{(s,r)}_{Y\to X} \mid_{t_0}^{T} < \infty$ hold. Observe that

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}}\left[KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right) \mathbb{1}_{\left\{ \left| KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right) - g \right| \geq \epsilon \right\} \cap B_{\Delta t}} \right] \\ &\leq \mathbb{E}_{\mathbb{P}}\left[\gamma \mathbb{1}_{\left\{ \left| KL\left(P^{(\cdot)}_{\Delta t} || M^{(\cdot)}_{\Delta t}\right) - g \right| \geq \epsilon \right\}} \right] \to 0 \text{ as } \Delta t \downarrow 0, \end{aligned}$$

since clearly $\gamma \in L_1(\Omega, \mathcal{F}, \mathbb{P})$. Since $\mathcal{EPT}^{(s,r)}_{Y\to X} \mid_{t_0}^{T} < \infty$, Lemma 4 implies that

$$KL\left(\prod_{i=0}^{\lfloor \frac{T}{\Delta t'} \rfloor - \lfloor \frac{t_0}{\Delta t'} \rfloor - 1} \mathbb{P}^{(\cdot),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t'} \middle|\middle| \prod_{i=0}^{\lfloor \frac{T}{\Delta t'} \rfloor - \lfloor \frac{t_0}{\Delta t'} \rfloor - 1} \mathbb{P}^{(\cdot),(k)}_{X|\overleftarrow{X},i,\Delta t'} \right) \in L_1(\Omega, \mathcal{F}, \mathbb{P})$$

for all $\Delta t'$ in a small enough neighborhood of 0, thus

$$\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\mathbb{1}_{\left\{\left|KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)-g\right|\geq\epsilon\right\}\cap\overline{B_{\Delta t}}}\right]\to 0,\ \text{as } \Delta t\downarrow 0$$

since $\mathbb{P}\left(\overline{B_{\Delta t}}\right)\to 0$ as $\Delta t\downarrow 0$. Now for any $\epsilon>0$,

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\left[\alpha_{\Delta t}^{\epsilon}\right]\\
&=\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\mathbb{1}_{\left\{\left|KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)-g\right|\geq\epsilon\right\}}\right]\\
&=\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\mathbb{1}_{\left\{\left|KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)-g\right|\geq\epsilon\right\}\cap B_{\Delta t}}\right]\\
&\quad+\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\mathbb{1}_{\left\{\left|KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)-g\right|\geq\epsilon\right\}\cap\overline{B_{\Delta t}}}\right]\\
&\to 0,\ \text{as } \Delta t\downarrow 0.
\end{aligned}$$

$\square$

We now provide an alternate version of our main theorem under different conditions. Instead of an a.s. bounding condition on the KL-divergence of $M_{\Delta t}^{(\cdot)}$ from $P_{\Delta t}^{(\cdot)}$, we impose a bounding condition on the transfer entropy itself and obtain a similar equivalence.

Theorem 6. Let $\mathbb{T}\subset\mathbb{R}_{\geq 0}$ be an interval and $[t_0,T)\subset\mathbb{T}$ and $s,r>0$ be such that $(t_0-\max(s,r),T)\subset\mathbb{T}$. Suppose $X:=\{X_t\}_{t\in\mathbb{T}}$ and $Y:=\{Y_t\}_{t\in\mathbb{T}}$ are stochastic processes adapted to the filtered probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\in\mathbb{T}},\mathbb{P})$ such that for each $t\in\mathbb{T}, X_t$ and Y_t are random variables taking values in the measurable state space $(\Sigma,\mathcal{X})$ and Y is $(s,r)-$SPL consistent upon X on $[t_0,T)$. If

1. $\forall\Delta t>0,\ KL\left(\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\cdot),(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)}\middle|\middle|\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\cdot),(\lfloor\frac{s}{\Delta t}\rfloor)}\right)\in L_1(\Omega,\mathcal{F},\mathbb{P})$,
   $\forall i=0,1,\ldots,\left\lfloor\frac{T}{\Delta t}\right\rfloor-\left\lfloor\frac{t_0}{\Delta t}\right\rfloor-1.$

2. $\mathcal{EPT}_{Y\to X}^{(s,r)}\mid_{t_0}^{T}<\infty$.

3. $\left\{KL\left(P_{\Delta t}^{(\cdot)}\middle\|M_{\Delta t}^{(\cdot)}\right)\right\}_{\Delta t>0}$ is a UI family.

then $\exists\delta>0$ such that for some $M>0$,

$$\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}_{Y\to X}^{(k,l),\Delta t}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-i\Delta t\right)\le M,\quad\forall\Delta t\in(0,\delta),$$

iff

$$\lim_{\Delta t\to 0}\left[\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}_{Y\to X}^{(k,l),\Delta t}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-i\Delta t\right)\right]=\mathcal{EPT}_{Y\to X}^{(s,r)}\mid_{t_0}^{T},$$

where $k=\lfloor\frac{s}{\Delta t}\rfloor$ and $l=\lfloor\frac{r}{\Delta t}\rfloor$.

Proof. $(\Rightarrow)$ Assume $\exists\delta>0$ such that for some $M>0$,

$$\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}_{Y\to X}^{(k,l),\Delta t}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-i\Delta t\right)\le M,\quad\forall\Delta t\in(0,\delta).$$

For each $\Delta t>0,\omega\in\Omega$, let $P^{(\omega)},M^{(\omega)}$ and $g(\omega)$ be as in the proof of Theorem 5. From Gibbs' inequality, g is a nonnegative random variable thus

$$\mathbb{E}_{\mathbb{P}}\left[g\right]+1>0.$$

For each $\Delta t>0,\omega\in\Omega$, define $h_{\Delta t}(\omega)$ by

$$h_{\Delta t}(\omega)=\begin{cases}KL\left(P_{\Delta t}^{(\omega)}||M_{\Delta t}^{(\omega)}\right) & \left|KL\left(P_{\Delta t}^{(\omega)}||M_{\Delta t}^{(\omega)}\right)-g(\omega)\right|\le\mathbb{E}_{\mathbb{P}}\left[g\right]+1\\ 0 & \text{otherwise}\end{cases}$$

and note that $h_{\Delta t}$ is a nonnegative random variable $\forall\Delta t>0$.

Let $\alpha>0$. Since $\left\{KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\right\}_{\Delta t>0}$ is UI, $\exists\ 0<K_\alpha<\infty$ such that

$$\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\mathbb{1}_{\left\{KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\ge K_\alpha\right\}}\right]<\alpha,$$

$\forall\Delta t>0$, thus $\{h_{\Delta t}(\cdot)\}_{\Delta t>0}$ is UI since

$$\mathbb{E}_{\mathbb{P}}\left[h_{\Delta t}\mathbb{1}_{\{h_{\Delta t}\ge K_\alpha\}}\right]\le\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\mathbb{1}_{\left\{KL\left(P_{\Delta t}^{(\cdot)}||M_{\Delta t}^{(\cdot)}\right)\ge K_\alpha\right\}}\right]<\alpha,$$

$\forall \Delta t > 0$. Furthermore, if $J_{\Delta t} = \left\{ \left| KL\left(P_{\Delta t}^{(\omega)} || M_{\Delta t}^{(\omega)}\right) - g(\omega) \right| \leq \mathbb{E}_{\mathbb{P}}[g] + 1 \right\}$, then

$$\begin{aligned}
\mathbb{P}\left(\{|h_{\Delta t} - g| > \alpha\}\right) &= \mathbb{P}\left(\{|h_{\Delta t} - g| > \alpha\} \cap J_{\Delta t}\right) + \mathbb{P}\left(\{|h_{\Delta t} - g| > \alpha\} \cap \overline{J_{\Delta t}}\right) \\
&\leq \mathbb{P}\left(\left\{ \left| KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) - g \right| > \alpha \right\}\right) + \mathbb{P}\left(\overline{J_{\Delta t}}\right) \\
&\to 0, \text{ as } \Delta t \downarrow 0 \text{ (due to (2.37))}
\end{aligned}$$

hence

$$h_{\Delta t} \xrightarrow{\mathbb{P}} g, \quad \text{as } \Delta t \downarrow 0. \tag{2.48}$$

From 1. and consistency conditions 2a and 2b, we have that,

$$\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right)\right] = \sum_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} \mathbb{T}_{Y \to X}^{(k,l)}\left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t\right).$$

as shown in the proof of Theorem 5 implying

$$\mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right)\right] \leq M, \quad \forall \Delta t \in (0, \delta),$$

thus

$$KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \in L_1\left(\Omega, \mathcal{F}, \mathbb{P}\right), \quad \forall \Delta t \in (0, \delta). \tag{2.49}$$

Now using (2.48) and the uniform integrability of $\{h_{\Delta t}\}_{\Delta t > 0}$, we can apply the Vitali Convergence Theorem to obtain

$$||h_{\Delta t} - g||_{L_1} \to 0, \text{ as } \Delta t \downarrow 0$$

which implies that

$$\mathbb{E}_{\mathbb{P}}\left[h_{\Delta t}\right] \to \mathbb{E}_{\mathbb{P}}\left[g\right] \text{ as } \Delta t \downarrow 0. \tag{2.50}$$

Furthermore, observe that for all $\Delta t > 0$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[h_{\Delta t}\right] &= \mathbb{E}_{\mathbb{P}}\left[h_{\Delta t} \mathbb{1}_{J_{\Delta t}}\right] + \mathbb{E}_{\mathbb{P}}\left[h_{\Delta t} \mathbb{1}_{\overline{J_{\Delta t}}}\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \mathbb{1}_{J_{\Delta t}}\right] + 0,
\end{aligned}$$

and that

$$
\begin{aligned}
0 &\leq \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) - KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \mathbb{1}_{J_{\Delta t}} \right] \\
&= \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \mathbb{1}_{\overline{J_{\Delta t}}} \right] \\
&= \lim_{\substack{\Delta t' \downarrow 0 \\ \Delta t' \in (0,\delta)}} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t'}^{(\cdot)} || M_{\Delta t'}^{(\cdot)}\right) \mathbb{1}_{\overline{J_{\Delta t}}} \right] \\
&\leq \lim_{\substack{\Delta t' \downarrow 0 \\ \Delta t' \in (0,\delta)}} \mathbb{P}\left(\overline{J_{\Delta t}}\right) \left(\sup_{\substack{\Delta t' > 0 \\ \Delta t' \in (0,\delta)}} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t'}^{(\cdot)} || M_{\Delta t'}^{(\cdot)}\right) \right] \right) \\
&= 0.
\end{aligned}
$$

Now

$$
\begin{aligned}
\lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \right] &= \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \mathbb{1}_{J_{\Delta t}} \right] \\
&= \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}} \left[h_{\Delta t} \right]
\end{aligned}
$$

which from (2.50) implies that

$$
\lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \right] = \mathbb{E}_{\mathbb{P}} \left[\lim_{\Delta t \downarrow 0} KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \right]. \tag{2.51}
$$

As in the proof of Theorem 4. the result follows as

$$
\begin{aligned}
\mathcal{EPT}_{Y \to X}^{(s,r)} \big|_{t_0}^{T} &= \mathbb{E}_{\mathbb{P}} \left[\mathbb{E}_{P^{(\cdot)}} \left[\log \frac{d\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^{T} \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^{T}\}](\cdot)}{d\mathbb{P}_{X}^{(s)}[X_{t_0}^{T} \mid X_{t_0-s}^{t_0}](\cdot)} \right] \right] \\
&= \mathbb{E}_{\mathbb{P}} \left[\lim_{\Delta t \downarrow 0} KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \right] \\
&= \lim_{\Delta t \downarrow 0} \mathbb{E}_{\mathbb{P}} \left[KL\left(P_{\Delta t}^{(\cdot)} || M_{\Delta t}^{(\cdot)}\right) \right] \\
&= \lim_{\Delta t \downarrow 0} \left[\sum_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} \mathbb{T}_{Y \to X}^{(k,l),\Delta t} \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t \right) \right].
\end{aligned}
$$

($\Leftarrow$) Conversely, if

$$
\lim_{\Delta t \downarrow 0} \left[\sum_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} \mathbb{T}_{Y \to X}^{(k,l),\Delta t} \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t \right) \right] = \mathcal{EPT}_{Y \to X}^{(s,r)} \big|_{t_0}^{T},
$$

then from 2. we have $\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T}<\infty$, thus $\exists\delta>0$ such that

$$\begin{aligned}
&\Bigg|\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-i\Delta t\right)-\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T}\Bigg|<1\\
&\implies\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-i\Delta t\right)<1+\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T}=:M,
\end{aligned}$$

$\forall\Delta t\in(0,\delta)$ and the proof is complete. $\square$

Chapter 3

The transfer entropy rate

The generalization of information theoretic measures to the framework of information rates is a common paradigm in information theory. In this section we address the topic of instantaneous information transfer between processes using our methodology. We first provide a definition of transfer entropy rate using EPT as follows. It should be noted that a similar definition appears in [48].

Definition 3.0.1. For $t \in [t_0, T)$, define the *transfer entropy rate* from Y to X at t, denoted $\mathbb{T}^{(s,r)}_{Y\to X}(t)$, by

$$\mathbb{T}^{(s,r)}_{Y\to X}(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \left(\mathcal{EPT}^{(s,r)}_{Y\to X} \Big|_t^{t+\Delta t} \right). \tag{3.1}$$

whenever the limit in (3.1) exists.

Remark 8. Suppose the hypotheses of Theorem 5 hold for processes X and Y. If $t \in [t_0, T)$ and $\exists \delta > 0$ such that $\mathcal{EPT}^{(s,r)}_{Y\to X} |_t^{t+dt} < \infty$, $\forall dt \in (t, t+\delta)$, then

$$\begin{aligned} \mathbb{T}^{(s,r)}_{Y\to X}(t) &= \lim_{dt \downarrow 0} \frac{1}{dt} \left(\mathcal{EPT}^{(s,r)}_{Y\to X} \Big|_t^{t+dt} \right) \\ &= \lim_{\substack{dt \downarrow 0 \\ \Delta t \downarrow 0}} \left[\frac{1}{dt} \sum_{i=0}^{\lfloor \frac{t+dt}{\Delta t} \rfloor - \lfloor \frac{t}{\Delta t} \rfloor - 1} \mathbb{T}^{(k,l),\Delta t}_{Y\to X} \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor \Delta t - i\Delta t \right) \right]. \end{aligned}$$

Assuming some smoothness, we can recover the expected pathwise transfer entropy at any time given the rate by using the following straightforward result.

