Uniform Random Number Generation from Markov Chains: Non-Asymptotic and Asymptotic Analyses

Masahito Hayashi *Senior Member, IEEE* and Shun Watanabe *Member, IEEE*

arXiv:1503.04371v2 [cs.IT] 2 Feb 2016

Abstract—In this paper, we derive non-asymptotic achievability and converse bounds on the random number generation with/without side-information. Our bounds are efficiently computable in the sense that the computational complexity does not depend on the block length. We also characterize the asymptotic behaviors of the large deviation regime and the moderate deviation regime by using our bounds, which implies that our bounds are asymptotically tight in those regimes. We also show the second order rates of those problems, and derive single letter forms of the variances characterizing the second order rates. Further, we address the relative entropy rate and the modified mutual information rate for these problems.

Index Terms—Markov Chain, Non-Asymptotic Analysis, Random Number Generation,

I. INTRODUCTION

A. Uniform random number generation (URNG)

Uniform random number generation is one of important tasks for information theory as well as secure communication. When a non-uniform random number is generated subject to independent and identical distribution and the source distribution is known to P_X, we can convert it to the uniform random number, whose optimal conversion rate is known to be the entropy $H(P_X)$ [2]. Vembu and Verdú [3] extended this problem to the general information source. Applying their result to the Markovian source, we find that the optimal conversion rate is the entropy rate.

On the other hand, many researchers in information theory are attracted by non-asymptotic analysis recently [4], [5], [6]. Since all of realistic situations are non-asymptotic, it is strongly desired to evaluate the performance of a protocol in the non-asymptotic setting. In the case of uniform random number generation, we need to consider two issues:

A1) How to *quantitatively* guarantee the security for finite block length n. As the criterion, we employ the variational distance criterion because it is universal composable[7].

A2) How to implement the extracting method efficiently.

Fortunately, the latter problem has been solved by employing universal$_2$ hash functions, which can be constructed by combination of Toeplitz matrix and the identity matrix [8]. This construction has small amount of complexity and was implemented in a real demonstration [9], [10]. Recently, the paper [11] proposed a new class of hash functions, ε-almost dual universal hash functions, and the paper [10] proposed more efficient hash functions belonging to this new class. Hence, it is needed to solve the first problem.

So far, with a huge size n, quantitative evaluation of the security has been done only for the i.i.d. source [8], [12]. However, the source is not necessarily i.i.d. in the real world, and it is necessary to develop a technique to evaluate the security for non i.i.d. source. As a first step of this direction of research, we consider the Markov source in this paper. In the following, we explain difficulties to extend the existing results for the i.i.d. source to the Markov source.

Although it is not stated explicitly in any literatures, we believe that there are two important criteria for non-asymptotic bounds:

B1) Computational complexity, and
B2) Asymptotic optimality.

Let us first consider the first criterion, i.e., the computational complexity. For example, Han [13] introduced lower and upper bounds for the variational distance criterion by using the inf-spectral entropy, which are called the inf-spectral entropy bounds. For i.i.d. sources, these bounds can be computed by numerical calculation packages. However, there is no known method to efficiently compute these bounds for Markov sources. Consequently, there is no bound that is efficiently computable for the Markov chain so far. The first purpose of this paper is to derive non-asymptotic bounds that are efficiently computable.

Next, let us consider the second criterion, i.e., asymptotic optimality. So far, three kinds of asymptotic regimes have been studied in the information theory:

B2-1) The large deviation regime in which the error probability ε asymptotically behaves like e^{-nr} for some $r > 0$ [14],

B2-2) The moderate deviation regime in which ε asymptotically behaves like $e^{-n^{1-2t}r}$ for some $r > 0$ and $t \in (0, 1/2)$ [15], [16], [17], and

B2-3) The second order regime in which ε is a constant [18], [4], [5], [6], [15], [16], [19].

We shall claim that a good non-asymptotic bound should be asymptotically optimal in at least one of the above mentioned three regimes.

Further, when the generation rate is too large, the variational distance is close to 1. In this case, we cannot measure how far from the uniform random number the generated random number is. Hence, we employ the relative entropy rate (RER).

B. Secure uniform random number generation (SURNG)

When the initial random number X is partially leaked to the third party Y, to guarantee the security, we need to convert the random number to the uniform random number that has almost no correlation with the third party. When a non-uniform random number is generated subject to independent and identical distribution of the joint distribution is known to $P_{X,Y}$, we can convert it to the uniform random number, whose optimal conversion rate is known to be the conditional entropy $H(X|Y)$ [20], [21].

Bennett et al. [22], [23] and Håstad et al. [24] proposed to use universal$_2$ hash functions for this purpose, and derived two universal hashing lemma, which provides an upper bound for leaked information based on Rényi entropy of order 2. The paper [11] proposed to use ε-almost dual universal hash functions [11] that includes the hash functions by [10]. Hence, the problem A2) has been solved by employing universal$_2$ hash functions.

Therefore, the remaining problem is the problem A1), i.e., to quantitatively guarantee the security for finite block length n under these hash functions. For the security criterion, we employ the variational distance between the true distribution and the ideal distribution because it satisfies the universal composable property [7]. To achieve the rate $H(X|Y)$ via two universal hashing lemma, Renner [25] attached the smoothing to min entropy[1], which is a lower bound on the above conditional Rényi entropy of order 2[2]. That is, he proposed to maximize the min-entropy among the sub-distributions whose variational distance to the true distribution is less than a given threshold. Using Renner's method, the paper [12] derived a lower bound of the exponential decreasing rate. Tomamichel and Hayashi [26] derived an upper bound of the universal composable quantity of extracted key with a finite block-length n by combining the Renner's method and the method of information spectrum by Han. Further, Watanabe and Hayashi [27] compared two approaches: the combination of the Renner's method and the method of information spectrum[3], and the exponential bounding approach of [12]. Further, the paper [28] showed that similar evaluations are possible even for ε-almost dual universal hash functions [11].

For convenience, let us call the bound derived by the former approach the *inf-spectral entropy bound*, and the bound derived by the latter approach the *exponential bound*. It turned out that the exponential bound is tighter than the inf-spectral entropy bound when the required security level ε is rather

---

[1]Bennett et al. [23] also employed a similar idea without use of the terminology of smoothing, and derived the conversion rate $H(X|Y)$.

[2]In [25], Renner also showed a quantum extension of the two universal hashing lemma.

[3]The approach to derive a bound in [27] is almost the same as that in [26], but it should be noted that the security criterion in [27] is based on the variational distance while that in [26] is based on the purified distance.

small. A bound that interpolate both approaches was also derived in [27], which we called the *hybrid bound*.

Similar to uniform random number generation, for i.i.d. sources, the inf-spectral entropy bound and the hybrid bound can be computed by numerical calculation packages. However, there is no known method to efficiently compute these bounds for Markov sources. The computational complexity of the exponential bound is $O(1)$ since the exponential bound is described by using the Gallager function, which is an additive quantity. However, this is not the case for Markov sources. Consequently, there is no bound that is efficiently computable for the Markov chain so far. Further, the first order results for Markov sources have not been revealed as long as the authors know, and they are clarified in this paper.

Further, when the generation key rate is too large, the variational distance is close to 1. In this case, we cannot measure how far from the secure uniform random number the generated random number is. Hence, we employ the relative entropy between the generated random number and the ideal random number, which was introduced by Csiszár-Narayan [29] and is called the modified mutual information rate. Indeed, when we surpass axiomatic conditions, the leaked information measure must be this quantity [28].

C. Main Contribution for Non-Asymptotic Analysis

Although there are several studies for finite-length analysis for URNG and SURNG, they did not discuss the Markovian chain. Indeed, while they derived several single-shot bounds, these bounds cannot be directly applied to the Markovian chain, because the bounds obtained by such applications are not computable at least in the the Markovian chain. Hence, we need to derive new finite-length bounds for the Markovian chain by modifying existing single-shot bounds. For this purpose, we adopt the structure similar to the paper [30], which addresses the source coding with Markov chain because this paper employs the common structure between the uniform random number generation and the source coding. Hence, the obtained results are also quite similar to those of the paper [30]. To derive non-asymptotic achievability bounds on the problems, we basically use the exponential type bounds for the single shot setting. When there is no information leakage, those exponential type bounds are described by the Rényi entropy. Thus, we need to evaluate Rényi entropy for the Markov chain. For this purpose, we introduce Rényi entropy for transition matrices, which is defined irrespective of initial distributions (cf. (27)). Then, we evaluate the Rényi entropy for the Markov chain in terms of the Rényi entropy for the transition matrix. From this evaluation, we can also find that the Rényi entropy rate for the Markov chain coincides with the Rényi entropy for the transition matrix. Note that the former is defined as the limit and the latter is single letter characterized.

When a part of information is leaked to the third party, to generate secure uniform random number, we consider two assumptions on transition matrices (see Assumption 1 and Assumption 2 of Section II). Although a computable form of the conditional entropy rate is not known in general, Assumption 1, which is less restrictive than Assumption 2,

enables us to derive a computable form of the conditional entropy rate.

In the problems with side-information, exponential type bounds are described by conditional Rényi entropies. There are several definitions of conditional Rényi entropies (see [31], [32] for extensive review), and we use the one defined in [8] and the one defined by Arimoto [33]. We shall call the former one the *lower conditional Rényi entropy* (cf. (3)) and the latter one the *upper conditional Rényi entropy* (cf. (8)). To derive non-asymptotic bounds, we need to evaluate these information measures for the Markov chain. For this purpose, under Assumption 1, we introduce the lower conditional Rényi entropy for transition matrices (cf. (27)). Then, we evaluate the lower conditional Rényi entropy for the Markov chain in terms of its transition matrix counterpart. This evaluation gives non-asymptotic bounds for secure uniform random number generation under Assumption 1. Under more restrictive assumption, i.e., Assumption 2, we also introduce the upper conditional Rényi entropy for a transition matrix (cf. (34)). Then, we evaluate the upper Rényi entropy for the Markov chain in terms of its transition matrix counterpart. This evaluation gives non-asymptotic bounds that are tighter than those obtained under Assumption 1.

We also derive converse bounds for every problem by using the change of measure argument developed by the authors in the accompanying paper on information geometry [34], [35]. When there is no information leakage, the converse bounds are described by the Rényi entropy for transition matrices. When a part of information is leaked to the third party, we further introduce two-parameter conditional Rényi entropy and its transition matrix counterpart (cf. (14) and (38)). This novel information measure includes the lower conditional Rényi entropy and the upper conditional Rényi entropy as special cases.

In the problem of SURNG, instead of the RER, we employ the modified mutual information rate (MMIR), which was introduced by Csiszár and Narayan [29] and whose axiomatic characterization was obtained in the paper [28]. When the uniformity is guaranteed, this quantity is given by the equivocation rate introduced by Wyner [36]. When there is no information leakage, our lower and upper bounds are given by using the Rényi entropy for the Markov chain in terms of its transition matrix counterpart. When there exists information leakage, our lower and upper bounds are given by using the lower conditional Rényi entropy for the Markov chain in terms of its transition matrix counterpart under Assumption 1.

Here, we would like to remark on terminologies. There are a few ways to express exponential type bounds. In statistics or the large deviation theory, we usually use the cumulant generating function (CGF) to describe exponents. In information theory, we use the Gallager function or the Rényi entropies. Although these three terminologies are essentially the same and are related by change of variables, the CGF and the Gallager function are convenient for some calculations since they have good properties such as convexity. However, they are merely mathematical functions. On the other hand, the Rényi entropies are information measures including Shannon's information measures as special cases. Thus, the Rényi entropies are intuitively familiar in the field of information theory. The Rényi entropies also have an advantage that two types of bounds (eg. (215) and (218)) can be expressed in a unified manner. For these reasons, we state our main results in terms of the Rényi entropies while we use the CGF and the Gallager function in the proofs. For readers' convenience, the relation between the Rényi entropies and corresponding CGFs are summarized in Appendix A.

Overall, we summarize the contributions for non-asymptotic analysis in comparison to existing results as follows.

(1) *Finite-length bound:* For URNG and SURNG, we derive finite-length bounds satisfying the conditions B1) and B2) for Markovian chain. Theorems in Subsections III-C and IV-C are classified to this type of results. All existing finite-length bounds with computable form are obtained with i.i.d. setting. Indeed, several single-shot bounds were obtained in a more general form. However, their computabilities have not been discussed in the Markovian case. At least, many of them, (e.g, Lemmas 16, 17, 18, 22, 23, 25, and 28) are not given in a computable form in the Markovian case.

(2) *Single-shot bound:* In this paper, we employ several existing single-shot bounds. However, many of them cannot be given in a useful form. These bounds cannot be easily calculated at least in the Markovian case. To apply them to the Markovian case, we loosen these bounds. Lemmas 21, 24, 29 and 32 fall in this case. Since these bounds have a much simpler form than existing bounds, they might be applied to other cases. This discussion for the simplification is quite different from the case of source coding [30]. That is, this part has the most serious technical hardness compared to the paper [30] because the discussion in this paper is specialized to random number generation.

D. Main Contribution for Asymptotic Analysis

Among authors' knowledge, there is no existing study for the asymptotic analysis with the Markovian chain with respect to URNG and SURNG except for the following. When the general sequence of single information sources, the asymptotic rate of URNG is characterized by Vembu and Verdú [3] and Han [13]. Since the asymptotic entropy rate of Markovian chain is known, we can calculate the asymptotic rate of URNG for the Markovian chain. However, further study with respect to URNG and SURNG has not been discussed for the Markovian chain nor the general sequence of information sources.

We can easily see that these non-asymptotic bounds yields the asymptotic optimal random number generation rate while the case with information leakage requires Assumption 1. For asymptotic analyses of the large deviation and the moderate deviation regimes, we derive the characterizations[4] by using our non-asymptotic achievability and converse bounds, which

---

[4]For the large deviation regime, we only derive the characterizations up to the critical rates.

TABLE I
SUMMARY OF ASYMPTOTIC RESULTS AND NON-ASYMPTOTIC BOUNDS TO DERIVE ASYMPTOTIC RESULTS

Problem	First Order	Large Deviation	Moderate Deviation	Second Order	RER/MMIR
URNG	Solved	Solved* (U2), $O(1)$	Solved, $O(1)$	Solved, Tail	Solved, $O(1)$
SURNG	Solved (Ass. 1)	Solved* (Ass. 2, U2), $O(1)$	Solved (Ass. 1), $O(1)$	Solved (Ass. 1), Tail	Solved (Ass. 1), $O(1)$

URNG is the uniform random number generation without information leakage. SURNG is the secure uniform random number generation when a part of information is leaked to the third party.

implies that our non-asymptotic bounds are tight in the large deviation regime and the moderate deviation regime.

We also derive the second order rate. It is also clarified that the reciprocal coefficient of the moderate deviation regime and the variance of the second order regime coincide. Furthermore, a single letter form of the variance is clarified[5].

The asymptotic results and the non-asymptotic results are summarized in Table I. As a part of the non-asymptotic results, the table focuses on the computational complexities of the non-asymptotic bounds. "Solved*" indicates that those problems are solved up to the critical rates. "Ass. 1" and "Ass. 2" indicate that those problems are solved under Assumption 1 or Assumption 2. "U2" indicates that the converse results are obtained only for the worst case of the universal two hash family (see (105) and (178)). "$O(1)$" indicates that both the achievability part and the converse part of those asymptotic results are derived from our non-asymptotic achievability bounds and converse bounds whose computational complexities are $O(1)$. "Tail" indicates that both the achievability part and the converse part of those asymptotic results are derived from the information-spectrum type achievability bounds and converse bounds whose computational complexities depend on the computational complexities of tail probabilities.

Exact computations of tail probabilities are difficult in general though it may be feasible for a simple case such as an i.i.d. case. One way to approximately compute tail probabilities is to use the Berry-Esséen theorem [39, Theorem 16.5.1] or its variant [40]. This direction of research is still continuing [41], [42], and an evaluation of the constant was done in [42] though it is not clear how much tight it is. If we can derive a tight Berry-Esséen type bound for the Markov chain, we can derive a non-asymptotic bound that is asymptotically tight in the second order regime. However, the approximation errors of Berry-Esséen type bounds converge only in the order of $1/\sqrt{n}$, and cannot be applied when ε is rather small. Even in the cases such that exact computations of tail probabilities are possible, the information-spectrum type bounds are looser than the exponential type bounds when ε is rather small, and we need to use appropriate bounds depending on the size of ε. In fact, this observation was explicitly clarified in [27] for the random number generation with side-information. Consequently, we believe that our exponential type non-asymptotic bounds are very useful.

Further, we derive the asymptotic leaked information rate. When there is no information leakage, we discuss the RER,

which is asymptotically given by the entropy rate. When there exists information leakage, we discuss the MMIR, which is asymptotically given by the conditional entropy rate under Assumption 1.

Overall, we summarize the contributions for asymptotic analysis in comparison to existing results as follows.

(1) *New bounds for Markovian case:* For URNG and SURNG, we derive the optimal asymptotic performances in Subsections III-D, III-E, III-F, 19, III-G, IV-D, IV-E, IV-F, and IV-G under the four regimes, the large deviation regimes, the moderate deviation regimes, the second order regimes, and the asymptotic relative entropy rate regime (the asymptotic modified mutual information rate regime) for Markovian chain (with suitable conditions for SURNG). Except for the information spectrum approach, all existing asymptotic analyses with these three regimes assume the i.i.d. source. Further, analyses with the information spectrum approach derived only the general formulas, which did not derive any computable asymptotic bounds for these three regimes for the Markovian chain.

