Beyond the Central Limit Theorem: Universal and Non-universal Simulations of Random Variables by General Mappings

Lei Yu
Department of Electrical and Computer Engineering
National University of Singapore, Singapore
Email: leiyu@nus.edu.sg

December 5, 2018

Abstract
Motivated by the Central Limit Theorem, in this paper, we study both universal and non-universal simulations of random variables with an arbitrary target distribution \( Q_Y \) by general mappings, not limited to linear ones (as in the Central Limit Theorem). We derive the fastest convergence rate of the approximation errors for such problems. Interestingly, we show that for discontinuous or absolutely continuous \( P_X \), the approximation error for the universal simulation is almost as small as that for the non-universal one; and moreover, for both universal and non-universal simulations, the approximation errors by general mappings are strictly smaller than those by linear mappings. Furthermore, we also generalize these results to simulation from Markov processes, and simulation of random elements (or general random variables).

Index terms: Universal simulation, random number generation, absolutely continuous distribution, total variation distance, Kolmogorov-Smirnov distance, squeezing periodic functions

2000 Mathematics subject classification: Primary 65C10; Secondary 60F05; 62E20; 62E17

1 Introduction
The Central Limit Theorem (CLT) states that for a sequence of i.i.d. real-valued random variables \( X^n \sim P_X \), the normalized sum \( \frac{1}{\sqrt{\text{Var}(X)}} \sum_{i=1}^n (X_i - E[X]) \) converges in distribution to a standard Gaussian random variable as \( n \) goes to infinity. This implies that an \( n \)-dimensional i.i.d. random vector \( X^n \) can be used to simulate a standard Gaussian random variable \( Y \) by the normalized sum so that the approximation error asymptotically vanishes under the Kolmogorov–Smirnov distance. Moreover, from the Berry-Esseen theorem [1, Sec. XVI.5], the approximation error vanishes in a rate of \( \frac{1}{\sqrt{n}} \). Note that here, the distribution \( P_X \) of \( X \) is arbitrary, and given the mean and variance, the linear function is independent of \( P_X \). Hence such a linear function can be considered as a universal linear function. The corresponding simulation problem can be considered as being universal. In this paper, we consider general universal simulation problems, in which general\(^1\) simulation functions, not limited to linear ones, are allowed. We are interested in the following question: What is the optimal convergence rate for such universal simulation problems? To know how important the knowledge of the distribution \( P_X \) is in a simulation, we are also interested in the optimal convergence rate for non-universal simulation problems (in which \( P_X \) is known). Is the optimal convergence rate for universal simulation as fast as, or strictly slower than, that for non-universal simulation?

The CLT is about universal simulation of a continuous random variable (more specifically, a Gaussian random variable). In addition to simulation of continuous random variables, there are a large number of works that consider universal simulation of a sequence of discrete (or atomic) random variables from another

\(^1\)We say a mapping is general if it is either linear or non-linear.
sequence of discrete random variables. In 1951, von Neumann [2] described a procedure for exactly generating a sequence of independent and identically distributed (i.i.d.) unbiased random coins from a sequence of i.i.d. biased random coins with an unknown distribution. To obtain unbiased outputs, two pairs of bits (0, 1) and (1, 0) (which have the same empirical distribution) are mapped to 0 and 1, respectively, and (0, 0) and (1, 1) are discarded. Elias [3] and Blum [4] considered a more general situation in which the process of the repeated coin tosses is subject to an unknown Markov process, instead of a traditional i.i.d. process, and then studied the efficiency of such a procedure measured according to the expected number of output coins per input coin. Knuth and Yao [5], Roche [6], Abrahams [7], and Han and Hoshi [8] considered another general simulation problem in which an arbitrary target distribution is generated by using an unbiased or biased M-coin (i.e., an M-sided coin) but with a known distribution. They showed that the minimum expected number of coin tosses required to generate the target distribution can be expressed in terms of the ratio of the entropy of the target distribution to that of the seed distribution. In all of the works above [2–8], simulators are defined as functions that map a variable-length input sequence to a fixed-length output sequence. Hence, to produce an output symbol, arbitrarily long delay or waiting time may be required.

To reduce delay, a direction of generalizing the random number generation problem is to require that an output must be generated for every k bits input from an unbiased or biased coin, for any fixed k, but at the same time, relax the requirement of exact generation to that of approximate generation. That is, we may require only that the target distribution should be generated approximately within a nonzero but arbitrarily small tolerance in terms of some suitable distance measures such as the total variation distance or divergences. Such a problem in the asymptotic context with known seed and target distributions has been formulated and studied by Han and Verdú [9]; its inverse problem has been investigated by Vembu and Verdú [10]; and a general version of these problems —– generating an i.i.d. sequence from another i.i.d. sequence with arbitrary known seed and target distributions —— has been studied in [11–13].

All of the works above only considered simulating a sequence of discrete random variables from another sequence of discrete random variables. In contrast, in this paper we consider approximately generating an arbitrary random variable (or a random element) from a sequence of random variables (or another random element) with arbitrary but unknown seed distribution.

Besides the CLT, this work is also motivated by the following questions. 1) Given a distribution Q (defined on (Ω, B_Ω)), is there a measurable function f : R → R such that P_f(X) = Q for all absolutely continuous distribution P_X? Here P_f(X) is the distribution of the image f(X) induced by P_X and the function f. 2) Given Q, is there a sequence of measurable functions f_k : R → R such that P_{f_k(X)} → Q as k → ∞ (under the total variation distance or other distance measures) for all absolutely continuous distribution P_X? By some simple derivations, it is easy to show that the answer to the first question is negative. So it is intuitive to conjecture the answer to the second one is also negative, since the second question reduces to the first question if the limit of the sequence {f_k} is set to the function f. However, the results in this paper show that this conjecture is not right, since the limit of the optimal sequence {f_k} does not exist and hence these two questions are not equivalent. Interestingly, we show that the answer to the second question is positive.

### 1.1 Problem Formulation

Before formulating our problem, we first introduce two statistical distances. For an arbitrary measurable space (Ω, B_Ω), we use P(Ω, B_Ω) to denote the set of all the probability measures (a.k.a. distributions) defined on (Ω, B_Ω). Given an arbitrary measurable space (Ω, B_Ω), the total variation (TV) distance between two probability measures P, Q ∈ P(Ω, B_Ω) is defined as

\[ |P − Q|_{TV} = \sup_{A ∈ B_Ω} |P(A) − Q(A)|. \]

The Kolmogorov–Smirnov (KS) distance between two probability measures P, Q ∈ P(ℝ, B_ℝ) is defined as

\[ |P − Q|_{KS} = \sup_{x ∈ ℝ} |F(x) − G(x)|, \]
where $F$ and $G$ respectively denote the CDFs (cumulative distribution functions) of $P$ and $Q$. For $P, Q \in \mathcal{P}(\mathbb{R}, \mathcal{B}_\mathbb{R})$, we have

$$0 \leq |P - Q|_{KS} \leq |P - Q|_{TV},$$

since $|P - Q|_{KS} = \sup_{A \in \mathcal{I}} |P(A) - Q(A)|$ with $\mathcal{I} := \{(-\infty, y] : y \in \mathbb{R}\} \subseteq \mathcal{B}_\mathbb{R}$. Furthermore, both $|P - Q|_{KS}$ and $|P - Q|_{TV}$ are metrics, and hence $|P - Q|_{KS} = 0 \iff P = Q$ and $|P - Q|_{TV} = 0 \iff P = Q$.

Based on these two distances, we next formulate our problem. In this paper, we consider the following problem: When we use an approximation error is measured by the TV distance or the KS distance. We term the Borel space the fastest convergence speed of the approximation error over all functions.

**Definition 1.1.** Given the seed Borel space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ and the target Borel space $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, a simulator is a measurable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Given a random vector $X^n \sim P_{X^n}$ and a target distribution $Q_Y$, we want to find an optimal simulator $Y = f(X^n)$ that minimizes the TV distance or the KS distance between the output distribution $P_Y := P_{X^n} \circ f^{-1}$ (the distribution of the output random variable $Y$) and the target distribution $Q_Y$. For such a simulation problem, we consider two different scenarios where $P_{X^n}$ is respectively known and unknown a priori.

As illustrated in Fig. 1 (a), if the seed distribution $P_{X^n}$ is unknown, but the class $\mathcal{P}_{X^n} \subseteq \mathcal{P}(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ that $P_{X^n}$ belongs to is known, we term such simulation problems as (universal) $(\mathcal{P}_{X^n}, Q_Y)$-simulation problems. Hence, the simulator $f : \mathbb{R}^n \rightarrow \mathbb{R}$ in the universal simulation problem may depend on everything including $Q_X$ and $P_{X^n}$, but except for $P_{X^n}$. That is, it is independent of $P_{X^n}$ given $\mathcal{P}_{X^n}$. Next we give a mathematical formulation for the universal simulation problem, which avoids ambiguous languages, like “$P_{X^n}$ is unknown”.

**Definition 1.2.** A function $g : \mathcal{P}_{X^n} \rightarrow \mathbb{R}$ is called TV-achievable (resp. KS-achievable) for the universal $(\mathcal{P}_{X^n}, Q_Y)$-simulation, if there exists a sequence of simulators $\{f_{n,k}\}_{k=1}^{\infty}$ such that

$$\limsup_{k \rightarrow \infty} |P_{Y_{n,k}} - Q_Y|_{\theta} \leq g(P_{X^n})$$

for all $P_{X^n} \in \mathcal{P}_{X^n}$, where $P_{Y_{n,k}} := P_{X^n} \circ f_{n,k}^{-1}$ and $\theta = TV$ (resp. $\theta = KS$).
Definition 1.3. The set of TV-achievable (resp. KS-achievable) functions for the universal \((P_{X^n}, Q_Y)\)-simulation is defined as
\[ E_\theta(P_{X^n}, Q_Y) := \{ g : P_{X^n} \to \mathbb{R} : g \text{ is } \theta\text{-achievable} \} \]
where \(\theta = \text{TV} \) (resp. \(\theta = \text{KS} \)).

According to Lebesgue’s decomposition theorem [14], the distributions of real-valued random variables can be partitioned into three classes\(^2\): discontinuous distributions (including discrete distributions and mixtures of discrete and continuous distributions), absolutely continuous distributions, and continuous but not absolutely continuous distributions (including singular continuous distributions and mixtures of singular continuous and absolutely continuous distributions). The sets of these distributions are respectively denoted as \(P_{dc}, P_{ac}, \text{ and } P_c \setminus P_{ac} \), where \(P_c = \mathcal{P}(\mathbb{R}, B_\mathbb{R}) \setminus P_{dc} \) denotes the set of continuous distributions on \((\mathbb{R}, B_\mathbb{R})\).

For the i.i.d. case, we define \(P_{X}^{(n)} := \{ P_X : P_X \in P_X \} \). In this paper, we want to characterize \(E_\theta(P_{X}^{(n)}, Q_Y)\) with \(P_X\) respectively set to \(P_{dc}, P_{ac}, \text{ or } P_c \setminus P_{ac} \). Note that \(E_\theta(P_{X}^{(n)}, Q_Y)\) is an upper set. For brevity, for such sets and a function \(g : P_X \to \mathbb{R}\), we denote \(E_\theta(P_{X}^{(n)}, Q_Y) \approx g \) if there exists a \(g' \in E_\theta(P_{X}^{(n)}, Q_Y)\) such that \(g'(P_X) \leq g(P_X), \forall P_X \in P_X\); and \(E_\theta(P_{X}^{(n)}, Q_Y) \preceq g \) if \(g'(P_X) \geq g(P_X), \forall P_X \in P_X\) for all \(g' \in E_\theta(P_{X}^{(n)}, Q_Y)\). In addition, \(E_\theta(P_{X}^{(n)}, Q_Y) \asymp g \) if and only if \(E_\theta(P_{X}^{(n)}, Q_Y) \preceq g \) and \(E_\theta(P_{X}^{(n)}, Q_Y) \succeq g \).

In general, there does not necessarily exist \(g\) such that \(E_\theta(P_{X}^{(n)}, Q_Y) \asymp g\). However, if it exists, then \(g(P_X) = \inf_{g' \in E_\theta(P_{X}^{(n)}, Q_Y)} g'(P_X)\) for \(P_X \in P_X\).