Lemma 7. If $[t_0, T] \ni t \mapsto \mathcal{EPT}^{(s,r)}_{Y\to X} |_{t_0}^t \in \mathcal{C}^1([t_0, T])$, then

$$\int_{t_0}^T \mathbb{T}^{(s,r)}_{Y\to X}(t) dt = \mathcal{EPT}^{(s,r)}_{Y\to X} \Big|_{t_0}^T .$$

Proof. From the Fundamental Theorem of Calculus, we have that

$$\begin{aligned}\int_{t_0}^{T} \mathbb{T}^{(s,r)}_{Y\to X}(t)dt &= \mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T} - \mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{t_0}\\ &= \mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T} - \mathbb{E}_{\mathbb{P}}\left[\log(1)\right]\\ &= \mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T}.\end{aligned}$$

□

Note that in the previous lemma we impose differentiablity not simply right-hand differentiability.

Lemma 8. Suppose t_0 and T are distinct elements of $\mathbb{T}$ and $r, s > 0$ satisfy $(t_0 - \max(s,r), T) \subset \mathbb{T}$. If Y is (s,r)-consistent upon X on $[t_0, T)$ and $\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{\cdot}$ is linear on $[t_0, T]$, then for any $t \in [t_0, T)$,

$$\mathbb{T}^{(s,r)}_{Y\to X}(t) = \frac{1}{T - t_0}\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T}.$$

Proof. From linearity $\mathbb{T}^{(s,r)}_{Y\to X}$ is constant, thus clearly $\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{\cdot} \in \mathcal{C}^1([t_0, T])$, thus from Lemma 7, we have

$$\begin{aligned}\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T} &= \int_{t_0}^{T} \mathbb{T}^{(s,r)}_{Y\to X}(t')dt'\\ &= (T - t_0)\,\mathbb{T}^{(s,r)}_{Y\to X}(t),\end{aligned}$$

for any $t \in [t_0, T)$ and the proof is complete. □

3.1 Application to stationary processes

Definition 3.1.1. Stochastic processes X and Y indexed over $\mathbb{T}$ are *conditionally stationary* if $\forall \omega \in \Omega, k \geq 1$, all collections of times $\{t_i\}_{0\leq i\leq k}$ of $\mathbb{T}$ such that $t_i < t_{i+1}$, and all $A \in \mathcal{X}$,

$$\begin{aligned}&\mathbb{P}\left(X_{t_{i+1}} \in A | X_{t_i}, \ldots X_{t_{i-k}}, Y_{t_i}, \ldots Y_{t_{i-k}}\right)(\omega) =\\ &\mathbb{P}\left(X_{t_{i+1}+\tau} \in A | X_{t_i+\tau}, \ldots X_{t_{i-k}+\tau}, Y_{t_i+\tau}, \ldots, Y_{t_{i-k}+\tau}\right)(\omega)\end{aligned} \tag{3.2}$$

for all $i \in [k-1], \tau > 0$.

Definition 3.1.2. Suppose k and l are positive integers. Stochastic processes X and Y on $\mathbb{T}$ are (k,l) - *order conditionally stationary processes* if $\forall\omega\in\Omega$, all collections of times $\{t_i\}_{0\le i\le \max(k,l)}$ of $\mathbb{T}$ such that $t_i<t_{i+1}$, and all $A\in\mathcal{X}$,

$$\begin{aligned}&\mathbb{P}\left(X_{t_{i+1}}\in A|X_{t_i},\dots X_{t_{i-k}},Y_{t_i},\dots Y_{t_{i-l}}\right)(\omega)=\\&=\mathbb{P}\left(X_{t_{i+1}+\tau}\in A|X_{t_i+\tau},\dots X_{t_{i-k}+\tau},Y_{t_i+\tau},\dots,Y_{t_{i-l}+\tau}\right)(\omega)\end{aligned}\tag{3.3}$$

for all $i\in[\max(k,l)-1],\tau>0$.

Observe that if X and Y are conditionally stationary processes, then they are by definition $(k,l)-$ order conditionally stationary for all $k,l\ge 1$. Moreover, if X and Y are stationary, then $\forall\Delta t>0$ and $s,r>0$ such that $[t_0-\max(s,r),T)\subset\mathbb{T}$, we have that X and Y are also $(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)-$ order conditionally stationary. We exploit this stationarity in the following key observation.

Observation 4. For any $\Delta t>0$ and $j=0,\cdots,\left\lfloor\frac{T}{\Delta t}\right\rfloor-\left\lfloor\frac{t_0}{\Delta t}\right\rfloor-1$, we have

$$\begin{aligned}&\sum_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-i\Delta t\right)\\
&=\mathbb{E}_{\mathbb{P}}\left[KL\left(\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}\middle\|\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor)}_{X|\overleftarrow{X},i,\Delta t}\right)\right]\\
&=\mathbb{E}_{\mathbb{P}}\left[KL\left(\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)}_{X|\overleftarrow{X},\overleftarrow{Y},j,\Delta t}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor-\left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\middle\|\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor)}_{X|\overleftarrow{X},j,\Delta t}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor-\left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\right)\right]\\
&=\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor-\left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\mathbb{E}_{\mathbb{P}}\left[KL\left(\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)}_{X|\overleftarrow{X},\overleftarrow{Y},j,\Delta t}\middle\|\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor)}_{X|\overleftarrow{X},j,\Delta t}\right)\right]\\
&=\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor-\left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\left\lfloor\frac{T}{\Delta t}\right\rfloor\Delta t-j\Delta t\right),\end{aligned}\tag{3.4}$$

where in the second to last equality we used that

$$\frac{d\left(c\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)}_{X|\overleftarrow{X},\overleftarrow{Y},j,\Delta t}\right)}{d\left(c\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor)}_{X|\overleftarrow{X},j,\Delta t}\right)}=\frac{d\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor,\lfloor\frac{r}{\Delta t}\rfloor)}_{X|\overleftarrow{X},\overleftarrow{Y},j,\Delta t}}{d\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor)}_{X|\overleftarrow{X},j,\Delta t}},\quad\mathbb{P}^{(\omega),(\lfloor\frac{s}{\Delta t}\rfloor)}_{X|\overleftarrow{X},j,\Delta t}\text{-a.s.}$$

for any $c \neq 0$ due to the a.s. uniqueness of the RN-derivative.

We can use Observation 3.4 to provide an expression for the transfer entropy rate for stationary processes that have (r, s) −consistency on subintervals of $[t_0, T)$ of the form $[t_0, t)$. It should be noted that a result similar to the statement of part 2 of the following corollary appears as a remark in [48] without proof.

Corollary 8.1. Suppose $[t_0, T) \subset \mathbb{T}$, $r, s > 0$ satisfy $(t_0 - \max(s, r), T) \in \mathbb{T}$ and X and Y are stationary processes such that Y is (s, r) - consistent upon X on $[t_0, t)$ and satisfies (2.31), $\forall t \in (t_0, T]$.

1. If $\forall t \in (t_0, T]$, $\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{T}^{(k,l),\Delta t}_{Y \to X}\left(\Delta t \left\lfloor \frac{t_1}{\Delta t} \right\rfloor\right)$ exists $\forall t_1 \in [t_0, t)$, then
$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{T}^{(k,l),\Delta t}_{Y \to X}\left(\Delta t \left\lfloor \frac{t_1}{\Delta t} \right\rfloor\right) = \frac{\mathcal{EPT}^{(s,r)}_{Y \to X} \mid_{t_0}^{t_1}}{t_1 - t_0}, \forall t_1 \in (t_0, t).$$
2. $\mathbb{T}^{(s,r)}_{Y \to X}(t) = \frac{1}{T - t_0} \mathcal{EPT}^{(s,r)}_{Y \to X} \mid_{t_0}^{T}$.

Proof. (Proof of 1.) Let $t_1 \in (t_0, t)$ with $t \in (t_0, T]$. Since $\lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{T}^{(k,l),\Delta t}_{Y \to X}\left(\Delta t \left\lfloor \frac{t_1}{\Delta t} \right\rfloor\right)$ exists, we have that

$$\lim_{\Delta t \to 0} \mathbb{T}^{(k,l),\Delta t}_{Y \to X}\left(\Delta t \left\lfloor \frac{t_1}{\Delta t} \right\rfloor\right) = \left(\lim_{\Delta t \downarrow 0} \Delta t\right)\left(\lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{T}^{(k,l),\Delta t}_{Y \to X}\left(\Delta t \left\lfloor \frac{t_1}{\Delta t} \right\rfloor\right)\right) = 0. \tag{3.5}$$

From Theorem 5 and (3.4) we have that

$$\begin{aligned} \infty > \ & \mathcal{EPT}^{(s,r)}_{Y \to X} \mid_{t_0}^{t_1} \\ & = \lim_{\Delta t \to 0} \sum_{i=0}^{\lfloor \frac{t_1}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{T}^{(k,l),\Delta t}_{Y \to X}\left(\Delta t \left\lfloor \frac{t_1}{\Delta t} \right\rfloor - i\Delta t\right) \qquad (3.6) \\ & = \lim_{\Delta t \to 0} \left(\left\lfloor \frac{t_1}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor\right) \mathbb{T}^{(k,l),\Delta t}_{Y \to X}\left(\Delta t \left\lfloor \frac{t_1}{\Delta t} \right\rfloor - j\Delta t\right), \end{aligned}$$

for any $j = 0, \cdots, \left\lfloor \frac{t_1}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$.
Note that for each $\Delta t > 0$, $\exists C_{\Delta t} \in (-2, 2)$ such that

$$\left\lfloor \frac{t_1}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor = \frac{t_1 - t_0}{\Delta t} + C_{\Delta t}.$$

Letting $j = 0$ in (3.6)

$$\begin{aligned}
&\lim_{\Delta t\to 0}\left(\left\lfloor\frac{t_1}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_1}{\Delta t}\right\rfloor\right)\\
&= \lim_{\Delta t\to 0}\left(\frac{t_1 - t_0}{\Delta t} + C_{\Delta t}\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_1}{\Delta t}\right\rfloor\right)\\
&= (t_1 - t_0)\lim_{\Delta t\to 0}\frac{1}{\Delta t}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_1}{\Delta t}\right\rfloor\right) + \lim_{\Delta t\to 0} C_{\Delta t}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_1}{\Delta t}\right\rfloor\right).
\end{aligned} \tag{3.7}$$

Since $C_{\Delta t}$ is bounded, $\lim_{\Delta t\to 0} C_{\Delta t}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_1}{\Delta t}\right\rfloor\right) = 0$. Now using (3.6) we get

$$(t_1 - t_0)\lim_{\Delta t\to 0}\frac{1}{\Delta t}\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_1}{\Delta t}\right\rfloor\right) = \mathcal{EPT}^{(s,r)}_{Y\to X}\big|_{t_0}^{t_1}$$

and the result follows from division by $t_1 - t_0$.

(Proof of 2.) Suppose t_1, t_2 are distinct elements of $[t_0, T]$. Without loss of generality, suppose $t_1 > t_2 \neq t_0$. Per assumption, X and Y are stationary processes such that Y is (s, r) - consistent upon X on $[t_0, t_1)$ and $[t_0, t_2)$. If

$$j' = \left\lfloor\frac{t_1}{\Delta t}\right\rfloor - \left\lfloor\frac{t_2}{\Delta t}\right\rfloor,$$

then from (3.4),

$$\begin{aligned}
&\mathcal{EPT}^{(s,r)}_{Y\to X}\big|_{t_0}^{t_1}\\
&= \lim_{\Delta t\to 0}\left(\left\lfloor\frac{t_1}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_1}{\Delta t}\right\rfloor - j'\Delta t\right)\\
&= \lim_{\Delta t\to 0}\left(\frac{t_1 - t_0}{\Delta t} + C_{\Delta t}\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_2}{\Delta t}\right\rfloor\right)\\
&= \lim_{\Delta t\to 0}\frac{t_1 - t_0}{t_2 - t_0}\left(\frac{t_1 - t_0 + \Delta t C_{\Delta t}}{(t_1 - t_0)\Delta t}\right)(t_2 - t_0)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_2}{\Delta t}\right\rfloor\right)\\
&= \frac{t_1 - t_0}{t_2 - t_0}\lim_{\Delta t\to 0}\left(\frac{\Delta t C_{\Delta t}}{(t_1 - t_0)}\right)\left(\left\lfloor\frac{t_2}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor - K_{\Delta t}\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_2}{\Delta t}\right\rfloor\right)\\
&+ \frac{t_1 - t_0}{t_2 - t_0}\lim_{\Delta t\to 0}\left(\left\lfloor\frac{t_2}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor - K_{\Delta t}\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_2}{\Delta t}\right\rfloor\right).
\end{aligned} \tag{3.8}$$

Per assumption, $\left(\left\lfloor\frac{t_2}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_2}{\Delta t}\right\rfloor\right)$ exists thus since $C_{\Delta t}$ and $K_{\Delta t}$ are bounded

$$\frac{t_1 - t_0}{t_2 - t_0}\lim_{\Delta t\to 0}\left(\frac{\Delta t C_{\Delta t}}{(t_1 - t_0)}\right)\left(\left\lfloor\frac{t_2}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right)\mathbb{T}^{(k,l),\Delta t}_{Y\to X}\left(\Delta t\left\lfloor\frac{t_2}{\Delta t}\right\rfloor\right) = 0$$

and

$$\frac{t_1 - t_0}{t_2 - t_0} \lim_{\Delta t \to 0} \left(\frac{\Delta t C_{\Delta t}}{(t_1 - t_0)} \right) K_{\Delta t} \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\Delta t \left\lfloor \frac{t_2}{\Delta t} \right\rfloor \right) = 0.$$

Now we have that

$$\begin{aligned} &\mathcal{EPT}^{(s,r)}_{Y \to X} \mid^{t_1}_{t_0} \\ &= \frac{t_1 - t_0}{t_2 - t_0} \lim_{\Delta t \to 0} \left(\frac{\Delta t C_{\Delta t}}{(t_1 - t_0)} \right) \left(\left\lfloor \frac{t_2}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - K_{\Delta t} \right) \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\Delta t \left\lfloor \frac{t_2}{\Delta t} \right\rfloor \right) \\ &+ \frac{t_1 - t_0}{t_2 - t_0} \lim_{\Delta t \to 0} \left(\left\lfloor \frac{t_2}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - K_{\Delta t} \right) \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\Delta t \left\lfloor \frac{t_2}{\Delta t} \right\rfloor \right) \\ &= \frac{t_1 - t_0}{t_2 - t_0} \lim_{\Delta t \to 0} \left(\left\lfloor \frac{t_2}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - K_{\Delta t} \right) \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\Delta t \left\lfloor \frac{t_2}{\Delta t} \right\rfloor \right) \end{aligned} \tag{3.9}$$

and since $\frac{t_1 - t_0}{t_2 - t_0} \lim_{\Delta t \to 0} K_{\Delta t} \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\Delta t \left\lfloor \frac{t_2}{\Delta t} \right\rfloor \right) = 0$, we have

$$\begin{aligned} &\mathcal{EPT}^{(s,r)}_{Y \to X} \mid^{t_1}_{t_0} = \frac{t_1 - t_0}{t_2 - t_0} \lim_{\Delta t \to 0} \left(\left\lfloor \frac{t_2}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor \right) \mathbb{T}^{(k,l),\Delta t}_{Y \to X} \left(\Delta t \left\lfloor \frac{t_2}{\Delta t} \right\rfloor \right) \\ &\implies \mathcal{EPT}^{(s,r)}_{Y \to X} \mid^{t_2}_{t_0} = \frac{t_2 - t_0}{t_1 - t_0} \mathcal{EPT}^{(s,r)}_{Y \to X} \mid^{t_1}_{t_0}, \end{aligned} \tag{3.10}$$

that is, $\mathcal{EPT}^{(s,r)}_{Y \to X} \mid^{t}_{t_0}$ is linear in $t - t_0$ and we get the result by applying Lemma 8. □

Simply put, Corollary 8.1 states that under stationarity in a rather strict sense, the TE rate is the average value of the expected pathwise transfer entropy.