(2) *New bound even for i.i.d. case:* Among the above asymptotic results, Theorem 30 is novel even for the i.i.d. case. This theorem gives the converse bound for large deviation for SURNG.

E. Two criteria

In this paper, to consider a practical issue, we employ two criteria. In the channel coding, such a practical issue is discussed as a coding theory in a form separate from the fundamental issue. However, in the random number generation case, we can discuss the performance of hash functions with a small construction complexity in the same way as the fundamental issue. Such a practical issue is also the target of this paper. Usually, when we discuss a fundamental aspect of the topic of information theory, we focus only on the minimum leaked information among all of hash function, which is denoted by $\Delta(M)$ in this paper, whose precise definition will be given in Subsections III-A and IV-A. However, when we take account into the complexity of construction of protocol, we need to restrict hash functions into hash functions with a small construction complexity. Hence, it is desired to minimize the leaked information among a class of hash functions with small calculation complexity for its construction. In this paper we focus on the family of two-universal hash functions, named by the two-universal hash family $\mathcal{F}$ because this family contains a hash function with a small construction complexity. However, this paper focuses on the worst leaked information

---

[5]An alternative way to derive a single letter characterization of the variance for the Markov chain was shown in [37, Lemma 20]. It should be also noted that a single letter characterization can be derived by using the fundamental matrix [38].

$\overline{\Delta}(M)$ among the two-universal hash family $\mathcal{F}$, which is more important from a practical view point than the best case due to the following two reasons.

(1) Usually, the optimal hash function depends on the source distribution. However, it is not easy to perfectly identify the source distribution. In such a case, instead of the optimal hash function, we need to choose a hash function that universally works well. If we apply a two-universal hash function, its leaked information is always better than the worst leaked information $\overline{\Delta}(M)$. Hence, if the quantity $\overline{\Delta}(M)$ is sufficiently close to the optimal case $\Delta(M)$, we can say that any two-universal hash function universally works well.

(2) Although the two-universal hash family $\mathcal{F}$ contains a hash function with a small calculation complexity for its construction, any two-universal hash function does not necessarily have a small calculation complexity. If the quantity $\overline{\Delta}(M)$ is sufficiently close to the optimal case $\Delta(M)$, we can take the priority to minimize the construction complexity among the two-universal hash family $\mathcal{F}$ over the optimization of the leaked information.

In this paper, we show that the worst leaked information $\overline{\Delta}(M)$ is close to the minimum leaked information $\Delta(M)$ in the moderate deviation and the second order. These results guarantee that any two-universal hash function has a sufficiently good performance. That is, they allow us to employ any two-universal hash function to achieve these asymptotic optimal performances. These results amplify our choice of hash function to achieve the asymptotically optimality.

F. Organization of Paper and Notations

As preparation, we explain information measures for single-shot setting in Subsection II-A. Then, we address conditional Rényi entropies for transition matrix in Subsection II-B, and discuss the relation between these information measures and Markov chain in Subsection II-C. These information measures and their properties will be used in the latter sections. These contents were obtained in the paper [30], and their proofs are available in the paper [30]. However, the paper [30] did not address the conditional min entropy, which corresponds to the order parameter ∞. So, in Subsections II-D and II-E, we discuss the relation between the limit of the conditional Rényi entropy and the conditional min entropy, which are new results and are shown in Appendix.

Section III addresses the uniform random number generation without information leakage. The obtained upper and lower bounds are numerically calculated in a typical example in this section. Then, Section IV proceeds to addresses the secure uniform random number generation with partial information leakage. As we mentioned above, we state our main result in terms of the Rényi entropies, and we use the CGFs and the Gallager function in the proofs. In Appendix A, the relation between the Rényi entropies and corresponding CGFs are summarized. The relation between the Rényi entropies and the Gallager function are explained as necessary. Proofs of some technical results are also shown in the rest of appendices.

A random variable is denoted by upper case letter, and its realization is denoted by lower case letter. The notation $\mathcal{P}(\mathcal{X})$ is the set of all distribution on alphabet $\mathcal{X}$. The notation $\bar{\mathcal{P}}(\mathcal{X})$ is the set of all non-negative sub-normalized functions on $\mathcal{X}$. $|\mathcal{X}|$ represent the cardinality of the set $\mathcal{X}$. The cumulative distribution function of the standard Gaussian random variable is denoted by

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] dx. \tag{1}$$

Throughout the paper, the base of the logarithm is e.

II. INFORMATION MEASURES

In this section, we introduce information measures that will be used in Section III and Section IV. All of lemmas and theorems in this section except for Lemmas 15 and 12 and Theorem 6 were shown in [30].

A. Information Measures for Single-Shot Setting

1) Conditional Rényi entropy relative to a general distribution: In this section, we introduce conditional Rényi entropies for the single-shot setting. For more detailed review of conditional Rényi entropies, see [32]. For a correlated random variable (X, Y) on $\mathcal{X} \times \mathcal{Y}$ with probability distribution P_{XY} and a marginal distribution Q_Y on $\mathcal{Y}$, we introduce the conditional Rényi entropy of order $1 + \theta$ relative to Q_Y as

$$H_{1+\theta}(P_{XY}|Q_Y) := -\frac{1}{\theta} \log \sum_{x,y} P_{XY}(x,y)^{1+\theta} Q_Y(y)^{-\theta}, \tag{2}$$

where $\theta \in (-1,0) \cup (0,\infty)$. The conditional Rényi entropy of order 0 relative to Q_Y is defined by the limit with respect to θ. When $\mathcal{Y}$ is singleton, it is nothing but the ordinary Rényi entropy, and it is denoted by $H_{1+\theta}(X) = H_{1+\theta}(P_X)$ throughout the paper.

2) Lower conditional Rényi entropy: One of important special cases of $H_{1+\theta}(P_{XY}|Q_Y)$ is the case with $Q_Y = P_Y$. We shall call this special case the *lower conditional Rényi entropy* of order $1 + \theta$ and denote[6]

$$H_{1+\theta}^{\downarrow}(X|Y) := H_{1+\theta}(P_{XY}|P_Y) \tag{3}$$
$$= -\frac{1}{\theta} \log \sum_{x,y} P_{XY}(x,y)^{1+\theta} P_Y(y)^{-\theta}. \tag{4}$$

The following property holds.

Lemma 1 We have

$$\lim_{\theta \to 0} H_{1+\theta}^{\downarrow}(X|Y) = H(X|Y) \tag{5}$$

and

$$\mathsf{V}(X|Y) := \mathrm{Var}\left[\log \frac{1}{P_{X|Y}(X|Y)}\right] \tag{6}$$
$$= \lim_{\theta \to 0} \frac{2\left[H(X|Y) - H_{1+\theta}^{\downarrow}(X|Y)\right]}{\theta}. \tag{7}$$

---

[6] This notation was first introduce in [43].

3) Upper conditional Rényi entropy: The other important special cases of $H_{1+\theta}(P_{XY}|Q_Y)$ is the measure maximized over Q_Y. We shall call this special case the *upper conditional Rényi entropy* of order $1 + \theta$ and denote[7]

$$
\begin{aligned}
&H_{1+\theta}^{\uparrow}(X|Y) \\
&:= \max_{Q_Y \in \mathcal{P}(\mathcal{Y})} H_{1+\theta}(P_{XY}|Q_Y) && (8)\\
&= H_{1+\theta}(P_{XY}|P_Y^{(1+\theta)}) && (9)\\
&= -\frac{1+\theta}{\theta} \log \sum_y P_Y(y) \left[\sum_x P_{X|Y}(x|y)^{1+\theta} \right]^{\frac{1}{1+\theta}} && (10)
\end{aligned}
$$

where the expression (10) is the same as Arimoto's proposal for the conditional Rényi entropy [33] and

$$
P_Y^{(1+\theta)}(y) := \frac{\left[\sum_x P_{XY}(x,y)^{1+\theta} \right]^{\frac{1}{1+\theta}}}{\sum_{y'} \left[\sum_x P_{XY}(x,y')^{1+\theta} \right]^{\frac{1}{1+\theta}}}. \tag{11}
$$

For this measure, we also have properties similar to Lemma 1.

Lemma 2 ([30], [45], [44]) We have

$$
\lim_{\theta \to 0} H_{1+\theta}^{\uparrow}(X|Y) = H(X|Y) \tag{12}
$$

and

$$
\lim_{\theta \to 0} \frac{2\left[H(X|Y) - H_{1+\theta}^{\uparrow}(X|Y) \right]}{\theta} = \mathsf{V}(X|Y). \tag{13}
$$

4) Properties of conditional Rényi entropies: When we derive converse bounds, we need to consider the case such that the order of the Rényi entropy and the order of conditioning distribution defined in (11) are different. For this purpose, we introduce two-parameter conditional Rényi entropy:

$$
\begin{aligned}
&H_{1+\theta, 1+\theta'}(X|Y) && (14)\\
&:= H_{1+\theta}(P_{XY}|P_Y^{(1+\theta')}) && (15)\\
&= -\frac{1}{\theta} \log \sum_y P_Y(y) \left[\sum_x P_{X|Y}(x|y)^{1+\theta} \right] \\
&\quad \cdot \left[\sum_x P_{X|Y}(x|y)^{1+\theta'} \right]^{\frac{\theta}{1+\theta'}} + \frac{\theta'}{1+\theta'} H_{1+\theta'}^{\uparrow}(X|Y).
\end{aligned}
$$

The measures defined above has the following properties:

Lemma 3 ([30], [45], [44])

1) For fixed Q_Y, $\theta H_{1+\theta}(P_{XY}|Q_Y)$ is a concave function of θ, and it is strict concave iff. $\mathrm{Var}\left[\log \frac{Q_Y(Y)}{P_{XY}(X,Y)} \right] > 0$.

2) For fixed Q_Y, $H_{1+\theta}(P_{XY}|Q_Y)$ is a monotonically decreasing[8] function of θ.

3) The function $\theta H_{1+\theta}^{\downarrow}(X|Y)$ is a concave function of θ, and it is strict concave iff. $\mathsf{V}(X|Y) > 0$.

4) $H_{1+\theta}^{\downarrow}(X|Y)$ is a monotonically decreasing function of θ, and it is strictly monotonically decreasing iff. $\mathsf{V}(X|Y) > 0$.

5) The function $\theta H_{1+\theta}^{\uparrow}(X|Y)$ is a concave function of θ, and it is strict concave iff. $\mathsf{V}(X|Y) > 0$.

6) $H_{1+\theta}^{\uparrow}(X|Y)$ is a monotonically decreasing function of θ, and it is strictly monotonically decreasing iff. $\mathsf{V}(X|Y) > 0$.

7) For every $\theta \in (-1, 0) \cup (0, \infty)$, we have $H_{1+\theta}^{\downarrow}(X|Y) \le H_{1+\theta}^{\uparrow}(X|Y)$.

8) For fixed θ', the function $\theta H_{1+\theta, 1+\theta'}(X|Y)$ is a concave function of θ, and it is strict concave iff. $\mathsf{V}(X|Y) > 0$.

9) For fixed θ', $H_{1+\theta, 1+\theta'}(X|Y)$ is a monotonically decreasing function of θ.

10) We have

$$
H_{1+\theta, 1}(X|Y) = H_{1+\theta}^{\downarrow}(X|Y). \tag{16}
$$

11) We have

$$
H_{1+\theta, 1+\theta}(X|Y) = H_{1+\theta}^{\uparrow}(X|Y). \tag{17}
$$

12) For every $\theta \in (-1, 0) \cup (0, \infty)$, $H_{1+\theta, 1+\theta'}(X|Y)$ is maximized at $\theta' = \theta$.

5) Functions related to lower conditional Rényi entropy: Since Item 5) of Lemma 3 guarantees that the function $\theta \mapsto \frac{d[\theta H_{1+\theta}^{\downarrow}(X|Y)]}{d\theta}$ is strictly monotone decreasing, we can define the inverse functions[9] $\theta(a) = \theta^{\downarrow}(a)$ and $a(R) = a^{\downarrow}(R)$ by

$$
\left. \frac{d[\theta H_{1+\theta}^{\downarrow}(X|Y)]}{d\theta} \right|_{\theta = \theta(a)} = a \tag{18}
$$

and

$$
(1 + \theta(a(R)))a(R) - \theta(a(R))H_{1+\theta(a(R))}^{\downarrow}(X|Y) = R, \tag{19}
$$

for $R(\underline{a}) < R \le H_0^{\downarrow}(X|Y)$, where $\underline{a} = \underline{a}^{\downarrow} := \lim_{\theta \to \infty} \frac{d[\theta H_{1+\theta}^{\downarrow}(X|Y)]}{d\theta}$.

6) Functions related to upper conditional Rényi entropy: For $\theta H_{1+\theta}^{\uparrow}(X|Y)$, we also introduce the inverse functions $\theta(a) = \theta^{\uparrow}(a)$ and $a(R) = a^{\uparrow}(R)$ by

$$
\left. \frac{d\theta H_{1+\theta}^{\uparrow}(X|Y)}{d\theta} \right|_{\theta = \theta(a)} = a \tag{20}
$$

and

$$
(1 + \theta(a(R)))a(R) - \theta(a(R))H_{1+\theta(a(R))}^{\uparrow}(X|Y) = R, \tag{21}
$$

for $R(\underline{a}) < R \le H_0^{\uparrow}(X|Y)$, where $\underline{a} = \underline{a}^{\uparrow} := \lim_{\theta \to \infty} \frac{d[\theta H_{1+\theta}^{\uparrow}(X|Y)]}{d\theta}$.

---

[7]For $-1 < \theta < 0$, (9) can be proved by using the Hölder inequality, and, for $0 < \theta$, (9) can be proved by using the reverse Hölder inequality [44, Lemma 8].

[8]Technically, $H_{1+\theta}(P_{XY}|Q_Y)$ is always non-increasing and it is monotonically decreasing iff. strict concavity holds in Statement 1. Similar remarks are also applied for other information measures throughout the paper.

[9]Throughout the paper, the notations $\theta(a)$ and $a(R)$ are reused for several inverse functions. Although the meanings of those notations are obvious from the context, we occasionally put superscript $\downarrow$ or $\uparrow$ to emphasize that those inverse functions are induced from corresponding conditional Rényi entropies. This definition is related to Legendre transform of the concave function $\theta \mapsto \theta H_{1+\theta}^{\downarrow}(X|Y)$. For its detail, see [30].

B. Information Measures for Transition Matrix

1) Conditions for transition matrices: Let $\{W(x,y|x',y')\}_{((x,y),(x',y'))\in(\mathcal{X}\times\mathcal{Y})^2}$ be an ergodic and irreducible transition matrix. The purpose of this section is to introduce transition matrix counterparts of those measures in Section II-A. For this purpose, we first need to introduce some assumptions on transition matrices:

Assumption 1 (Non-Hidden [30], [34], [35]) We say that a transition matrix W is *non-hidden* (with respect to $\mathcal{Y}$) if

$$\sum_x W(x,y|x',y') = W_Y(y|y') \tag{22}$$

for every $x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}^{10}$.

Assumption 2 (Strongly Non-Hidden) We say that a transition matrix W is *strongly non-hidden* (with respect to $\mathcal{Y}$) if, for every $\theta \in (-1,\infty)$ and $y, y' \in \mathcal{Y}$,

$$W_{Y,\theta}(y|y') := \sum_x W(x,y|x',y')^{1+\theta} \tag{23}$$

is well defined, i.e., the right hand side of (23) is independent of x'.

Assumption 1 requires (23) to hold only for $\theta = 0$, and thus Assumption 2 implies Assumption 1. However, Assumption 2 is strictly stronger condition than Assumption 1. For example, let consider the case such that the transition matrix is a product form, i.e., $W(x,y|x',y') = W_X(x|x')W_Y(y|y')$. In this case, Assumption 1 is obviously satisfied. However, Assumption 2 is not satisfied in general.

Assumption 1 means that we can decompose $W(x,y|x',y')$ as

$$W(x,y|x',y') = W_Y(y|y')W_{X|X',Y',Y}(x|x',y',y). \tag{24}$$

Thus, Assumption 2 can be rephrased as

$$\sum_x W_{X|X',Y',Y}(x|x',y',y)^{1+\theta} \tag{25}$$

does not depend on x'. By taking θ sufficiently large, we find that the largest value of $W_{X|X',Y',Y}(x|x',y',y)$ does not depend on x'. By repeating this argument for the second largest value of $W_{X|X',Y',Y}(x|x',y',y)$ and so on, we eventually find that Assumption 2 is satisfied iff., for every $x' \neq \tilde{x}'$, there exists a permutation π on $\mathcal{X}$ such that $W_{X|X',Y',Y}(x|x',y',y) = W_{X|X',Y',Y}(\pi(x)|\tilde{x}',y',y)$.

Non-trivial examples satisfying Assumption 1 and Assumption 2 are given in [30].

---

[10] The reason of the name "non-hidden" is the following. In general, the random variable Y is subject to a hidden Markov process. However, when the condition (22) holds, the random variable Y is subject to a Markov process. Hence, we call the condition (22) non-hidden.