Similarly, we write \(E_\theta(P_{X}^{(n)}, Q_Y) \gtrsim g_n\) if there exists a \(g'_n \in E_\theta(P_{X}^{(n)}, Q_Y)\) such that \(g'_n(P_X) \leq g_n(P_X), \forall P_X \in P_X\); and \(E_\theta(P_{X}^{(n)}, Q_Y) \gtrsim g\) if \(g'_n(P_X) \geq g_n(P_X), \forall P_X \in P_X\) for all \(g'_n \in E_\theta(P_{X}^{(n)}, Q_Y)\). In addition, \(E_\theta(P_{X}^{(n)}, Q_Y) \gtrsim g\) if and only if \(E_\theta(P_{X}^{(n)}, Q_Y) \gtrsim g\) and \(E_\theta(P_{X}^{(n)}, Q_Y) \gtrsim g\). Furthermore, \(E_\theta(P_{X}^{(n)}, Q_Y) = e^{-\Omega(g_n)}\) (resp. \(E_\theta(P_{X}^{(n)}, Q_Y) = e^{-\omega(g_n)}\)) if there exists a \(g'_n \in E_\theta(P_{X}^{(n)}, Q_Y)\) such that \(g'_n(P_X) = e^{-\Omega(g_n(P_X))}\) (resp. \(g'_n(P_X) = e^{-\omega(g_n(P_X))}\)) for all \(P_X \in P_X\).

Conversely, as illustrated in Fig. 1 (b), if \(P_{X^n}\) is known, we term such problems as \((\text{non-universal}) (P_{X^n}, Q_Y)\)-simulation problems. The simulator \(f : \mathbb{R}^n \to \mathbb{R}\) in the non-universal simulation problem may depend on all of \(P_{X^n}, Q_X, \text{ etc.}\).

Definition 1.4. The optimal TV-achievable (resp. KS-achievable) approximation error for the non-universal \((P_{X^n}, Q_Y)\)-simulation is defined as
\[ E_\theta(P_{X^n}, Q_Y) := \inf_{f_n : \mathbb{R}^n \to \mathbb{R}} |P_{Y^n} - Q_Y|_\theta, \]
where \(P_{Y^n} := P_{X^n} \circ f_n^{-1}\) and \(\theta = \text{TV} \) (resp. \(\theta = \text{KS} \)).

The non-universal \((P_{X^n}, Q_Y)\)-simulation problem can be seen as a special universal \((P_{X^n}, Q_Y)\)-simulation problem with \(P_{X^n}\) set to \(\{ P_X \} \). Hence \(E_\theta(P_{X^n} \setminus Q_Y) = E_\theta(P_{X^n}, Q_Y)\). On the other hand, by definitions, the set \(E_\theta(P_{X^n}, Q_Y)\) for the universal \((P_{X^n}, Q_Y)\)-simulation with \(\theta \in \{ \text{KS, TV} \}\) must satisfy \(E_\theta(P_{X^n}, Q_Y) \geq E_\theta(P_{X^n}, Q_Y)\). That is, the approximation errors for non-universal simulation problems are not larger than those for universal simulation problems.

In general, simulating a continuous random variable is more difficult than simulating a discontinuous one, as stated in the following lemma. Hence in this paper, sometimes we only provide upper bounds on the approximation errors for simulating continuous random variables. It should be understood that those upper bounds are also upper bounds for simulating any other random variables (e.g., discrete random variables). Furthermore, to make our results easier to follow, we summarize them in Table 1.

\(^2\)We say a distribution (or a probability measure) \(P\) is discrete (or atomic) if it is purely atomic; continuous if it does not have any atoms; discontinuous if it has at least one atom; absolutely continuous if it is absolutely continuous with respect to Lebesgue measure, i.e., having a probability density function; and singular continuous if it is continuous, and meanwhile, singular with respect to Lebesgue measure.

\(^3\)Throughout this paper, for two positive sequences \(f(n), g(n)\), we write \(f(n) \leq g(n)\) or \(g(n) \geq f(n)\) if
Table 1: Summary of our results. Here $\mathcal{P}_X$ and $\mathcal{Q}_Y$ respectively denote the classes that $P_X$ and $Q_Y$ belong to.

| Table 1: Summary of our results. Here $\mathcal{P}_X$ and $\mathcal{Q}_Y$ respectively denote the classes that $P_X$ and $Q_Y$ belong to. |
|---------------------------------------------------------------|
| $(P_X, Q_Y)$ | **Non-universal Simulation** |
| (continuous, arbitrary) | $E_\theta(P_X, Q_Y) = 0$ for $\theta \in \{\text{KS, TV}\}$ (Prop. 2.1) |
| (discontinuous, arbitrary) | $E_{\text{KS}}(P^n_X, Q_Y) \leq \frac{1}{2} (\max_x P_X(x))^n$ (Cor. 2.1 & Lem. 1.1) |
| **Special Case 1:** (discontinuous, continuous) | $E_{\text{KS}}(P^n_X, Q_Y) = \frac{1}{2} (\max_x P_X(x))^n$ (Cor. 2.1) |
| **Special Case 2:** (discrete with finite alphabet, discrete with finite alphabet) | $E_{\text{KS}}(P^n_X, Q_Y) \leq (\min_x P_X(x))^n$ (Prop. 2.3) |
| $(P_X, Q_Y)$ | **Universal Simulation** |
| (absolutely continuous, arbitrary) | $\mathcal{E}_\theta(P_X, Q_Y) = 0$ for $\theta \in \{\text{KS, TV, Renyi}\}$ (Thm. 3.1 & 3.2) |
| (discontinuous, arbitrary) | $\mathcal{E}_{\text{KS}}(P^n_X, Q_Y) \leq (\max_x P_X(x))^n$ (Cor. 3.1 & Lem. 1.1) |
| **Special Case:** (discontinuous, continuous) | $\mathcal{E}_{\text{KS}}(P^n_X, Q_Y) \leq (\max_x P_X(x))^n$ (Cor. 3.1) |
| (continuous but not absolutely continuous, arbitrary) | $\mathcal{E}_{\text{KS}}(P^n_X, Q_Y) \approx e^{-\omega(n)}$ (Cor. 3.2) |
| **Special Case:** ($F_X$ is Hölder continuous with exponent $\alpha$ where $0 < \alpha \leq 1$, arbitrary) | $\mathcal{E}_{\text{KS}}(P^n_X, Q_Y) \approx e^{-\alpha H(n \log n)}$ (Cor. 3.2) |

| Simulating a Random Variable from a Stationary Memoryless Process |
|---------------------------------------------------------------|
| $(\mathcal{P}_X, \mathcal{Q}_Y)$ | **Non-universal Simulation** |
| (a Markov chain of order $k$ with finite state space $\mathcal{X}$ and initial state $x^0_{-k+1}$, arbitrary) | $E_{\text{KS}}(\mathcal{P}_X^n, \mathcal{Q}_Y) \leq e^{-n H_\infty(x^0_{-k+1}, P_{X_{k+1}|X^k})}$ (Cor. 4.1 & Lem. 1.1) |
| **Special Case:** (a Markov chain of order $k$ with finite state space $\mathcal{X}$ and initial state $x^0_{-k+1}$, continuous) | $E_{\text{KS}}(\mathcal{P}_X^n, \mathcal{Q}_Y) \geq e^{-n H_\infty(x^0_{-k+1}, P_{X_{k+1}|X^k})}$ (Cor. 4.1) |
| $(\mathcal{P}_X, \mathcal{Q}_Y)$ | **Universal Simulation** |
| (a Markov chain of order $k$ with finite state space $\mathcal{X}$ and initial state $x^0_{-k+1}$, arbitrary) | $E_{\text{KS}}(\mathcal{P}_X^n, \mathcal{Q}_Y) \leq e^{-n H_\infty(x^0_{-k+1}, P_{X_{k+1}|X^k})}$ (Thm. 4.1 & Lem. 1.1) |
| **Special Case:** (a Markov chain of order $k$ with finite state space $\mathcal{X}$ and initial state $x^0_{-k+1}$, continuous) | $E_{\text{KS}}(\mathcal{P}_X^n, \mathcal{Q}_Y) \approx e^{-n H_\infty(x^0_{-k+1}, P_{X_{k+1}|X^k})}$ (Thm. 4.1) |

| Simulating a Random Element from another Random Element |
|---------------------------------------------------------------|
| $(\mathcal{P}_X, \mathcal{Q}_Y)$ | **Non-universal Simulation** |
| (continuous, arbitrary) | $E_{\text{TV}}(\mathcal{P}_X, \mathcal{Q}_Y) = 0$ (Thm. 5.1) |
| **Special Case:** (continuous random variable, arbitrary random vector) | $E_\theta(\mathcal{P}_X, \mathcal{Q}_Y) = 0$ for $\theta \in \{\text{KS, TV}\}$ (Cor. 5.1) |
| $(\mathcal{P}_X, \mathcal{Q}_Y)$ | **Universal Simulation** |
| (absolutely continuous respect to a continuous distribution, arbitrary) | $E_{\text{TV}}(\mathcal{P}_X, \mathcal{Q}_Y) = 0$ (Thm. 5.2) |
| **Special Case:** (absolutely continuous random variable, arbitrary random vector) | $E_\theta(\mathcal{P}_X, \mathcal{Q}_Y) = 0$ for $\theta \in \{\text{KS, TV}\}$ (Cor. 5.2) |
Lemma 1.1. Assume $Q_Y$ and $Q_Z$ are two distributions defined on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, and moreover, $Q_Y$ is continuous. Then the approximation errors for non-universal and universal simulations satisfy

\begin{align}
E_\theta(P_{X^n}, Q_Z) &\leq E_\theta(P_{X^n}, Q_Y), \\
E_\theta(P_{X^n}, Q_Z) &\geq E_\theta(P_{X^n}, Q_Y),
\end{align}

for any $P_{X^n}$ and $P_{X^n}$, where $\theta \in \{\text{KS}, \text{TV}\}$.

Proof. Proposition 2.1 (which is given in the next section) states that there exists a non-decreasing mapping $z = g(y) : \mathbb{R} \to \mathbb{R}$ such that $Z = g(Y) \sim Q_Z$, where $Y \sim Q_Y$. Observe that for any $P_Z$,

$$|P_Z - Q_Z|_{\text{KS}} = \sup_{A \in \mathcal{I}} |P_Z(A) - Q_Z(A)|$$

$$= \sup_{A \in \mathcal{I}} |P_Y(g^{-1}(A)) - Q_Y(g^{-1}(A))|$$

$$\leq \sup_{B \in \mathcal{I}} |P_Y(B) - Q_Y(B)|$$

$$= |P_Y - Q_Y|_{\text{KS}},$$

where $\mathcal{I} := \{(-\infty, y) : y \in \mathbb{R}\}$. Hence (1) and (2) hold for $\theta = \text{KS}$.

By similar steps but with $\mathcal{I}$ replaced by $\mathcal{B}_\mathbb{R}$, we can easily obtain that (1) and (2) also hold for $\theta = \text{TV}$. \qed

2 Non-universal Simulation from a Stationary Memoryless Process

In this section, we consider non-universal simulation of a real-valued random variable. If the seed distribution $P_X$ is continuous and the target distribution $Q_Y$ is arbitrary, then we can simulate a random variable $Y$ that exactly follows the distribution $Q_Y$. The following is a well-known result for such a case. Hence the proof is omitted.

Proposition 2.1. For a continuous distribution $P_X$ and an arbitrary distribution $Q_Y$, using the inverse transform sampling function $y = G_Y^{-1}(F_X(x))$, we obtain $P_Y = Q_Y$, where $G_Y^{-1}(t) := \min\{y : G_Y(y) \geq t\}$ denotes the quantile function (generalized inverse distribution function) of $G_Y$. That is, $E_\theta(P_X, Q_Y) = 0$ for $\theta \in \{\text{KS}, \text{TV}\}$.

Next we consider the case $P_X$ is discontinuous. For this case, exact simulation cannot be obtained.

Proposition 2.2. Assume $P_X$ is discontinuous and $Q_Y$ is continuous. Then for the non-universal $(P_X, Q_Y)$-simulation problem,\footnote{For two positive sequences $f(n), g(n)$, we write $f(n) = \Omega(g(n))$ (resp. $f(n) = \omega(g(n))$) if $\liminf_{n \to \infty} \frac{f(n)}{g(n)} > 0$ (resp. $\limsup_{n \to \infty} \frac{f(n)}{g(n)} < 0$).}

$$E_{\text{KS}}(P_X, Q_Y) = \frac{1}{2} \max_x P_X(x).$$

Proof. Denote $A$ as the set of discontinuity points of $F_X$. Then for each $x \in \mathbb{R}\setminus A$, map $x$ to $G_Y^{-1}(F_X(x))$. For each $x \in A$, map $x$ to $G_Y^{-1}(\lim_{t \uparrow x} F_X(\bar{x})) + \frac{1}{2} P_X(x)$. For such mapping, we have $|P_Y - Q_Y|_{\text{KS}} = \frac{1}{2} \max_x P_X(x)$. Furthermore, the converse is obvious. \qed

Applying Proposition 2.2 to the vector case, we get the following corollary.