3.2 Sufficient conditions for PT and EPT continuity

This section is devoted to the establishment of sufficient conditions for continuity of the pathwise and expected pathwise transfer entropy in time. Suppose $t \in [t_0, T), \omega \in \Omega$. Let

$$\mathbb{A}_\omega \left(x^t_{t_0} \right) = \left\{ \frac{\mathbb{P}^{(s,r)}_{X|X,\{Y\}} \left[A \middle| X^{t_0}_{t_0 - s}, \{Y^t_{t_0 - r}\} \right] (\omega)}{\mathbb{P}^{(s)}_X \left[A \middle| X^{t_0}_{t_0 - s} \right] (\omega)} \;\middle|\; A \in \mathcal{F}^{[t_0,t)}_X \text{and } x^t_{t_0} \in A \right\}$$

and let us denote by $a_\omega(x^t_{t_0})$ the limit point of $\mathbb{A}_\omega(x^t_{t_0})$, if it exists. From here on in, we will say that processes X and Y satisfy the *Piccioni condition* if $a_\omega(x^t_{t_0})$ exists and is unique. Due to [41], there exists a version of $\mathcal{PT}^{(s,r)}_{Y \to X} \mid^t_{t_0} (\omega, \cdot)$ which is continuous at $x_{t_0}^t$ and for which the equality $\mathcal{PT}_{Y\to X}^{(s,r)}\mid_{t_0}^t(\omega, x_{t_0}^t) = a_\omega(x_{t_0}^t)$ holds, implying that

$$\begin{aligned}
\mathcal{PT}_{Y\to X}^{(s,r)}\mid_{t_0}^t(\omega, x_{t_0}^t) &= \log\left(\frac{d\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^T \mid X_{t_o-s}^{t_0}, \{Y_{t_0-r}^T\}](\omega)}{d\mathbb{P}_X^{(s)}[X_{t_0}^T \mid X_{t_o-s}^{t_0}](\omega)}(x_{t_0}^T)\right) \\
&= \log\left(\lim_{\substack{\epsilon\downarrow 0\\ \Delta t\to 0}} \prod_{i=0}^{\lfloor \frac{T}{\Delta t}\rfloor - \lfloor \frac{t_0}{\Delta t}\rfloor - 1} \frac{\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(s,r)}\left(\pi_{\lfloor \frac{T}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^T,\epsilon\right)\right)\right)}{\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(s)}\left(\pi_{\lfloor \frac{T}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^T,\epsilon\right)\right)\right)}\right) \\
&= \lim_{\substack{\epsilon\downarrow 0\\ \Delta t\downarrow 0}} \sum_{i=0}^{\lfloor \frac{T}{\Delta t}\rfloor - \lfloor \frac{t_0}{\Delta t}\rfloor - 1} \log\left(\frac{\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(s,r)}\left(\pi_{\lfloor \frac{T}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^T,\epsilon\right)\right)\right)}{\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(s)}\left(\pi_{\lfloor \frac{T}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^T,\epsilon\right)\right)\right)}\right).
\end{aligned}$$

The following lemma proves continuity of the RN-derivative in time as opposed to continuity on path space $\Omega_X^{[t_0,T)}$.

Lemma 9. For $t \in [t_0, T)$ and $x_{t_0}^T \in \Omega_X^{[t_0,T)}$, let $C_t\left(x_{t_0}^T\right) = x_{t_0}^T\Big|_{[t_0,t]}$. If X and Y satisfy the Piccioni condition and all of their respective sample paths are elements of $\mathcal{C}^0([t_0,T))$, then for each $\omega\in\Omega$, there exists a version of $\mathcal{PT}_{Y\to X}^{(s,r)}\mid_{t_0}^{\cdot}(\omega,\cdot)$ such that $\forall x_{t_0}^T \in \Omega_X^{[t_0,T)}$,

$$t \mapsto \mathcal{PT}_{Y\to X}^{(s,r)}\mid_{t_0}^t\left(\omega, C_t\left(x_{t_0}^T\right)\right) \in \mathcal{C}^0([t_0,t)).$$

Proof. Fix $t\in[t_0,T)$, $x_{t_0}^T \in \Omega_X^{[t_0,T)}, \omega\in\Omega$ and note that $\forall dt > 0$, we have

$$\begin{aligned}
&\left|\mathcal{PT}_{Y\to X}^{(s,r)}\mid_{t_0}^{t+dt}\left(\omega, C_{t+dt}\left(x_{t_0}^T\right)\right) - \mathcal{PT}_{Y\to X}^{(s,r)}\mid_{t_0}^{t}\left(\omega, C_{t}\left(x_{t_0}^T\right)\right)\right| \\
&= \left|\log\left(\frac{d\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^{t+dt} \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^{t+dt}\}](\omega)}{d\mathbb{P}_X^{(s)}[X_{t_0}^{t+dt} \mid X_{t_0-s}^{t_0}](\omega)}\left(x_{t_0}^{t+dt}\right)\right)\right. \\
&\qquad \left. - \log\left(\frac{d\mathbb{P}_{X|X,\{Y\}}^{(s,r)}[X_{t_0}^{t} \mid X_{t_0-s}^{t_0}, \{Y_{t_0-r}^{t}\}](\omega)}{d\mathbb{P}_X^{(s)}[X_{t_0}^{t} \mid X_{t_0-s}^{t_0}](\omega)}\left(x_{t_0}^{t}\right)\right)\right| \\
&= \log\left(\frac{\lim\limits_{\substack{\epsilon\downarrow 0\\ \Delta t\downarrow 0}} \prod\limits_{i=0}^{\lfloor \frac{t+dt}{\Delta t}\rfloor - \lfloor \frac{t_0}{\Delta t}\rfloor - 1} \frac{\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(s,r)}\left(\pi_{\lfloor \frac{t+dt}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^{t+dt},\epsilon\right)\right)\right)}{\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(s)}\left(\pi_{\lfloor \frac{t+dt}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^{t+dt},\epsilon\right)\right)\right)}}{\lim\limits_{\substack{\epsilon\downarrow 0\\ \Delta t\downarrow 0}} \prod\limits_{i=0}^{\lfloor \frac{t}{\Delta t}\rfloor - \lfloor \frac{t_0}{\Delta t}\rfloor - 1} \frac{\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(s,r)}\left(\pi_{\lfloor \frac{t}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^{t},\epsilon\right)\right)\right)}{\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(s)}\left(\pi_{\lfloor \frac{t}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t_0}^{t},\epsilon\right)\right)\right)}}\right) \\
&= \log\left(\lim_{\substack{\epsilon\downarrow 0\\ \Delta t\downarrow 0}} \prod_{i=0}^{\lfloor \frac{t+dt}{\Delta t}\rfloor - \lfloor \frac{t}{\Delta t}\rfloor - 1} \frac{\mathbb{P}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}^{(\omega),(s,r)}\left(\pi_{\lfloor \frac{t+dt}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t}^{t+dt},\epsilon\right)\right)\right)}{\mathbb{P}_{X|\overleftarrow{X},i,\Delta t}^{(\omega),(s)}\left(\pi_{\lfloor \frac{t}{\Delta t}\rfloor \Delta t - i\Delta t}\left(\mathcal{B}\left(x_{t}^{t+dt},\epsilon\right)\right)\right)}\right).
\end{aligned}$$

Observe that since the sample paths of X are continuous, the sample x_t^{t+dt} is uniformly continuous on $[t, t+dt]$, thus for each $\epsilon > 0, \exists\delta(\epsilon) > 0$ s.t. if $\Delta t < \delta(\epsilon)$, then

$$\frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}\left(\pi_{\lfloor\frac{t+dt}{\Delta t}\rfloor\Delta t - i\Delta t}\left(\mathcal{B}\left(x_t^{t+dt},\epsilon\right)\right)\right)}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t}\left(\pi_{\lfloor\frac{t}{\Delta t}\rfloor\Delta t - i\Delta t}\left(\mathcal{B}\left(x_t^{t+dt},\epsilon\right)\right)\right)} = 1,$$

$\forall i = 0, \ldots, \left\lfloor\frac{t+dt}{\Delta t}\right\rfloor - \left\lfloor\frac{t}{\Delta t}\right\rfloor - 1$, hence

$$\begin{aligned}
&\lim_{dt\downarrow 0}\left|\mathcal{PT}^{(s,r)}_{Y\to X}\,|^{t+dt}_{t_0}\left(\omega, C_{t+dt}\left(x^T_{t_0}\right)\right) - \mathcal{PT}^{(s,r)}_{Y\to X}\,|^{t}_{t_0}\left(\omega, C_{t}\left(x^T_{t_0}\right)\right)\right| \\
&= \lim_{dt\downarrow 0}\log\left(\lim_{\substack{\epsilon\downarrow 0\\ \Delta t\downarrow 0}}\prod_{i=0}^{\lfloor\frac{t+dt}{\Delta t}\rfloor - \lfloor\frac{t}{\Delta t}\rfloor - 1}\frac{\mathbb{P}^{(\omega),(s,r)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}\left(\pi_{\lfloor\frac{t+dt}{\Delta t}\rfloor\Delta t - i\Delta t}\left(\mathcal{B}\left(x_t^{t+dt},\epsilon\right)\right)\right)}{\mathbb{P}^{(\omega),(s)}_{X|\overleftarrow{X},i,\Delta t}\left(\pi_{\lfloor\frac{t}{\Delta t}\rfloor\Delta t - i\Delta t}\left(\mathcal{B}\left(x_t^{t+dt},\epsilon\right)\right)\right)}\right) \\
&= \lim_{dt\downarrow 0}\log\left(\prod_{i=0}^{\lfloor\frac{t+dt}{\Delta t}\rfloor - \lfloor\frac{t}{\Delta t}\rfloor - 1} 1\right) \\
&= 0,
\end{aligned}$$

proving continuity. $\square$

A natural question arising from Lemma 9 is the question of when the expected pathwise transfer entropy is continuous in time. The following lemma provides an answer.

Lemma 10. For each $t \in [t_0, T) \subset \mathbb{T}$ and $\omega \in \Omega$, let

$$KL\left(t,\omega\right) = KL\left(\mathbb{P}^{(s,r)}_{X|X,\{Y\}}[X^t_{t_0} \mid X^{t_0}_{t_o-s}, \{Y^t_{t_0-r}\}](\omega)\middle\|\mathbb{P}^{(s)}_{X|X}[X^t_{t_0} \mid X^{t_0}_{t_o-s}](\omega)\right).$$

If

1. $t \mapsto KL(t,\omega)$ is continuous in t for a.e. $\omega \in \Omega$,

2. $\{KL(t,\cdot)\}_{t\in[t_0,T)}$ is a UI family.

then $t \mapsto \mathcal{EPT}^{(s,r)}_{Y\to X}\,|^t_{t_0}$ is continuous on $[t_0, T)$.

Proof. Let $t \in [t_0, T)$. It suffices to show that

$$\mathcal{EPT}^{(s,r)}_{Y\to X}\,|^{t_n}_{t_0} \to \mathcal{EPT}^{(s,r)}_{Y\to X}\,|^t_{t_0}$$

for any sequence $\{t_n\}_{n\geq 1}$ converging to t as $n \to \infty$. Suppose $t_n \to t$, as $n \to \infty$, $\epsilon > 0$ and for $n \geq 1$ define

$$S_{n,\epsilon} = \{\omega \in \Omega \quad | \quad |KL(t_n, \omega) - KL(t, \omega)| \geq \epsilon\}.$$

From 1., for a.e. $\omega \in \Omega, \exists N \geq 1$ such that $\omega \in \overline{S_{n,\epsilon}}, \forall n \geq N$. Thus

$$1 = \mathbb{P}(\Omega \setminus B) \leq \mathbb{P}\left(\bigcup_{n\geq 1} \overline{S_{n,\epsilon}}\right) \leq 1 \implies \mathbb{P}\left(\bigcup_{n\geq 1} \overline{S_{n,\epsilon}}\right) = 1,$$

where B is the $\mathbb{P}$-null set such that $t \mapsto KL(t, \omega)$ is discontinuous on $[t_0, T)$ for any $\omega \in B$. Observe that $\{\overline{S_{n,\epsilon}}\}_{n\geq 1}$ is an increasing sequence of events, thus

$$\lim_{n\to\infty} \mathbb{P}\left(\overline{S_{n,\epsilon}}\right) = \mathbb{P}\left(\bigcup_{n\geq 1} \overline{S_{n,\epsilon}}\right) = 1$$

and so

$$\mathbb{P}(S_{n,\epsilon}) \to 0, \text{ as } n \to \infty.$$

Thus $\forall t \in [t_0, T)$,

$$KL(t_n, \cdot) \xrightarrow{\mathbb{P}} KL(t, \cdot)$$

for all sequences $t_n \to t$, as $n \to \infty$. Applying the Vitali Convergence Theorem (Theorem 7.29 in [16]), we get that

$$\mathbb{E}_{\mathbb{P}}[KL(t_n, \omega) - KL(t, \omega)] \leq \mathbb{E}_{\mathbb{P}}[|KL(t_n, \omega) - KL(t, \omega)|] \to 0 \text{ as } n \to \infty$$

proving continuity. $\square$

While Lemma 9 and Lemma 10 provide sufficient conditions for continuity of PT and EPT, differentiability of these functions is an open problem.

Chapter 4

Cadlag processes

This section is devoted to the investigation of EPT between cadlag processes, a scenario ubiquitous in the literature concerning the application of TE to neural spike trains common to neuroscience. To this end, we define a cadlag process as follows.

Definition 4.0.1. A stochastic process X is *cadlag* if its sample paths are right-continuous with left limits with probability one.