2) Lower conditional Rényi entropy $H_{1+\theta}^{\downarrow,W}(X|Y)$: First, we introduce information measures under Assumption 1. In order to define a transition matrix counterpart of (3), let us introduce the following tilted matrix:

$$\tilde{W}_\theta(x,y|x',y') := W(x,y|x',y')^{1+\theta}W_Y(y|y')^{-\theta}. \tag{26}$$

Here, we should notice that the tilted matrix $\tilde{W}_\theta$ is not normalized, i.e., is not a transition matrix. Let λ_θ be the Perron-Frobenius eigenvalue and $\tilde{P}_{\theta,XY}$ be its normalized eigenvector. Then, we define the lower conditional Rényi entropy for W by

$$H_{1+\theta}^{\downarrow,W}(X|Y) := -\frac{1}{\theta}\log\lambda_\theta, \tag{27}$$

where $\theta \in (-1,0) \cup (0,\infty)$. For $\theta = 0$, we define the lower conditional Rényi entropy for W by

$$H_1^{\downarrow,W}(X|Y) := \lim_{\theta\to 0} H_{1+\theta}^{\downarrow,W}(X|Y). \tag{28}$$

When we define the conditional entropy $H^W(X|Y)$ for W by using the stationary distribution $P_{0,XY}$ as

$$H^W(X|Y)$$
$$:= -\sum_{x',y'} P_{0,XY}(x',y')\sum_{x,y} W(x,y|x',y')\log\frac{W(x,y|x',y')}{W_Y(y|y')},$$

as shown below, we have

$$H^W(X|Y) = H_1^{\downarrow,W}(X|Y). \tag{29}$$

Taking the derivative with respect to θ, we can show (29) as follows

$$H_1^{\downarrow,W}(X|Y) = \frac{d\theta H_\theta^{\downarrow,W}(X|Y)}{d\theta}\Big|_{\theta=0} = -\frac{d\lambda_\theta}{d\theta}\Big|_{\theta=0}$$

$$= -\frac{d}{d\theta}\sum_{x,y,x',y'}\tilde{W}_\theta(x,y|x',y')\tilde{P}_{\theta,XY}(x',y')\Big|_{\theta=0}$$

$$= \sum_{x,y,x',y'} -\frac{d}{d\theta}\tilde{W}_\theta(x,y|x',y')\Big|_{\theta=0}\tilde{P}_{0,XY}(x',y')$$

$$\quad - \sum_{x,y,x',y'}\tilde{W}_0(x,y|x',y')\frac{d}{d\theta}\tilde{P}_{\theta,XY}(x',y')\Big|_{\theta=0}$$

$$= \sum_{x,y,x',y'}\tilde{P}_{0,XY}(x',y')W(x,y|x',y')\log\frac{W(x,y|x',y')}{W_Y(y|y')}$$

$$\quad - \frac{d}{d\theta}\sum_{x,y,x',y'} W(x,y|x',y')\tilde{P}_{\theta,XY}(x',y')\Big|_{\theta=0}$$

$$= H^W(X|Y),$$

where the final equation follows from the relation $\sum_{x,y,x',y'} W(x,y|x',y')\tilde{P}_{\theta,XY}(x',y') = 1$.

As a counterpart of (7), we also define

$$\mathsf{V}^W(X|Y) := \lim_{\theta\to 0}\frac{2\left[H^W(X|Y) - H_{1+\theta}^{\downarrow,W}(X|Y)\right]}{\theta}. \tag{30}$$

Remark 1 When a transition matrix W satisfies Assumption 2, $H_{1+\theta}^{\downarrow,W}(X|Y)$ can be written as

$$H_{1+\theta}^{\downarrow,W}(X|Y) = -\frac{1}{\theta}\log\lambda'_\theta, \tag{31}$$

where λ'_θ is the Perron-Frobenius eigenvalue of $W_{Y,\theta}(y|y')W_Y(y|y')^{-\theta}$. In fact, for the left Perron-Frobenius eigenvector $\hat{Q}_\theta$ of $W_{Y,\theta}(y|y')W_Y(y|y')^{-\theta}$, we have

$$\sum_{x,y} \hat{Q}_\theta(y) W(x,y|x',y')^{1+\theta} W_Y(y|y')^{-\theta} = \lambda'_\theta Q_\theta(y'), \quad (32)$$

which implies that λ'_θ is the Perron-Frobenius eigenvalue of $\tilde{W}_\theta$. Consequently, we can evaluate $H_{1+\theta}^{\downarrow,W}(X|Y)$ by calculating the Perron-Frobenius eigenvalue of $|\mathcal{Y}| \times |\mathcal{Y}|$ matrix instead of $|\mathcal{X}||\mathcal{Y}| \times |\mathcal{X}||\mathcal{Y}|$ matrix when W satisfies Assumption 2.

3) Upper conditional Rényi entropy $H_{1+\theta}^{\uparrow,W}(X|Y)$: Next, we introduce information measures under Assumption 2. In order to define a transition matrix counterpart of (8), let us introduce the following $|\mathcal{Y}| \times |\mathcal{Y}|$ matrix:

$$K_\theta(y|y') := W_{Y,\theta}(y|y')^{\frac{1}{1+\theta}}, \quad (33)$$

where $W_{Y,\theta}$ is defined by (23). Let κ_θ be the Perron-Frobenius eigenvalue of K_θ. Then, we define the upper conditional Rényi entropy for W by

$$H_{1+\theta}^{\uparrow,W}(X|Y) := -\frac{1+\theta}{\theta} \log \kappa_\theta, \quad (34)$$

where $\theta \in (-1,0) \cup (0,\infty)$.

Lemma 4 ([30, Lemma 5]) We have

$$\lim_{\theta \to 0} H_{1+\theta}^{\uparrow,W}(X|Y) = H^W(X|Y) \quad (35)$$

and

$$\lim_{\theta \to 0} \frac{2\left[H^W(X|Y) - H_{1+\theta}^{\uparrow,W}(X|Y)\right]}{\theta} = \mathsf{V}^W(X|Y). \quad (36)$$

Now, let us introduce a transition matrix counterpart of (14). For this purpose, we introduce the following $|\mathcal{Y}| \times |\mathcal{Y}|$ matrix:

$$N_{\theta,\theta'}(y|y') := W_{Y,\theta}(y|y')W_{Y,\theta'}(y|y')^{\frac{-\theta}{1+\theta'}}. \quad (37)$$

Let $\nu_{\theta,\theta'}$ be the Perron-Frobenius eigenvalue of $N_{\theta,\theta'}$. Then, we define the two-parameter conditional Rényi entropy by

$$H_{1+\theta,1+\theta'}^W(X|Y) := -\frac{1}{\theta} \log \nu_{\theta,\theta'} + \frac{\theta'}{1+\theta'} H_{1+\theta'}^{\uparrow,W}(X|Y). \quad (38)$$

Remark 2 Although we defined $H_{1+\theta}^{\downarrow,W}(X|Y)$ and $H_{1+\theta}^{\uparrow,W}(X|Y)$ by (27) and (34) respectively, we can alternatively define these measures in the same spirit as the single-shot setting by introducing a transition matrix counterpart of $H_{1+\theta}(P_{XY}|Q_Y)$ as follows. For the marginal $W_Y(y|y')$ of $W(x,y|x',y')$, let $\mathcal{Y}_{W_Y}^2 := \{(y,y') : W(y|y') > 0\}$. For another transition matrix $\overline{W}_Y$ on $\mathcal{Y}$, we define $\mathcal{Y}_{\overline{W}_Y}^2$ in a similar manner. For $\overline{W}_Y$ satisfying $\mathcal{Y}_{W_Y}^2 \subset \mathcal{Y}_{\overline{W}_Y}^2$, we define[11]

$$H_{1+\theta}^{W|\overline{W}_Y}(X|Y) := -\frac{1}{\theta} \log \lambda_\theta^{W|\overline{W}_Y} \quad (39)$$

[11]Although we can also define $H_{1+\theta}^{W|\overline{W}_Y}(X|Y)$ even if $\mathcal{Y}_{W_Y}^2 \subset \mathcal{Y}_{\overline{W}_Y}^2$ is not satisfied (see [34] for the detail), for our purpose of defining $H_{1+\theta}^{\downarrow,W}(X|Y)$ and $H_{1+\theta}^{\uparrow,W}(X|Y)$, other cases are irrelevant.

for $\theta \in (-1,0) \cup (0,\infty)$, where $\lambda_\theta^{W|\overline{W}_Y}$ is the Perron-Frobenius eigenvalue of

$$W(x,y|x',y')^{1+\theta}\overline{W}_Y(y|y')^{-\theta}. \quad (40)$$

By using this measure, we obviously have

$$H_{1+\theta}^{\downarrow,W}(X|Y) = H_{1+\theta}^{W|W_Y}(X|Y). \quad (41)$$

Furthermore, under Assumption 2, the relation

$$H_{1+\theta}^{\uparrow,W}(X|Y) = \max_{\overline{W}_Y} H_{1+\theta}^{W|\overline{W}_Y}(X|Y) \quad (42)$$

holds [30, (62)], where the maximum is taken over all transition matrices satisfying $\mathcal{Y}_{W_Y}^2 \subset \mathcal{Y}_{\overline{W}_Y}^2$.

4) Properties of conditional Rényi entropies: The information measures introduced in this section have the following properties:

Lemma 5 ([30, Lemma 6])
1) The function $\theta H_{1+\theta}^{\downarrow,W}(X|Y)$ is a concave function of θ, and it is strict concave iff. $\mathsf{V}^W(X|Y) > 0$.
2) $H_{1+\theta}^{\downarrow,W}(X|Y)$ is a monotonically decreasing function of θ, and it is strictly monotonically decreasing iff. $\mathsf{V}(X|Y) > 0$.
3) The function $\theta H_{1+\theta}^{\uparrow,W}(X|Y)$ is a concave function of θ, and it is strict concave iff. $\mathsf{V}^W(X|Y) > 0$.
4) $H_{1+\theta}^{\uparrow,W}(X|Y)$ is a monotonically decreasing function of θ, and it is strictly monotonically decreasing iff. $\mathsf{V}(X|Y) > 0$.
5) For every $\theta \in (-1,0) \cup (0,\infty)$, we have $H_{1+\theta}^{\downarrow,W}(X|Y) \leq H_{1+\theta}^{\uparrow,W}(X|Y)$.
6) For fixed θ', the function $\theta H_{1+\theta,1+\theta'}^W(X|Y)$ is a concave function of θ, and it is strict concave iff. $\mathsf{V}^W(X|Y) > 0$.
7) For fixed θ', $H_{1+\theta,1+\theta'}^W(X|Y)$ is a monotonically decreasing function of θ.
8) We have

$$H_{1+\theta,1}^W(X|Y) = H_{1+\theta}^{\downarrow,W}(X|Y). \quad (43)$$

9) We have

$$H_{1+\theta,1+\theta}^W(X|Y) = H_{1+\theta}^{\uparrow,W}(X|Y). \quad (44)$$

10) For every $\theta \in (-1,0) \cup (0,\infty)$, $H_{1+\theta,1+\theta'}^W(X|Y)$ is maximized at $\theta' = \theta$, i.e.,

$$\left. \frac{dH_{1+\theta,1+\theta'}^W(X|Y)}{d\theta'} \right|_{\theta'=\theta} = 0. \quad (45)$$

5) Functions related to $H_{1+\theta}^{\downarrow,W}(X|Y)$: From Statement 1 of Lemma 5, $\frac{d[\theta H_{1+\theta}^{\downarrow,W}(X|Y)]}{d\theta}$ is monotonically decreasing. Thus, we can define the inverse function $\theta(a) = \theta^\downarrow(a)$ of $\frac{d[\theta H_{1+\theta}^{\downarrow,W}(X|Y)]}{d\theta}$ by

$$\left. \frac{d[\theta H_{1+\theta}^{\downarrow,W}(X|Y)]}{d\theta} \right|_{\theta=\theta(a)} = a \quad (46)$$

for $\underline{a} < a \leq \overline{a}$, where $\underline{a} = \underline{a}^\downarrow := \lim_{\theta \to \infty} \frac{d[\theta H_{1+\theta}^{\downarrow,W}(X|Y)]}{d\theta}$ and $\overline{a} = \overline{a}^\downarrow := \lim_{\theta \to -1} \frac{d[\theta H_{1+\theta}^{\downarrow,W}(X|Y)]}{d\theta}$. Then, due to the definition

(46), we have the following lemma because the function $\theta \mapsto \theta H_{1+\theta}^{\downarrow,W}(X|Y)$ is concave.

Lemma 6 The function $\theta(R)$ defined in (46) satisfies that

$$\theta(R)H_{1+\theta(R)}^{\downarrow,W}(X|Y) - \theta(R)R = \sup_{0 \leq \theta}(\theta H_{1+\theta}^{\downarrow,W}(X|Y) - \theta R).$$
(47)

Next, let

$$R^{\downarrow}(a) := (1 + \theta(a))a - \theta(a)H_{1+\theta(a)}^{\downarrow,W}(X|Y).$$ (48)

Since

$$\frac{dR^{\downarrow}}{da}(a) = 1 + \theta(a),$$ (49)

$R(a)$ is a monotonic increasing function of $\underline{a} < a < R(\overline{a})$. Thus, we can define the inverse function $a(R) = a^{\downarrow}(R)$ of $R(a)$ by

$$(1 + \theta(a(R)))a(R) - \theta(a(R))H_{1+\theta(a(R))}^{\downarrow,W}(X|Y) = R$$ (50)

for $R(\underline{a}) < R < H_0^{\downarrow,W}(X|Y)$, where $H_0^{\downarrow,W}(X|Y) := \lim_{\theta \to -1} H_{1+\theta}^{\downarrow,W}(X|Y)$.

Due to (30), when $\theta(a)$ is close to 0, we have

$$\theta(a)H_{1+\theta(a)}^{\downarrow,W}(X|Y)$$
$$= \theta(a)H^W(X|Y) - \frac{1}{2}V^W(X|Y)\theta(a)^2 + o(\theta(a)^2).$$ (51)

Taking the derivative, (46) implies that

$$a = H^W(X|Y) - V^W(X|Y)\theta(a) + o(\theta(a)).$$ (52)

Hence, when R is close to $H^W(X|Y)$, we have

$$R = (1 + \theta(a(R)))a(R) - \theta H_{1+\theta(a(R))}^{\downarrow,W}(X|Y)$$
$$= H^W(X|Y) - (1 + \frac{\theta(a(R))}{2})\theta(a(R))V^W(X|Y)$$
$$+ o(\theta(a(R))),$$ (53)

i.e.,

$$\theta(a(R)) = -\frac{R - H^W(X|Y)}{V^W(X|Y)} + o(\frac{R - H^W(X|Y)}{V^W(X|Y)}).$$ (54)

Further, Eqs. (51) and (52) imply

$$-\theta(a(R))a(R) + \theta(a(R))H_{1+\theta(a(R))}^{\downarrow,W}(X|Y)$$
$$= V^W(X|Y)\frac{\theta(a(R))^2}{2} + o(\theta(a(R))^2)$$
$$= \frac{V^W(X|Y)}{2}(\frac{R - H^W(X|Y)}{V^W(X|Y)})^2 + o((\frac{R - H^W(X|Y)}{V^W(X|Y)})^2).$$
(55)

6) Functions related to $H_{1+\theta}^{\uparrow,W}(X|Y)$: For $\theta H_{1+\theta}^{\uparrow,W}(X|Y)$, by the same reason, we can define the inverse function $\theta(a) = \theta^{\uparrow}(a)$ by

$$\frac{d[\theta H_{1+\theta,1+\theta(a)}^W(X|Y)]}{d\theta}\bigg|_{\theta=\theta(a)}$$
$$= \frac{d[\theta H_{1+\theta}^{\uparrow,W}(X|Y)]}{d\theta}\bigg|_{\theta=\theta(a)} = a$$ (56)

for $\underline{a} < a \leq \overline{a}$, where $\underline{a} = \underline{a}^{\uparrow} := \lim_{\theta \to \infty}\frac{d[\theta H_{1+\theta}^{\uparrow,W}(X|Y)]}{d\theta}$ and $\overline{a} = \overline{a}^{\uparrow} := \lim_{\theta \to -1}\frac{d[\theta H_{1+\theta}^{\uparrow,W}(X|Y)]}{d\theta}$. Here, the first equation in (56) follows from (45). We also define the inverse function $a(R) = a^{\uparrow}(R)$ of

$$R^{\uparrow}(a) := (1 + \theta(a))a - \theta(a)H_{1+\theta(a)}^{\uparrow,W}(X|Y)$$ (57)

by

$$(1 + \theta(a(R)))a(R) - \theta(a(R))H_{1+\theta(a(R))}^{\uparrow,W}(X|Y) = R$$ (58)

for $R(\underline{a}) < R < H_0^{\uparrow,W}(X|Y)$, where $H_0^{\uparrow,W}(X|Y) := \lim_{\theta \to -1} H_{1+\theta}^{\uparrow,W}(X|Y)$. Then, we can show the following lemma in the same way as Lemma 8 of [30].