Corollary 2.1. Assume $P_X$ is discontinuous and $Q_Y$ is continuous. Then for the non-universal $(P_X^n, Q_Y)$-simulation problem, $E_{\text{KS}}(P_X^n, Q_Y) = \frac{1}{2} \left( \max_x P_X(x) \right)^n$.

\[
\limsup_{n \to \infty} \frac{1}{n} \log \frac{f(n)}{g(n)} \leq 0. \quad \text{In addition, } f(n) \doteq g(n) \text{ if and only if } f(n) \preceq g(n) \text{ and } f(n) \succeq g(n).
\]

\footnote{Here the minimum exists since CDFs are right-continuous.}

\[
\lim_{n \to \infty} \frac{f(n)}{g(n)} = \infty.
\]

\footnote{For simplicity, we denote $P_X(\{x\})$ for $x \in \mathbb{R}$ as $P_X(x)$. Hence for a discrete random variable $X$, $P_X(x)$ is the probability mass function of $X$.}
The result above shows that if the seed and target distributions are respectively discontinuous and continuous, then the optimal approximation error vanishes exponentially fast. We next show that if the seed and target distributions are both discrete, the optimal approximation error vanishes faster. The proof of Proposition 2.3 is provided in Appendix A.

**Proposition 2.3.** Assume both $P_X$ and $Q_Y$ are discrete with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Then for the non-universal $(P^n_X, Q^n_Y)$-simulation problem, $E_{\mathrm{KS}}(P^n_X, Q^n_Y) \leq (\min_x P_X(x))^n$.

**Remark 2.1.** More specifically, we can prove that for $n \geq |\mathcal{Y}| \max_{1 \leq i \leq |\mathcal{X}|-1} \frac{P_X(x_i)}{P_X(x_{i+1})}$,

$$E_{\mathrm{KS}}(P^n_X, Q_Y) \leq \frac{1}{2} P_X(x_{|\mathcal{X}|-1}) \left( P_X(x_{|\mathcal{X}|}) \right)^{n-1},$$

where $x_1, x_2, ..., x_{|\mathcal{X}|}$ is a resulting sequence after sorting the elements in $\mathcal{X}$ such that $P_X(x_1) \geq P_X(x_2) \geq ... \geq P_X(x_{|\mathcal{X}|})$.

### 3 Universal Simulation from a Stationary Memoryless Process

In this section, we consider universal simulation of a real-valued random variable. For universal simulation, we divide the seed distributions into three kinds: absolutely continuous, discontinuous, as well as continuous but not absolutely continuous distributions.

#### 3.1 Absolutely continuous seed distributions

We first consider absolutely continuous seed distributions, and show an impossibility result for this case.

**Proposition 3.1.** For non-degenerate distribution $Q_Y$, there is no simulator (measurable function) $Y = f(X) : (\mathbb{R}, B_\mathbb{R}) \to (\mathbb{R}, B_\mathbb{R})$ such that $P_Y = Q_Y$ for any absolutely continuous $P_X$.

**Proof.** Suppose that there exists a measurable function $f : (\mathbb{R}, B_\mathbb{R}) \to (\mathbb{R}, B_\mathbb{R})$ such that $P_Y = Q_Y$ for any absolutely continuous $P_X$.

Case 1: Suppose that there exists a set $A \in B_\mathbb{R}$ such that both $f^{-1}(A)$ and $\mathbb{R}\setminus f^{-1}(A)$ have positive Lebesgue measures. Then $P_Y(A) = P_X(f^{-1}(A))$. For two absolutely continuous measures $P_X$ and $\tilde{P}_X$ such that $P_X(f^{-1}(A)) \neq \tilde{P}_X(f^{-1}(A))$, then $P_Y(A) \neq \tilde{P}_Y(A)$, where $\tilde{P}_Y$ is the distribution induced by $\tilde{P}_X$ through the mapping $f$. This implies that $P_Y(A) \neq Q_Y(A)$ or $\tilde{P}_Y(A) \neq Q_Y(A)$. This contradicts with the assumption that $P_Y = Q_Y$ for any absolutely continuous $P_X$.

Case 2: Suppose that either $f^{-1}(A)$ or $\mathbb{R}\setminus f^{-1}(A)$ has zero Lebesgue measure for all $A \in B_\mathbb{R}$. Then for any absolutely continuous $P_X$, we have $P_Y(A) = P_X(f^{-1}(A)) = 0$ or $P_Y(\mathbb{R}\setminus A) = 1 - P_Y(A) = 1 - P_X(f^{-1}(A)) = P_X(f^{-1}(\mathbb{R}\setminus A)) = 1$ for all $A \in B_\mathbb{R}$. That is, $P_Y(A) = 0$ or 1 for all $A \in B_\mathbb{R}$. However for any non-degenerate measure $Q_Y$, there exists an $A \in B_\mathbb{R}$ such that $0 < Q_Y(A) < 1$. This contradicts with the assumption that $P_Y = Q_Y$.

Combining the two cases above, we have Proposition 3.1. \qed

The theorem above implies that for any simulator $f : \mathbb{R} \to \mathbb{R}$, we always have $|P_Y - Q_Y|_{\mathrm{TV}} > 0$ for some absolutely continuous $P_X$. However, we can prove that there exists a sequence of simulators that make the TV-approximation error $|P_Y - Q_Y|_{\mathrm{TV}}$ arbitrarily close to zero for any absolutely continuous $P_X$.

**Theorem 3.1.** Assume $P_X = P_{\mathrm{ac}}$ and $Q_Y$ is arbitrary. Then for the universal $(P_X, Q_Y)$-simulation problem, $\mathcal{E}_\theta(P_X, Q_Y) \to 0$ for $\theta \in \{\mathrm{KS}, \mathrm{TV}\}$.

Given Proposition 3.1, I think this result is rather surprising and counter-intuitive. If $f$ is a differentiable bijective function, then the input distribution is determined by $f$ and the output distribution, since $p_X(x) = p_Y(f(x))f'(x)$, where $p_X$ and $p_Y$ are respectively the PDFs (probability density functions, i.e., Radon–Nikodym derivatives with respect to the Lebesgue measure) of $P_X$ and $P_Y$. Hence given $f$ and the output
distribution, the input distribution is unique. However, in our case, we consider a sequence of non-bijective mappings $f_n$. Hence given $f_n$ and the output distribution, the input distribution is not unique. The essence of our proof of this theorem is that any PDF $p_X$ can be approximated within any level of approximation error by a sequence of step functions, and on the other hand, such step functions can be used to generate any distribution in a universal way. Hence $P_X$ can be used to simulate any distribution within any level of approximation error.

Proof. We first restrict our attention to the case that $Q_Y$ is absolutely continuous. Let $p_X$, $p_Y$, and $q_Y$ be the PDFs of $P_X$, $P_Y$, and $Q_Y$ respectively.

**Universal Mapping:** Partition the real line into intervals with the same length $\Delta$, i.e., $\bigcup_{i=-\infty}^{\infty} (i\Delta, (i+1)\Delta]$. We first simulate a uniform distribution on $[a,b]$ by mapping each interval $(i\Delta, (i+1)\Delta]$ into $[a,b]$ using the linear function $x \mapsto a + \frac{b-a}{\Delta}(x-i\Delta)$. We then transform the output distribution to the target distribution $Q_Y$, by using function $x \mapsto G^{-1}_Y \left( \frac{x-a}{b-a} \right)$, where $G^{-1}_Y(t) := \min \{ y : G_Y(y) \geq t \}$. Therefore, each $x \in (i\Delta, (i+1)\Delta]$ is mapped to $G^{-1}_Y \left( \frac{x}{\Delta} \right)$. Hence the final mapping is

$$f_\Delta(x) := \sum_{i=-\infty}^{\infty} G^{-1}_Y \left( \frac{1}{\Delta} (x-i\Delta) \right) \mathbb{1} \{ x \in (i\Delta, (i+1)\Delta] \},$$

which is shown in Fig. 2. Furthermore, some properties on squeezing such a periodic function are provided in Appendix D.1.

The PDF of the output of this mapping (respect to the input distribution $P_X$) is denoted as $P_Y$. However, if the input distribution is $\hat{P}_X$ with PDF $\hat{p}(x) := \sum_{i=-\infty}^{\infty} \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p(x') dx' \cdot \mathbb{1} \{ x \in (i\Delta, (i+1)\Delta] \}$, then the output distribution becomes the one with CDF $\hat{G}_Y(y)$ satisfying

$$\hat{G}_Y(y) = \int_{\{ x : f_\Delta(x) \leq y \}} \sum_{i=-\infty}^{\infty} \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p(x') dx' \cdot \mathbb{1} \{ x \in (i\Delta, (i+1)\Delta] \} dx$$

$$= \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{i\Delta+\Delta} \hat{G}_Y(y) dx \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p(x) dx$$

$$= \sum_{i=-\infty}^{\infty} \Delta G_Y(y) \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p(x) dx$$

Figure 2: Illustration of the universal mapping.
where in (3), swapping the integral and the sum follows from Fubini’s theorem. Hence the output distribution induced by inputting \( \hat{P}_X \) results in \( Q_Y \) exactly.

Based on the above observations, we have

\[
|P_Y - Q_Y|_{TV} \leq |P_X - \hat{P}_X|_{TV} = \int_{-\infty}^{\infty} |p(x) - \hat{p}(x)|\, dx \\
= 2 \int_{-\infty}^{\infty} [p(x) - \hat{p}(x)]^+\, dx, \tag{4}
\]

where \( [z]^+ = \max\{z, 0\} \), and (4) follows from the data processing inequality on the total variable distance, i.e., \( |P_Y - Q_Y|_{TV} \leq |P_X - Q_X|_{TV} \) with \( P_Y (\cdot) := \int P_Y_{|X}(\cdot|x)\,dP_X(x) \) and \( Q_Y (\cdot) := \int P_Y_{|X}(\cdot|x)\,dQ_X(x) \).

Observe that \( [p(x) - \hat{p}(x)]^+ \leq p(x) \) and \( p(x) \) is integrable, and moreover,

\[
\lim_{\Delta \to 0} \hat{p}(x) = \lim_{\Delta \to 0} \sum_{i=-\infty}^{\infty} \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p(x') dx' \{ x \in (i\Delta, (i+1)\Delta) \} \\
= \lim_{\Delta \to 0} \frac{1}{\Delta} \int_{\lfloor \frac{x}{\Delta} \rfloor \Delta}^{\lfloor \frac{x+1}{\Delta} \rfloor \Delta} p(x') dx' \\
= p(x) \text{ a.e.}, \tag{6}
\]

where (6) follows by Lebesgue’s differentiation theorem [15, Thm. 7.7]. Hence by Lebesgue’s dominated convergence theorem [15, Thm. 1.34],

\[
\lim_{\Delta \to 0} \int_{-\infty}^{\infty} [p(x) - \hat{p}(x)]^+\, dx = \int_{-\infty}^{\infty} \lim_{\Delta \to 0} [p(x) - \hat{p}(x)]^+\, dx \\
= 0. \tag{7}
\]

Therefore, combining (5) with (7) yields

\[
\lim_{\Delta \to 0} |P_Y - Q_Y|_{TV} = 0.
\]

That is, \( \mathcal{E}_{TV}(P_X, Q_Y) \sim 0 \).

If \( Q_Y \) is not absolutely continuous, we can first simulate an absolutely continuous random variable \( Z \) with distribution \( Q_Z \) and then use it to simulate \( Y \sim Q_Y \). As stated in Lemma 1.1, this will result in a smaller TV-approximation error for \( (P_X, Q_Y) \)-simulation problem than that for \( (P_X, Q_Z) \)-simulation problem. Hence the TV-approximation error for this case also approaches to zero as \( \Delta \to 0 \).

For the universal mapping proposed in the proof of Theorem 3.1, the induced approximation error \( |P_Y - Q_Y|_{TV} \) depends on the interval length \( \Delta \), and converges to zero as \( \Delta \to 0 \). We next investigate how fast the approximation error converges to zero as \( \Delta \to 0 \).