Examples of such processes are Levy processes and Poisson processes. Suppose now that X and Y are cadlag processes. We can specify a sample path of either process by providing its transition times and states, specifically, for any realization $x_{t_0}^T$ of $X_{t_0}^T$, there exists $t_0 \leq t_1 < \ldots < T$ such that we can write

$$x_{t_0}^T = \{\{t_i, x_{t_i}\}_{i=0}^{N_X^{[t_0,T)}(x_{t_0}^T)}\}, \tag{4.1}$$

where $N_X^{[t_0,T)}(x_{t_0}^T) = \left| Range\left(x_{t_0}^T\right) \right| - 1$ and $x_{t_i} = x\left(t_i\right)$.
Furthermore, we define conditional escape and transition rates similar to those in [48] as follows.

Definition 4.0.2. For cadlag processes $X = (X_t)_{t\in[t_0,T)}$ and $Y = (Y_t)_{t\in[t_0,T)}$, with Σ countable, define for all $\omega \in \Omega, t \in [t_0, T), r, s > 0$, and $x' \in \Sigma$ the *conditional transition rate of X given X and Y of x' at t*, denoted $\psi\left[x'\middle|\overleftarrow{X}, \overleftarrow{Y}\right](t,\omega)$ by

$$\psi\left[x'\middle|\overleftarrow{X}, \overleftarrow{Y}\right](t,\omega) =$$
$$\lim_{\Delta t\downarrow 0} \frac{1}{\Delta t}\mathbb{P}\left(\{\omega' \in \Omega \mid X_{t'}(\omega') = x', \text{ for some } t' \in [t, t+\Delta t)\} \mid X_{t^- -s}^{t^-}, Y_{t^- -r}^{t^-}\right)(\omega), \tag{4.2}$$

the *conditional transition rate of X given $\overleftarrow{X}$ of x' at t*, denoted $\psi\left[x'\middle|\overleftarrow{X}\right](t,\omega)$ by

$$\begin{aligned}&\psi\left[x'\middle|\overleftarrow{X}\right](t,\omega) = \\ \lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}\mathbb{P}&\left(\left\{\omega'\in\Omega\middle|X_{t'}(\omega')=x',\ \text{for some } t'\in[t,t+\Delta t)\right\}\mid X_{t^--s}^{t^-}\right)(\omega),\end{aligned} \tag{4.3}$$

and the *conditional escape rates* $\lambda^{(s)}_{X|X}(t,\omega)$ and $\lambda^{(s,r)}_{X|X,Y}(t,\omega)$ by

$$\lambda^{(s)}_{X|X}(t,\omega) = \sum_{x'\in\Sigma, x'\neq x_t^-}\psi\left[x'\middle|\overleftarrow{X}\right](t,\omega) \tag{4.4}$$

and

$$\lambda^{(s,r)}_{X|X,Y}(t,\omega) = \sum_{x'\in\Sigma, x'\neq x_t^-}\psi\left[x'\middle|\overleftarrow{X},\overleftarrow{Y}\right](t,\omega). \tag{4.5}$$

In the forthcoming, we will sometimes regard the conditional transition rates defined above as measures on the space $(\Sigma,\mathcal{X})$ for fixed $\omega\in\Omega, t\in\mathbb{T}$ in agreeance with standard definitions of transition kernels (see Section 1.2 of [25]).

Notation 6. for $t\in[t_0,T),\omega\in\Omega$,and $s,r>0$, let

$$\Delta\lambda^{(s,r)}(t,\omega) = \lambda^{(s)}_{X|X}(t,\omega) - \lambda^{(s,r)}_{X|X,Y}(t,\omega).$$

We restrict our attention to TE between time-homogeneous Markov processes.

Definition 4.0.3. Suppose $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space, $\mathbb{T}\subset\mathbb{R}_{\geq 0}$ is a bounded and closed interval, Σ is a countable set, and $\mathcal{X}$ is a $\sigma-$algebra of subsets of Σ containing all singletons of Σ. A stochastic process $X=(X_t)_{t\in\mathbb{T}}$ is a *time-homogeneous Markov jump process* if all of its sample paths are piecewise constant and right-continuous and $\forall n\geq 1$, times $t_0<t_1<\cdots<t_{n-1}$ and sets A_i with $t_i\in\mathbb{T}$, $A_i\in\mathcal{X}$, $\forall 0\leq i\leq n$,

$$\begin{aligned}&\mathbb{P}_{t_{n-1}+\tau}\left[X_{t_{n-1}+\tau}\in A_{n-1}\middle|X_{t_{n-2}+\tau},\cdots,X_{t_0+\tau}\right](\omega)\\ &=\mathbb{P}_{t_{n-1}+\tau}\left[X_{t_{n-1}+\tau}\in A_{n-1}\middle|X_{t_{n-2}+\tau}\right](\omega)\\ &=\mathbb{P}_{t_{n-1}}\left[X_{t_{n-1}}\in A_{n-1}\middle|X_{t_{n-2}}\right](\omega),\end{aligned}$$

for each $\omega\in\Omega$ and all $\tau\geq 0$ such that $t_{i-1+\tau}\in\mathbb{T}$ for $0\leq i\leq n$.

We now present a Girsanov formula for the pathwise transfer entropy when the destination process is a time-homogeneous Markov jump process and the source process is any cadlag process.

Theorem 11. Suppose X and Y are cadlag processes on $\mathbb{T}$ with $[t_0, T) \subset \mathbb{T}$ and Σ countable, where X is a time-homogeneous Markov process with conditional transition rates given by (4.2) and (4.3) and conditional escape rates given by (4.5) and (4.4). If

1. $\forall \omega \in \Omega$, $\psi\left[x_{t_0} \middle| \overleftarrow{X}, \overleftarrow{Y}\right](t_0, \omega) = \psi\left[x_{t_0} \middle| \overleftarrow{X}\right](t_0, \omega)$.
2. The conditional escape rates are bounded and positive.
3. $\psi\left[\cdot \middle| \overleftarrow{X}, \overleftarrow{Y}\right](t, \omega) \ll \psi\left[\cdot \middle| \overleftarrow{X}\right](t, \omega), \forall \omega \in \Omega$ and $t \in [t_0, T)$.

Then $\forall \omega \in \Omega$, we have

$$\mathcal{PT}^{(s,r)}_{Y \to X} \big|_{t_0}^{T} \left(\omega, x_{t_0}^T\right) = \sum_{i=1}^{N_X^{[t_0,T)}(x_{t_0}^T)} \log\left[\frac{\psi\left[x_{\tau_i} \middle| \overleftarrow{X}, \overleftarrow{Y}\right](\tau_i, \omega)}{\psi\left[x_{\tau_i} \middle| \overleftarrow{X}\right](\tau_i, \omega)}\right] + \int_{t_0}^{T} \left(\Delta\lambda^{(s,r)}(t, \omega)\right) dt. \tag{4.6}$$

for every realization $x_{t_0}^T$ of the process $X_{t_0}^T$.

Proof. Since X is Markov, there exists an increasing sequence of finite random jump times $\{\tau_n\}_{n\geq 0}$ such that $\tau_0 = t_0$, X_{τ_n} is constant on $[\tau_n, \tau_{n+1})$, and $X_{\tau_n^-} \neq X_{\tau_n}$. Furthermore, from the Markov assumption, conditionally on $\{X_{\tau_n}\}_{n\geq 0}$, the variables $\{\tau_{n+1} - \tau_n\}_{n\geq 0}$ are independent and exponentially distributed.
We first need to show that for arbitrary measures $P \ll Q$ on the path space of cadlag sample paths of X with transition probabilities $p_P(\cdot,\cdot), p_Q(\cdot,\cdot)$ and escape rates γ_P, γ_Q , that for every realization $x_{t_0}^T$ of the process $X_{t_0}^T$,

$$\frac{dP}{dQ}\left(x_{t_0}^T\right) = \sum_{i=0}^{N_X^{[t_0,T)}(x_{t_0}^T)} \log\left[\frac{\gamma_P(x_{\tau_i}^-) p_P\left(x_{\tau i}^-, x_{\tau i}\right)}{\gamma_Q(x_{\tau_i}^-) p_Q\left(x_{\tau i}^-, x_{\tau_i}\right)}\right] + \int_{t_0}^{T} \left(\gamma_Q(x_t) - \gamma_P(x_t^-)\right) dt, \tag{4.7}$$

where $\{\tau_i\}_{i=0}^{N_X^{[t_0,T)}}$ is the sequence of jump times of the realization $x_{t_0}^T$. A proof of (4.7) is given in Appendix 1, Proposition 2.6 of [27].

Now letting P and Q be the measures in (4.2) and (4.3), respectively, using assumption 1., and noting that

$$\frac{\psi\left[x_{\tau_i}\middle|\overleftarrow{X},\overleftarrow{Y}\right](\tau_i,\omega)}{\lambda^{(s,r)}_{X|X,Y}(\tau_i,\omega)} = p_{X|X,Y}(x_{\tau_i},x_{\tau_i^-},Y_{\tau_i^-}(\omega))$$

and

$$\frac{\psi\left[x_{\tau_i}\middle|\overleftarrow{X}\right](\tau_i,\omega)}{\lambda^{(s)}_{X|X}(\tau_i,\omega)} = p_{X|X}(x_{\tau_i},x_{\tau_i^-})$$

where $p_{X|X,Y}$ and $p_{X|X}$ denote conditional transition probabilities, we get that

$$\begin{aligned}
&\mathcal{PT}^{(s,r)}_{Y\to X}\Big|_{t_0}^{T}\left(\omega,x_{t_0}^T\right) = \\
&\sum_{i=0}^{N_X^{[t_0,T)}\left(x_{t_0}^T\right)} \log\left[\frac{\left(\lambda^{(s,r)}_{X|X,Y}(\tau_i,\omega)\right)\left(p_{X|X,Y}(x_{\tau_i},x_{\tau_i^-},Y_{\tau_i^-}(\omega))\right)}{\left(\lambda^{(s)}_{X|X}(\tau_i,\omega)\right)\left(p_{X|X}(x_{\tau_i},x_{\tau_i^-})\right)}\right] + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\omega)\right)dt \\
&= \sum_{i=0}^{N_X^{[t_0,T)}\left(x_{t_0}^T\right)} \log\left[\frac{\psi\left[x_{\tau_i}\middle|\overleftarrow{X},\overleftarrow{Y}\right](\tau_i,\omega)}{\psi\left[x_{\tau_i}\middle|\overleftarrow{X}\right](\tau_i,\omega)}\right] + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\omega)\right)dt \\
&= \log\left[\frac{\psi\left[x_0\middle|\overleftarrow{X},\overleftarrow{Y}\right](\tau_0,\omega)}{\psi\left[x_0\middle|\overleftarrow{X}\right](\tau_0,\omega)}\right] + \sum_{i=1}^{N_X^{[t_0,T)}\left(x_{t_0}^T\right)} \log\left[\frac{\psi\left[x_{\tau_i}\middle|\overleftarrow{X},\overleftarrow{Y}\right](\tau_i,\omega)}{\psi\left[x_{\tau_i}\middle|\overleftarrow{X}\right](\tau_i,\omega)}\right] \\
&\quad + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\omega)\right)dt \\
&= \sum_{i=1}^{N_X^{[t_0,T)}(x_{t_0}^T)} \log\left[\frac{\psi\left[x_{\tau_i}\middle|\overleftarrow{X},\overleftarrow{Y}\right](\tau_i,\omega)}{\psi\left[x_{\tau_i}\middle|\overleftarrow{X}\right](\tau_i,\omega)}\right] + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\omega)\right)dt.
\end{aligned} \tag{4.8}$$

□

The conclusion of Theorem 11 holds for Feller processes (See Theorem 3.13 of [18].) under some conditions that imply absolute continuity.

Observation 5. Note that under the hypotheses of Theorem 11,

$$
\begin{aligned}
&\mathcal{PT}^{(s,r)}_{Y\to X}\,|_{t_0}^{T}\left(\omega, x_{t_0}^T\right) = \\
&\sum_{i=1}^{N_X^{[t_0,T)}(\omega)} \log\left[\frac{\psi\left[x_{\tau_i}\middle|\overleftarrow{X},\overleftarrow{Y}\right](\tau_i,\omega)}{\psi\left[x_{\tau_i}\middle|\overleftarrow{X}\right](\tau_i,\omega)}\right] + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\omega)\right)dt \\
&= \int_{t_0}^{T} \log\left[\frac{\psi\left[x_t\middle|\overleftarrow{X},\overleftarrow{Y}\right](t,\omega)}{\psi\left[x_t\middle|\overleftarrow{X}\right](t,\omega)}\right] dN_X^{[t_0,t)}(\omega) + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\omega)\right)dt \\
&= \int_{t_0}^{T}\left(\lambda^{(s,r)}_{X|X,Y}(t,\omega)\log\left[\frac{\psi\left[x_t\middle|\overleftarrow{X},\overleftarrow{Y}\right](t,\omega)}{\psi\left[x_t\middle|\overleftarrow{X}\right](t,\omega)}\right]\right)dt + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\omega)\right)dt \\
&= \int_{t_0}^{T}\left[\left(\lambda^{(s,r)}_{X|X,Y}(t,\omega)\right)\left(\log\left[\frac{\psi\left[x_t\middle|\overleftarrow{X},\overleftarrow{Y}\right](t,\omega)}{\psi\left[x_t\middle|\overleftarrow{X}\right](t,\omega)}\right]-1\right) + \lambda^{(s)}_{X|X}(t,\omega)\right]dt,
\end{aligned}
\tag{4.9}
$$

where the second to last equality comes from the observation that the process $\left(N_X^{[t_0,t)}(\cdot) - \int_{t_0}^{T}\lambda_{X|X,Y}(t,\cdot)\right)_{t\in[t_0,T)}$ is a mean-zero martingale from Watanabe's well-known martingale characterization of Poisson processes (see pp. 225 - 235 of [9]). Furthermore,

$$
\begin{aligned}
&\mathcal{EPT}^{(s,r)}_{Y\to X}\,|_{t_0}^{T} = \\
&\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}^{(s,r)}_{X|X,\{Y\}(\cdot)}}\left[\int_{t_0}^{T}\left(\lambda^{(s,r)}_{X|X,Y}(t,\cdot)\log\left[\frac{\psi\left[x_t\middle|\overleftarrow{X},\overleftarrow{Y}\right](t,\cdot)}{\psi\left[x_t\middle|\overleftarrow{X}\right](t,\cdot)}\right] + \int_{t_0}^{T}\left(\Delta\lambda^{(s,r)}(t,\cdot)\right)\right)dt\right]\right]
\end{aligned}
\tag{4.10}
$$

Corollary 11.1. If X is a cadlag process on $[t_0, T)$, such that the hypotheses of Theorem 11 hold, then $\forall t \in [t_0, T)$, the transfer entropy rate, $\mathbb{T}^{(s,r)}_{Y\to X}(t)$, is given by

$$
\begin{aligned}
&\mathbb{T}^{(s,r)}_{Y\to X}(t) = \\
&\quad\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}^{(s,r)}_{X|X,\{Y\}(\cdot)}}\left[\left(\lambda^{(s,r)}_{X|X,Y}(t,\cdot)\right)\left(\log\left[\frac{\psi\left[x_t\middle|\overleftarrow{X},\overleftarrow{Y}\right](t,\cdot)}{\psi\left[x_t\middle|\overleftarrow{X}\right](t,\cdot)}\right]-1\right)+\lambda^{(s)}_{X|X}(t,\cdot)\right]\right].
\end{aligned}
\tag{4.11}
$$

Proof. Let $\widetilde{\psi}_{t,\omega} = \psi\left[x_t\middle|\overleftarrow{X},\overleftarrow{Y}\right](t,\omega)$ and $\bar{\psi}_{t,\omega} = \psi\left[x_t\middle|\overleftarrow{X}\right](t,\omega)$ for each $t \in [t_0, T)$

and $\omega \in \Omega$. From Theorem 11 and Observation 5,

$$\mathbb{T}^{(s,r)}_{Y \to X}(t) =$$

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{E}_{\mathbb{P}} \left[\mathbb{E}_{\mathbb{P}^{(s,r)}_{X|X,\{Y\}(\omega)}} \left[\int_t^{t+\Delta t} \left[\left(\lambda^{(s,r)}_{X|X,Y}(t', \omega) \right) \left(\log \left[\frac{\widetilde{\psi}_{t',\omega}}{\bar{\psi}_{t',\omega}} \right] - 1 \right) + \lambda^{(s)}_{X|X}(t', \omega) \right] dt' \right] \right]$$
$$= \mathbb{E}_{\mathbb{P}} \left[\mathbb{E}_{\mathbb{P}^{(s,r)}_{X|X,\{Y\}(\omega)}} \left[\left(\lambda^{(s,r)}_{X|X,Y}(t, \omega) \right) \left(\log \left[\frac{\widetilde{\psi}_{t,\omega}}{\bar{\psi}_{t,\omega}} \right] - 1 \right) + \lambda^{(s)}_{X|X}(t, \omega) \right] \right] \tag{4.12}$$

where the last equality comes from Theorem A16.1 in [53].