Lemma 7 For $R(\underline{a}) < R < H_0^{\uparrow,W}(X|Y)$, we have

$$\sup_{\theta \geq 0}\frac{-\theta R + \theta H_{1+\theta}^{\uparrow,W}(X|Y)}{1 + \theta}$$
$$= -\theta(a(R))a(R) + \theta(a(R))H_{1+\theta(a(R))}^{\uparrow,W}(X|Y).$$ (59)

When the rate R is larger than the critical rate R_{cr} defined by

$$R_{\mathrm{cr}} := R\left(\frac{d[\theta H_{1+\theta}^{\uparrow,W}(X|Y)]}{d\theta}\bigg|_{\theta=1}\right),$$ (60)

the definition (57) of $R(a) = R^{\uparrow}(a)$ yields

$$\sup_{0 \leq \theta \leq 1}\frac{-\theta R + \theta H_{1+\theta}^{\uparrow,W}(X|Y)}{1 + \theta}$$
$$= -\theta(a(R))a(R) + \theta(a(R))H_{1+\theta(a(R))}^{\uparrow,W}(X|Y).$$ (61)

Remark 3 As we can find from (29), (30), and Lemma 4, both the conditional Rényi entropies expand as

$$H_{1+\theta}^{\downarrow,W}(X|Y) = H^W(X|Y) - \frac{1}{2}V^W(X|Y)\theta + o(\theta)$$ (62)

$$H_{1+\theta}^{\uparrow,W}(X|Y) = H^W(X|Y) - \frac{1}{2}V^W(X|Y)\theta + o(\theta)$$ (63)

around $\theta = 0$. Thus, the difference of these measures significantly appear only when $|\theta|$ is rather large.

Remark 4 When $\mathcal{Y}$ is singleton, $H_{1+\theta}^{\downarrow,W}(X|Y)$ coincides with $H_{1+\theta}^{\uparrow,W}(X|Y)$. So, they are simply called the Rényi entropy and denoted by $H_{1+\theta}^W(X)$ for W. $\theta^{\downarrow}(a)$, $a^{\downarrow}(R)$, $R^{\downarrow}(a)$, $\underline{a}^{\downarrow}$, and $\overline{a}^{\downarrow}$ coincide with $\theta^{\uparrow}(a)$, $a^{\uparrow}(R)$, $R^{\uparrow}(a)$, $\underline{a}^{\uparrow}$, and $\overline{a}^{\uparrow}$. They are simplified to $\theta(a)$, $a(R)$, and $R(a)$, $\underline{a}$, and $\overline{a}$.

C. Information Measures for Markov Chain

Let $(\mathbf{X}, \mathbf{Y})$ be the Markov chain induced by a transition matrix W and some initial distribution $P_{X_1 Y_1}$. Now, we show how information measures introduced in Section II-B are related to the conditional Rényi entropy rates. First, we introduce the following lemma, which gives finite upper and lower bounds on the lower conditional Rényi entropy.

Lemma 8 ([30, Lemma 9]) Suppose that a transition matrix W satisfies Assumption 1. Let v_θ be the eigenvector of $\tilde{W}_\theta^T$ with respect to the Perron-Frobenius eigenvalue λ_θ such that[12]

$$\min_{x,y} v_\theta(x,y) = 1. \qquad (64)$$

Let $w_\theta(x,y) := P_{X_1 Y_1}(x,y)^{1+\theta} P_{Y_1}(y)^{-\theta}$. Then, we have

$$(n-1)\theta H_{1+\theta}^{\downarrow,W}(X|Y) + \underline{\delta}(\theta) \le \theta H_{1+\theta}^{\downarrow}(X^n|Y^n)$$
$$\le (n-1)\theta H_{1+\theta}^{\downarrow,W}(X|Y) + \overline{\delta}(\theta), \qquad (65)$$

where

$$\overline{\delta}(\theta) := -\log\langle v_\theta|w_\theta\rangle + \log\max_{x,y} v_\theta(x,y), \qquad (66)$$

$$\underline{\delta}(\theta) := -\log\langle v_\theta|w_\theta\rangle < 0, \qquad (67)$$

and $\langle v_\theta|w_\theta\rangle$ is defined as $\sum_{x,y} v_\theta(x,y) w_\theta(x,y)$.

From Lemma 8, we have the following.

Theorem 1 ([30, Theorem 1]) Suppose that a transition matrix W satisfies Assumption 1. For any initial distribution, we have

$$\lim_{n\to\infty} \frac{1}{n} H_{1+\theta}^{\downarrow}(X^n|Y^n) = H_{1+\theta}^{\downarrow,W}(X|Y), \qquad (68)$$

$$\lim_{n\to\infty} \frac{1}{n} H(X^n|Y^n) = H^W(X|Y). \qquad (69)$$

We also have the following asymptotic evaluation of the variance:

Theorem 2 ([30, Theorem 2]) Suppose that the transition matrix W satisfies Assumption 1. For any initial distribution, we have

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{V}(X^n|Y^n) = \mathsf{V}^W(X|Y). \qquad (70)$$

Theorem 2 is practically important since the limit of the variance can be described by a single letter characterized quantity. A method to calculate $\mathsf{V}^W(X|Y)$ can be found in [35].

Next, we show the lemma that gives finite upper and lower bound on the upper conditional Rényi entropy in terms of the upper conditional Rényi entropy for the transition matrix.

Lemma 9 ([30, Lemma 10]) Suppose that a transition matrix W satisfies Assumption 2. Let v_θ be the eigenvector of K_θ^T with respect to the Perron-Frobenius eigenvalue κ_θ such that $\min_y v_\theta(y) = 1$. Let $w_{Y,\theta}$ be the $|\mathcal{Y}|$-dimensional vector defined by

$$w_{Y,\theta}(y) := \left[\sum_x P_{X_1 Y_1}(x,y)^{1+\theta}\right]^{\frac{1}{1+\theta}}. \qquad (71)$$

Then, we have

$$(n-1)\frac{\theta}{1+\theta} H_{1+\theta}^{\uparrow,W}(X|Y) + \underline{\xi}(\theta) \le \frac{\theta}{1+\theta} H_{1+\theta}^{\uparrow}(X^n|Y^n)$$
$$\le (n-1)\frac{\theta}{1+\theta} H_{1+\theta}^{\uparrow,W}(X|Y) + \overline{\xi}(\theta), \qquad (72)$$

where

$$\overline{\xi}(\theta) := -\log\langle v_\theta|w_{Y,\theta}\rangle + \log\max_y v_\theta(y), \qquad (73)$$

$$\underline{\xi}(\theta) := -\log\langle v_\theta|w_{Y,\theta}\rangle. \qquad (74)$$

From Lemma 9, we have the following.

Theorem 3 ([30, Theorem 3]) Suppose that a transition matrix W satisfies Assumption 2. For any initial distribution, we have

$$\lim_{n\to\infty} \frac{1}{n} H_{1+\theta}^{\uparrow}(X^n|Y^n) = H_{1+\theta}^{\uparrow,W}(X|Y). \qquad (75)$$

Finally, we show the lemma that gives finite upper and lower bounds on the two-parameter conditional Rényi entropy in terms of the two-parameter conditional Rényi entropy for the transition matrix.

Lemma 10 ([30, Lemma 11]) Suppose that a transition matrix W satisfies Assumption 2. Let $v_{\theta,\theta'}$ be the eigenvector of $N_{\theta,\theta'}^T$ with respect to the Perron-Frobenius eigenvalue $\nu_{\theta,\theta'}$ such that $\min_y v_{\theta,\theta'}(y) = 1$. Let $w_{\theta,\theta'}$ be the $|\mathcal{Y}|$-dimensional vector defined by

$$w_{\theta,\theta'}(y) := \left[\sum_x P_{X_1 Y_1}(x,y)^{1+\theta}\right] \left[\sum_x P_{X_1 Y_1}(x,y)^{1+\theta'}\right]^{\frac{-\theta}{1+\theta'}}. \qquad (76)$$

Then, we have

$$(n-1)\theta H_{1+\theta,1+\theta'}^W(X|Y) + \underline{\zeta}(\theta,\theta') \le \theta H_{1+\theta,1+\theta'}(X^n|Y^n)$$
$$\le (n-1)\theta H_{1+\theta,1+\theta'}^W(X|Y) + \overline{\zeta}(\theta,\theta'), \qquad (77)$$

where

$$\overline{\zeta}(\theta,\theta') := -\log\langle v_{\theta,\theta'}|w_{\theta,\theta'}\rangle + \log\max_y v_{\theta,\theta'}(y) + \theta\overline{\xi}(\theta'), \qquad (78)$$

$$\underline{\zeta}(\theta,\theta') := -\log\langle v_{\theta,\theta'}|w_{\theta,\theta'}\rangle + \theta\underline{\xi}(\theta') \qquad (79)$$

for $\theta > 0$ and

$$\overline{\zeta}(\theta,\theta') := -\log\langle v_{\theta,\theta'}|w_{\theta,\theta'}\rangle + \log\max_y v_{\theta,\theta'}(y) + \theta\underline{\xi}(\theta'), \qquad (80)$$

$$\underline{\zeta}(\theta,\theta') := -\log\langle v_{\theta,\theta'}|w_{\theta,\theta'}\rangle + \theta\overline{\xi}(\theta') \qquad (81)$$

for $\theta < 0$.

From Lemma 10, we have the following.

Theorem 4 ([30, Theorem 4]) Suppose that a transition matrix W satisfies Assumption 2. For any initial distribution, we have

$$\lim_{n\to\infty} \frac{1}{n} H_{1+\theta,1+\theta'}(X^n|Y^n) = H_{1+\theta,1+\theta'}^W(X|Y). \qquad (82)$$

---

[12]Since the eigenvector corresponding to the Perron-Frobenius eigenvalue for an irreducible non-negative matrix has always strictly positive entries[46, Theorem 8.4.4, p. 508], we can choose the eigenvector v_θ satisfying (64).

D. Analysis with $\theta = \infty$: One-terminal case

To close this section, we address the case $\theta = \infty$, which was not discussed in the paper [30]. Since the conditional Rényi entropy is monotonically decreasing for θ, the conditional Rényi entropy with the case $\theta = \infty$ is often called the conditional min entropy. To avoid difficulty, we first consider the case when $\mathcal{Y}$ is singleton.

For a single-shot random variable, we have

$$\lim_{\theta \to \infty} H_{1+\theta}(X) = H_\infty(X) \qquad (83)$$
$$:= -\log \max_x P_X(x), \qquad (84)$$

which is usually called min-entropy. For each $x \in \mathcal{X}$, let $\mathcal{C}_x$ be the set of all Hamilton cycle from x to itself. For a path $c = (x_1, x_2, \ldots, x_k)$, we define the set $\hat{c} := \{(x_i, x_{i+1})\}_{i=1}^{k-1}$ and the number $|c|$ to be the number of edges in cycle c, which is the number of elements in the set $\hat{c}$. Then, we define the min-entropy for W by

$$H_\infty^W(X) := -\log \max_{\bar{x} \in \mathcal{X}} \max_{c \in \mathcal{C}_{\bar{x}}} \left(\prod_{(x_a, x_b) \in \hat{c}} W(x_b | x_a) \right)^{1/|c|}, \qquad (85)$$

which is characterized as follows.

Lemma 11 We have

$$\lim_{\theta \to \infty} H_{1+\theta}^W(X) = H_\infty^W(X). \qquad (86)$$

Proof: See Appendix C. ∎

We also have the following lemma.

Lemma 12 For (x, x'), let $\mathcal{C}_{x,x'}$ be the set of all Hamilton paths from x to x'. Then, let

$$A := \min_{\substack{(\bar{x}, \bar{x}') \\ \bar{x} \neq \bar{x}'}} \max_{c \in \mathcal{C}_{\bar{x}, \bar{x}'}} \prod_{(x_a, x_b) \in \hat{c}} W(x_b | x_a). \qquad (87)$$

Furthermore, let x^* and $c^* \in \mathcal{C}_{x^*}$ be such that $H_\infty^W(X)$ is achieved in (85). Then, we have

$$(n-1)H_\infty^W(X) + \underline{\delta}_\infty \leq H_\infty(X^n)$$
$$\leq (n-1)H_\infty^W(X) + \overline{\delta}_\infty, \qquad (88)$$

where

$$\overline{\delta}_\infty := |c^*| H_\infty^W(X) - \log \max_x P_{X_1}(x)$$
$$- \log \min(A, e^{-H_\infty^W(X)}), \qquad (89)$$
$$\underline{\delta}_\infty := -\log \max_x P_{X_1}(x) + \log A. \qquad (90)$$

Proof: See Appendix B. ∎

From Lemma 12, we can derive the following.

Theorem 5 For any initial distribution, we have

$$\lim_{n \to \infty} \frac{1}{n} H_\infty(X^n) = H_\infty^W(X). \qquad (91)$$

E. Analysis with $\theta = \infty$: Two-terminal case

Next, we proceed to the two-terminal case. For single-shot random variables X and Y, we can derive the following.

Lemma 13 ([32]) We have

$$\lim_{\theta \to \infty} H_{1+\theta}^\uparrow(X|Y) = H_\infty^\uparrow(X|Y) \qquad (92)$$
$$:= -\log \sum_y P_Y(y) \max_x P_{X|Y}(x|y), \qquad (93)$$
$$\lim_{\theta \to \infty} H_{1+\theta}^\downarrow(X|Y) = H_\infty^\downarrow(X|Y) \qquad (94)$$
$$:= -\log \max_{\substack{x \in \mathcal{X} \\ y \in \mathrm{supp}(P_Y)}} P_{X|Y}(x|y). \qquad (95)$$

We define the lower min-entropy for W by

$$H_\infty^{\downarrow, W}(X|Y)$$
$$:= -\log \max_{(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}} \max_{c \in \mathcal{C}_{(\bar{x}, \bar{y})}} \left(\prod_{((x', y'), (x, y)) \in \hat{c}} W_{X|X', Y', Y}(x|x', y', y) \right)^{1/|c|}. \qquad (96)$$

Then, similar to Lemma 11, we can show the following lemma.

Lemma 14 We have

$$\lim_{\theta \to \infty} H_{1+\theta}^{\downarrow, W}(X|Y) = H_\infty^{\downarrow, W}(X|Y). \qquad (97)$$

Next, we consider the upper min-entropy for W. When W satisfies Assumption 2, we note that

$$T(y|y') := \max_x W_{X|X', Y', Y}(x|x', y', y) \qquad (98)$$

is well defined, i.e., the right hand side of (98) is independent of x'. Let κ_∞ be the Perron-Frobenius eigenvalue of $W_Y(y|y')T(y|y')$. Then, we define

$$H_\infty^{\uparrow, W}(X|Y) := -\log \kappa_\infty. \qquad (99)$$

Lemma 15 We have

$$\lim_{\theta \to \infty} H_{1+\theta}^{\uparrow, W}(X|Y) = H_\infty^{\uparrow, W}(X|Y). \qquad (100)$$

Proof: See Appendix D. ∎

Theorem 6 Suppose that a transition matrix W satisfies Assumption 1. For any initial distribution, we have

$$\lim_{n \to \infty} \frac{1}{n} H_\infty^\downarrow(X^n|Y^n) = H_\infty^{\downarrow, W}(X|Y). \qquad (101)$$

Suppose that a transition matrix W satisfies Assumption 2. For any initial distribution, we have

$$\lim_{n \to \infty} \frac{1}{n} H_\infty^\uparrow(X^n|Y^n) = H_\infty^{\uparrow, W}(X|Y). \qquad (102)$$

Proof: See Appendix E. ∎

TABLE II
SUMMARY OF THE BOUNDS FOR THE UNIFORM RANDOM NUMBER GENERATION.

Ach./Conv.	Markov	Single Shot	$\Delta,\overline{\Delta},D,\overline{D}$	Complexity	Large Deviation	Moderate Deviation	Second Order	RER Rate
Achievability	Theorem 10	Lemma 19	$\overline{\Delta}$	$O(1)$	✓*	✓		
	Lemma 18		$\overline{\Delta}$	Tail		✓	✓	
	Theorem 13	Theorem 9	$\overline{D}$	$O(1)$				✓
Converse	Theorem 11	Theorem 7	Δ	$O(1)$		✓		
	Theorem 12	Theorem 8	$\overline{\Delta}$	$O(1)$	✓*	✓		
	Lemma 21		Δ	Tail		✓	✓	
	Theorem 14	Proposition 1	D	$O(1)$				✓

III. UNIFORM RANDOM NUMBER GENERATION

In this section, we investigate the uniform random number generation when there is no information leakage. Then, we discuss the single terminal Markov chain. In this case, as is explained in Remark 4, all quantities with the superscript ↓ equal those with the superscript ↑, and these the superscripts are omitted. We start this section by showing the problem setting in Section III-A. Then, we review and introduce some single-shot bounds in Section III-B. We derive non-asymptotic bounds for the Markov chain in Section III-C. Then, in Sections III-D and III-E, we show the asymptotic characterization for the large deviation regime and the moderate deviation regime by using those non-asymptotic bounds. We also derive the second order rate in Section III-F.

The results shown in this section are summarized in Table II. The checkmarks ✓ indicate that the tight asymptotic bounds (large deviation, moderate deviation, and second order) can be obtained from those bounds. The marks ✓* indicate that the large deviation bound can be derived up to the critical rate. The computational complexity "Tail" indicates that the computational complexities of those bounds depend on the computational complexities of tail probabilities.

In Table II, we didn't call the bounds of Lemmas 19 and 18 as theorems due to the following reason. In Subsection I-A, we listed the requirement for the finite-length bounds. Hence, we give a status of Theorem only for a non-asymptotic bound with a computable form. However, Lemmas 19 and 18 require the calculation of the tail probability whose calculation complexity is not $O(1)$ at least in the Markovian case. Hence, Lemmas 19 and 18 are not given the status of Theorem although they derive the asymptotic tight bounds.