**Proposition 3.2** (Convergence Rate as \( \Delta \to 0 \)). Assume \( P_X \) is an absolutely continuous distribution with an a.e. (almost everywhere) continuously differentiable PDF \( p_X \) such that \( |p_X'(x)| \) is bounded, and \( Q_Y \) is an arbitrary distribution. Then the TV-approximation error induced by the universal mapping \( x \mapsto \sum_{i=-\infty}^{\infty} G_{\Delta}^{-1} \left( \frac{1}{\Delta} (x - i\Delta) \right) 1 \{ x \in (i\Delta, (i+1)\Delta) \} \) satisfies \( \limsup_{\Delta \to 0} \frac{1}{\Delta} |P_Y - Q_Y|_{TV} \leq \int |p_X'(x)|\, dx \), where \( P_{\Delta} \) denotes the distribution of the output \( Y \).
Proof: The mean value theorem implies there exists \( x_i \in (i\Delta, (i+1)\Delta) \) such that \( P_X(x_i) = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p_X(x)dx \). By Taylor’s theorem, we have \( p_X(x) = p_X(x_i) + p_X'(x_i, \Delta)(x - x_i) \) for some \( x_i, \Delta \in (i\Delta, (i+1)\Delta) \). Therefore, from Remark D.3, we have

\[
\frac{1}{\Delta} |P_Y - Q_Y|_{TV} \leq \frac{1}{\Delta} \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} \left| p_X(x) - \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p_X(x)dx \right| dx
\]

\[
= \frac{1}{\Delta} \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} |p_X(x_i) + p_X'(x_i, \Delta)(x - x_i) - p_X(x_i)| dx
\]

\[
\leq \sum_{i=-\infty}^{\infty} |p_X'(x_i, \Delta)| \Delta.
\]

Since \( |p_X'(x)| \) is continuous a.e. and bounded, \( |p_X'(x)| \) is Riemann-integrable on every interval \([a, b]\) with \( a < b \). Then

\[
\limsup_{\Delta \to 0} \sum_{i=-\infty}^{\infty} |p_X'(x_i, \Delta)| \Delta = \text{Riemann-} \int |p_X'(x)| dx,
\]

where \( \text{Riemann-} \int \) denotes the Riemann-integral. Furthermore, for any non-negative Riemann-integrable function, its Riemann-integral and Lebesgue-integral are the same. That is

\[
\text{Riemann-} \int |p_X'(x)| dx = \int |p_X'(x)| dx.
\]

Therefore, \( \limsup_{\Delta \to 0} \frac{1}{\Delta} |P_Y - Q_Y|_{TV} \leq \int |p_X'(x)| dx. \)

Proposition 3.2 implies that if the total variation \( \int |p_X'(x)| dx \) of \( p_X \) is finite, then the approximation error \( |P_Y - Q_Y|_{TV} \) converges to zero at least linearly fast as \( \Delta \to 0 \). If \( \int |p_X'(x)| dx \) is infinity, then it is not easy to obtain a general bound on \( |P_Y - Q_Y|_{TV} \). However, for some special cases, e.g., \( p_X = -\log x, x \in (0, 1) \) or \( p_X = (1 - r)x^{-r}, x \in (0, 1], r \in (0, 1) \), we provide upper bounds as follows.

**Example 3.1 (Convergence Rate as \( \Delta \to 0 \) for \( \int |p_X'(x)| dx = \infty \)).** [16] For the PDF \( p_X(x) = -\log x, x \in (0, 1] \), \( |P_Y - Q_Y|_{TV} \leq \frac{\Delta}{2} \ln 2\pi \frac{1}{\Delta}. \) For the PDF \( p_X(x) = (1 - r)x^{-r}, x \in (0, 1], r \in (0, 1) \), \( |P_Y - Q_Y|_{TV} \leq C\Delta^{1-r} \) for some constant \( C. \)

Universal simulations in Theorem 3.1 for the uniform distribution on \([0, 1]\) for different \( P_X \) are illustrated in Fig. 3. For Gaussian and exponential distributions, the total variations of their PDFs are finite. Hence the approximation errors for these two distributions decay linearly in \( \Delta \). For logarithmic and polynomial-like distributions, the total variations of their PDFs are infinite. As stated in Proposition 3.1, the approximation errors for these two distributions decay respectively in order of \( \Delta \ln \frac{1}{\Delta} \) and \( \Delta^{1-r} \).

### 3.1.1 Relative Entropy and Rényi Divergence Measures

Next we extend Theorem 3.1 to the relative entropy and Rényi divergence measures. The relative entropy and Rényi divergence are two information measures that quantify the “distance” between probability measures.

Fix distributions \( P_X, Q_X \in \mathcal{P}(\mathbb{R}, \mathcal{B}_\mathbb{R}). \) The relative entropy and the Rényi divergence of order \( \alpha \in (0, 1) \cup (1, \infty) \) are respectively defined as

\[
D(P_X||Q_X) := \int \left( \frac{dP_X}{dQ_X} \log \frac{dP_X}{dQ_X} \right) dQ_X
\]

\[
D_\alpha(P_X||Q_X) := \frac{1}{\alpha - 1} \log \int \left( \frac{dP_X}{dQ_X} \right)^\alpha dQ_X,
\]

10
Figure 3: Universal simulations in Theorem 3.1 for the uniform distribution on $[0, 1]$ by the mapping $x \mapsto \sum_{i=-\infty}^{\infty} \frac{1}{\Delta} (x - i \Delta) 1 \{ x \in (i \Delta, (i + 1) \Delta) \}$ with $\Delta = 0.01$. The seed distributions $P_X$ are illustrated in top figures, which are respectively the standard Gaussian distribution $\mathcal{N}(0, 1)$, exponential distribution $\text{Exp}(1)$, logarithmic distribution $p_X(x) = -\log x, x \in (0, 1]$, and polynomial-like distribution $p_X(x) = (1 - r) x^{-r}, x \in (0, 1], r = 0.5$. The generated distributions $P_Y$ are illustrated in bottom figures.

and the conditional versions are respectively defined as

$$D(P_{Y|X} \parallel Q_{Y|X} | P_X) := D(P_X P_{Y|X} \parallel P_X Q_{Y|X})$$

$$D_\alpha(P_{Y|X} \parallel Q_{Y|X} | P_X) := D_\alpha(P_X P_{Y|X} \parallel P_X Q_{Y|X}).$$

The Rényi divergence of order $\alpha \in \{0, 1, \infty\}$ is defined by continuous extension. Throughout, log and exp are to the natural base $e$. It is known that $D_1(P_X \parallel Q_X) := \lim_{\alpha \to 1} D_\alpha(P_X \parallel Q_X) = D(P_X \parallel Q_X)$ so a special case of the Rényi divergence (resp. the conditional version) is the usual relative entropy (resp. the conditional version). The Rényi divergence of infinity order is defined as

$$D_{\infty}(P_X \parallel Q_X) := \lim_{\alpha \to \infty} D_\alpha(P_X \parallel Q_X) = \log \text{ess sup}_{P_X} \frac{dP_X}{dQ_X}.$$

**Definition 3.1.** A function $g : \mathcal{P}_X \to \mathbb{R}$ is called $\alpha$-Rényi-achievable for the universal $(\mathcal{P}_X, Q_Y)$-simulation, if there exists a sequence of simulators $\{f_k\}_{k=1}^\infty$ such that

$$\lim_{k \to \infty} \sup_{P_X} D_\alpha(P_{Y_k} \parallel Q_Y) \leq g(P_X) \quad (8)$$

for all $P_X \in \mathcal{P}_X$, where $P_{Y_k} := P_X \circ f_k^{-1}$.

**Definition 3.2.** The set of $\alpha$-Rényi-achievable functions for the universal $(\mathcal{P}_X, Q_Y)$-simulation is defined as

$$\mathcal{E}_{\text{Rényi}}^{(\alpha)}(\mathcal{P}_X, Q_Y) := \{ g : \mathcal{P}_X \to \mathbb{R} : g \text{ is } \alpha\text{-Rényi-achievable} \}.$$  

For $(\mathcal{A}, \mathcal{B}_A) \subset (\mathbb{R}, \mathcal{B}_R)$, define

$$\mathcal{P}_{\text{ac}}^{(\alpha)}(\mathcal{A}, \mathcal{B}_A) := \left\{ P_X \in \mathcal{P}_{\text{ac}}(\mathcal{A}, \mathcal{B}_A) : \lim_{\Delta \to 0} \sup_{\Delta \leq 1} D_\alpha(P_X|\{A_i\} \parallel P_X) = 0 \right\},$$

where the supremum is taken over all partitions $\{\{A_i\} : A_i \text{ are intervals, and } \bigcup_i A_i = A_i, A_i \cap A_j = 0, \forall i \neq j\}$ of $A$, $|A_i|$ is the length of the interval $A_i$, and $P_X|\{A_i\} := \frac{1}{|A_i|} P_X(A_i)$. Based on the notations above, we have the following theorem.
Theorem 3.2. Assume \((\mathcal{A}, \mathcal{B}_A) \subset (\mathbb{R}, \mathcal{B}_\mathbb{R})\),
\[
\mathcal{P}_X = \begin{cases} 
\mathcal{P}_{ac} & \alpha \in [0,1) \\
\mathcal{P}_{ac}^{(\alpha)} (\mathcal{A}, \mathcal{B}_A) & \alpha \in [1,\infty] 
\end{cases},
\]
and \(Q_Y\) is arbitrary. Then for the universal \((\mathcal{P}_X, Q_Y)\)-simulation problem, \(\mathcal{E}_{\text{Rényi}}^{(\alpha)}(\mathcal{P}_X, Q_Y) \approx 0\).

Remark 3.1. For \(0 < a \leq b < \infty\),
\[
\mathcal{P}^{[a,b]}_{ac} (\mathcal{A}, \mathcal{B}_A) := \{P_X \in \mathcal{P}_{ac} : a \leq p(x) \leq b \text{ a.e. in } \mathcal{A}, \text{ and } P_X(\mathbb{R}\setminus\mathcal{A}) = 0\}
\subset \mathcal{P}^{(\alpha)}_{ac} (\mathcal{A}, \mathcal{B}_A), \alpha \in [1,\infty]
\]
Hence Theorem 3.2 still holds for \(\mathcal{P}_X = \mathcal{P}^{[a,b]}_{ac} (\mathcal{A}, \mathcal{B}_A)\) and \(\alpha \in [1,\infty]\). Furthermore, Theorem 3.2 also holds if in (8) \(D_\alpha(P_Y \| P_X)\) is replaced with \(D_\alpha(Q_Y \| P_X)\), \(\alpha \in [0,\infty]\).

Proof. It is easy to verify that (4) with the total variable distance replaced by the Rényi divergence still holds. This implies the cases \(\alpha \in [1,\infty]\) in Theorem 3.2. The cases \(\alpha \in [0,1)\) are implied by combining Theorem 3.1 with the following lemma which shows the “equivalence” between the Rényi divergence of order \(\alpha \in (0,1)\) and the TV distance, in the sense that \(D_\alpha(P_X \| Q_X) \rightarrow 0\) if and only if \(|P_X - Q_X|_{TV} \rightarrow 0\).

Lemma 3.1. For \(\alpha \in (0,1)\),
\[
\frac{1}{\alpha - 1} \log \left( 1 + \frac{1}{2} |P_X - Q_X|_{TV} \right) \leq D_\alpha(P_X \| Q_X) \leq \frac{1}{\alpha - 1} \log \left( 1 - \frac{1}{2} |P_X - Q_X|_{TV} \right).
\]

Proof. This lemma follows from the following two inequalities. Define \(A := \{x : P_X(x) \geq Q_X(x)\}\). Then
\[
D_\alpha(P_X \| Q_X) = \frac{1}{\alpha - 1} \log \left( \int_A \left( \frac{dP_X}{dQ_X} \right) \alpha \ dQ_X + \int_{\mathbb{R}\setminus A} \left( \frac{dQ_X}{dP_X} \right)^{1-\alpha} dP_X \right)
\leq \frac{1}{\alpha - 1} \log \left( \int_A dQ_X + \int_{\mathbb{R}\setminus A} dP_X \right)
= \frac{1}{\alpha - 1} \log \left( \int_A (P_X(A) - Q_X(A)) \right)
= \frac{1}{\alpha - 1} \log \left( 1 - \frac{1}{2} |P_X - Q_X|_{TV} \right).
\]
Similarly,
\[
D_\alpha(P_X \| Q_X) \geq \frac{1}{\alpha - 1} \log \left( 1 + \frac{1}{2} |P_X - Q_X|_{TV} \right).
\]

3.2 Discontinuous seed distributions

Next we consider the case \(P_X \subseteq \mathcal{P}_{ac}\). We first derive a discontinuous version of Theorem 3.1. Since in the previous subsection, Theorem 3.1 is proven only for absolutely continuous seed distributions, one may doubt the effectiveness of the proposed universal mapping in Fig. 2 when the seed distributions are discontinuous or even discrete. In the following, we prove that our proposed universal mapping still works well for discontinuous seed distributions, as long as the CDF of seed distribution is smooth enough.
Figure 4: Universal simulations in Proposition 3.3 for the uniform distribution on $[0,1]$ by the mapping $x \mapsto \sum_{i=-\infty}^{\infty} \frac{1}{\Delta}(x-i\Delta)1 \{x \in (i\Delta, (i+1)\Delta]\}$ with $\Delta = 0.01$. The blue and red curves in the left figure correspond to the cases in which the seed distributions $P_X$ are respectively the standard Gaussian distribution $\mathcal{N}(0,1)$ and its quantized version. The blue and red curves in the right figure correspond to the cases in which the seed distributions $P_X$ are respectively the logarithmic distribution $p_X(x) = -\log x, x \in [0,1]$ and its quantized version. For both of these two cases, the quantized versions are generated by using the same quantization step $\frac{1}{n} = 0.0001$.