□

4.1 Thinned Poisson point process

In this section we present an expression for the PT and EPT from a *time-homogeneous point process* to a *thinned* version of the process. The following definitions make these notions precise.

Definition 4.1.1. A point process $\Psi = (T_n)_{n \geq 1}$ on a nonempty set $\mathbb{A}$ is a *time - homogeneous Poisson point process (THPPP) with intensity* λ if and only if $T_j - T_{j-1} \sim exp(\lambda), \forall j \geq 1$ and the random variables $T_1, T_2 - T_1, \ldots T_i - T_{i-1}, \ldots$ are independent.

Definition 4.1.2. Suppose $\Psi = (T_n)_{n \geq 1}$ is a THPPP with intensity λ on a nonempty set $\mathbb{A}$. The *Counting Process of* Ψ is the process $(X_t)_{t \in \mathbb{A}}$, where X_t is the random variable defined by

$$X_t(\omega) = \sum_{n \geq 1} \mathbb{1}_{\{T_n \in (0,t)\}}(\omega)$$

Definition 4.1.3. For any given time-homogeneous Poisson point process (THPPP) $\Psi_1 = \left(T_n^{\Psi_1}\right)_{n \geq 1}$ and $p \in (0, 1)$, the process $\Psi_2 = \left(T_n^{\Psi_2}\right)_{n \geq 1}$ is called a *p-thinning* of Ψ_1 if

1. every arrival (point) that occurs in Ψ_2 also occurs in Ψ_1 a.s.

2. every arrival (point) that occurs in Ψ_2 also occurs in Ψ_1 with probability p independently of Ψ_2

We now give a result for TE between the counting processes of a THPPP and a thinned version of said THPPP if history windows are the same.

Corollary 11.2. Suppose Y is the counting process of a time-homogeneous Poisson point process Ψ with intensity λ on $[t_0, T)$ and X is the counting process of a p - thinning of Ψ for some $p \in (0,1)$. If $r = s$ and $X_{t_0} = Y_{t_0}$ a.s., then $\forall \omega \in \Omega$ and all realizations $x_{t_0}^T$ of $X_{t_0}^T$,

$$\mathcal{PT}_{Y \to X}^{(s,r)} \mid_{t_0}^{T} \left(\omega, x_{t_0}^T\right) = \log(p) N_X^{[t_0,T)}(x_{t_0}^T) + (1-p) \int_{t_0}^{T} \left(\lambda_{X|X}^{(s)}(t,\omega)\right) dt.$$

Furthermore,

$$\mathcal{EPT}_{Y \to X}^{(s,r)} \mid_{t_0}^{T} = \lambda \log(p) \left(T - t_0\right) + (1-p)\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}_{X|X,Y}}\left[\int_{t_0}^{T} \left(\lambda_{X|X}^{(s)}(t,\omega)\right) dt.\right]\right]$$

Proof. Note that any p-thinning of an intensity λ THPPP Ψ is also a THPPP with intensity $p\lambda$, thus both X and Y are time-homogeneous Markov processes. From Exercise 6.2.12 in [30], we have that $\forall t \in [t_0, T), \omega \in \Omega, x' \in \Sigma$,

$$\psi\left[x' \middle| \overleftarrow{X}, \overleftarrow{Y}\right](t,\omega) = p\psi\left[x' \middle| \overleftarrow{X}\right](t,\omega) \tag{4.13}$$

Applying Theorem 11, we get that

$$\begin{aligned} \mathcal{PT}_{Y \to X}^{(s,r)} \mid_{t_0}^{T} \left(\omega, x_{t_0}^T\right) &= \sum_{i=1}^{N_X^{[t_0,T)}(x_{t_0}^T)} \log\left[\frac{\psi\left[x_{\tau_i} \middle| \overleftarrow{X}, \overleftarrow{Y}\right](\tau_i,\omega)}{\psi\left[x_{\tau_i} \middle| \overleftarrow{X}\right](\tau_i,\omega)}\right] + \int_{t_0}^{T} \left(\Delta\lambda^{(s,r)}(t,\omega)\right) dt \\ &= \sum_{i=1}^{N_X^{[t_0,T)}(x_{t_0}^T)} \log[p] + \int_{t_0}^{T} \left(\Delta\lambda^{(s,r)}(t,\omega)\right) dt \\ &= \log(p) N_X^{[t_0,T)}(x_{t_0}^T) + \int_{t_0}^{T} \left(\Delta\lambda^{(s,r)}(t,\omega)\right) dt \\ &= \log(p) N_X^{[t_0,T)}(x_{t_0}^T) + \int_{t_0}^{T} \left(\lambda_{X|X}^{(s)}(t,\omega) - p\lambda_{X|X}^{(s)}(t,\omega)\right) dt \\ &= \log(p) N_X^{[t_0,T)}(x_{t_0}^T) + (1-p)\int_{t_0}^{T} \left(\lambda_{X|X}^{(s)}(t,\omega)\right) dt, \end{aligned} \tag{4.14}$$

where we have used (4.13) to get the second to last equality. Note that

$$\mathbb{E}_{\mathbb{P}}\left[N_X^{[t_0,T)}\right] = \lambda\left(T - t_0\right)$$

since X is a Poisson Process, thus

$$\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T} = \lambda \log(p)\left(T - t_0\right) + (1-p)\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}_{X|X,Y}}\left[\int_{t_0}^{T}\left(\lambda^{(s)}_{X|X}(t,\omega)\right)dt.\right]\right].$$

□

Remark 9. From Corollary 11.2, we obtain the TE rate in this case by applying the definition for any $t \in [t_0, T)$ and get

$$\begin{aligned}
&\mathbb{T}^{(s,r)}_{Y\to X}(t)\\
&= \lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}\left(\lambda \log(p)\left(t+\Delta t - t\right) + (1-p)\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}_{X|X,Y}}\left[\int_{t}^{t+\Delta t}\left(\lambda^{(s)}_{X|X}(t',\omega)\right)dt'.\right]\right]\right)\\
&= \lambda\log(p) + (1-p)\lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}_{X|X,Y}}\left[\int_{t}^{t+\Delta t}\left(\lambda^{(s)}_{X|X}(t',\omega)\right)dt'\right]\right].
\end{aligned} \tag{4.15}$$

4.2 Application: Lagged Poisson point process

In the forthcoming, we provide an example of two processes which satisfy (2.46) for some $\gamma > 0$ in a particular case. In the following example, we consider TE from a time-lagged version of the counting process of a given THPPP to itself, a case through which we demonstrate the applicability of our results.

Example 6. Suppose $[t_0, T) \subset \mathbb{T} \subset \mathbb{R}, X = (X_t)_{t\in\mathbb{T}}$ is the counting process of a THPPP with intensity λ. Suppose further that $\epsilon > 0$ and $Y = (Y_t)_{t\in\mathbb{T}}\, Y_t = X_{t+\epsilon}, \forall t \geq -\epsilon$. If X is the counting process with intensity $\lambda > 0$ of a THPPP $\psi := (T_n)_{n\geq 1}$, then Y is also a counting process of a THPPP with intensity $\lambda > 0$, specifically that of the point process $\psi' := (T_n - \epsilon)_{n\geq 1})$. Note that the state space of X_t is the natural numbers for any $t \in [t_0, T)$; a Polish space with discrete metrics.

For any $\omega \in \Omega$ and $\Delta t > 0$ we can calculate for any $i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$,

$$\begin{aligned}
&\mathbb{P}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \middle| \left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+k+1)\Delta t}^{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t} \right) \right) (\omega) \left(b_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \right) \\
&= \mathbb{P}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \middle| X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t} \right) (\omega) \left(b_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \right) \\
&= \mathbb{P} \left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} - X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t} = b_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} - X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}(\omega) \right) \\
&= e^{-\lambda \Delta t} \left(\frac{(\lambda \Delta t)^{b_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} - X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}(\omega)}}{\left(b_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} - X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}(\omega) \right)!} \right) \\
&= pois \left(\lambda \Delta t; b_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} - X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}(\omega) \right),
\end{aligned} \tag{4.16}$$

where $pois\,(x, n) = \frac{e^{-x} x^n}{n!}$, for $x > 0$ and integers $n \geq 0$.

Suppose that $[t_0 - \max(\epsilon, s), T) \subset \mathbb{T}$ and $0 < r < \epsilon$. Then $\exists \Delta t^\star > 0$ such that

$$0 < j\Delta t^\star < \epsilon, \forall j = 1, 2, \cdots, \left\lfloor \frac{r}{\Delta t^\star} \right\rfloor.$$

Letting $L = \left\lfloor \frac{r}{\Delta t^\star} \right\rfloor$ we get that

$$\begin{aligned}
&\mathbb{P}_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - i\Delta t^\star} \left(X_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - i\Delta t^\star} \middle| \left(X_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - (i+k+1)\Delta t^\star}^{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - (i+1)\Delta t^\star} \right), \left(Y_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - (i+L+1)\Delta t^\star}^{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - (i+1)\Delta t^\star} \right) \right) (\omega) \\
&= \mathbb{P}_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - i\Delta t^\star} \left(X_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - i\Delta t^\star} \middle| X_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - (i+1)\Delta t^\star}, X_{\lfloor \frac{T}{\Delta t^\star} \rfloor \Delta t^\star - (i+L)\Delta t^\star + \epsilon} \right) (\omega)(\cdot) \\
&= \frac{pois \left(\lambda(\epsilon - L\Delta t^\star); X_{T-(i+L)\Delta t^\star + \epsilon}(\omega) - b_{T - i\Delta t^\star} \right) \cdot p_{\Delta t^\star, i, \omega}}{pois \left(\lambda \left((1-L)\Delta t^\star + \epsilon \right); X_{T-(i+L)\Delta t^\star + \epsilon}(\omega) - X_{T-(i+1)\Delta t^\star}(\omega) \right)} \\
&=: f_{\epsilon, \lambda, \omega}(\Delta t^\star, i, b_{T - i\Delta t^\star}),
\end{aligned} \tag{4.17}$$

where $p_{\Delta t^\star, i, \omega} = pois \left(\lambda \Delta t^\star; b_{T-i\Delta t^\star} - X_{T-(i+1)\Delta t^\star}(\omega) \right)$.

Let $a_{\omega,i} = X_{T-(i+1)\Delta t^\star}(\omega)$ and $c_{\omega,i} = X_{T-(i+L)\Delta t^\star + \epsilon}(\omega)$ and observe that for any $i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t^\star} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t^\star} \right\rfloor - 1$,

$$
\begin{aligned}
&KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t^\star} \;\middle\|\; \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t^\star}\right)\\
&= \sum_{b\in Range\left(X_{T-i\Delta t^\star}\right)} f_{\epsilon,\lambda,\omega}(\Delta t^\star, i, b)\log\frac{f_{\epsilon,\lambda,\omega}(\Delta t^\star, i, b)}{pois\left(\lambda\Delta t^\star; b - X_{\lfloor \frac{T}{\Delta t^\star}\rfloor \Delta t^\star-(i+1)\Delta t^\star}(\omega)\right)}\\
&= \sum_{a_\omega\le b\le c_\omega} f_{\epsilon,\lambda,\omega}(\Delta t^\star, i, b)\log\frac{f_{\epsilon,\lambda,\omega}(\Delta t^\star, i, b)}{pois\left(\lambda\Delta t^\star; b - X_{\lfloor \frac{T}{\Delta t^\star}\rfloor \Delta t^\star-(i+1)\Delta t^\star}(\omega)\right)}\\
&= \sum_{0\le b\le c_\omega-a_\omega} f_{\epsilon,\lambda,\omega}(\Delta t^\star, i, a_\omega+b)\log\frac{f_{\epsilon,\lambda,\omega}(\Delta t^\star, i, a_\omega+b)}{pois\left(\lambda\Delta t^\star; b + \left(a_\omega - X_{\lfloor \frac{T}{\Delta t^\star}\rfloor \Delta t^\star-(i+1)\Delta t^\star}(\omega)\right)\right)}\\
&= \sum_{0\le b\le c_\omega-a_\omega} f_{\epsilon,\lambda,\omega}(\Delta t^\star, i, a_\omega+b)\log\frac{pois\left(\lambda(\epsilon-L\Delta t^\star); c_\omega-a_\omega-b\right)}{pois\left(\lambda\left((1-L)\Delta t^\star+\epsilon\right); c_\omega-a_\omega\right)}\\
&= \sum_{0\le b\le c_\omega-a_\omega}\binom{c_\omega-a_\omega}{b}\left(\frac{\Delta t^\star}{\epsilon-L\Delta t^\star}\right)^b\left(\frac{\epsilon-L\Delta t^\star}{\epsilon+(1-L)\Delta t^\star}\right)^{c_\omega-a_\omega}\times\\
&\qquad\left[\lambda\Delta t^\star+\log\left((c_\omega-a_\omega)^{\underline{b}}\right)-b\log(\lambda(\epsilon-L\Delta t^\star))-(c_\omega-a_\omega)\log\left(1+\frac{\Delta t^\star}{\epsilon-L\Delta t^\star}\right)\right]\\
&=\left[\eta\left(\left(\frac{\epsilon-L\Delta t^\star}{\epsilon+(1-L)\Delta t^\star}\right)^{c_\omega-a_\omega}\right)+\lambda\Delta t^\star\left(\frac{\epsilon-L\Delta t^\star}{\epsilon+(1-L)\Delta t^\star}\right)^{c_\omega-a_\omega}\right]\sum_{0\le b\le c_\omega-a_\omega}\zeta_{\Delta t^\star}(b)\\
&\qquad+\left(\left(\frac{\epsilon-L\Delta t^\star}{\epsilon+(1-L)\Delta t^\star}\right)^{c_\omega-a_\omega}\right)\sum_{0\le b\le c_\omega-a_\omega}\zeta_{\Delta t^\star}(b)\log\left(\frac{(c_\omega-a_\omega)^{\underline{b}}}{\lambda^b(\epsilon-\lambda\Delta t^\star)^b}\right),
\end{aligned}
$$

where $\zeta_{\Delta t^\star}(b)=\binom{c_\omega-a_\omega}{b}\left(\frac{\Delta t^\star}{\epsilon-L\Delta t^\star}\right)^b$, for $0\le b\le c_\omega-a_\omega$, $\eta(x)=x\log(x)$, for $x>0$ and $x^{\underline{b}}$ denotes the b-*th falling factorial of* x.