A. Problem Formulation

We first present the problem formulation by the single shot setting. Let X be a source whose distribution is P. A random number generator is a function $f : \mathcal{X} \to \{1,\ldots,M\}$. The approximation error is defined by

$$\Delta[f] := \frac{1}{2}\|P_{f(X)} - P_{\overline{U}}\|_1, \qquad (103)$$

where $\overline{U}$ is the uniform random variable on $\{1,\ldots,M\}$. For notational convenience, we introduce the infimum of approximation error under the condition that the range size is M:

$$\Delta(M) := \inf_f \Delta[f]. \qquad (104)$$

When we construct a random number generator, we often use a two-universal hash family $\mathcal{F}$ and a random function F on $\mathcal{F}$. Then, we bound the approximation error averaged over the random function by only using the property of two-universality. As explained in Subsection I-E, to take into the practical aspects, we introduce the worst leaked information:

$$\overline{\Delta}(M) := \sup_F \mathbb{E}[\Delta[F]], \qquad (105)$$

where the supremum is taken over all two-universal hash families from $\mathcal{X}$ to $\{1,\ldots,M\}$. From the definition, we obviously have $\Delta(M) \leq \overline{\Delta}(M)$. When we consider n-fold extension, the random number generator and related quantities are denoted with subscript n. Instead of evaluating the approximation error $\Delta(M_n)$ (or $\overline{\Delta}(M_n)$) for given M_n, we are also interested in evaluating

$$M(n,\varepsilon) := \sup\{M_n : \Delta(M_n) \leq \varepsilon\}, \qquad (106)$$
$$\overline{M}(n,\varepsilon) := \sup\{M_n : \overline{\Delta}(M_n) \leq \varepsilon\} \qquad (107)$$

for given $0 \leq \varepsilon < 1$.

When the output size M is too large, $\Delta(M)$ and $\overline{\Delta}(M)$ are close to 1. So, the criteria $\Delta(M)$ and $\overline{\Delta}(M)$ do not work as proper security measures. In this case, to quantify the performance of the output random number, according to Wyner [36], to discuss the imperfectness of the generated random number, we focus on the difference between the entropies of the generated random number and the ideal uniform random number, which is given as

$$\log M - H(P_{f(X)})$$
$$= \log M - \sum_z \Big(\sum_{x\in f^{-1}(z)} P_X(x)\Big) \log \Big(\sum_{x\in f^{-1}(z)} P_X(x)\Big)$$
$$= D(P_{f(X)}\|P_{\overline{U}}), \qquad (108)$$

where $D(P\|Q)$ is the divergence between two distributions P and Q. When the block size is n, we call the quantity $\frac{1}{n}D(P_{f(X)}\|P_{\overline{U}})$ the relative entropy rate. Then, we focus on the following quantities.

$$D(M) := \inf_f D(P_{f(X)}\|P_{\overline{U}}) \qquad (109)$$

$$\overline{D}(M) := \sup_F \mathbb{E}[D(P_{F(X)}\|P_{\overline{U}})], \qquad (110)$$

where the supremum is taken over all two-universal hash families from $\mathcal{X}$ to $\{1, \ldots, M\}$. Due to the same reason for $\overline{\Delta}(M)$, we consider the criterion $\overline{D}(M)$ in addition to the criterion $D(M)$.

B. Single Shot Bounds

In this section, we review existing single shot bounds and also show novel converse bounds. For the information measures used below, see Remark 4 in Section II, which explains the information measures when $\mathcal{Y}$ is singleton. Furthermore, we need to introduce other information measures. For $P \in \overline{\mathcal{P}}(\mathcal{X})$, let

$$H_{\min}(P) := \log \frac{1}{\max_x P(x)} \quad (111)$$

be the min-entropy. Then, let

$$H_{\min}^\varepsilon(P) := \max_{P' \in \mathcal{B}^\varepsilon(P)} H_{\min}(P') \quad (112)$$

and

$$\overline{H}_{\min}^\varepsilon(P) := \max_{P' \in \overline{\mathcal{B}}^\varepsilon(P)} H_{\min}(P') \quad (113)$$

be smooth min-entropies, where

$$\mathcal{B}^\varepsilon(P) := \left\{ P' \in \mathcal{P}(\mathcal{X}) : \frac{1}{2}\|P - P'\|_1 \leq \varepsilon \right\}, \quad (114)$$

$$\overline{\mathcal{B}}^\varepsilon(P) := \left\{ P' \in \overline{\mathcal{P}}(\mathcal{X}) : \frac{1}{2}\|P - P'\|_1 \leq \varepsilon \right\}, \quad (115)$$

and $\mathcal{P}(\mathcal{X})$ ($\overline{\mathcal{P}}(\mathcal{X})$) is the set of distributions (sub-distributions) over the set $\mathcal{X}$.

First, we have the following achievability bound.

Lemma 16 (Lemma 2.1.1 of [13]) We have

$$\Delta(M) \leq \inf_{\gamma \geq 0} \left[P_X \left\{ \log \frac{1}{P_X(X)} < \gamma \right\} + \frac{M}{e^\gamma} \right]. \quad (116)$$

By using the two-universal hash family, we can derive the following bound.

Lemma 17 ([25]) We have

$$\overline{\Delta}(M) \leq \inf_{0 \leq \varepsilon \leq 1} \left[2\varepsilon + \frac{1}{2}\sqrt{M e^{-\overline{H}_{\min}^\varepsilon(P_X)}} \right]. \quad (117)$$

However, the bound in Lemma 17 cannot be directly calculated in the Markovian chain. To resolve this problem, we slightly loosen Lemma 17 as follows.

Lemma 18 We have

$$\overline{\Delta}(M) \leq \inf_{\gamma \geq 0} \left[P_X \left\{ \log \frac{1}{P_X(X)} < \gamma \right\} + \frac{1}{2}\sqrt{\frac{M}{e^\gamma}} \right]. \quad (118)$$

We also have the following achievability bound.

Lemma 19 (Theorem 1 of [12]) We have

$$\overline{\Delta}(M) \leq \inf_{0 \leq \theta \leq 1} \frac{3}{2} M^{\frac{\theta}{1+\theta}} e^{-\frac{\theta}{1+\theta} H_{1+\theta}(X)}. \quad (119)$$

We also have the following converse bound, which is a special case of Lemma 28 ahead for the more general non-singleton case.

Lemma 20 We have

$$\Delta(M) \geq \min_{H_{\min}^\varepsilon(P) \geq \log M} \varepsilon. \quad (120)$$

Similar to Lemma 17, the bound in Lemma 20 cannot be directly calculated in the Markovian chain. To resolve this problem, we slightly loosen Lemma 20 as follows.

Lemma 21 We have

$$\Delta(M) \geq \max_{\gamma \geq 0} \left[P_X \left\{ \log \frac{1}{P_X(X)} < \gamma \right\} \left(1 - \frac{e^\gamma}{M} \right) \right]. \quad (121)$$

Proof: Fix arbitrary $\gamma \geq 0$. Then, from Lemma 20, there exists $P' \in \mathcal{B}^\varepsilon(P)$ such that

$$\Delta(M) \geq \frac{1}{2}\|P_X - P'\|_1, \quad (122)$$

$$\log \frac{1}{\max_x P'(x)} \geq \log M. \quad (123)$$

Then, we have

$$\frac{1}{2}\|P_X - P'\|_1 = \max_{S \subset \mathcal{X}}(P_X(S) - P'(S)) \quad (124)$$

$$\geq P_X \left\{ x : \log \frac{1}{P_X(x)} < \gamma \right\} - P' \left\{ x : \log \frac{1}{P_X(x)} < \gamma \right\} \quad (125)$$

$$\geq P_X \left\{ x : \log \frac{1}{P_X(x)} < \gamma \right\} - \frac{1}{M}\left| \left\{ x : \log \frac{1}{P_X(x)} < \gamma \right\} \right| \quad (126)$$

$$\geq P_X \left\{ x : \log \frac{1}{P_X(x)} < \gamma \right\} - \frac{1}{M} \sum_{x : \log \frac{1}{P_X(x)} < \gamma} P_X(x)e^\gamma \quad (127)$$

$$= P_X \left\{ \log \frac{1}{P_X(X)} < \gamma \right\} \left(1 - \frac{e^\gamma}{M} \right), \quad (128)$$

where (126) follows from (123). (122) and (128) yield (121). ∎

Although Lemma 21 is useful for the large deviation regime and the moderate deviation regime, it is not useful for the second order regime. To resolve this problem, we loosen Lemma 21 as follows.

Lemma 22 (Lemma 2.1.2 of [13]) We have

$$\Delta(M) \geq \max_{\gamma \geq 0} \left[P_X \left\{ \log \frac{1}{P_X(X)} < \gamma \right\} - \frac{e^\gamma}{M} \right]. \quad (129)$$

This fact implies that Lemma 21 is better than the previous bound given in Lemma 22.

Furthermore, by using a property of the strong universal hash family introduced in [12], we can derive the following converse[13].

---

[13] The paper [12] introduced the strong universal hash family as a special case of a two-universal hash family. Theorem 2 of [12] shows that the strong universal hash family F satisfies $\mathbb{E}[\Delta[F]] \geq \left(1 - \frac{|\Omega|}{M} \right)^2 P_X(\Omega)$.

Lemma 23 (Theorem 2 of [12]) For any subset $\Omega \subset \mathcal{X}$ such that $|\Omega| \leq M$, we have

$$\overline{\Delta}(M) \geq \left(1 - \frac{|\Omega|}{M}\right)^2 P_X(\Omega). \tag{130}$$

Similar to Lemmas 17 and 20, the bound in Lemma 23 cannot be directly calculated in the Markovian chain. To resolve this problem, we modify Lemma 23 as follows.

Lemma 24 For any $0 < \nu < 1$, we have

$$\overline{\Delta}(M) \geq (1-\nu)^2 P_X \left\{ \log \frac{1}{P_X(X)} \leq a(R) \right\}, \tag{131}$$

where $R = \log(M\nu)$, and $a(R)$ is the inverse function defined by (50).

Proof: See Appendix F. ∎

To derive a converse bound for $\Delta(M)$ based on the Rényi entropy, we substitute the formula in Proposition 3 in Appendix A into the bound in Lemma 21 for $a = \gamma = \log(M/2)$. So, we have the following.

Theorem 7 We have

$$- \log \Delta(M)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta}>\theta(a)}} \frac{1}{s} \left[(1+s)\tilde{\theta}\left(H_{1+\tilde{\theta}}(X) - H_{1+(1+s)\tilde{\theta}}(X)\right) \right.$$
$$\left. - (1+s)\log\left(1 - e^{(\theta(a)-\tilde{\theta})a - \theta(a)H_{1+\theta(a)}(X) + \tilde{\theta}H_{1+\tilde{\theta}}(X)}\right) \right]$$
$$+ \log 2, \tag{132}$$

where $a = \log(M/2)$ and $\theta(a)$ is the inverse function defined in (46).

Proof: We evaluate $-\log \Delta(M)$ by using Lemma 21. To evaluate the probability $P_X \left\{ \log \frac{1}{P_X(X)} < a \right\} = P_X \left\{ \log P_X(X) > -a \right\}$, we apply Proposition 3 in Appendix A to the random variable $\log P_X(X)$ whose cumulant generating function $\phi(\rho)$ is $-\theta H_{1+\theta}(X)$. Then, $\rho(-a) = \theta(a)$. Hence,

$$- \log P_X \left\{ \log P_X(X) > -a \right\}$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta}>\theta(a)}} \frac{1}{s} \left[(1+s)\tilde{\theta}\left(H_{1+\tilde{\theta}}(X) - H_{1+(1+s)\tilde{\theta}}(X)\right) \right.$$
$$\left. - (1+s)\log\left(1 - e^{(\theta(a)-\tilde{\theta})a - \theta(a)H_{1+\theta(a)}(X) + \tilde{\theta}H_{1+\tilde{\theta}}(X)}\right) \right]. \tag{133}$$

Since $1 - \frac{e^\gamma}{M} = \frac{1}{2}$, we obtain (132). ∎

To derive a converse bound for $\overline{\Delta}(M)$ based on the Rényi entropy, we substitute the formula in Proposition 3 in Appendix A into the bound in Lemma 24 for $\nu = \frac{1}{2}$. So, we have the following.

Theorem 8 We have

$$- \log \overline{\Delta}(M)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta}>\theta(a(R))}} \frac{1}{s} \left[(1+s)\tilde{\theta}\left(H_{1+\tilde{\theta}}(X) - H_{1+(1+s)\tilde{\theta}}(X)\right) \right.$$
$$- (1+s)\log\left(1 \right.$$
$$\left.\left. - e^{(\theta(a(R))-\tilde{\theta})a(R) - \theta(a(R))H_{1+\theta(a(R))}(X) + \tilde{\theta}H_{1+\tilde{\theta}}(X)}\right) \right]$$
$$+ 2\log 2, \tag{134}$$

where $R = \log(M/2)$, and $\theta(a)$ and $a(R)$ are the inverse functions defined in (46) and (50).

Proof: We evaluate $-\log \overline{\Delta}(M)$ by using Lemma 24 with $\nu = \frac{1}{2}$. The probability $P_X \left\{ \log \frac{1}{P_X(X)} < a(R) \right\} = P_X \left\{ \log P_X(X) > -a(R) \right\}$ can be evaluated by (133). Since $(1-\nu)^2 = \frac{1}{2^2}$, we obtain (134). ∎

Finally, we address the relative entropy rate. As the direct part, we have the following theorem.

Theorem 9 The relative entropy $\overline{D}(M)$ is evaluated as

$$\overline{D}(M) \leq \frac{1}{\theta} \log(1 + M^\theta e^{-\theta H_{1+\theta}(X)}). \tag{135}$$

Proof: Lemma 10 of [47] shows that any two-universal hash function F satisfies the relation

$$\mathbb{E}[M^\theta e^{-\theta H_{1+\theta}(F(X))}] \leq 1 + M^\theta e^{-\theta H_{1+\theta}(X)}, \tag{136}$$

which implies that $\mathbb{E}[\log M - H(F(X))] \leq \mathbb{E}[\log M - H_{1+\theta}(F(X))] = \mathbb{E}\frac{1}{\theta}\log(M^\theta e^{-\theta H_{1+\theta}(F(X))}) \leq \frac{1}{\theta}\log \mathbb{E}(M^\theta e^{-\theta H_{1+\theta}(F(X))}) \leq \frac{1}{\theta}\log(1 + M^\theta e^{-\theta H_{1+\theta}(X)})$. ∎

As the converse part, we have the following theorem.

Proposition 1

$$D(M) \geq \log M - H(P_X) \tag{137}$$

Proof: Inequality (137) follows from the inequality $H(P_X) \geq H(P_{f(X)})$. ∎

C. Finite-Length Bounds for Markov Source

In this subsection, we derive several finite-length bounds for Markovian source with a computable form. Unfortunately, it is not easy to evaluate how tight these bounds are only with their formula. Their tightness will be discussed by considering the asymptotic limit in the remaining subsections of this section. Since we assume the irreducibility for the transition matrix describing the Markovian chain, the following bounds hold with any initial distribution.

To lower bound $-\log \overline{\Delta}(M_n)$ by the Rényi entropy of transition matrix, we substitute the formula for the Rényi entropy given in Lemma 8 into the bound in Lemma 19, we have the following bound.

Theorem 10 Let $R := \frac{1}{n}\log M_n$. Then we have

$$- \log \overline{\Delta}(M_n)$$
$$\geq \sup_{0 \leq \theta \leq 1} \frac{-\theta n R + (n-1)\theta H_{1+\theta}^W(X) + \underline{\delta}(\theta)}{1+\theta} - \log(3/2).$$
$$(138)$$

To upper bound $-\log \Delta(M_n)$ by the Rényi entropy of transition matrix, we substitute the formula for the tail probability given in and Proposition 4 with $a = R$ into the bound in Lemma 21 with $\gamma = nR$, we have the following bound.

Theorem 11 Let $R = \frac{1}{n}\log(M_n/2)$. If $\underline{a} < R < H^W(X)$, then we have

$$- \log \Delta(M_n)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta} > \theta(R)}} \frac{1}{s}\left[(n-1)(1+s)\tilde{\theta}\left[H_{1+\tilde{\theta}}^W(X) - H_{1+(1+s)\tilde{\theta}}^W(X)\right] + \delta_1\right.$$
$$- (1+s)\log\left(1\right.$$
$$\left. - e^{(n-1)[(\theta(R)-\tilde{\theta})R - \theta(R)H_{1+\theta(R)}^W(X) + \tilde{\theta}H_{1+\tilde{\theta}}^W(X)] + \delta_2}\right)\bigg]$$
$$+ \log 2, \qquad (139)$$

where $\theta(a)$ is the inverse function defined in (46), and

$$\delta_1 = (1+s)\overline{\delta}(\tilde{\theta}) - \underline{\delta}((1+s)\tilde{\theta}), \qquad (140)$$
$$\delta_2 = (\theta(R) - \tilde{\theta})R + \overline{\delta}(\tilde{\theta}) - \underline{\delta}(\theta(R)). \qquad (141)$$

Proof: Theorem 11 can be shown by the same way as Theorem 7 with replacing the role of Proposition 3 in Appendix A by Proposition 4. ∎

To upper bound $-\log \overline{\Delta}(M_n)$ by the Rényi entropy of transition matrix, we substitute the formula for the tail probability given in and Proposition 4 with $a = R$ into the bound in Lemma 23, we have the following bound.