Assume $\{\Delta_n\}$ is a sequence of non-increasing positive numbers. Assume $\{P_{X_n}\}_{n=1}^{\infty}$ is a sequence of distribution sets such that for every sequence $\{P_{X_n}\} \in \{P_{X_n}\}$, its CDFs $\{F_{X_n}\}$ satisfy

$$\lim_{n \to \infty} \sup_{x_1:F_{X_n}(x_1+\Delta_n)>F_{X_n}(x_1)} \sup_{x \in (x_1,x_1+\Delta_n)} \left| \frac{F_{X_n}(x_1+x) - F_{X_n}(x_1)}{F_{X_n}(x_1+\Delta_n) - F_{X_n}(x_1)} - \frac{x}{\Delta_n} \right| = 0. \tag{9}$$

We obtain a discontinuous version of Theorem 3.1. The proof of Proposition 3.3 is provided in Appendix B.

**Proposition 3.3.** Assume $\{P_{X_n}\}$ is the sequence of distribution sets defined above. Then there exists a sequence of universal mappings $Y_n = f_n(X_n)$ (which are dependent on $\{\Delta_n\}$) such that $\lim_{n \to \infty} |P_{Y_n} - Q_Y|_{KS} = 0, \forall \{P_{X_n}\} \in \{P_{X_n}\}$. That is, $\lim_{n \to \infty} \varepsilon_{KS}(P_{X_n}, Q_Y) \approx 0$.

The proposition above implies that if the CDF of seed random variable $X_n$ gets more and more smooth as $n \to \infty$ in the sense that it can be approximated by a linear function for every small interval $(x_1, x_1 + \Delta_n]$ as (9), then we can find a sequence of universal mappings that achieve vanishing KS-approximation error. Here is a simple example.

**Example 3.2.** Assume $X$ is an absolutely continuous random variable with a bounded PDF. We define $X_n := \left\lfloor \frac{nX}{\Delta_n} \right\rfloor$ as a quantized version of $X$ with quantization step $\frac{1}{n}$, and $\Delta_n$ is set to $\frac{1}{\sqrt{n}}$, then $\{\{X_n\}, \{\Delta_n\}\}$ satisfies (9).

The example above with $\frac{1}{n} = 0.0001$, $\Delta = 0.01$, and $P_X$ to be the standard Gaussian distribution or logarithmic distribution, is illustrated in Fig. 4.

If the seed is a sequence of i.i.d. discrete random vectors, then the approximation error decays exponentially fast. Given a Borel subset $(X, \mathcal{B}_X) \subseteq (\mathbb{R}, \mathcal{B}_\mathbb{R})$ with $\mathcal{X}$ countable, $\mathcal{P}(X, \mathcal{B}_X)$ denotes the set of distributions on $(X, \mathcal{B}_X)$.

**Theorem 3.3.** Assume $P_X = P(X, \mathcal{B}_X)$ and $Q_Y$ is continuous. Then for the universal $(P_{X}^{(n)}, Q_Y)$-simulation problem, $\varepsilon_{KS}(P_{X}^{(n)}, Q_Y) \approx (\max_x P_X(x))^n$. 

13
Proof. We first consider the case in which \( \mathcal{X} \) is a finite set. We use a type-based mapping scheme to prove Theorem 3.3. Here we adopt the notation from [17]. We use \( T_{x^n} (x) := \frac{1}{n} \sum_{i=1}^{n} 1 \{ x_i = x \} \) to denote the type (empirical distribution) of a sequence \( x^n \), and \( T_X \) to denote a type of sequences in \( \mathcal{X}^n \), where the indicator function \( 1 \{ A \} \) equals 1 if the clause \( A \) is true and 0 otherwise. For a type \( T_X \), the type class (set of sequences having the same type \( T_X \)) is denoted by \( \mathcal{T}_X \). The set of types of sequences in \( \mathcal{X}^n \) is denoted as \( P_n (\mathcal{X}) := \{ T_{x^n} : x^n \in \mathcal{X}^n \} \). It has been shown that \( |P_n (\mathcal{X})| \leq (n + 1)^{|\mathcal{X}|} \) in [17].

For any i.i.d. \( X^n \), all sequences in a type class have an equal probability. That is, under the condition \( X^n \in \mathcal{T}_X \), it is uniformly distributed over the type class \( \mathcal{T}_X \), regardless of \( P_X \). Now we construct a mapping \( f \) that maps the uniform random vector on \( \mathcal{T}_X \) to a random variable such that \( \sup_{y \in \mathbb{R}} |F_Y (y) - G_Y (y)| \) is minimized. Here \( F_Y (y|T_X) \) denotes the CDF of the output random variable for the type \( T_X \). Since the probability values of uniform random vectors are all equal to \( |\mathcal{T}_X|^{-1} \), \( \sup_{y \in \mathbb{R}} |F_Y (y|T_X) - G_Y (y)| = \frac{1}{2} |\mathcal{T}_X|^{-1} \). Therefore, the output distribution induced by \( f \) is

\[
F_Y (y) = \sum_{T_X} \sum_{x^n \in \mathcal{T}_X} P_X^n (x^n) 1 \{ f(x^n) \leq y \}
\]

\[
= \sum_{T_X} P_X^n (\mathcal{T}_X) \sum_{x^n \in \mathcal{T}_X} \frac{P_X^n (x^n)}{P_X (\mathcal{T}_X)} 1 \{ f(x^n) \leq y \}
\]

\[
= \sum_{T_X} P_X^n (\mathcal{T}_X) \sum_{x^n \in \mathcal{T}_X} \frac{1}{|T_X|} 1 \{ f(x^n) \leq y \}
\]

\[
= \sum_{T_X} P_X^n (\mathcal{T}_X) F_Y (y|T_X)
\]

\[
\in \sum_{T_X} P_X^n (\mathcal{T}_X) \left( G_Y (y) + \left[ - \frac{1}{2} |\mathcal{T}_X|^{-1}, \frac{1}{2} |\mathcal{T}_X|^{-1} \right] \right)
\]

\[
= G_Y (y) + \sum_{T_X} P_X^n (\mathcal{T}_X) \left[ - \frac{1}{2} |\mathcal{T}_X|^{-1}, \frac{1}{2} |\mathcal{T}_X|^{-1} \right].
\]

Using this equation we obtain

\[
|F_Y (y) - G_Y (y)| \leq \frac{1}{2} \sum_{T_X} P_X^n (\mathcal{T}_X) |\mathcal{T}_X|^{-1}
\]

\[
= \frac{1}{2} \sum_{T_X} e^n \sum_{x} T_X (x) \log P_X (x)
\]

\[
\leq \frac{1}{2} \left( n + 1 \right)^{|\mathcal{X}|} \max_{T_X} e^n \sum_{x} T_X (x) \log P_X (x)
\]

\[
\leq \frac{1}{2} e^n \left( \log \max_x P_X (x) + |\mathcal{X}| \log (n + 1) \right)
\]

\[
= e^n \log \max_x P_X (x)
\]

\[
\left( \max_x P_X (x) \right)^n
\]

(11)

We next consider the case in which \( \mathcal{X} \) is countably infinite. For brevity, we assume \( \mathcal{X} = \mathbb{Z} \). We partition\( \mathbb{Z} \) into \( 2k + 1 \) intervals\(^7\) \( \mathcal{U}_{-k} := [ -\infty : -k ] \), \( \mathcal{U}_{-(k-1)} := [ -(k-1) : k-1 ] \), \( \ldots \), \( \mathcal{U}_{k-1} := [ k-1 : k ] \), \( \mathcal{U}_k := [ k : \infty ] \). Denote \( Z_k = f_{1,k} (X) \in \mathbb{Z} := [ -k : k ] \) as the index that \( X \in \mathcal{U}_{Z_k} \). Hence \( P_{Z_k} \) is defined on the finite set \( \mathbb{Z} \). Now we use \( Z_k \) to simulate \( Y \sim Q_Y \). By the derivation above, we have that there exists a universal mapping \( Y_k = f_{2,k} (Z_k) : \mathbb{Z} \to \mathcal{Y} \) such that \( |P_{Y_k} - Q_Y|_{KS} \leq (\max_z P_{Z_k} (z))^n \). Furthermore,

\(^7\)Sometimes, we use \( [a : b] \) to denote \( \mathbb{Z} \cap [a, b] \).
as \( k \to \infty \), \( \max_x P_{Z_k}(z) \to \max_x P_X(x) \). Therefore, the universal mappings \( f_{2,k} \circ f_{1,k}, k \in \mathbb{Z} \) satisfy

\[
\limsup_{k \to \infty} |P_{Y_k} - Q_Y|_{KS} \leq (\max_x P_X(x))^n.
\]

The converse part follows from Theorem 2.3, since even non-universal simulation cannot make the approximation error decay faster than \( (\max_x P_X(x))^n \), hence universal simulation cannot as well.

Now we consider a discontinuous \( P_X \). We partition the real line into intervals \( U_k := ((k-1)\Delta, k\Delta], k \in \mathbb{Z} \) with \( \Delta = \frac{1}{2n} \). Denote \( Z_k = f_{1,k}(X) \in \mathbb{Z} \) as the index that \( X \in U_{Z_k} \). Hence \( P_{Z_k} \) is defined on the set \( \mathbb{Z} \). Now we use \( Z_k \) to simulate \( Y \sim Q_Y \). By Theorem 3.3, we have that there exists a universal mapping \( Y_k = f_{2,k}(Z_k) : \mathbb{Z} \to \mathcal{Y} \) such that \( |P_{Y_k} - Q_Y|_{KS} \leq (\max_x P_{Z_k}(z))^n \). Furthermore, as \( k \to \infty \), we have \( \max_x P_{Z_k}(z) \to \max_x P_X(x) \). Therefore, the universal mappings \( f_{2,k} \circ f_{1,k}, k \in \mathbb{Z} \) satisfy

\[
\limsup_{k \to \infty} |P_{Y_k} - Q_Y|_{KS} \leq (\max_x P_X(x))^n.
\]

We therefore have the following result.

**Corollary 3.1.** Assume \( P_X = P_{dc} \) and \( Q_Y \) is continuous. Then for the universal \( (\mathcal{P}_X^n, Q_Y) \)-simulation problem, \( \mathcal{E}_{KS}(\mathcal{P}_X^n, Q_Y) \approx (\max_x P_X(x))^n \).

### 3.3 Continuous but not absolutely continuous seed distributions

Next we consider continuous but not absolutely continuous \( P_X \). For this case, we have \( \max_x P_X(x) = 0 \). Hence the approximation error decays super-exponentially fast. To provide a better bound, we assume \( F_X \) is Hölder continuous with exponent \( \alpha \), where \( 0 < \alpha \leq 1 \). That is, \( L = \sup_{x_1 \neq x_2} \frac{F_X(x_2) - F_X(x_1)}{|x_2 - x_1|^\alpha} \) is finite. Consider the following mapping. We partition the real line into \( 2k + 2 \) intervals \( U_{-k} := (-\infty, -k\Delta], U_{-(k-1)} := (-k\Delta, -(k-1)\Delta], \ldots, U_k := ((k-1)\Delta, k\Delta], U_{k+1} := (k\Delta, \infty) \). Denote \( Z = f_1(X) \in \mathbb{Z} := [-k : k + 1] \) as the index that \( X \in U_Z \). Now we use \( Z \) to simulate \( Y \sim Q_Y \). By the derivation till (11), we have that there exists a universal mapping \( Y_n = f_2(Z) : \mathbb{Z} \to \mathcal{Y} \) such that \( |P_Y - Q_Y|_{KS} \leq 2 \cdot e^{n(\log \max_x P_X(z) + (2k+2)\log(n+1)/n)} \). Set \( \Delta = \frac{\log n}{n}, k = \frac{n}{\log n} \). Then \( \Delta \to 0 \) and \( k\Delta = \sqrt{\log n} \to \infty \). Since \( F_X \) is Hölder continuous with exponent \( \alpha \), we have \( \max_x P_Z(z) \leq \max_x \{F_X(x + \Delta) - F_X(x)\} \leq L\Delta^n \). Hence the universal mapping \( Y_n = f_2 \circ f_1(X^n) \) satisfies \( |P_{Y_n} - Q_Y|_{KS} \leq e^{-\alpha n \log n} \). Therefore, we have the following result.