We suppose now that $\forall\omega\in\Omega, \exists\Delta t_\omega>0$ such that $X_{t+\Delta t_\omega}(\omega)-X_t(\omega)\le 1, \forall t\in[t_0,T)$, that is, there is no more than one event in any interval of length Δt_ω. From this, we have that $\forall\omega\in\Omega$ and $0<\Delta t<\min\{\Delta t_\omega,\Delta t^\star\}$,

$$
\begin{aligned}
&KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \;\middle\|\; \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right)\\
&= \sum_{a_{\omega,i}\le b\le e_{\omega,i}}\left[f_{\epsilon,\lambda,\omega}(\Delta t, i, b)\log\left(\frac{f_{\epsilon,\lambda,\omega}(\Delta t, i, b)}{pois\left(\lambda\Delta t; b-X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t}(\omega)\right)}\right)\right]\\
&=\left[\eta\left(\left(\frac{\epsilon-L\Delta t}{\epsilon+(1-L)\Delta t}\right)^{d_\omega}\right)+\lambda\Delta t\left(\frac{\epsilon-L\Delta t}{\epsilon+(1-L)\Delta t}\right)^{d_\omega}\right]\sum_{0\le b\le d_\omega}\binom{d_\omega}{b}\left(\frac{\Delta t}{\epsilon-L\Delta t}\right)^b\\
&\qquad+\left(\left(\frac{\epsilon-L\Delta t}{\epsilon+(1-L)\Delta t}\right)^{d_\omega}\right)\sum_{0\le b\le d_\omega}\binom{d_\omega}{b}\left(\frac{\Delta t}{\epsilon-L\Delta t}\right)^b\log\left(\frac{(d_\omega)^{\underline{b}}}{\lambda^b(\epsilon-\lambda\Delta t)^b}\right),
\end{aligned}
$$

where $e_{\omega,i} \in \{a_{\omega,i}, a_{\omega,i}+1\}$ and $d_{\omega,i} \in \{0,1\}$.

For any $i = 0, 1, \cdots \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$, if $d_{\omega,i} = 0$, then

$$KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \;\middle\|\; \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right) = \lambda \Delta t$$

and if $d_{\omega,i} = 1$, then

$$\begin{aligned}
&KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \;\middle\|\; \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right) \\
&= \lambda \Delta t \left(\frac{\epsilon - L\Delta t}{\epsilon + (1-L)\Delta t}\right) + \eta\left(\left(\frac{\epsilon - L\Delta t}{\epsilon + (1-L)\Delta t}\right)\right) \\
&\quad + \frac{\lambda(\Delta t)^2 - \log(\lambda)\Delta t}{\epsilon + (1-L)\Delta t} + \Delta t \eta \left(\frac{1}{\epsilon + (1-L)\Delta t}\right) \\
&=: S(\lambda, \Delta t).
\end{aligned}$$

Recall that

$$KL\left(P^{(\omega)}_{\Delta t} \;\middle\|\; M^{(\omega)}_{\Delta t}\right) = \sum_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \;\middle\|\; \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right)$$

from the proof of Theorem 4 and let $Q_{\omega,\Delta t} = \sum_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} d_{\omega,i}$.

Then $\forall \omega \in \Omega$,

$$\begin{aligned}
&KL\left(\prod_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} \mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \;\middle\|\; \prod_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right) \\
&= \sum_{i=0}^{\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1} KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \;\middle\|\; \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right) \\
&= \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - Q_{\omega,\Delta t}\right) \lambda \Delta t + Q_{\omega,\Delta t} S(\lambda, \Delta t) \\
&= \lambda \Delta t \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor\right) + Q_{\omega,\Delta t}\left(S(\lambda, \Delta t) - \lambda \Delta t\right) \\
&\leq \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor\right) S(\lambda, \Delta t).
\end{aligned}$$

Since whenever $0 < r < \epsilon$

$$\lim_{\Delta t \downarrow 0} \left(\left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor\right) S(\lambda, \Delta t) = (T - t_0)\left(\lambda - \frac{\log\left(\lambda(\epsilon - r)\right)}{\epsilon - r}\right), \tag{4.18}$$

$KL\left(\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}\;\middle\|\;\prod_{i=0}^{\lfloor\frac{T}{\Delta t}\rfloor-\lfloor\frac{t_0}{\Delta t}\rfloor-1}\mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right)$ is bounded in a sufficiently small neighborhood of 0. Note that this limit is independent of ω. For illustration, Figure 4.1 shows the bound established in (6) as a function of Δt for specific parameters.

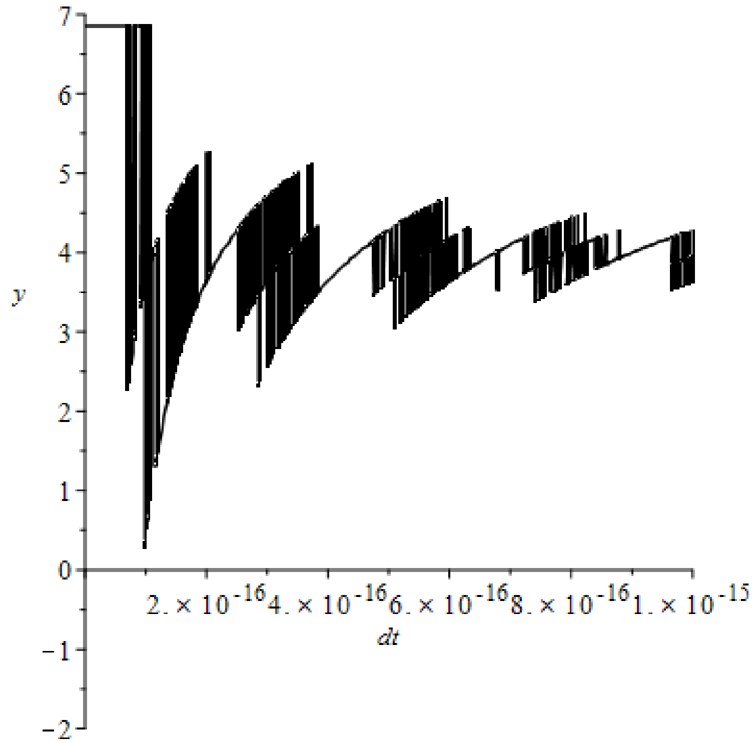

Figure 4.1: KL bound for lagged PPP
$y = \left(\left\lfloor\frac{T}{\Delta t}\right\rfloor - \left\lfloor\frac{t_0}{\Delta t}\right\rfloor\right) S(\lambda, \Delta t)$ plotted as a function of Δt with $r = 0.5, \epsilon = 1, \lambda = 0.2, T = 2$, and $t_0 = 1$. It should be noted that there is clear numerical error as the function is not constant near 0.

For each $\Delta t > 0$, let $A_{\Delta t} = \{\omega \in \Omega \mid X_{t+\Delta t}(\omega) - X_t(\omega) \leq 1, \forall t \in [t_0, T)\}$ and $B_{\Delta t,\gamma}$ be as in Corollary 5.2, that is,

$$B_{\Delta t,\gamma} = \left\{\omega \in \Omega \;\middle|\; \Delta t' \in (0, \Delta t) \implies KL\left(P^{(\omega)}_{\Delta t} \;\middle\|\; M^{(\omega)}_{\Delta t}\right) \leq \gamma\right\}.$$

Fix $\gamma > (T - t_0)\left(\lambda - \frac{\log(\lambda(\epsilon-r))}{\epsilon-r}\right)$. We have now shown that for all $\Delta t > 0$, there exists $0 < \widetilde{\Delta t} < \Delta t$ such that $A_{\Delta t} \subset B_{\widetilde{\Delta t},\gamma}$. Furthermore, since $(B_{\Delta t,\gamma})_{\Delta t>0}$ is a decreasing collection of sets,

$$\mathbb{P}\left(A_{\Delta t}\right) \leq \mathbb{P}\left(B_{\widetilde{\Delta t},\gamma}\right) \leq \mathbb{P}\left(B_{\Delta t',\gamma}\right) \text{ for all } 0 < \Delta t' < \widetilde{\Delta t}. \tag{4.19}$$

From properties of the Poisson point process,

$$\mathbb{P}\left(A_{\Delta t}\right) = 1 - o(\Delta t),$$

thus $\mathbb{P}\left(A_{\Delta t}\right) \to 1$ as $\Delta t \downarrow 0$. Now due to (4.19), we have that $\mathbb{P}\left(B_{\Delta t,\gamma}\right) \to 1$ as $\Delta t \downarrow 0$, which establishes the existence of processes that satisfy (2.46).

Chapter 5

Future Directions

5.1 Alternate Definition of EPT

Motivated by [55], we present the following alternate definition of EPT. This definition defines EPT as a limsup of conditional mutual information over sub-partitions of the interval $[t_0, T)$. This approach has practical relevance as implementing a non-uniform partitioning of time has been used in [51] and [31]. We begin by defining sub-partitions of an interval of the form $[t_0, T)$.

Definition 5.1.1. A *sub-partition* P of an interval $[t_0, T) \subset \mathbb{R}$ is a set of real numbers $t_0, t_1, \ldots, t_n$ such that

$$t_0 < t_1 < \cdots < t_n < T.$$

Definition 5.1.2. Let $P_{[t_0,T)}$ be the set of sub-partitions of the interval $[t_0, T) \subset \mathbb{T}$ and let $||P||$ denote the *mesh* of a sub-partition $P \in P_{[t_0,T)}$, defined by

$$||P|| = \max_{\substack{t_i \in P \\ i \geq 1}} |t_i - t_{i-1}|\,.$$

For all $P \in P_{[t_0,T)}, r, s > 0$, such that $(t_0 - \max(r, s), T] \subset \mathbb{T}$, define the *sub-partitioned expected pathwise transfer entropy* of the sub-partition P, denoted $\mathcal{EPT}_{Y \to X}^{(s,r),P} |_{t_0}^{T}$, by

$$\mathcal{EPT}_{Y \to X}^{(s,r),P} |_{t_0}^{T} = \sum_{i=1}^{||P||} I\left(X_{t_{i-1}}^{t_i}; Y_{t_i - r}^{t_i} \mid X_{t_{i-1}-s}^{t_{i-1}}\right). \tag{5.1}$$

Definition 5.1.3.

$$\begin{aligned}
\mathcal{EPT}^{(s,r)}_{Y\to X}\mid_{t_0}^{T} &= \limsup_{\substack{\Delta t\downarrow 0\\ P\in P_{[t_0,T)},||P||\le\Delta t}} \mathcal{EPT}^{(s,r),P}_{Y\to X}\mid_{t_0}^{T} \\
&= \limsup_{\substack{\Delta t\downarrow 0\\ P\in P_{[t_0,T)},||P||\le\Delta t}} \sum_{i=1}^{||P||} I\left(X^{t_i}_{t_{i-1}}; Y^{t_i}_{t_i-r}\mid X^{t_{i-1}}_{t_{i-1}-s}\right).
\end{aligned} \tag{5.2}$$

Remark 10. The mutual information in 5.1 can be expressed as a supremum of conditional mutual information between discrete random variables over partitions of the sigma-algebra generated by the path spaces $\Omega^{\mathbb{T}}_X$ and $\Omega^{\mathbb{T}}_Y$ due to Wyner's definition of conditional mutual information presented in [55].

Specifically, suppose $\{A_1,\cdots,A_m\}$ and $\{B_1,\cdots,B_n\}$ are finite partitions of the path spaces $\mathcal{F}^{\mathbb{T}}_X$ and $\mathcal{F}^{\mathbb{T}}_Y$, respectively. Now define discrete random variables $\tilde{X},\tilde{Y}$ by

$$\tilde{X}(\omega)=i,\ \text{if } \omega\in A_i \text{ and } \tilde{Y}(\omega)=j,\ \text{if } \omega\in B_j. \tag{5.3}$$

From Theorem 1.6.1 in [23] we have that

$$I(X;Y)=\sup_{P_X,P_Y} I(\tilde{X};\tilde{Y}),$$

where P_X and P_Y denote the set of all finite partitions of $\Omega^{\mathbb{T}}_X$ and $\Omega^{\mathbb{T}}_Y$, respectively. With this along with equation 2.6 a. in [55], we can deduce that for any $P\in P_{[t_0,T)}$ and $\forall i\in[|\ P\ |]$, we have

$$I\left(X^{t_i}_{t_{i-1}}; Y^{t_i}_{t_i-r}\mid X^{t_{i-1}}_{t_{i-1}-s}\right)=\sup_{P_{X^{t_i}_{t_{i-1}}},\,P_{Y^{t_i}_{t_i-r}}} I(\tilde{X}^{t_i}_{t_{i-1}}; \tilde{Y}^{t_i}_{t_i-r}\mid X^{t_{i-1}}_{t_{i-1}-s}).$$

Note that $\tilde{X}^{t_i}_{t_{i-1}}$ and $\tilde{Y}^{t_i}_{t_i-r}$ are discrete random variables which are generally easier to deal with than the RN-derivatives in the previous definition of pathwise transfer entropy. This alternate definition of pathwise transfer entropy allows us to express pathwise transfer entropy as a limit of discrete time transfer entropy as in the Theorem 5, but without having to satisfy the rather strict SPL conditions.

We prove the following proposition which establishes time-dilation invariance of the EPT as defined in Definition 5.1.3.