Theorem 12 Let R be such that

$$(n-1)R + \{(1+\theta(a(R)))a(R) - \underline{\delta}(\theta(a(R)))\}$$
$$= \log(M_n/2). \qquad (142)$$

If $R(\underline{a}) < R < H^W(X)$, then we have

$$- \log \overline{\Delta}(M_n)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta} > \theta(a(R))}} \frac{1}{s}\left[(n-1)(1+s)\tilde{\theta}\left(H_{1+\tilde{\theta}}^W(X) - H_{1+(1+s)\tilde{\theta}}^W(X)\right)\right.$$
$$\left. + \delta_1 - (1+s)\log\left(1 - e^{C_{1,n}}\right)\right] + 2\log 2,$$
$$(143)$$

where $\theta(a)$ and $a(R)$ are the inverse functions defined in (46) and (50), and

$$C_{1,n} := (n-1)\left[(\theta(a(R)) - \tilde{\theta})a(R)\right.$$
$$\left. - \theta(a(R))H_{1+\theta(a(R))}^W(X) + \tilde{\theta}H_{1+\tilde{\theta}}^W(X)\right] + \delta_2,$$
$$\delta_1 := (1+s)\overline{\delta}(\tilde{\theta}) - \underline{\delta}((1+s)\tilde{\theta}),$$
$$\delta_2 := (\theta(a(R)) - \tilde{\theta})a(R) + \overline{\delta}(\tilde{\theta}) - \underline{\delta}(\theta(a(R))).$$

Proof: See Appendix G. ∎

To upper bound $\overline{D}(e^{nR})$ by the Rényi entropy of transition matrix, we substitute the formula for the Rényi entropy given in Lemma 8 into the bound in Theorem 9, we have the following bound for the relative entropy rate $\frac{1}{n}\overline{D}(e^{nR})$.

Theorem 13 When $R - H_{1+\theta}^W(X) \geq 0$, for $\theta \in [0,1]$, we have

$$\frac{1}{n}\overline{D}(e^{nR}) \leq R - \frac{n-1}{n}H_{1+\theta}^W(X) + \frac{1}{\theta n}(\log 2 - \underline{\delta}(\theta)).$$
$$(144)$$

Proof: Theorem 9 and Lemma 8 yield (a) and (b), respectively, in the following way.

$$D(e^{nR})$$
$$\overset{(a)}{\leq} \frac{1}{\theta}\log(1 + e^{\theta(nR - H_{1+s}(X^n))})$$
$$\overset{(b)}{\leq} \frac{1}{\theta}\log(1 + e^{\theta(nR - (n-1)H_{1+\theta}^W(X)) - \underline{\delta}(\theta)})$$
$$= n(R - H_{1+\theta}^W(X))$$
$$\quad + \frac{1}{\theta}\log(e^{n\theta(H_{1+\theta}^W(X) - R)} + e^{\theta H_{1+\theta}^W(X) - \underline{\delta}(\theta)})$$
$$\leq n(R - H_{1+\theta}^W(X)) + \frac{1}{\theta}\log(1 + e^{\theta H_{1+\theta}^W(X) - \underline{\delta}(\theta)})$$
$$\leq n(R - H_{1+\theta}^W(X)) + \frac{1}{\theta}\log(2e^{\theta H_{1+\theta}^W(X) - \underline{\delta}(\theta)})$$
$$= n(R - H_{1+\theta}^W(X)) + \frac{1}{\theta}(\log 2 + \theta H_{1+\theta}^W(X) - \underline{\delta}(\theta))$$
$$= nR - (n-1)H_{1+\theta}^W(X) + \frac{1}{\theta}(\log 2 - \underline{\delta}(\theta)). \qquad (145)$$

∎

To lower bound $\overline{D}(e^{nR})$ by the Rényi entropy of transition matrix, we substitute the other formula for the Rényi entropy given in Lemma 8 into the bound in Proposition 1, we have the following bound for the relative entropy rate $\frac{1}{n}\overline{D}(e^{nR})$.

Theorem 14 For $\theta \in [0,1]$, we have

$$\frac{1}{n}D(e^{nR}) \geq R - \frac{n-1}{n}H_{1-\theta}^W(X) + \frac{\delta(-\theta)}{\theta n} \qquad (146)$$

Proof: Lemma 8 implies that

$$H(X^n) \leq H_{1-\theta}(X^n) \leq (n-1)H_{1-\theta}^W(X) - \frac{\delta(-\theta)}{\theta}. \qquad (147)$$

Hence, using Proposition 1, we obtain (146). ∎

D. Large Deviation

Taking the limit in the formulas in Theorems 10 and 12, we have the following.

Theorem 15 For $R < H^W(X)$, we have

$$\liminf_{n \to \infty} -\frac{1}{n}\log \overline{\Delta}(e^{nR}) \geq \sup_{0 \leq \theta \leq 1} \frac{-\theta R + \theta H_{1+\theta}^W(X)}{1+\theta}. \qquad (148)$$

On the other hand, for $R(\underline{a}) < R < H^W(X)$, we have

$$\limsup_{n\to\infty} -\frac{1}{n}\log\overline{\Delta}(e^{nR})$$

$$\leq -\theta(a(R))a(R) + \theta(a(R))H^W_{1+\theta(a(R))}(X) \quad (149)$$

$$= \sup_{0\leq\theta} \frac{-\theta R + \theta H^W_{1+\theta}(X)}{1+\theta}. \quad (150)$$

Due to Lemma 7, the lower bound (148) and the upper bound (150) coincide when R is not less than the critical rate R_{cr} given in (60).

Proof: (138) yields (148). Lemma 7 guarantees (150). So, we will prove (149) as follows.

We fix $s > 0$ and $\tilde{\theta} > \theta(a(R))$. Then, (143) implies that

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\Delta}(M_n) \leq \frac{1+s}{s}\tilde{\theta}\left(H^W_{1+\tilde{\theta}}(X) - H^W_{1+(1+s)\tilde{\theta}}(X)\right). \quad (151)$$

Taking the limit $s \to 0$ and $\tilde{\theta} \to \theta(a(R))$, we have

$$\frac{1+s}{s}\tilde{\theta}\left\{H^W_{1+\tilde{\theta}}(X) - H^W_{1+(1+s)\tilde{\theta}}(X)\right\}$$

$$= \frac{1}{s}\left(\tilde{\theta}H^W_{1+\tilde{\theta}}(X) - (1+s)\tilde{\theta}H^W_{1+(1+s)\tilde{\theta}}(X)\right) + \tilde{\theta}H^W_{1+\tilde{\theta}}(X)$$

$$\to -\tilde{\theta}\frac{d\theta H^W_{1+\theta}(X)}{d\theta}\bigg|_{\theta=\tilde{\theta}} + \tilde{\theta}H^W_{1+\tilde{\theta}}(X) \quad (\text{as } s \to 0)$$

$$\to -\theta(a(R))\frac{d\theta H^W_{1+\theta}(X)}{d\theta}\bigg|_{\theta=\theta(a(R))} + \theta(a(R))H^W_{1+\theta(a(R))}(X)$$

$$\qquad\qquad (\text{as } \tilde{\theta} \to \theta(a(R)))$$

$$\stackrel{(a)}{=} \theta(a(R))a + \theta(a(R))H^W_{1+\theta(a(R))}(X), \quad (152)$$

where (a) follows from (56). Hence, (152) and (151) imply that

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{\Delta}(M_n) \leq \theta(a(R))a + \theta(a(R))H^W_{1+\theta(a(R))}(X), \quad (153)$$

which implies (149). ∎

For the general class of functions, we can derive the following converse bound from Theorem 11.

Theorem 16 For $\underline{a} < R < H^W(X)$, we have

$$\limsup_{n\to\infty} -\frac{1}{n}\log\Delta(e^{nR}) \leq -\theta(R)R + \theta(R)H^W_{1+\theta(R)}(X). \quad (154)$$

E. Moderate Deviation

Taking the limit with $R = H^W(X) - n^{-t}\delta$ in Theorem 10 and Theorem 11 (or Theorem 12), we have the following.

Theorem 17 For arbitrary $t \in (0, 1/2)$ and $\delta > 0$, we have

$$\lim_{n\to\infty} -\frac{1}{n^{1-2t}}\log\Delta\left(e^{nH^W(X)-n^{1-t}\delta}\right)$$

$$= \lim_{n\to\infty} -\frac{1}{n^{1-2t}}\log\overline{\Delta}\left(e^{nH^W(X)-n^{1-t}\delta}\right)$$

$$= \frac{\delta^2}{2V^W(X)}. \quad (155)$$

Proof: We apply Theorem 10 and Theorem 11 to the case with $R = H^W(X) - n^{-t}\delta$, i.e., $\theta(a(R)) = -n^{-t}\frac{\delta}{V^W(X)} + o(n^{-t})$. Eqs. (54) and (138) in Theorem 10 imply that

$$-\log\overline{\Delta}(M_n)$$

$$\geq \sup_{0\leq\theta\leq1} \frac{-\theta nR + (n-1)\theta H^W_{1+\theta}(X)}{1+\theta}$$

$$\quad + \inf_{0\leq\theta\leq1} \frac{\underline{\delta}(\theta)}{1+\theta} - \log(3/2)$$

$$\geq n^{1-2t}\frac{\delta^2}{2V^W(X)} + o(n^{1-2t}). \quad (156)$$

We fix an arbitrary $s > 0$. Since $\theta(R) = -n^{-t}\frac{\delta}{V^W(X)} + o(n^{-t})$, we can choose $\tilde{\theta} > \theta(R)$ such that $\tilde{\theta} = -n^{-t}\frac{\delta}{V^W(X)} + o(n^{-t})$. Then, (139) implies that

$$\lim_{n\to\infty} -\frac{1}{n^{1-2t}}\log\Delta(M_n)$$

$$\leq \lim_{n\to\infty} n^{2t}\frac{1+s}{s}\tilde{\theta}\left\{H^W_{1+\tilde{\theta}}(X) - H^W_{1+(1+s)\tilde{\theta}}(X)\right\}$$

$$= \lim_{n\to\infty} n^{2t}\frac{1+s}{s}s\tilde{\theta}^2\frac{dH^W_{1+\theta}(X)}{d\theta}\bigg|_{\theta=\tilde{\theta}} = (1+s)\frac{\delta^2}{2V^W(X)}. \quad (157)$$

Taking the limit $s \to 0$, we obtain the desired argument. ∎

F. Second Order

By applying the central limit theorem to Lemmas 18 and 22, and by using Theorem 2, we have the following.

Theorem 18 For arbitrary $\varepsilon \in (0, 1)$, we have

$$\lim_{n\to\infty} \frac{\log M(n, \varepsilon) - nH^W(X)}{\sqrt{n}}$$

$$= \lim_{n\to\infty} \frac{\log\overline{M}(n, \varepsilon) - nH^W(X)}{\sqrt{n}} = \sqrt{V^W(X)}\Phi^{-1}(\varepsilon). \quad (158)$$

Proof: The central limit theorem for Markovian process [41], [48], [49] [35, Corollary 6.2.] guarantees that the random variable $(-\log P_{X^n}(X^n) - nH^W(X))/\sqrt{n}$ asymptotically obeys the normal distribution with the average 0 and the variance $V^W(X)$. Let $R = \sqrt{V^W(X)}\Phi^{-1}(\varepsilon)$. Substituting $M = e^{nH^W(X)+\sqrt{n}R}$ and $\gamma = nH^W(X) + \sqrt{n}R + n^{\frac{1}{4}}$ in Lemma 18, we have

$$\lim_{n\to\infty} \overline{\Delta}(e^{nH^W(X)+\sqrt{n}R}) \leq \epsilon. \quad (159)$$

Also, substituting $M = e^{nH^W(X)+\sqrt{n}R}$ and $\gamma = nH^W(X) + \sqrt{n}R - n^{\frac{1}{4}}$ in Lemma 22, we have

$$\lim_{n\to\infty} \Delta(e^{nH^W(X)+\sqrt{n}R}) \geq \epsilon. \quad (160)$$
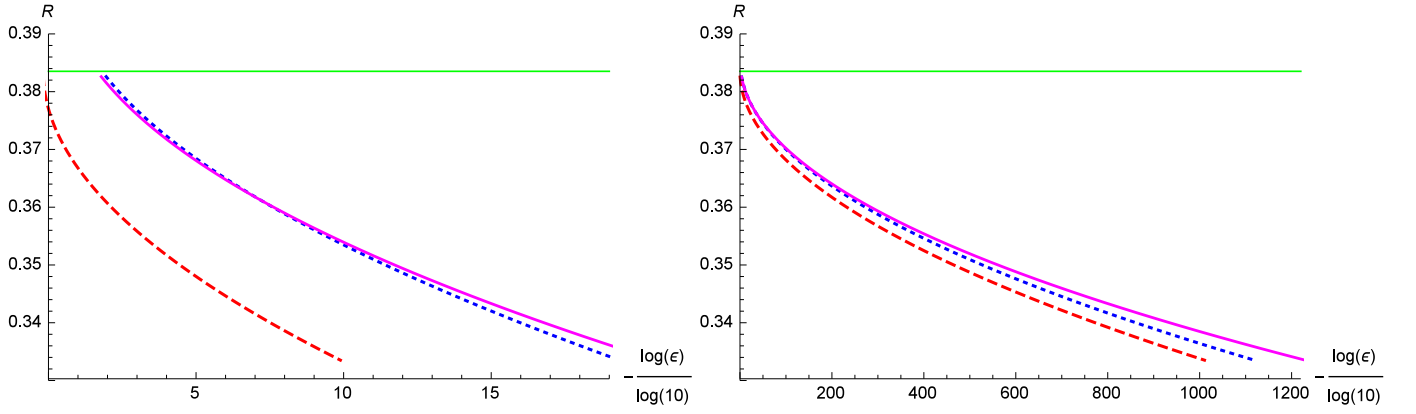
Combining (159) and (160), we obtain (158). ∎

Fig. 1. Comparisons of the bounds for $p = 0.1$ and $q = 0.2$. The left and right graphs express the cases with $n = 10000$ and 1000000, respectively. The horizontal axis is $-\log_{10}(\varepsilon)$, and the vertical axis is the rate R (nats). The red dashed curve is the achievability bound in Theorem 10. The blue dotted curve is the converse bound in Theorem 12. The purple thick curve is the converse bound in Theorem 11. The green normal horizontal line is the entropy $H^W(X)$.

G. Relative Entropy Rate (RER)

Taking the limit in Theorems 13 and 14, we have the following.

Theorem 19 The relative entropy rate (RER) is asymptotically calculated as

$$\lim_{n\to\infty} \frac{1}{n} D(e^{nR}) = \lim_{n\to\infty} \frac{1}{n} \overline{D}(e^{nR}) = [R - H^W(X)]_+, \tag{161}$$

where $[x]_+ := \max(x, 0)$.

Proof: When $R \geq H^W_{1+\theta}(X)$, (144) of Theorem 13 implies that

$$\lim_{n\to\infty} \frac{1}{n} \overline{D}(e^{nR}) \leq R - H^W_{1+\theta}(X) \tag{162}$$

for $\theta \in (0, 1)$. Since $\overline{D}(e^{nR}) \geq \overline{D}(e^{nR'})$ for $R \geq R'$, (162) implies that

$$\lim_{n\to\infty} \frac{1}{n} \overline{D}(e^{nR}) \leq [R - H^W_{1+\theta}(X)]_+ \tag{163}$$

for $\theta \in (0, 1)$ and any R.

Also, (146) of Theorem 14 implies that

$$\lim_{n\to\infty} \frac{1}{n} D(e^{nR}) \geq R - H^W_{1-\theta}(X) \tag{164}$$

for $\theta \in (0, 1)$ and any R. Since $D(e^{nR}) \geq 0$, we have

$$\lim_{n\to\infty} \frac{1}{n} D(e^{nR}) \geq [R - H^W_{1-\theta}(X)]_+ \tag{165}$$

for $\theta \in (0, 1)$ and any R. Taking the limit $\theta \to 0$, we have (161). ∎
[b]

H. Numerical Example

In this section, we numerically evaluate the achievability bound in Theorem 10 and the converse bounds in Theorems 11 and 12. As shown in Theorem 15, the finite-length bounds in Theorems 10 and 12 achieve the optimal rate in the sense of Large deviation when R is larger than the critical rate. Hence,

we can expect that the converse bounds in Theorem 12 is better than that in Theorem 11. Now, we numerically demonstrate how the converse bounds in Theorem 12 is better than that in Theorem 11. Note that the single-shot bounds for second order in Lemmas 18 and 22 are not given in a computable form with Markovian case. So, we compare the bounds given in Theorems 10, 11 and 12.
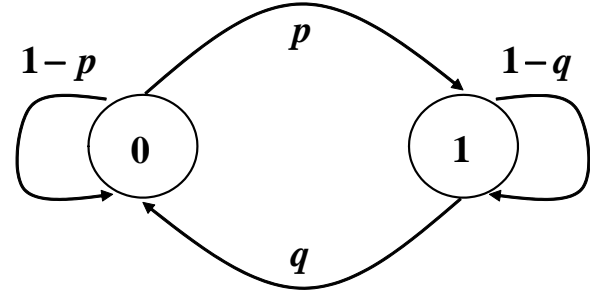

Fig. 2. The description of the transition matrix.