**Corollary 3.2.** Assume \( P_X = P_c \setminus P_{uc} \) and \( Q_Y \) is arbitrary. Then for the universal \( (\mathcal{P}_X^n, Q_Y) \)-simulation problem, \( \mathcal{E}_{KS}(\mathcal{P}_X^n, Q_Y) \approx e^{-\omega(n)} \). That is, there exists a sequence of simulators such that \( |P_Y(n) - Q_Y|_{KS} \) decays super-exponentially fast as \( n \to \infty \) for any \( P_X \). Moreover, if \( P_X = \{P_X : F_X \) is Hölder continuous with exponent \( \alpha \} \) with \( 0 < \alpha \leq 1 \), then \( \mathcal{E}_{KS}(\mathcal{P}_X^n, Q_Y) \approx e^{-\alpha n \log n} \).

### 4 Simulating a Random Variable from a Markov Process

In the preceding sections, we consider simulation of a random variable from a stationary memoryless process. Next we extend Theorem 3.3 to Markov processes of order \( k \geq 1 \).

**Definition 4.1.** Given a Markov chain \( X = \{X_n : n \in \mathbb{N}\} \) of order \( k \geq 1 \) with finite state space \( \mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\} \), initial state \( x^0_{-k+1} := (x_{-k+1}, x_{-k+2}, \ldots, x_0) \), and transition probability \( P_{X_{k+1}|X^k} \), the min-entropy (\( \infty \)-order Rényi entropy) rate of \( X \) is defined as

\[
H_\infty(X) = -\lim_{n \to \infty} \frac{1}{n} \max_{P(x^n)} \log P(x^n).
\]

Since the distribution of \( X^n \) is determined by the initial state \( x^0_{-k+1} \) and transition probability \( P_{X_{k+1}|X^k} \), hence sometimes we also use \( H_\infty(x^0_{-k+1}, P_{X_{k+1}|X^k}) \) to denote \( H_\infty(X) \).

Given a state space \( \mathcal{X} \) of any Markov chain \( \{X_n : n \in \mathbb{N}\} \) of order \( k = 1 \) with transition probability \( P_{X_{k+1}|X^k} \), a loop is a sequence of distinct states of the chain \((i_1, i_2, \ldots, i_l)\) with \( l \geq 1 \) such that \( P_{i_{l-1}i_l} > 0 \)

\^8The existence of the limit is guaranteed by Fekete’s subadditive lemma.
for $s = 1, 2, \ldots, l$ where $i_{s+1} = i_1$. (If $P_{i,i} > 0$, then $(i)$ is a loop.) The set of all loops of length $l$ is denoted by $C_l(P)$.

Let $P$ be the transition matrix of an ergodic Markov chain of order $k = 1$ on a finite alphabet $\mathcal{X}$. The min-entropy rate of this Markov chain is given by [18]

$$H_\infty(X) = \min_{1 \leq l \leq |\mathcal{X}|} \frac{1}{l} \sum_{s=1}^{l} \log \frac{1}{P_{t_s+1,t_s}},$$

where the inner minimum is taken over all loops $(i_1, i_2, \ldots, i_l) \in C_l(P)$, and $t_{X_1}(j|i) := \log \frac{1}{P_{j|i}}$.

### 4.1 Non-universal Simulation from a Markov Process

As a direct consequence of Proposition 2.1, we can obtain the approximation error for non-universal simulation from a Markov process.

**Corollary 4.1.** Assume $X = \{X_n : n \in \mathbb{N}\}$ is a Markov chain of order $k$ with finite state space $\mathcal{X}$, initial state $x_0^{k+1}$, and transition probability $P_{X_{k+1}|X^k}$, and $Q_Y$ is a continuous distribution. Then for the non-universal $(P_{X^n}, Q_Y)$-simulation problem, $E_{KS}(P_{X^n}, Q_Y) \geq e^{-nH_\infty(x_0^{k+1}, P_{X_{k+1}|X^k})}$.

### 4.2 Universal Simulation from a Markov Process

Given a finite state space $\mathcal{X}$, a order $k \geq 1$, and a initial state $x_0^{k+1}$, we denote the set of all possible distributions of Markov chains with these parameters as $\mathcal{P}_{X_{k+1}|X^k} := \{\prod_{n=1}^{k+1} P_{X_{k+1}|X^k} : P_{X_{k+1}|X^k} \in \mathcal{P}_{X_{k+1}|X^k}\}$, where $P_{X_{k+1}|X^k}$ denotes the set of all possible transition probability $P_{X_{k+1}|X^k}$ (from $x^k$ to $\mathcal{X}$).

We next consider universal simulation from a Markov process. We generalize Theorem 3.3 to this case. The proof of Theorem 4.1 is provided in Appendix C.

**Theorem 4.1.** Assume $Q_Y$ is a continuous distribution. Then given a finite state space $\mathcal{X}$, an order $k$, and an initial state $x_0^{k+1}$, for the universal $(\mathcal{P}^{(n)}_{X_{k+1}|X^k}, Q_Y)$-simulation problem, we have $E_{KS}(\mathcal{P}^{(n)}_{X_{k+1}|X^k}, Q_Y) \geq e^{-nH_\infty(x_0^{k+1}, P_{X_{k+1}|X^k})}$.

### 5 Simulating a Random Element from another Random Element

#### 5.1 Non-universal Simulation

Next we show that an arbitrary continuous random element (or general random variable) is sufficient to simulate another arbitrary random element. Here random elements is a generalization of random variable, which may be defined on any non-empty Borel set in a separable metric space.

**Theorem 5.1.** Assume $P_X$ and $Q_Y$ are two distributions respectively defined on any non-empty Borel sets $(\mathcal{X}, \mathcal{B}_X)$ and $(\mathcal{Y}, \mathcal{B}_Y)$ in two Polish spaces (complete separable metric spaces). If $P_X$ is continuous, then there exists a measurable mapping $Y = f(X)$ such that $P_Y = Q_Y$. That is, $E_{TV}(P_X, Q_Y) = 0$.

**Proof.** For any two Borel subsets of Polish spaces, they are Borel-isomorphic if and only if they have the same cardinality, which moreover is either finite, countable, or $\mathfrak{c}$ (the cardinal of the continuum, that is, of $[0,1]$) [14]. Hence for any measurable space $(\mathcal{X}, \mathcal{B}_X)$, we can always find a Borel subset $(\mathcal{W}, \mathcal{B}_W)$ of $([0,1], \mathcal{B}_{[0,1]})$ such that $(\mathcal{X}, \mathcal{B}_X)$ and $(\mathcal{W}, \mathcal{B}_W)$ are Borel-isomorphic. Suppose $\varphi$ is a Borel isomorphism from $(\mathcal{X}, \mathcal{B}_X)$ to $(\mathcal{W}, \mathcal{B}_W)$. Denote $P_W := P_X \circ \varphi^{-1}$. Since $P_X$ is continuous (or atomless), $P_W$ must be continuous as well. This is because if $P_X(\varphi^{-1}(w)) = P_W(w) > 0$ for some $w \in [0,1]$, then it contradicts with

---

9Here the definition of $\preceq$ is analogous to that for stationary memoryless processes.
Theorem 5.2. Assume $R_X$ and $Q_Y$ are two distributions respectively defined on any non-empty Borel sets $(X, \mathcal{B}_X)$ and $(Y, \mathcal{B}_Y)$ in two Polish spaces, and moreover, $R_X$ is continuous. $P_X(R_X)$ denotes the set of all absolutely continuous distributions (defined on $(X, \mathcal{B}_X)$) respect to $R_X$. Then for the universal $(P_X(R_X),Q_Y)$-simulation problem, $\mathcal{E}_{TV}(P_X(R_X),Q_Y) > 0$.

Proof. Since $R_X$ is continuous, as shown in the proof of Theorem 5.1, there exists a Borel isomorphism $\varphi$ from $(X, \mathcal{B}_X)$ to a Borel subset $(W, \mathcal{B}_W)$ of $([0,1], \mathcal{B}_{[0,1]})$ such that the output distribution $R_X \circ \varphi^{-1}$ is continuous. Denote $R_Z$ as the uniform distribution (which is also the Lebesgue measure) on $([0,1], \mathcal{B}_{[0,1]})$. Then by Proposition 2.1, we know that there exists a measurable mapping $\eta : (W, \mathcal{B}_W) \to ([0,1], \mathcal{B}_{[0,1]})$ such that

$$R_X \circ \varphi^{-1} \circ \eta^{-1} = R_Z. \quad (13)$$

Hence a random element $\tilde{X} \sim R_X$ is mapped to a uniform random variable $\tilde{Z} \sim R_Z$ through the mapping $\tilde{Z} = \eta \circ \varphi(\tilde{X})$. We define $P_Z := P_X \circ \varphi^{-1} \circ \eta^{-1}$, which denotes the distribution of $Z = \eta \circ \varphi(X)$ where $X \sim P_X$. Since $P_X$ is absolutely continuous respect to $R_X$, we have that $P_Z$ is absolutely continuous respect to $R_Z$ (or the Lebesgue measure). This is because on one hand, by (13), we have $R_X(\varphi^{-1} \circ \eta^{-1}(A)) = 0$ for any $A$ such that $R_Z(A) = 0$; on the other hand, $P_X$ is absolutely continuous respect to $R_X$, hence $P_X(\varphi^{-1} \circ \eta^{-1}(A)) = 0$, i.e., $P_Z(A) = 0$.

Since $P_Z$ is absolutely continuous, by Theorem 3.1, we know that there exists a sequence of universal mappings $\tau_k : ([0,1], \mathcal{B}_{[0,1]}) \to ([0,1], \mathcal{B}_{[0,1]})$ (independent of $P_Z$) such that the resulting approximation error

$$\lim_{k \to \infty} |P_Z \circ \tau_k^{-1} - R_Z|_{TV} = 0.$$

Observe that $P_Z = P_X \circ \varphi^{-1} \circ \eta^{-1}$ and $\eta, \varphi$ (only depend on $R_X$) are independent of $P_X$. Hence the universal mappings $\tau_k \circ \eta \circ \varphi$ satisfy

$$\lim_{k \to \infty} |P_X \circ \varphi^{-1} \circ \eta^{-1} \circ \tau_k^{-1} - R_Z|_{TV} = 0. \quad (14)$$

Since $R_Z$ is continuous, by Theorem 5.1, we know that then there exists a measurable mapping $\kappa : ([0,1], \mathcal{B}_{[0,1]}) \to (Y, \mathcal{B}_Y)$ such that $R_Z \circ \kappa^{-1} = Q_Y$.

Now consider the universal mappings $Y = \kappa \circ \tau_k \circ \eta \circ \varphi(X)$. We have

$$|P_X \circ \varphi^{-1} \circ \eta^{-1} \circ \tau_k^{-1} \circ \kappa^{-1} - Q_Y|_{TV} = \sup_{A \in \mathcal{B}_Y} |P_X \circ \varphi^{-1} \circ \eta^{-1} \circ \tau_k^{-1} \circ \kappa^{-1}(A) - Q_Y(A)|$$

$$= \sup_{A \in \mathcal{B}_Y} |P_X \circ \varphi^{-1} \circ \eta^{-1} \circ \tau_k^{-1}(\kappa^{-1}(A)) - R_Z(\kappa^{-1}(A))|$$

$$\leq \sup_{B \in \mathcal{B}_{[0,1]}} |P_X \circ \varphi^{-1} \circ \eta^{-1} \circ \tau_k^{-1}(B) - R_Z(B)|$$
\begin{equation*}
= |P_X \circ \varphi^{-1} \circ \eta^{-1} \circ \tau_k^{-1} - R|_{\text{TV}}.
\end{equation*}
Combining this with \((14)\) yields
\begin{equation*}
\lim_{k \to \infty} |P_X \circ \varphi^{-1} \circ \eta^{-1} \circ \tau_k^{-1} \circ \kappa^{-1} - Q_Y|_{\text{TV}} = 0.
\end{equation*}
\hfill \Box

Note that a random vector defined on \((\mathbb{R}^n, \mathcal{B}_{\mathbb{R}}^n)\), \(n \in \mathbb{Z}^+\) is a special case of such a random element. Furthermore, for any absolutely continuous (respect to the Lebesgue measure) \(P_{X^n}\) defined on \((\mathbb{R}^n, \mathcal{B}_{\mathbb{R}}^n)\), \(n \in \mathbb{Z}^+\), it must be absolutely continuous respect to the \(n\)-dimensional standard Gaussian distribution (since its PDF is positive for every point in \(\mathbb{R}^n\)). Hence we have the following corollary.

**Corollary 5.2.** For the set \(\mathcal{P}_{ac}\) of absolutely continuous distributions on \((\mathbb{R}, \mathcal{B}_{\mathbb{R}})\) and an arbitrary \(Q_{Y^n}\) on \((\mathbb{R}^n, \mathcal{B}_{\mathbb{R}}^n)\), \(n \in \mathbb{Z}^+\), the approximation errors for the universal \((\mathcal{P}_{ac}, Q_{Y^n})\)-simulation problem satisfies \(\mathcal{E}_\theta(\mathcal{P}_{ac}, Q_{Y^n}) \approx 0\) for \(\theta \in \{\text{KS}, \text{TV}\}\).