Proposition 1. Suppose ϕ is linear and monotone increasing. If $\tilde{X}_{\phi(t)} = X_t$ and $\tilde{Y}_{\phi(t)} = Y_t, \forall t \in [t_0, T)$, then

$$\mathcal{EPT}^{(s,r)}_{Y \to X} \mid^{T}_{t_0} = \mathcal{EPT}^{(\phi(s),\phi(r))}_{\tilde{Y} \to \tilde{X}} \mid^{\phi(T)}_{\phi(t_0)} .$$

Proof. Mutual information is invariant to injective transformations, thus for any $P \in P_{[t_0,T)}$, we have

$$\mathcal{EPT}^{(s,r),P}_{Y \to X} \mid^{T}_{t_0} = \sum_{i=1}^{||P||} I\left(X^{t_i}_{t_{i-1}}; Y^{t_i}_{t_i - r} \mid X^{t_{i-1}}_{t_{i-1}-s}\right) = \sum_{i=1}^{||\phi(P)||} I\left(\tilde{X}^{\phi(t_i)}_{\phi(t_{i-1})}; \tilde{Y}^{\phi(t_i)}_{\phi(t_i - r)} \mid \tilde{X}^{\phi(t_{i-1})}_{(\phi(t_{i-1}-s))}\right),$$

where the partition $\phi(P)$ of $[\phi(t_0), \phi(T))$ is defined by

$$\phi(P) = \{\phi(t^\star) \mid t^\star \in P\}.$$

From the continuity and monotonicity of ϕ we have that

$$\{P \mid P \in P_{[\phi(t_0),\phi(T))}\} = \{\phi(P) \mid P \in P_{[t_0,T)}\}$$

and so

$$\begin{aligned}
\mathcal{EPT}^{(s,r)}_{Y \to X} \mid^{T}_{t_0} &= \inf_{\substack{n \geq 1 \\ ||P||=n \\ P \in P_{[t_0,T)}}} \mathcal{EPT}^{(s,r),P}_{Y \to X} \mid^{T}_{t_0} \\
&= \inf_{\substack{n \geq 1 \\ ||P||=n \\ P \in P_{[t_0,T)}}} \sum_{i=1}^{n} I\left(X^{t_i}_{t_{i-1}}; Y^{t_i}_{t_i - r} \mid X^{t_{i-1}}_{t_{i-1}-s}\right) \\
&= \inf_{\substack{n \geq 1 \\ ||\phi(P)||=n \\ \phi(P) \in P_{[\phi(t_0),\phi(T))}}} \sum_{i=1}^{n} I\left(\tilde{X}^{\phi(t_i)}_{\phi(t_{i-1})}; \tilde{Y}^{\phi(t_i)}_{\phi(t_i - r)} \mid \tilde{X}^{\phi(t_{i-1})}_{(\phi(t_{i-1}-s))}\right) \qquad (5.4) \\
&= \inf_{\substack{n \geq 1 \\ ||\phi(P)||=n \\ \phi(P) \in P_{[\phi(t_0),\phi(T))}}} \sum_{i=1}^{n} I\left(\tilde{X}^{\phi(t_i)}_{\phi(t_{i-1})}; \tilde{Y}^{\phi(t_i)}_{\phi(t_i) - \phi(r)} \mid \tilde{X}^{\phi(t_{i-1})}_{(\phi(t_{i-1}) - \phi(s))}\right) \\
&= \mathcal{EPT}^{(\phi(s),\phi(r))}_{\tilde{Y} \to \tilde{X}} \mid^{\phi(T)}_{\phi(t_0)},
\end{aligned}$$

where the second to last equality if from the linearity of ϕ. □

Question 2. *Is this definition advantageous or even equivalent to Definition 2.18?*

We have mentioned some advantages of this definition earlier; however, they are quite simplistic. There is a much richer collection of literature involving Wyner's definition, in some vicinity, of conditional mutual information than the approach we used to define the EPT. Thus, it is likely that defining EPT as in Definition 5.2 makes EPT easier to use and calculate. However, a rigorous exploration of this matter is not performed here and is left as an open question.

Question 3. *What other processes satisfy (2.31) or (2.46) other than the deterministically lagged counting process of a THPPP?*

In Appendix A, we provide a calculation for $KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}\middle|\middle|\mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right)$ where Y is a time-lagged version of a Wiener process X. However, there is no calculation of $KL\left(P^{(\omega)}_{\Delta t}\middle|\middle|M^{(\omega)}_{\Delta t}\right)$ for these processes or for any other process other than that presented in example 6. There is a wealth of transformations one could perform on a process to yield another: thinning, superimposition, deterministic lagging, random lagging, bump convolution, etc. Each of these transformations yields a new process that is not independent of the original process; thus, there should be a nonzero TE. Compound Poisson processes (CPP) are of particular relevance to the continuous-time framework presented in this work and are widely used to model neural spike trains, thus showing that either (2.31) or (2.46) hold for a transformed CPP (using the aforementioned transformations) would be a fruitful discovery.

5.2 Differentiability of EPT and estimators

In Section 2.3, we provided sufficient conditions for continuity of PT and EPT in time. However, there are no sufficient conditions for the existence of the limit in Definition 3.1; thus, differentiability of these functions is still an open topic.

Question 4. *Do there exist nontrivial sufficient conditions for the differentiability of the EPT function?*

One of the main contributions of this thesis is a definition of the TE rate native to continuous-time processes. However, our methodology does not present any practical means of measuring it, only the theoretical formulation.

Question 5. *Do there exist practical estimators of the EPT and the TE rate?*

The transfer entropy estimator presented in [28] is of practical utility for discrete-time processes. Can it be generalized to appropriately measure TE using the measure theoretical approach taken in this work? If so, what are its properties? There is a wealth of questions one could propose pertaining to such an estimator, e.g., is this estimator biased or asymptotically biased/unbiased? Is it an efficient estimator and how is its speed performance? Does there exist an appropriate model class under which an MLE for TE exists? How does this estimator compare with binning and partitioning based estimators used in the literature referenced in Section 1.4?

If there is no such estimator that can be used in a general setting, does there exist one when the destination and source process are a particular type of continuous-time stochastic process? Providing estimators for TE rate and EPT amongst a pair of non-homogeneous PPPs, compound Poisson processes, or Brownian motions with various effects appear to be the types of processes for which an estimator with appealing properties would be of most interest as these processes are encountered or considered in many applications in which causality in real-time data is held in high regard.

Bibliography

[1] Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis: a Hitchhiker's Guide*. Berlin; London: Springer, 2006.

[2] Stefan Ankirchner and Peter Imkeller. "Financial markets with asymmetric information: information drift, additional utility and entropy". In: *Stochastic processes and applications to mathematical finance*. World Scientific, 2007, pp. 1–21.

[3] Juliana Martins de Assis and Francisco Marcos de Assis. "Estimation of transfer entropy between discrete and continuous random processes". In: *Journal of Communication and Information Systems* 33.1 (2018).

[4] Rami Atar and Tsachy Weissman. "Mutual information, relative entropy, and estimation in the Poisson channel". In: *IEEE Transactions on Information theory* 58.3 (2012), pp. 1302–1318.

[5] Nasir Ahmad Aziz. "Transfer entropy as a tool for inferring causality from observational studies in epidemiology". In: *bioRxiv* (2017), p. 149625.

[6] Lionel Barnett, Adam B Barrett, and Anil K Seth. "Granger causality and transfer entropy are equivalent for Gaussian variables". In: *Physical review letters* 103.23 (2009), p. 238701.

[7] Demian Battaglia et al. "Dynamic effective connectivity of inter-areal brain circuits". In: *PLoS Computational Biology* 8.3 (2012), e1002438.

[8] Daniel A Butts et al. "Temporal precision in the neural code and the timescales of natural vision". In: *Nature* 449.7158 (2007), p. 92.

[9] Erhan Çınlar. *Probability and stochastics*. Vol. 261. Springer Science & Business Media, 2011.

[10] Emanuele Crosato et al. "Informative and misinformative interactions in a school of fish". In: *Swarm Intelligence* 12.4 (2018), pp. 283–305.

[11] Łukasz Dębowski. “A general definition of conditional information and its application to ergodic decomposition”. In: *Statistics & Probability Letters* 79.9 (2009), pp. 1260–1268.

[12] Thomas Dimpfl and Franziska Julia Peter. “Using transfer entropy to measure information flows between financial markets”. In: *Studies in Nonlinear Dynamics and Econometrics* 17.1 (2013), pp. 85–102.

[13] Ernst-Erich Doberkat. “Measures for probabilistic systems”. In: *Special Topics in Mathematics for Computer Scientists: Sets, Categories, Topologies and Measures*. Springer International Publishing, 2015, pp. 427–683.

[14] Rick Durrett. *Probability: theory and examples*. Cambridge University Press, 2010.

[15] E Erten et al. “Criticality and information dynamics in epidemiological models”. In: *Entropy* 19.5 (2017), p. 194.

[16] Ionut Florescu. *Probability and stochastic processes*. John Wiley & Sons, 2014.

[17] Clive WJ Granger. “Investigating causal relations by econometric models and cross-spectral methods”. In: *Econometrica: Journal of the Econometric Society* (1969), pp. 424–438.

[18] Stefan Grosskinsky. “Large-scale dynamics of stochastic particle systems”. In: *Lecture notes, go. warwick. ac. uk/SGrosskinsky/teaching/tcc. html, Warwick* (2017).

[19] Bernard Harris. *The statistical estimation of entropy in the non-parametric case*. Tech. rep. 1975.

[20] Saike He et al. “Identifying peer influence in online social networks using transfer entropy”. In: *Pacific-Asia Workshop on Intelligence and Security Informatics*. Springer. 2013, pp. 47–61.

[21] Sebastian Herzog, Christian Tetzlaff, and Florentin Wörgötter. “Transfer entropy-based feedback improves performance in artificial neural networks”. In: *arXiv preprint arXiv:1706.04265* (2017).

[22] Katerina Hlaváčková-Schindler et al. “Causality detection based on information-theoretic approaches in time series analysis”. In: *Physics Reports* 441.1 (2007), pp. 1–46.

[23] Shunsuke Ihara. *Information theory for continuous systems*. Vol. 2. World Scientific, 1993.

[24] Shinya Ito et al. “Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model”. In: *PloS one* 6.11 (2011), e27431.

[25] Martin Jacobsen. *Point process theory and applications: marked point and piecewise deterministic processes*. Springer Science & Business Media, 2006.

[26] A Kaiser and T Schreiber. “Information transfer in continuous processes”. In: *Physica D, v.166, 43-62 (2002)* 166 (June 2002).

[27] Claude Kipnis and Claudio Landim. *Scaling limits of interacting particle systems*. Vol. 320. Springer Science & Business Media, 2013.

[28] LF Kozachenko and Nikolai N Leonenko. “Sample estimate of the entropy of a random vector”. In: *Problemy Peredachi Informatsii* 23.2 (1987), pp. 9–16.

[29] Okyu Kwon and J-S Yang. “Information flow between stock indices”. In: *EPL (Europhysics Letters)* 82.6 (2008), p. 68003.

[30] Gunter Last and Andreas Brandt. *Marked point processes on the real line: the dynamical approach*. Springer Science & Business Media, 1995.

[31] Joon Lee et al. “Transfer entropy estimation and directional coupling change detection in biomedical time series”. In: *Biomedical Engineering Online* 11.1 (2012), p. 19.

[32] X Liang. “The Liang-Kleeman information flow: theory and applications”. In: *Entropy* 15.1 (2013), pp. 327–360.

[33] Michael Lindner et al. “TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy”. In: *BMC Neuroscience* 12.1 (2011), p. 119.

[34] Faezeh Marzbanrad et al. “Quantifying the interactions between maternal and fetal heart rates by transfer entropy”. In: *PloS one* 10.12 (2015), e0145672.

[35] Shamim Nemati et al. “Respiration and heart rate complexity: effects of age and gender assessed by band-limited transfer entropy”. In: *Respiratory Physiology & Neurobiology* 189.1 (2013), pp. 27–33.

[36] Sunny Nigam et al. “Rich-club organization in effective connectivity among cortical neurons”. In: *Journal of Neuroscience* 36.3 (2016), pp. 670–684.

[37] Oliver Obst, Joschka Boedecker, and Minoru Asada. “Improving recurrent neural network performance using transfer entropy”. In: *International Conference on Neural Information Processing*. Springer. 2010, pp. 193–200.

[38] N Orange and N Abaid. “A transfer entropy analysis of leader-follower interactions in flying bats”. In: *The European Physical Journal Special Topics* 224.17-18 (2015), pp. 3279–3293.

[39] Prakash Panangaden. *Labelled Markov Processes*. World Scientific, 2009.

[40] Liam Paninski. “Estimation of entropy and mutual information”. In: *Neural computation* 15.6 (2003), pp. 1191–1253.

[41] M Piccioni. “Continuous versions of Radon-Nikodym derivatives as likelihood ratios”. In: *Systems & Control Letters* 2.6 (1983), pp. 369–374.

[42] Alberto Porta et al. “Conditional self-entropy and conditional joint transfer entropy in heart period variability during graded postural challenge”. In: *PLoS One* 10.7 (2015), e0132851.

[43] MM Rao. “Projective limits of probability spaces”. In: *Journal of Multivariate Analysis* 1.1 (1971), pp. 28–57.

[44] Leonidas Sandoval. “Structure of a global network of financial companies based on transfer entropy”. In: *Entropy* 16.8 (2014), pp. 4443–4482.

[45] Thomas Schreiber. “Measuring information transfer”. In: *Phys. Rev. Lett.* 85 (2 2000), pp. 461–464.

[46] Ahmet Sensoy et al. “Effective transfer entropy approach to information flow between exchange rates and stock markets”. In: *Chaos, Solitons & Fractals* 68 (2014), pp. 180–185.

[47] Claude Elwood Shannon. “A mathematical theory of communication”. In: *Bell System Technical Journal* 27.3 (1948), pp. 379–423.

[48] Richard E Spinney, Mikhail Prokopenko, and Joseph T Lizier. “Transfer entropy in continuous time, with applications to jump and neural spiking processes”. In: *Physical Review E* 95.3 (2017), p. 032319.

[49] Nicholas M Timme et al. “High-degree neurons feed cortical computations”. In: *PLoS computational biology* 12.5 (2016), e1004858.

[50] Thai Quang Tung et al. “Inferring gene regulatory networks from microarray time series data using transfer entropy”. In: *Twentieth IEEE International Sym-*

posium on Computer-Based Medical Systems (CBMS'07). IEEE. 2007, pp. 383–388.

[51] Greg Ver Steeg and Aram Galstyan. "Information transfer in social media". In: *Proceedings of the 21st international conference on World Wide Web*. ACM. 2012, pp. 509–518.

[52] Michael Wibral, R Vicente, and Michael Lindner. "Transfer entropy in neuroscience". In: *Understanding Complex Systems* (Jan. 2014), pp. 3–36.