We consider a binary transition matrix W given by Fig. 2, i.e.,

$$W = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}. \tag{166}$$

In this case, the stationary distribution is

$$\tilde{P}(0) = \frac{q}{p+q}, \tag{167}$$

$$\tilde{P}(1) = \frac{p}{p+q}. \tag{168}$$

The entropy is

$$H^W(X) = \frac{q}{p+q} h(p) + \frac{p}{p+q} h(q), \tag{169}$$

where $h(\cdot)$ is the binary entropy function. The tilted transition matrix is

$$\tilde{W}_\theta = \begin{bmatrix} (1-p)^{1+\theta} & q^{1+\theta} \\ p^{1+\theta} & (1-q)^{1+\theta} \end{bmatrix}. \tag{170}$$

TABLE III
SUMMARY OF THE BOUNDS FOR UNIFORM RANDOM NUMBER GENERATION WITH SIDE-INFORMATION.

Ach./Conv.	Markov	Single Shot	$\Delta,\overline{\Delta},D,\overline{D}$	Complexity	Large Deviation	Moderate Deviation	Second Order	MMIR Rate
Achievability	Theorem 23 (Ass. 1)	(Lemma 27)	$\overline{\Delta}$	$O(1)$		✓		
	Theorem 25 (Ass. 2)	Lemma 27	$\overline{\Delta}$	$O(1)$	✓*	✓		
	Lemma 26		$\overline{\Delta}$	Tail		✓	✓	
	Theorem 27 (Ass. 1)	Theorem 22	$\overline{D}$	$O(1)$				✓
Converse	Theorem 24 (Ass. 1)	Theorem 20	Δ	$O(1)$		✓		
	Theorem 26 (Ass. 2)	Theorem 21	$\overline{\Delta}$	$O(1)$	✓*	✓		
	Lemma 29		Δ	Tail		✓	✓	
	Theorem 28 (Ass. 1)	Proposition 2	D	$O(1)$				✓

The Perron-Frobenius eigenvalue is

$$\lambda_\theta = \frac{(1-p)^{1+\theta} + (1-q)^{1+\theta}}{2} + \frac{\sqrt{\{(1-p)^{1+\theta} - (1-q)^{1+\theta}\}^2 + 4p^{1+\theta}q^{1+\theta}}}{2} \quad (171)$$

and its normalized eigenvector is

$$\tilde{P}_\theta(0) = \frac{q^{1+\theta}}{\lambda_\theta - (1-p)^{1+\theta} + q^{1+\theta}}, \quad (172)$$

$$\tilde{P}_\theta(1) = \frac{\lambda_\theta - (1-p)^{1+\theta}}{\lambda_\theta - (1-p)^{1+\theta} + q^{1+\theta}}. \quad (173)$$

The eigenvector of $\tilde{W}_\theta^T$ satisfying (64) is also given by

$$v_\theta(0) = \frac{q^{1+\theta}}{\min(\lambda_\theta - (1-p)^{1+\theta}, q^{1+\theta})}, \quad (174)$$

$$v_\theta(1) = \frac{\lambda_\theta - (1-p)^{1+\theta}}{\min(\lambda_\theta - (1-p)^{1+\theta}, q^{1+\theta})}. \quad (175)$$

From these calculations, we can evaluate the bounds in Theorems 10, 11, and 12. When the initial distribution is given as $P_X(0) = 1$ and $P_X(1) = 0$, for $p = 0.1$, $q = 0.2$, we plotted the bounds in Fig. 1 for fixed block length $n = 10000$ and $n = 1000000$ and varying $\varepsilon = \Delta(M)$ or $\overline{\Delta}(M)$. The two bounds in Theorems 11 and 12 have similar values in the left of Fig. 1. However, the bound in Theorem 12 has a clear advantage in the right of Fig. 1. That is, to clarify the advantage of Theorem 12, we need a very huge size n and a very small ϵ. Although one may consider that $n = 1000000$ is too large to realize, this size is realizable as follows. A typical two-universal hash family can be realized by using Toeplitz matrix. This kind two-universal hash family with $n = 10^8$ was realized efficiently by using a typical personal computer [10, Appendix B][9].

IV. SECURE UNIFORM RANDOM NUMBER GENERATION

In this section, we investigate the secure random number generation with partial information leakage, which is also known as the privacy amplification. We start this section by showing the problem setting in Section IV-A. Then, we review and introduce some single-shot bounds in Section IV-B. We derive non-asymptotic bounds for the Markov chain in Section IV-C. Then, in Sections IV-D and IV-E, we show the

asymptotic characterization for the large deviation regime and the moderate deviation regime by using those non-asymptotic bounds. We also derive the second order rate in Section IV-F.

The results shown in this section are summarized in Table III. The checkmarks ✓ indicate that the tight asymptotic bounds (large deviation, moderate deviation, and second order) can be obtained from those bounds. The marks ✓* indicate that the large deviation bound can be derived up to the critical rate. The computational complexity "Tail" indicates that the computational complexities of those bounds depend on the computational complexities of tail probabilites. It should be noted that Theorem 23 is derived from a special case ($Q_Y = P_Y$) of Lemma 27. The asymptotically optimal choice is $Q_Y = P_Y^{(1+\theta)}$, which corresponds to (190) of Lemma 27. Under Assumption 1, we can derive the bound of the Markov case only for that special choice of Q_Y, while under Assumption 2, we can derive the bound of the Markov case for the optimal choice of Q_Y. Here, we didn't call several lemmas as theorems although they derive the asymptotic tight bound. This is because they are not computable form as explained in the beginning of Section III.

A. Problem Formulation

The privacy amplification is conducted by a function $f: \mathcal{X} \to \{1, \ldots, M\}$. The security of the generated key is evaluated by

$$\Delta[f] := \frac{1}{2}\|P_{f(X)Y} - P_{\overline{U}} \times P_Y\|_1, \quad (176)$$

where $\overline{U}$ is the uniform random variable on $\{1, \ldots, M\}$ and $\|\cdot\|_1$ is the variational distance. For notational convenience, we introduce the infimum of the security criterion under the condition that the range size is M:

$$\Delta(M) := \inf_f \Delta[f]. \quad (177)$$

When we construct a function for the privacy amplification, we often use a two-universal hash family $\mathcal{F}$ and a random function F on $\mathcal{F}$. Then, we bound the security criterion averaged over the random function by only using the property of two-universality. As explained in Subsection I-E, to take into the practical aspects, we introduce the worst leaked information:

$$\overline{\Delta}(M) := \sup_F \mathbb{E}[\Delta[F]], \quad (178)$$

where the supremum is taken over all two-universal hash families from $\mathcal{X}$ to $\{1, \ldots, M\}$. From the definition, we obviously have

$$\Delta(M) \leq \overline{\Delta}(M). \tag{179}$$

When we consider n-fold extension, the security criteria are denoted by $\Delta(M_n)$ or $\overline{\Delta}(M_n)$. As in the single-terminal case, we also introduce the quantities $M(n, \varepsilon)$ and $\overline{M}(n, \varepsilon)$ (cf. (106) and (107)).

Remark 5 Note that the security definition in (176) implies the universal composable security criterion [50], [51]. A slightly weaker security criterion defined by

$$\inf_{Q_Y} \frac{1}{2} \|P_{f(X)Y} - P_{\overline{U}} \times Q_Y\|_1 \tag{180}$$

also implies the universal composable security criterion. In fact some literatures employs this kinds of security criteria [52], [26], [53]. Since the triangle inequality and the information processing inequality $\|Q_Y - P_Y\|_1 \leq \|P_{\overline{U}} \times Q_Y - P_{f(X)Y}\|_1$ imply

$$\frac{1}{2} \|P_{f(X)Y} - P_{\overline{U}} \times P_Y\|_1$$
$$\leq \frac{1}{2} \|P_{f(X)Y} - P_{\overline{U}} \times Q_Y\|_1 + \frac{1}{2} \|P_{\overline{U}} \times Q_Y - P_{\overline{U}} \times P_Y\|_1$$
$$= \frac{1}{2} \|P_{f(X)Y} - P_{\overline{U}} \times Q_Y\|_1 + \frac{1}{2} \|Q_Y - P_Y\|_1$$
$$\leq \frac{1}{2} \|P_{f(X)Y} - P_{\overline{U}} \times Q_Y\|_1 + \frac{1}{2} \|P_{\overline{U}} \times Q_Y - P_{f(X)Y}\|_1,$$

we have

$$\frac{1}{2} \|P_{f(X)Y} - P_{\overline{U}} \times P_Y\|_1 \leq \|P_{f(X)Y} - P_{\overline{U}} \times Q_Y\|_1 \tag{181}$$

holds for any Q_Y. Thus, the two criteria differ only in constant factor, which means that the asymptotic behaviors of the large deviation regime and the moderate deviation regime are not affected by the choice of the security criteria.

For the second order regime, the same fact can be shown as follows. The achievability part (Lemma 26 given in Subsection IV-B) can be used without modification since the optimization over Q_Y is already incorporated into the bound. For the converse part, we need to replace $H_{\min}^\varepsilon(P_{XY}|P_Y)$ with $H_{\min}^\varepsilon(P_{XY}|Q_Y)$ in Lemma 28 given in Subsection IV-B. Then, the converse bound in Lemma 29 given in Subsection IV-B is modified accordingly, i.e.,

$$\Delta(M) \geq \inf_{Q_Y} \max_{\gamma \geq 0} \left[P_{XY} \left\{ \log \frac{Q_Y(y)}{P_{XY}(x,y)} < \gamma \right\} \left(1 - \frac{e^\gamma}{M} \right) \right].$$

However, by noting the inequality

$$P_{XY} \left\{ \log \frac{Q_Y(y)}{P_{XY}(x,y)} < \gamma \right\}$$
$$\geq P_{XY} \left\{ \log \frac{1}{P_{X|Y}(x|y)} < \gamma - \nu \right\} - P_{XY} \left\{ \log \frac{Q_Y(y)}{P_Y(y)} > \nu \right\}$$
$$\geq P_{XY} \left\{ \log \frac{1}{P_{X|Y}(x|y)} < \gamma - \nu \right\} - e^{-\nu} \tag{182}$$

for any $\nu > 0$, the choice $Q_Y = P_Y$ turns out to be the optimal choice asymptotically up to $o(\sqrt{n})$. Thus, the asymptotic

behavior of the second order regime is also not affected by the choice of the security criteria.

When the output size M is too large, $\Delta(M)$ is close to 1 anymore. In this case, to quantify the performance of the output random number, according to Csiszár-Narayan [29], we focus on the relative entropy between the generated random number and the ideal random number as follows.

$$D(P_{f(X)Y} \| P_{\overline{U}} \times P_Y) = \log M - H(f(X)|Y)$$
$$= I(f(X); Y) + D(P_{f(X)} \| P_{\overline{U}}). \tag{183}$$

Since this quantity can be regarded as a modification of the mutual information $I(f(X); Y)$, we call it the modified mutual information. This quantity is naturally given under axiomatic conditions [28]. Then, we address the following quantities.

$$D(M) := \inf_f D(P_{f(X)Y} \| P_{\overline{U}} \times P_Y) \tag{184}$$
$$\overline{D}(M) := \sup_F \mathbb{E}[D(P_{F(X)YF} \| P_{\overline{U}} \times P_Y)]$$
$$= D(P_{F(X)YF} \| P_{\overline{U}} \times P_Y \times P_F) \tag{185}$$

where the supremum is taken over all two-universal hash families from $\mathcal{X}$ to $\{1, \ldots, M\}$. The reason why we consider such a supremum is the same as the case of $\overline{\Delta}(M)$.

B. Single Shot Bounds

In this section, we review existing single shot bounds, and show a novel converse bound. For the information measures used below, see Section II. We also introduce the following information measures. For $P_{XY} \in \overline{\mathcal{P}}(\mathcal{X} \times \mathcal{Y})$ and $Q_Y \in \mathcal{P}(\mathcal{Y})$[14], let

$$H_{\min}(P_{XY}|Q_Y) := -\log \max_{x,y} \frac{P_{XY}(x,y)}{Q_Y(y)} \tag{186}$$

be the conditional min-entropy. Then, for $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, let

$$H_{\min}^\varepsilon(P_{XY}|Q_Y) := \max_{P'_{XY} \in \mathcal{B}^\varepsilon(P_{XY})} H_{\min}(P'_{XY}|Q_Y) \tag{187}$$

and

$$\overline{H}_{\min}^\varepsilon(P_{XY}|Q_Y) := \max_{P'_{XY} \in \overline{\mathcal{B}}^\varepsilon(P_{XY})} H_{\min}(P'_{XY}|Q_Y) \tag{188}$$

be the smooth min-entropy, where

$$\mathcal{B}(P_{XY}) := \left\{ P'_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \frac{1}{2} \|P_{XY} - P'_{XY}\|_1 \leq \varepsilon \right\},$$
$$\overline{\mathcal{B}}(P_{XY}) := \left\{ P'_{XY} \in \overline{\mathcal{P}}(\mathcal{X} \times \mathcal{Y}) : \frac{1}{2} \|P_{XY} - P'_{XY}\|_1 \leq \varepsilon \right\}.$$

By using the two-universal hash family, we can derive the following bound.

Lemma 25 ([25]) For any $Q_Y \in \mathcal{P}(\mathcal{Y})$, we have

$$\overline{\Delta}(M) \leq 2\varepsilon + \frac{1}{2} \sqrt{M e^{-\overline{H}_{\min}^\varepsilon(P_{XY}|Q_Y)}}.$$

---

[14]Technically, we restrict Q_Y to be such that $\text{supp}(P_Y) \subset \text{supp}(Q_Y)$.

However, the bound in Lemma 25 cannot be directly calculated in the Markovian chain. To resolve this problem, we slightly loosen Lemma 25 as follows. (cf. [28, Theorem 23] or [27, Lemma 3]).

Lemma 26 For any $Q_Y \in \mathcal{P}(\mathcal{Y})$, we have

$$\overline{\Delta}(M) \leq \inf_{\gamma \geq 0} \left[P_{XY} \left\{ \log \frac{Q_Y(y)}{P_{XY}(x,y)} < \gamma \right\} + \frac{1}{2}\sqrt{\frac{M}{e^\gamma}} \right].$$

We also have the following exponential bound.

Lemma 27 ([12]) We have

$$\overline{\Delta}(M)$$
$$\leq \min_{Q_Y \in \mathcal{P}(\mathcal{Y})} \inf_{0 \leq \theta \leq 1} \frac{3}{2} M^{\frac{\theta}{1+\theta}} e^{-\frac{\theta}{1+\theta} H_{1+\theta}(P_{XY}|Q_Y)} \quad (189)$$
$$= \inf_{0 \leq \theta \leq 1} \frac{3}{2} M^{\frac{\theta}{1+\theta}} e^{-\frac{\theta}{1+\theta} H_{1+\theta}^\uparrow(X|Y)}. \quad (190)$$

For the converse bound, the following is known[15].

Lemma 28 ([25]) We have

$$\Delta(M) \geq \min_{H_{\min}^\varepsilon(P_{XY}|P_Y) \geq \log M} \varepsilon. \quad (191)$$

Similar to Lemma 25, the bound in Lemma 28 cannot be directly calculated in the Markovian chain. To resolve this problem, we slightly loosen Lemma 28 as follows.

Lemma 29 We have

$$\Delta(M) \geq \max_{\gamma \geq 0} \left[P_{XY} \left\{ \log \frac{1}{P_{X|Y}(x|y)} < \gamma \right\} \left(1 - \frac{e^\gamma}{M} \right) \right]. \quad (192)$$

Proof: The proof is exactly the same as Lemma 21. ■

Although Lemma 29 is useful for the large deviation regime and the moderate deviation regime, it is not useful for the second order regime. To resolve this problem, we loosen Lemma 29 as follows.

Lemma 30 We have

$$\Delta(M) \geq \sup_{\gamma \geq 0} \left[P_{XY} \left\{ \log \frac{1}{P_{X|Y}(x|y)} < \gamma \right\} - \frac{e^\gamma}{M} \right]. \quad (193)$$

Furthermore, by using a property of the strong universal hash family, we can derive the following converse as a generalization of Lemma 23.

Lemma 31 For $\{\Omega_y\}_{y \in \mathcal{Y}}$ such that $|\Omega_y| \leq N \leq M$ for every $y \in \mathcal{Y}$, let $\Omega = \cup_{y \in \mathcal{Y}} \Omega_y \times \{y\}$. Then, we have

$$\overline{\Delta}(M) \geq \left(1 - \frac{N}{M} \right)^2 P_{XY}(\Omega). \quad (194)$$

Proof: We apply Lemma 23 to each $P_{X|Y}(\cdot|y)$ and take average over y. Then, we can derive the lemma since $|\Omega_y| \leq N$ by the assumption. ■

[15]See also [27] for a proof that is specialized for the classical case.

Similar to Lemmas 25 and 28, the bound in Lemma 31 cannot be directly calculated in the Markovian chain. To resolve this problem, we slightly loosen Lemma 31 as follows.

Lemma 32 For any $0 < \nu < 1$, we have

$$\overline{\Delta}(M) \geq (1-\nu)^2 P_{XY} \left\{ \log \frac{P_Y^{(1+\theta(a(R)))}(y)}{P_{XY}(x,y)} \leq a(R) \right\}, \quad (195)$$

where $R = \log(M\nu)$, and $\theta(a)$ and $a(R)$ are the inverse functions $\theta^\uparrow(a)$ and $a^\uparrow(R)$ defined by (20) and (21) respectively.