### 6 Concluding Remarks

In this paper, motivated by the CLT and other universal simulation problems in the literature, we consider both universal and non-universal simulations of random variables with an arbitrary target distribution \(Q_Y\) by general mappings. We investigate the fastest convergence rate of the approximation error for such a problem. One of our interesting results is that under universal simulation, an absolutely continuous random element (or a general random variable, including random vectors) respect to some continuous distribution is sufficient to simulate another random element arbitrarily well. This requirement is a little stronger than that for non-universal simulation, since under non-universal simulation, a continuous random element is sufficient to exactly simulate another random element. Another interesting result is that when we use a stationary memoryless process or a Markov process to simulate a random variable by a universal mapping, the approximation error decays at least exponentially fast with rate \(H_\infty(P_X) := -\log \max_x P_X(x)\) as the dimension \(n\) of \(X^n\) goes to infinity. Furthermore, as a byproduct, we also obtain a property on uncorrelation between a squeezed periodic function and any other integrable function. We think this topic is of independent interest, and expect it to be further applied in other problems in the future.

As for application aspects of our results, although practical digital computers have finite precision for processing or storing datum, as indicated by Proposition 3.3, our proposed universal mapping in Fig. 2 still works well on such digital computers, as long as they have sufficiently high precision; see the illustration in Fig. 4.

### A Proof of Proposition 2.3

Sort the sequences in \(X^n\) as \(x_1^n, x_2^n, \ldots, x_{|X|^n}\) such that \(P_X^n(x_1^n) \geq P_X^n(x_2^n) \geq \ldots \geq P_X^n(x_{|X|^n})\). Map \(x_1^n\) to \(y_1 := \arg \max_{y \in Y} Q_Y(y)\); map \(x_2^n\) to \(y_2 := \arg \max_{y \in Y} \{Q_Y(y) - P_X^n(x_1^n)1\{y = y_1\}\}; \ldots;\) map \(x_{|X|^n-|Y|-1}^n\) to \(y_{|X|^n-|Y|-1} := \arg \max_{y \in Y} \{Q_Y(y) - \sum_{i=1}^{i=|X|^n-|Y|-1} P_X^n(x_i^n)1\{y = y_i\}\}\). Map the remaining \(|Y| + 1\) sequences \(x_j^n, |X|^n - |Y| \leq j \leq |X|^n\) to sequences in \(\mathcal{Y}\) in a similar way as in the proof of Proposition 2.2, such that \(\sup_{y \in \mathcal{Y}} \sum_{i=1}^j P_X^n(x_i^n)1\{y' = y_i\} - G_Y(y)\) is minimized for \(|X|^n - |Y| \leq j \leq |X|^n\).

For this mapping, observe that for any \(1 \leq j \leq |X|^n\),
\begin{equation*}
\max_{y \in \mathcal{Y}} \left\{Q_Y(y) - \sum_{i=1}^j P_X^n(x_i^n)1\{y = y_i\}\right\} \geq \frac{\sum_{y \in \mathcal{Y}} \left\{Q_Y(y) - \sum_{i=1}^j P_X^n(x_i^n)1\{y = y_i\}\right\}}{|\mathcal{Y}|} = \frac{1 - \sum_{i=1}^j P_X^n(x_i^n)}{|\mathcal{Y}|}
\end{equation*}
Next we prove the following claim.

Claim A.1. If \( P^n_X(x^n_{j+1}) \leq \frac{\sum_{i=1}^{j+1} P^n_Y(x^n_i)}{|Y|} \) holds for \( 1 \leq j \leq m \) (for some integer \( m \)), then \( \sum_{i=1}^{m+1} P^n_X(x^n_i)1 \{ y_j = y_i \} \leq Q_Y(y) \) for all \( y \in Y \).

We split the proof into two cases.

- **Case 1** (\( 1 \leq j \leq |X|^n - |Y|^2 \)): For \( 1 \leq j \leq |X|^n - n - 2 \), denote \( k \) as the smallest integer such that \( k \geq j + 2 \) and \( T^n_{x^n_k} \neq T^n_{x^n_{j+1}} \). Then we have

\[
P^n_X(x^n_{j+1}) \leq P^n_X(x^n_k) \max_{1 \leq i \leq |X| - 1} \frac{P_X(x_i)}{P_X(x_{i+1})}.
\]

This is because if \( x^n_k \) and \( x^n_{j+1} \) are different in only one component, say the \( i \)th components \( x_{k,i} \) and \( x_{j+1,i} \), then by the generation process of \( x^n_1, x^n_2, ..., x^n_{|X|,n} \), we have \( \frac{P_X(x_{i+1})}{P_X(x_{i+1})} \geq \min_{1 \leq i \leq |X| - 1} \frac{P_X(x_k)}{P_X(x_{k+1})} \). Hence

\[
P^n_X(x^n_{j+1}) = P^n_X(x^n_k) \frac{P_X(x_{j+1,i})}{P_X(x_{k,i})} \leq P^n_X(x^n_k) \max_{1 \leq i \leq |X| - 1} \frac{P_X(x_i)}{P_X(x_{i+1})}.
\]

If \( x^n_k \) and \( x^n_{j+1} \) are different in more than one components, then by the generation process of \( x^n_1, x^n_2, ..., x^n_{|X|,n} \), we have \( P^n_X(x^n_{j+1}) \geq P^n_X(x^n_k) \geq P^n_X(\tilde{x}^n_k) \) for any sequence \( \tilde{x}^n_k \) such that \( \tilde{x}^n_k \) are different from \( x^n_{j+1} \) in only one component. Hence by the same argument above, we obtain

\[
P^n_X(x^n_{j+1}) \leq P^n_X(\tilde{x}^n_k) \max_{1 \leq i \leq |X| - 1} \frac{P_X(x_i)}{P_X(x_{i+1})} \leq P^n_X(x^n_k) \max_{1 \leq i \leq |X| - 1} \frac{P_X(x_i)}{P_X(x_{i+1})}.
\]

Therefore, (15) holds. Next, we prove (15) implies Claim A.2.

\[
\frac{P^n_X(x^n_{j+1})}{\sum_{i=1}^{|X|^n} P^n_X(x^n_i)} \leq \frac{P^n_X(x^n_k)}{\sum_{i=k}^{|X|^n} P^n_X(x^n_i)} \leq \frac{P^n_X(x^n_k) \max_{1 \leq i \leq |X| - 1} \frac{P_X(x_i)}{P_X(x_{i+1})}}{\sum_{i=k}^{|X|^n} P^n_X(x^n_i)} \leq \frac{\max_{1 \leq i \leq |X| - 1} \frac{P_X(x_i)}{P_X(x_{i+1})}}{n} \leq \frac{1}{|Y|}.
\]

For \( n \geq |Y| \max_{1 \leq i \leq |X| - 1} \frac{P_X(x_i)}{P_X(x_{i+1})} \), we have (16) \( \leq \frac{1}{|Y|} \). Therefore, \( P^n_X(x^n_{j+1}) \leq \frac{\sum_{i=1}^{|X|^n} P^n_Y(x^n_i)}{|Y|} \).
• **Case 2** ($|\mathcal{X}|^n - n - 1 \leq j \leq |\mathcal{X}|^n - |\mathcal{Y}| - 2$): For $|\mathcal{X}|^n - n - 1 \leq j \leq |\mathcal{X}|^n - |\mathcal{Y}| - 2$, we know $x_{j+1}^n, ..., x_{n-1}^n$ belong to the same type class, $P(x_{j+1}^n) = ... = P(x_{n-1}^n) = P_X(x_{[X]}^{n-1})$, and $P(x_{n}^n) = (P_X(x_{[X]}))^{n-1}$. Hence we have

$$P^n_{X}(x_{j+1}^n) \leq \frac{P^n_{X}(x_{j+1})}{|\mathcal{Y}|^{n-1}} = \frac{1}{|\mathcal{X}|^n}.$$  

Combining the two cases above, we have Claim A.2. By Claim A.1, we further have $\sum_{i=1}^{n+1} P^n_{X}(x_i^1) \{ y = y_i \} \leq Q_Y(y)$ for all $y \in \mathcal{Y}$ with $m = |\mathcal{X}|^n - |\mathcal{Y}| - 2$.

Since the remaining $|\mathcal{Y}| + 1$ sequences are mapped to sequences in $\mathcal{Y}$ in a similar way as in the proof of Proposition 2.2, similar to Proposition 2.2, here we can show that the output measure $P_Y$ induced by the mapping satisfies $|P_Y - Q_Y|_{\text{KS}} \leq \frac{1}{2} P_X(x_{[X]}^{n-1}) (P_X(x_{[X]}))^{n-1}$.

**B Proof of Proposition 3.3**

**Universal Mapping:** For $X_n$, partition the real line into intervals with the same length $\Delta_n$, i.e., $\bigcup_{i=-\infty}^{\infty} (i \Delta_n, (i+1) \Delta_n)$. We first simulate a uniform distribution on $[a, b]$ by mapping each interval $(i \Delta_n, (i+1) \Delta_n]$ into $[a, b]$ using the linear function $x \mapsto a + \frac{b - a}{\Delta_n} (x - i \Delta_n)$. That is, the function used here is

$$f_1(x) := \sum_{i=-\infty}^{\infty} \left( a + \frac{b - a}{\Delta_n} (x - i \Delta_n) \right) \{ x \in (i \Delta_n, (i+1) \Delta_n) \}.$$  

We then transform the output distribution to the target distribution $Q_Y$, by using function $x \mapsto G_Y^{-1} \left( \frac{x - a}{b - a} \right)$, where $G_Y^{-1}(t) := \min \{ y : G_Y(y) \geq t \}$. Therefore, each $x \in (i \Delta_n, (i+1) \Delta_n]$ is mapped to $G_Y^{-1} \left( \frac{1}{\Delta_n} (x - i \Delta) \right)$. Hence the final mapping is

$$f(x) := \sum_{i=-\infty}^{\infty} G_Y^{-1} \left( \frac{1}{\Delta_n} (x - i \Delta_n) \right) \{ x \in (i \Delta_n, (i+1) \Delta_n) \}.$$  

For such a universal simulator, we have

$$|P_Y - Q_Y|_{\text{KS}} \leq |P_{f_1(X_n)} - \text{Unif}([a, b])|_{\text{KS}}$$  

$$= \sup_{x \in [a, b]} \left| \sum_{i=-\infty}^{\infty} F_X(x + \Delta_n \frac{z - a}{b - a} - F_X(i \Delta_n) \right| - \frac{z - a}{b - a} \right|$$  

$$= \sup_{x \in [0, \Delta_n]} \left| \sum_{i=-\infty}^{\infty} [F_X(x + \Delta_n) - F_X(i \Delta_n)] \left( \frac{F_X(i \Delta_n + x) - F_X(i \Delta_n)}{F_X((i+1) \Delta_n) - F_X(i \Delta_n)} - \frac{x}{\Delta_n} \right) \right|$$  

$$\leq \sum_{i=-\infty}^{\infty} [F_X((i+1) \Delta_n) - F_X(i \Delta_n)] \sup_{x \in [0, \Delta_n]} \left| \frac{F_X(i \Delta_n + x) - F_X(i \Delta_n)}{F_X((i+1) \Delta_n) - F_X(i \Delta_n)} - \frac{x}{\Delta_n} \right|$$  

$$\leq \sup_{x_1 : F_X(x_1 + \Delta_n) > F_X(x_1)} \sup_{x \in (x_1, x_1 + \Delta_n]} \left| \frac{F_X(x_1 + x) - F_X(x_1)}{F_X(x_1 + \Delta_n) - F_X(x_1)} - \frac{x}{\Delta_n} \right| \to 0,$$  

as $n \to \infty$,

where (17) follows from Lemma 1.1.
C Proof of Theorem 4.1

We still use a type-based mapping scheme (similar to that used in the proof of Theorem 3.3) to prove Theorem 4.1.

Here we adopt the notation from [20]. Assume the Markov process $X$ starts from the fixed initial state $(x_{-k+1}, x_{-k+2}, ..., x_0)$. The $k$-th order Markov type class $T_k$ is the set of all sequences $x^n \in \mathcal{X}^n$ that have the same type $T_k$. Obviously, all sequences in a type class have a equal probability, i.e., $P(x^n) = P(\hat{x}^n)$ for all $x^n, \hat{x}^n \in T_k$. That is, under the condition $X^n \in T_k$, it is uniformly distributed over the type class $T_k$, regardless of the distribution of the Markov process. Furthermore, the set of $k$-th order Markov types of sequences in $\mathcal{X}^n$ is denoted as $\mathcal{P}_{k,n}(\mathcal{X}) := \{T_{x^n} : x^n \in \mathcal{X}^n\}$. It has been shown that $|\mathcal{P}_{k,n}(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|+1}$ in [20].