[53] David Williams. *Probability with martingales*. Cambridge University Press, 1991.

[54] Patricia Wollstadt et al. "Efficient transfer entropy analysis of non-stationary neural time series". In: *PloS one* 9.7 (2014), e102833.

[55] Aaron D Wyner. "A definition of conditional mutual information for arbitrary ensembles". In: *Information and Control* 38.1 (1978), pp. 51–59.

[56] Li Xie, Valery A Ugrinovskii, and Ian R Petersen. "A duality relationship for regular conditional relative entropy". In: *IFAC Proceedings Volumes* 38.1 (2005), pp. 248–253.

[57] Jie Zhu et al. "Contribution to transfer entropy estimation via the k-nearest-neighbors approach". In: *Entropy* 17.6 (2015), pp. 4173–4201.

Appendix A

A lagged Weiner process calculation

Suppose $\epsilon > 0$ (non- random) and $Y_t = X_{t+\epsilon}$, $\forall t \geq -\epsilon$. If X is a Weiner process, then Y is also a Weiner process and for fixed $\Delta t > 0$, we can calculate for any $\omega \in \Omega$ and any Borel set, the conditional probabilities for (2.11) and (2.12).

From incremental independence of the Poisson counting process, we have that

$$
\begin{aligned}
&\prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \Big((\mathbb{P}^{X}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}\Big(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \mid X^{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+k+1)\Delta t}\Big)(\omega)\Big) \Big(B_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}\Big) \\
&= \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \Big((\mathbb{P}^{X}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}\Big(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \mid X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}\Big)(\omega)\Big) \Big(B_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}\Big) \\
&= \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \left(\frac{1}{\sqrt{2\pi \Delta t}} \int_{B_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}} \exp\left(- \frac{\left(x - X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}(\omega)\right)^2}{2\Delta t} \right) dx \right).
\end{aligned} \tag{A.1}
$$

Let $B_{\Delta t,i} = B_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}$ for $i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$. If $0 < \Delta t < \epsilon$, then

$$
\begin{aligned}
&\prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t}\left(B_{\Delta t,i}\right) \\
&= \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}^{X}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}\Big(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \mid X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}, Y^{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}_{\lfloor \frac{T}{\Delta t} \Delta t - (i+l+1)\Delta t}\Big)(\omega)\left(B_{\Delta t,i}\right) \\
&= \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}^{X}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}\Big(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \mid X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}, Y_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}\Big)(\omega)\left(B_{\Delta t,i}\right) \\
&= \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \mathbb{P}^{X}_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t}\Big(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - i\Delta t} \mid X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t}, X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t - (i+1)\Delta t + \epsilon}\Big)(\omega)\left(B_{\Delta t,i}\right) \\
&= \prod_{i=0}^{\lfloor \frac{T}{\Delta t} \rfloor - \lfloor \frac{t_0}{\Delta t} \rfloor - 1} \left(\int_{B_{\Delta t,i}} \mathcal{N}\left(x; \mu_{i,\Delta t}, \sigma^2_{i,\Delta t}\right) dx \right),
\end{aligned}
$$

where

$$\mu_{i,\Delta t} = X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t+\epsilon}(\omega)\frac{\Delta t}{\epsilon} + X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t}(\omega)\frac{\epsilon - \Delta t}{\epsilon},$$

$$\sigma^2_{i,\Delta t} = \left(\frac{\epsilon - \Delta t}{\epsilon^2}\right)\Delta t,$$

and the last equality comes from the observation that if X is a SBM then for any $t_0 < t_1 < t_2$ we have

$$X_{t_1|X_{t_0}=x, X_{t_2}=y} \sim \mathcal{N}\left(\frac{t_2 - t_1}{t_2 - t_0}x + \frac{t_1 - t_0}{t_2 - t_0}y, \frac{(t_2 - t_1)(t_1 - t_0)}{t_2 - t_0}\right).$$

Note that for any two Gaussian distributions, say p and q with means μ_p, μ_q and variances σ_p, σ_q, respectively, we get after some calculations that

$$KL(p \mid\mid q) = \log\left(\frac{\sigma_q}{\sigma_p}\right) + \frac{\sigma_p + (\mu_p - \mu_q)^2}{2\sigma_q^2} - \frac{1}{2}.$$

Thus for any $\omega \in \Omega$ and $\Delta t > 0$, we can calculate for any $i = 0, 1, \ldots, \left\lfloor \frac{T}{\Delta t} \right\rfloor - \left\lfloor \frac{t_0}{\Delta t} \right\rfloor - 1$,

$$\begin{aligned}
&KL\left(\mathbb{P}^{(\omega),(k,l)}_{X|\overleftarrow{X},\overleftarrow{Y},i,\Delta t} \mid\mid \mathbb{P}^{(\omega),(k)}_{X|\overleftarrow{X},i,\Delta t}\right) \\
&= \log\left(\frac{\Delta t}{\left(\frac{\epsilon-\Delta t}{\epsilon^2}\right)\Delta t}\right) + \frac{\left(\frac{\epsilon-\Delta t}{\epsilon^2}\right)\Delta t}{2\left(\Delta t\right)^2} - \frac{1}{2} \\
&\quad + \frac{\left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t+\epsilon}(\omega)\frac{\Delta t}{\epsilon} + X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t}(\omega)\frac{\epsilon-\Delta t}{\epsilon} - X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t}(\omega)\right)^2}{2\left(\Delta t\right)^2} \\
&= \log\left(\frac{\epsilon^2}{\epsilon - \Delta t}\right) \\
&\quad + \frac{\epsilon - \Delta t\left(1+\epsilon^2\right)}{2\Delta t\epsilon^2} + \left(\frac{X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t+\epsilon}(\omega)\frac{\Delta t}{\epsilon} - \left(\frac{\Delta t}{\epsilon}\right)\left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t}(\omega)\right)}{\sqrt{2}\Delta t}\right)^2 \\
&= \log\left(\frac{\epsilon^2}{\epsilon - \Delta t}\right) + \frac{\epsilon - \Delta t\left(1+\epsilon^2\right)}{2\Delta t\epsilon^2} \\
&\quad + \left(\frac{X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t+\epsilon}(\omega) - \left(X_{\lfloor \frac{T}{\Delta t} \rfloor \Delta t-(i+1)\Delta t}(\omega)\right)}{\sqrt{2}\epsilon}\right)^2.
\end{aligned} \tag{A.2}$$

and the TE at any time $i\Delta t$ as

$$
\begin{aligned}
&\mathbb{T}^{(k,l)}_{Y\to X}(i\Delta t)\\
&= \mathbb{E}_{\mathbb{P}}\Bigg[\log\left(\frac{\epsilon^2}{\epsilon-\Delta t}\right)+\frac{\epsilon-\Delta t\,(1+\epsilon^2)}{2\Delta t\epsilon^2}\\
&\qquad+\left(\frac{X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t+\epsilon}(\omega)-X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t}(\omega)}{\sqrt{2}\epsilon}\right)^2\Bigg]\\
&= \log\left(\frac{\epsilon^2}{\epsilon-\Delta t}\right)+\frac{\epsilon-\Delta t\,(1+\epsilon^2)}{2\Delta t\epsilon^2}\\
&\qquad+\mathbb{E}_{\mathbb{P}}\left[\left(\frac{X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t+\epsilon}(\omega)-X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t}(\omega)}{\sqrt{2}\epsilon}\right)^2\right]\\
&= \log\left(\frac{\epsilon^2}{\epsilon-\Delta t}\right)+\frac{\epsilon-\Delta t\,(1+\epsilon^2)}{2\Delta t\epsilon^2}+\frac{1}{2\epsilon^2}\mathbb{E}_{\mathbb{P}}\left[\left(X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t+\epsilon}\right)^2\right]\\
&\qquad-\frac{1}{\epsilon^2}\mathbb{E}_{\mathbb{P}}\left[\left(X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t+\epsilon}\right)\left(X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t}\right)\right]+\frac{1}{2\epsilon^2}\mathbb{E}_{\mathbb{P}}\left[\left(X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t}\right)^2\right]\\
&= \log\left(\frac{\epsilon^2}{\epsilon-\Delta t}\right)+\frac{\epsilon-\Delta t\,(1+\epsilon^2)}{2\Delta t\epsilon^2}+\frac{1}{2\epsilon^2\left(\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t+\epsilon\right)^2}\\
&\qquad-\frac{1}{\epsilon^2}\mathbb{E}_{\mathbb{P}}\left[\left(X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t+\epsilon}\right)\left(X_{\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t}\right)\right]+\frac{1}{2\epsilon^2\left(\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t\right)^2}\\
&= \log\left(\frac{\epsilon^2}{\epsilon-\Delta t}\right)+\frac{\epsilon-\Delta t\,(1+\epsilon^2)}{2\Delta t\epsilon^2}+\frac{1}{2\epsilon^2\left(\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t+\epsilon\right)^2}\\
&\qquad+\frac{1}{2\epsilon^2\left(\lfloor\frac{T}{\Delta t}\rfloor\Delta t-(i+1)\Delta t\right)^2}.
\end{aligned}
$$

Appendix B

Lagged Poisson calculation

In Example 6, the source process is constructed as a time lagged version of the destination process. In what follows, we regard TE as a function of said lag and investigate its behavior, after applying a binning strategy, for different values of history length windows.

Suppose that X and Y are as in Example 6 and $\epsilon = s$. Let $n \geq 1$ be an integer and let

$$X_{n,\Delta t} = H\left(\left|(n\Delta t, (n+1)\Delta t) \cap (\psi_m)_{m\geq 1}\right|\right),$$

where H denotes the heaviside function. It should be noted that these random variables are Bernoulli random variables with a mean of $1-e^{-\lambda\Delta t}$. We utilize the fact that the value of these random variables is either 0 or 1 to calculate the probabilities in (1.1) by calculating the appropriate probabilities (those that appear in (1.1)) for all possible outcomes of these processes. For example, if $2\Delta t \leq s \leq 3\Delta t$, then

$$\mathbb{P}\left(\{X_{n,\Delta t} = 1\} \cap \{X_{n-1,\Delta t} = 1\} \cap \{X_{n-2,\Delta t} = 0\} \cap \{Y_{n-1,\Delta t} = 0\} \cap \{Y_{n-2,\Delta t} = 0\}\right) \tag{B.1}$$

can be easily calculated using the incremental independence property of the Poisson process as $\alpha^3 (1-\alpha)\left(1 - \frac{\alpha}{\alpha^2\beta}\right)$, where $\alpha = e^{-\lambda\Delta t}$ and $\beta = 1-\alpha$. The remaining 31 joint probabilities can be obtained similarly for this choice of k and l and each of the conditional probabilities in (1.1) can be obtained as a quotient of joint probabilities. Upon the aforementioned calculations, we obtain finally that the joint and conditional probabilities in the case of the lagged PPP with $k = l = 2$ are polynomials in α and β. For $n \geq 1$, Figure B.1 shows a graph of $\mathbb{T}^{(2,2)}_{Y_{j,\Delta t}\to X_{j,\Delta t}}$ as a function of s for a particular process intensity λ and bin width Δt. Upon similar calculations, the joint and conditional probabilities in the case that $k = l = 1$ are also polynomials in α and β and $\mathbb{T}^{(1,1)}_{Y_{j,\Delta t} \to X_{j,\Delta t}}$ is graphed as a function of s in Figure B.2, again with particular values of λ and Δt.

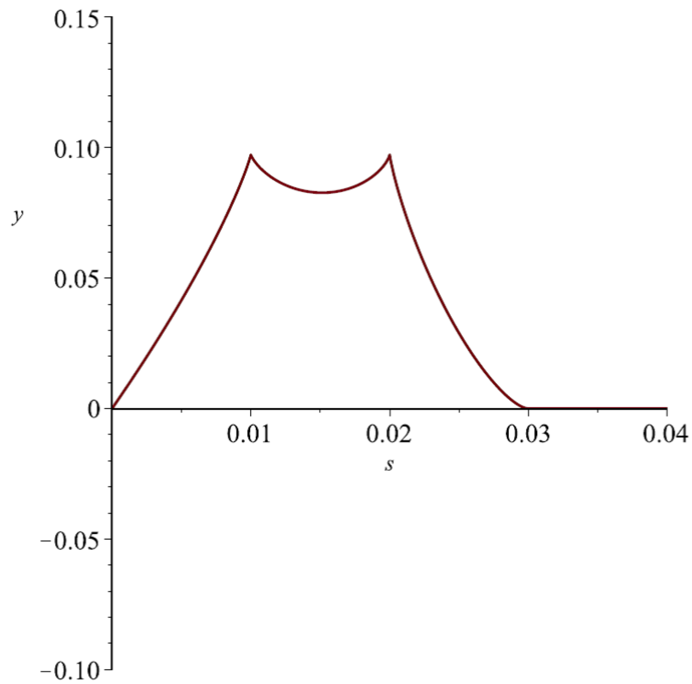

Figure B.1: Lagged PPP calculation with $k = l = 2$, $\lambda = 2, \Delta t = 0.2$.

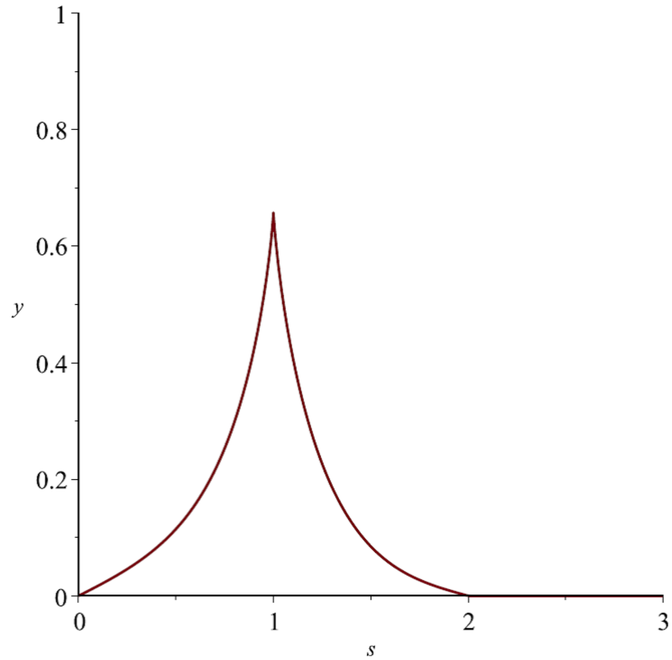

Figure B.2: Lagged PPP calculation with $k = l = 1$, $\lambda = 1$, $\Delta t = 1$.

The source code containing the calculation of $\mathbb{T}^{(2,2)}_{Y_{j,\Delta t} \to X_{j,\Delta t}}$ and $\mathbb{T}^{(1,1)}_{Y_{j,\Delta t} \to X_{j,\Delta t}}$ can be found at `https://github.com/edgarcde/transfer_entropy`.