Now we construct a mapping $f$ that maps the uniform random vector on $T_k$ to a random variable such that $\sup_{y \in \mathbb{R}} |F_Y(y|T_k) - G_Y(y)|$ is minimized. Here $F_Y(y|T_k)$ denotes the CDF of the output random variable for the uniform random vector on $T_k$. Since the probability values of uniform random vectors are all equal to $|T_k|^{-1}$, $\sup_{y \in \mathbb{R}} |F_Y(y|T_k) - G_Y(y)| = \frac{1}{2} |T_k|^{-1}$. Following the same steps as (10)-(12), we obtain

$$|F_Y(y) - G_Y(y)| \leq \max_{x^n} P(x^n).$$

The converse part follows from Theorem 2.3, since even non-universal simulation cannot make the approximation error decay faster than $\max_{x^n} P(x^n)$, hence universal simulation cannot as well.

D Properties on Squeezing Periodic Functions

For a function $g : [a, b] \to \mathbb{R}$ and a number $\Delta > 0$, we define a periodic function $g_{\Delta}$ induced by $g$ as

$$g_{\Delta}(x) := \sum_{i=-\infty}^{\infty} g\left(a + \frac{b-a}{\Delta}(x-i\Delta)\right) 1\{x \in (i\Delta, (i+1)\Delta]\}.$$

Now we squeeze this periodic function in the $x$-axis direction by letting $\Delta \to 0$. It is easy to see that the limit $\lim_{\Delta \to 0} g\left(a + \frac{b-a}{\Delta}(x-i\Delta)\right)$ of function $g$ does not exist. However, the integral $\lim_{\Delta \to 0} \int f(x)g_{\Delta}(x)dx$ for any integrable function $f$ exists. Moreover, this limit is equal to the product of the integral of $f(x)$ and the normalized integral of $g_{\Delta}(x)$ (or $g(x)$). That is, $f(x)$ and $g_{\Delta}(x)$ are asymptotically uncorrelated as $\Delta \to 0$.

Define

$$L_\Delta := \int_{-\infty}^{\infty} f(x)g_{\Delta}(x)dx$$

and

$$L := \frac{1}{b-a} \int_{-\infty}^{\infty} f(x)dx \int_{a}^{b} g(x')dx'.$$

Lemma D.1. Assume $f(x)$ and $g(x)$ are arbitrary integrable functions, and $|g(x)|$ is bounded a.e. Then $f(x)$ and $g_{\Delta}(x)$ are asymptotically uncorrelated as $\Delta \to 0$. That is,

$$\lim_{\Delta \to 0} L_\Delta = L.$$
Remark D.1. More specifically, it holds that
\[
|L_\Delta - L| \leq \text{ess sup}_x |g(x)| \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} |f(x) - \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)dx| \, dx.
\]

Remark D.2. If \( f(x) \) is bounded and continuous a.e. on an interval \([c, d]\) (i.e., Riemann-integrable), and \( f(x) = 0, x \in [c, d]\), then the condition \( \sup_x |g(x)| \) is finite can be relaxed to that \( \int_a^b |g(x)| \, dx \) is finite. Furthermore, for this case, Lemma D.1 also holds for \( g \) that \( \int_a^b |g(x)| \, dx + \sum_{i=-\infty}^{\infty} c_i \) is finite, where \( \delta(\cdot) \) denotes the Dirac delta function.

The proof of Lemma D.1 is provided in Appendix D.1.

Now we generalize Lemma D.1 by considering \( g : [a, b] \times \mathbb{R} \to \mathbb{R} \) to be a real multivariate function. For such \( g \) and a number \( \Delta > 0 \), we define a periodic function \( g_\Delta \) induced by \( g \) as
\[
g_\Delta(x, y) := \sum_{i=-\infty}^{\infty} g\left(a + \frac{b-a}{\Delta}(x - i\Delta), y\right) 1\{x \in (i\Delta, (i+1)\Delta]\}.
\]

Define
\[
L_\Delta(y) := \int_{-\infty}^{\infty} f(x)g_\Delta(x, y) \, dx
\]
and
\[
L(y) := \frac{1}{b-a} \int_{-\infty}^{\infty} f(x)dx \int_{a}^{b} g(x', y) \, dx'.
\]

Lemma D.2. Assume \( f(x) \) and \( g(x, y) \) are arbitrary integrable functions, and \( \int_{-\infty}^{\infty} |g(x, y)| \, dy \) is bounded a.e. Then \( f(x) \) and \( g_\Delta(x, y) \) are asymptotically uncorrelated under the \( L_1 \)-norm distance as \( \Delta \to 0 \). That is,
\[
\lim_{\Delta \to 0} \int_{-\infty}^{\infty} |L_\Delta(y) - L(y)| \, dy = 0.
\]
Furthermore, (18) also holds for \( g(x, y) = g_1(x, y) + g_3(x)\delta(y - g_2(x)) \) such that \( g_2(x) \) is a differentiable a.e. function and \( \frac{d^2}{dx^2}g_2(x) \neq 0 \) for almost every \( x \in [a, b] \) and \( \sup_x \int_{-\infty}^{\infty} |g(x, y)| \, dy = \sup_x \left\{ \int_{-\infty}^{\infty} |g_1(x, y)| \, dy + |g_3(x)| \right\} \) is finite.

Remark D.3. More specifically, it holds that
\[
\int_{-\infty}^{\infty} |L_\Delta(y) - L(y)| \, dy
\]
\[
\leq \text{ess sup}_x \int_{-\infty}^{\infty} |g(x, y)| \, dy \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} |f(x) - \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)dx| \, dx. \tag{19}
\]
If \( g(x, y) = g_1(x, y) + g_3(x)\delta(y - g_2(x)) \), then (19) still holds with \( \text{ess sup}_x \int_{-\infty}^{\infty} |g(x, y)| \, dy \) replaced by \( \text{ess sup}_x |g_3(x)| \).

Remark D.4. If \( f(x) \) is bounded and continuous a.e. on an interval \([c, d]\), and \( f(x) = 0, x \in [c, d]\), then the condition \( \int_{-\infty}^{\infty} |g(x, y)| \, dy \) is bounded a.e. can be relaxed to that \( \int_{-\infty}^{\infty} \int_{a}^{b} |g(x, y)| \, dx \, dy \) is finite. Furthermore, for this case, Lemma D.2 also holds for \( g(x, y) = g_1(x, y) + g_3(x)\delta(y - g_2(x)) \) such that \( g_2(x) \) is a differentiable a.e. function and \( g_2(x) \neq 0 \) for almost every \( x \in [a, b] \) and \( \int_{-\infty}^{\infty} \int_{a}^{b} |g(x, y)| \, dx \, dy = \int_{-\infty}^{\infty} \int_{a}^{b} |g_1(x, y)| \, dx \, dy + \int_{a}^{b} |g_3(x)| \, dx \) is finite.

The proof of Lemma D.2 is similar to that of Lemma D.1, and hence omitted. Furthermore, Lemmas D.1 and D.2 can be extended to multivariate function cases.
D.1 Proof of Lemma D.1

Define

\[ L_\Delta = \int_{-\infty}^{\infty} f(x)g_\Delta(x)\,dx \]

and

\[ \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} \left( \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)\,dx \right) g \left( a + \frac{b-a}{\Delta} (x - i\Delta) \right) \,dx \]

Then we bound \( |L_\Delta - L| \) as follows.

\[
|L_\Delta - L| \\
= \left| \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} f(x)g \left( a + \frac{b-a}{\Delta} (x - i\Delta) \right) \,dx \\
- \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} \left( \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)\,dx \right) g \left( a + \frac{b-a}{\Delta} (x - i\Delta) \right) \,dx \right| \\
= \left| \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} \left( f(x) + \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)\,dx \right) g \left( a + \frac{b-a}{\Delta} (x - i\Delta) \right) \,dx \right| \\
\leq \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} \left| f(x) - \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)\,dx \right| g \left( a + \frac{b-a}{\Delta} (x - i\Delta) \right) \,dx \\
\leq \operatorname{ess sup}_x |g(x)| \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} \left| f(x) - \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)\,dx \right| \,dx \\
= \operatorname{ess sup}_x |g(x)| \sum_{i=-\infty}^{\infty} \int_{i\Delta}^{(i+1)\Delta} 2 \left( f(x) - \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)\,dx \right)^+ \,dx \\
= \operatorname{ess sup}_x |g(x)| \int_{-\infty}^{\infty} 2 |f(x) - f_\Delta(x)|^+ \,dx,
\]

where \( [z]^+ = \max\{z, 0\} \), and

\[
f_\Delta(x) := \sum_{i=-\infty}^{\infty} \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x)\,dx 1\{x \in (i\Delta, (i+1)\Delta]\}.
\]
Observe that \([f(x) - f_{\Delta}(x)]^+ \leq f(x)\) and \(f(x)\) is integrable, and moreover,
\[
\lim_{\Delta \to 0} \sum_{i=-\infty}^{\infty} \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f(x) dx 1\{x_0 \in (i\Delta, (i+1)\Delta]\}
= \lim_{\Delta \to 0} \frac{1}{\Delta} \int_{\lfloor \frac{x_0}{\Delta} \rfloor \Delta}^{\lfloor \frac{x_0}{\Delta} \rfloor + 1} f(x) dx
= f(x_0), \text{ a.e.,}
\] (21)
where (21) follows by Lebesgue’s differentiation theorem [15, Thm. 7.7]. Hence by Lebesgue’s dominated convergence theorem [15, Thm. 1.34], (20) converges to zero as \(\Delta \to 0\).

Therefore, \(\lim_{\Delta \to 0} L_{\Delta}\) exists and equals \(L\).

Acknowledgments

The author is deeply indebted to Prof. Vincent Y. F. Tan for making helpful comments which greatly improve this paper. The author also acknowledge Prof. Hideki Yagi for discussing on the extension to Rényi divergence measures. The author is supported in part by a Singapore National Research Foundation (NRF) National Cybersecurity R&D Grant (R-263-000-C74-281 and NRF2015NCR-NCR003-006), and in part by a National Natural Science Foundation of China (NSFC) under Grant (61631017).

References

[1] W. Feller. An introduction to probability theory and its applications, volume 2. John Wiley & Sons, 2008.
[2] J. Von Neumann. Various techniques used in connection with random digits. Applied Math Series, 12(36-38):1, 1951.
[3] P. Elias. The efficient construction of an unbiased random sequence. The Annals of Mathematical Statistics, pages 865–870, 1972.
[4] M. Blum. Independent unbiased coin flips from a correlated biased source – a finite state markov chain. Combinatorica, 6(2):97–108, 1986.
[5] D. E. Knuth and A. C. Yao. The complexity of nonuniform random number generation. Algorithm and Complexity, New Directions and Results, pages 357–428, 1976.
[6] J. R. Roche. Efficient generation of random variables from biased coins. In Information Theory, 1991 (papers in summary form only received), Proceedings. 1991 IEEE International Symposium on (Cat. No. 91CH3003-1), pages 169–169. IEEE, 1991.
[7] J. Abrahams. Generation of discrete distributions from biased coins. In International Symposium on Information Theory and Its Applications, page 1181. Institution of Engineers, Australia, 1994.
[8] T. S. Han and M. Hoshi. Interval algorithm for random number generation. IEEE Trans. Inf. Theory, 43(2):599–611, 1997.
[9] T. Han and S. Verdú. Approximation theory of output statistics. IEEE Trans. Inf. Theory, 39(3):752–772, 1993.
[10] S. Vembu and S. Verdú. Generating random bits from an arbitrary source: Fundamental limits. IEEE Trans. Inf. Theory, 41(5):1322–1332, 1995.
[11] T. S. Han. Information-spectrum methods in information theory. Springer, 2003.
[12] W. Kumagai and M. Hayashi. Second-order asymptotics of conversions of distributions and entangled states based on rayleigh-normal probability distributions. *IEEE Trans. Inf. Theory*, 63(3):1829–1857, 2017.

[13] L. Yu and V. Y. F. Tan. Simulation of random variables under Rényi divergence measures of all orders. *arXiv preprint arXiv:1805.12451*, 2018.

[14] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.

[15] W. Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 2006.

[16] L. Owens. Exploring the rate of convergence of approximations to the riemann integral. *https://www.whitman.edu/Documents/Academics/Mathematics/2014/owensla.pdf*, 2014.

[17] I. Csiszar and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[18] S. Kamath and S. Verdú. Estimation of entropy rate and rényi entropy rate for markov chains. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 685–689. IEEE, 2016.

[19] V. I. Bogachev. *Measure theory*, volume 2. Springer, 2007.

[20] Á. Martín, N. Merhav, G. Seroussi, and M. J. Weinberger. Twice-universal simulation of markov sources and individual sequences. *IEEE Trans. Inf. Theory*, 56(9):4245–4255, 2010.