Proof: See Appendix H. ■

To derive a converse bound for $\Delta(M)$ based on the conditional Rényi entropy, we substitute the formula in Proposition 3 in Appendix A into the bound in Lemma 29 for $a = \gamma = \log(M/2)$. So, we have the following.

Theorem 20 We have

$$-\log \Delta(M)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta} > \theta(a)}} \frac{1}{s} \left[(1+s)\tilde{\theta} \left(H_{1+\tilde{\theta}}^\downarrow(X|Y) - H_{1+(1+s)\tilde{\theta}}^\downarrow(X|Y) \right) \right.$$
$$- (1+s) \log \left(1 \right.$$
$$\left. \left. - e^{(\theta(a)-\tilde{\theta})a - \theta(a)H_{1+\theta(a)}^\downarrow(X|Y) + \tilde{\theta}H_{1+\tilde{\theta}}^\downarrow(X|Y)} \right) \right]$$
$$+ \log 2, \quad (196)$$

where $a = \log(M/2)$, and $\theta(a)$ is the inverse function $\theta^\downarrow(a)$ defined by (18).

Proof: Theorem 20 can be shown by the same way as Theorem 11 with replacing the role of Lemma 21 by Lemma 20. ■

To derive a converse bound for $\overline{\Delta}(M)$ based on the conditional Rényi entropy, we substitute the formula in Proposition 3 in Appendix A into the bound in Lemma 21 for $\nu = \frac{1}{2}$. So, we have the following.

Theorem 21 We have

$$-\log \overline{\Delta}(M)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta} > \theta(a(R))}} \frac{1}{s} \left[(1+s)\tilde{\theta} \left(H_{1+\tilde{\theta},1+\theta(a(R))}(X|Y) \right. \right.$$
$$\left. - H_{1+(1+s)\tilde{\theta},1+\theta(a(R))}(X|Y) \right) - (1+s)\log \left(1 - e^{C_{2,n}} \right) \right]$$
$$+ 2\log 2, \quad (197)$$

where $R = \log(M/2)$,

$$C_{2,n} := [\theta(a(R)) - \tilde{\theta}]a(R) - \theta(a(R))H_{1+\theta(a(R))}^\uparrow(X|Y)$$
$$+ \tilde{\theta}H_{1+\tilde{\theta},1+\theta(a(R))}(X|Y),$$

and $\theta(a)$ and $a(R)$ are the inverse functions $\theta^\uparrow(a)$ and $a^\uparrow(R)$ defined by (20) and (21) respectively.

Proof: Theorem 21 can be shown by the same way as Theorem 8 with replacing the role of Lemma 24 by Lemma 32. ∎

Finally, we address the modified mutual information rate (MMIR). As the direct part, we have the following theorem.

Theorem 22 The maximum modified mutual information $\overline{D}(M)$ among two-universal hash family is bounded as

$$\overline{D}(M) \leq \frac{1}{\theta} \log(1 + M^\theta e^{-\theta H_{1+\theta}(X|Y)}). \tag{198}$$

Proof: Lemma 10 of [47] shows that any two-universal hash function F satisfies the relation

$$\mathbb{E}(M^\theta e^{-\theta H_{1+\theta}(F(X|Y))}) \leq 1 + M^\theta e^{-\theta H_{1+\theta}(X|Y)}, \tag{199}$$

which implies that $\mathbb{E}[\log M - H(F(X|Y))] \leq \mathbb{E}[\log M - H_{1+s}(F(X)|Y)] \leq \frac{1}{s} \log \mathbb{E}(M^s e^{-sH_{1+s}(F(X)|Y)}) \leq \frac{1}{s} \log(1 + M^s e^{-sH_{1+s}(X|Y)})$. ∎

As the converse part, we have the following theorem.

Proposition 2

$$D(M) \geq \log M - H(P_X) \tag{200}$$

Proof: Inequality (200) follows from the inequality $H(X|Y) \geq H(f(X)|Y)$. ∎

C. Finite-Length Bounds for Markov Source

Since we assume the irreducibility for the transition matrix describing the Markovian chain, the following bounds hold with any initial distribution. To lower bound $-\log \overline{\Delta}(M_n)$ by the lower conditional Rényi entropy of transition matrix, we substitute the formula for the lower conditional Rényi entropy given in Lemma 8 into the bound in Lemma 27 for $Q_{Y^n} = P_{Y^n}$, we have the following achievability bound.

Theorem 23 Suppose that a transition matrix W satisfies Assumption 1. Let $R := \frac{1}{n} \log M_n$. Then we have

$$-\log \overline{\Delta}(M_n)$$
$$\geq \sup_{0 \leq \theta \leq 1} \frac{-\theta n R + (n-1)\theta H_{1+\theta}^{\downarrow,W}(X|Y) + \underline{\delta}(\theta)}{1+\theta} - \log(3/2). \tag{201}$$

To upper bound $-\log \Delta(M_n)$ by the lower conditional Rényi entropy of transition matrix, we substitute the formula for the tail probability given in and Proposition 4 with $a = R$ into the bound in Lemma 29 with $\gamma = nR$, we have the following converse bound.

Theorem 24 Suppose that a transition matrix W satisfies Assumption 1. Let $R := \frac{1}{n} \log(M_n/2)$. For any $\underline{a} < R < H^W(X|Y)$, we have

$$-\log \Delta(M_n)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta} > \theta(a)}} \frac{1}{s} \Bigg[(n-1)(1+s)\tilde{\theta} \Big(H_{1+\tilde{\theta}}^{\downarrow,W}(X|Y)$$
$$- H_{1+(1+s)\tilde{\theta}}^{\downarrow,W}(X|Y) \Big) + \delta_1 - (1+s) \log \left(1 - e^{C_{3,n}} \right) \Bigg]$$
$$+ \log 2, \tag{202}$$

where $\theta(a)$ is the inverse function $\theta^\downarrow(a)$ defined by (46), and

$$C_{3,n} := (n-1) \Bigg((\theta(R) - \tilde{\theta})R - \theta(R) H_{1+\theta(R)}^{\downarrow,W}(X|Y)$$
$$+ \tilde{\theta} H_{1+\tilde{\theta}}^{\downarrow,W}(X|Y) \Bigg) + \delta_2, \tag{203}$$

$$\delta_1 := (1+s)\overline{\delta}(\tilde{\theta}) - \underline{\delta}((1+s)\tilde{\theta}), \tag{204}$$

$$\delta_2 := (\theta(R) - \tilde{\theta})R - \underline{\delta}(\theta(R)) + \overline{\delta}(\tilde{\theta}). \tag{205}$$

Proof: Theorem 24 can be shown by the same way as Theorem 11 with replacing the roles of Lemma 21 and Proposition 3 in Appendix A by Lemma 20 and Proposition 4. ∎

Next, we derive tighter bounds under Assumption 2. To lower bound $-\log \overline{\Delta}(M_n)$ by the upper conditional Rényi entropy of transition matrix, we substitute the formula for the upper conditional Rényi entropy given in Lemma 9 into the bound in Lemma 27, we have the following achievability bound.

Theorem 25 Suppose that a transition matrix W satisfies Assumption 2. Let $R := \frac{1}{n} \log M_n$. Then we have

$$-\log \overline{\Delta}(M_n)$$
$$\geq \sup_{0 \leq \theta \leq 1} \frac{-\theta n R + (n-1)\theta H_{1+\theta}^{\uparrow,W}(X|Y)}{1+\theta} + \underline{\xi}(\theta) - \log(3/2). \tag{206}$$

To upper bound $-\log \Delta(M_n)$ by the upper conditional Rényi entropy of transition matrix, we substitute the formula for the tail probability given in and Proposition 3 in Appendix A into the bound in Lemma 31[16], we have the following converse bound.

Theorem 26 Suppose that a transition matrix W satisfies Assumption 2. Let R be such that

$$(n-1)R + \Bigg((1 + \theta(a(R)))(a(R) - \underline{\xi}(\theta(a(R)))) \Bigg)$$
$$= \log(M_n/2). \tag{207}$$

If $R(\underline{a}) < R < H^W(X|Y)$, then we have

$$-\log \overline{\Delta}(M_n)$$
$$\leq \inf_{\substack{s>0 \\ \tilde{\theta} > \theta(a(R))}} \frac{1}{s} \Bigg[(n-1)(1+s)\tilde{\theta} \Big(H_{1+\tilde{\theta},1+\theta(a(R))}^W(X|Y)$$
$$- H_{1+(1+s)\tilde{\theta},1+\theta(a(R))}^W(X|Y) \Big) + \delta_1$$
$$- (1+s) \log \left(1 - e^{C_{4,n}} \right) \Bigg] + 2\log 2, \tag{208}$$

---

[16]We cannot apply Proposition 4 here since we cannot apply Lemma 34 for $\phi(\tilde{\rho}; P_{X^n Y^n} | Q_{Y^n}^{(1-\rho)})$. Instead, we need to apply Lemma 10.

where $\theta(a)$ and $a(R)$ are the inverse functions $\theta^{\uparrow}(a)$ and $a^{\uparrow}(R)$ defined by (56) and (58) respectively,

$$
\begin{aligned}
C_{4,n} := (n-1)\Big[&(\theta(a(R)) - \tilde{\theta})(a(R)) \\
&- \theta(a(R))H^{\uparrow,W}_{1+\theta(a(R))}(X|Y) \\
&+ \tilde{\theta}H^{W}_{1+\tilde{\theta},1+\theta(a(R))}(X|Y)\Big] + \delta_2 \quad (209)
\end{aligned}
$$

$$
\delta_1 := (1+s)\overline{\zeta}(\tilde{\theta},\theta(a(R))) - \underline{\zeta}((1+s)\tilde{\theta},\theta(a(R))), \quad (210)
$$

$$
\begin{aligned}
\delta_2 := &(\theta(a(R)) - \tilde{\theta})(a(R)) - \underline{\zeta}(\theta(a(R)),\theta(a(R))) \\
&+ \overline{\zeta}(\tilde{\theta},\theta(a(R))). \quad (211)
\end{aligned}
$$

Proof: See Appendix I. ∎

We derive finite-length bounds for modified mutual information rate under Assumption 1 by substituting the formula for the lower conditional Rényi entropy given in Lemma 8 into the bound in Theorem 22.

Theorem 27 When $R - H^{\downarrow,W}_{1+\theta}(X|Y) \geq 0$, for $\theta \in [0,1]$, we have

$$
\overline{D}(e^{nR}) \leq nR - (n-1)H^{\downarrow,W}_{1+\theta}(X|Y)) + \frac{1}{\theta}(\log 2 - \underline{\delta}(\theta))). \quad (212)
$$

Proof: Theorem 27 can be shown as the same way as Theorem 13 by replacing $H^{W}_{1+\theta}(X)$ and Theorem 9 by $H^{\downarrow,W}_{1+\theta}(X|Y)$ and Theorem 22, respectively. ∎

To lower bound $\overline{D}(e^{nR})$ by the lower conditional Rényi entropy of transition matrix, we substitute the other formula for the lower conditional Rényi entropy given in Lemma 8 into the bound in Proposition 2, we have the following bound.

Theorem 28 For $\theta \in [0,1]$, we have

$$
D(e^{nR}) \geq nR - (n-1)H^{\downarrow,W}_{1-\theta}(X) + \frac{\underline{\delta}(-\theta)}{\theta} \quad (213)
$$

Proof: Theorem 28 can be shown as the same way as Theorem 14 by replacing $H^{W}_{1-\theta}(X)$ and Proposition 1 by $H^{\downarrow,W}_{1-\theta}(X|Y)$ and Proposition 2, respectively. ∎

D. Large Deviation

We can show the following theorem in the same way as Theorem 15 by taking the limit in Theorems 23 and 24 with use of Lemma 6.

Theorem 29 Suppose that a transition matrix W satisfies Assumption 1. For $R < H^{W}(X|Y)$, we have

$$
\liminf_{n \to \infty} -\frac{1}{n}\log \overline{\Delta}\left(e^{nR}\right) \geq \sup_{0 \leq \theta \leq 1} \frac{-\theta R + \theta H^{\downarrow,W}_{1+\theta}(X|Y)}{1+\theta} \quad (214)
$$

On the other hand, for $\underline{a} < R < H^{W}(X|Y)$, we have

$$
\begin{aligned}
&\limsup_{n \to \infty} -\frac{1}{n}\log \Delta\left(e^{nR}\right) \\
&\leq -\theta(R)R + \theta(R)H^{\downarrow,W}_{1+\theta(R)}(X|Y) \quad (215) \\
&= \sup_{0 \leq \theta} -\theta R + \theta H^{\downarrow,W}_{1+\theta}(X|Y), \quad (216)
\end{aligned}
$$

where $\theta(a)$ is the inverse function $\theta^{\downarrow}(a)$ defined by (46).

Under Assumption 2, taking the limit in Theorems 25 and 26, we have the following tighter bound.

Theorem 30 Suppose that a transition matrix W satisfies Assumption 2. For $R < H^{W}(X|Y)$, we have

$$
\liminf_{n \to \infty} -\frac{1}{n}\log \overline{\Delta}\left(e^{nR}\right) \geq \sup_{0 \leq \theta \leq 1} \frac{-\theta R + \theta H^{\uparrow,W}_{1+\theta}(X|Y)}{1+\theta} \quad (217)
$$

On the other hand, for $R(\underline{a}) < R < H^{W}(X|Y)$, we have

$$
\begin{aligned}
&\limsup_{n \to \infty} -\frac{1}{n}\log \overline{\Delta}\left(e^{nR}\right) \\
&\leq -\theta(a(R))a(R) + \theta(a(R))H^{\uparrow,W}_{1+\theta(a(R))}(X|Y) \quad (218) \\
&= \sup_{0 \leq \theta} \frac{-\theta R + \theta H^{\uparrow,W}_{1+\theta}(X|Y)}{1+\theta}, \quad (219)
\end{aligned}
$$

where $\theta(a)$ and $a(R)$ are the inverse functions $\theta^{\uparrow}(a)$ and $a^{\uparrow}(R)$ defined by (56) and (58) respectively.

Due to Lemma 7, the lower bound (217) and the upper bound (218) coincide when R is not less than the critical rate R_{cr}.

Proof: (206) in Theorem 25 yields (217). Lemma 7 guarantees (219). So, we will prove (218).

We fix $s > 0$ and $\tilde{\theta} > \theta(a(R))$. Then, (208) implies that

$$
\begin{aligned}
&\lim_{n \to \infty} -\frac{1}{n}\log \overline{\Delta}(M_n) \\
&\leq \frac{1+s}{s}\tilde{\theta}\left(H^{W}_{1+\tilde{\theta},1+\theta(a(R))}(X|Y) - H^{W}_{1+(1+s)\tilde{\theta},1+\theta(a(R))}(X|Y)\right) \quad (220)
\end{aligned}
$$

Similar to (152), taking the limits $s \to 0$ and $\tilde{\theta} \to \theta(a(R))$, we have

$$
\begin{aligned}
&\frac{1+s}{s}\tilde{\theta}\left(H^{W}_{1+\tilde{\theta},1+\theta(a(R))}(X|Y) \right. \\
&\qquad\qquad\qquad \left. - H^{W}_{1+(1+s)\tilde{\theta},1+\theta(a(R))}(X|Y)\right) \\
&\to -\tilde{\theta}\frac{d\theta H^{W}_{1+\theta,1+\theta(a(R))}(X|Y)}{d\theta}\bigg|_{\theta=\tilde{\theta}} \\
&\qquad + \tilde{\theta}H^{W}_{1+\tilde{\theta},1+\theta(a(R))}(X|Y) \quad (\text{as } s \to 0) \\
&\to -\theta(a(R))\frac{d\theta H^{W}_{1+\theta,1+\theta(a(R))}(X|Y)}{d\theta}\bigg|_{\theta=\theta(a(R))} \\
&\qquad + \theta(a(R))H^{\uparrow,W}_{1+\theta(a(R))}(X|Y) \quad (\text{as } \tilde{\theta} \to \theta(a(R))) \\
&\overset{(a)}{=} \theta(a(R))a + \theta(a(R))H^{\uparrow,W}_{1+\theta(a(R))}(X|Y). \quad (221)
\end{aligned}
$$

where (a) follows from (56). Hence, (221) and (220) imply that

$$
\lim_{n \to \infty} -\frac{1}{n}\log \overline{\Delta}(M_n) \leq \theta(a(R))a + \theta(a(R))H^{\uparrow,W}_{1+\theta(a(R))}(X|Y), \quad (222)
$$

which implies (218). ∎

E. Moderate Deviation

Taking the limit with $R = H^W(X|Y) - n^{-t}\delta$ in Theorem 23 and Theorem 24, we have the following.

Theorem 31 Suppose that a transition matrix W satisfies Assumption 1. For arbitrary $t \in (0, 1/2)$ and $\delta > 0$, we have

$$
\lim_{n \to \infty} -\frac{1}{n^{1-2t}} \log \Delta \left(e^{nH^W(X|Y) - n^{1-t}\delta} \right)
$$
$$
= \lim_{n \to \infty} -\frac{1}{n^{1-2t}} \log \overline{\Delta} \left(e^{nH^W(X|Y) - n^{1-t}\delta} \right) = \frac{\delta^2}{2\mathsf{V}^W(X|Y)}.
$$
(223)

Proof: This theorem can be shown by the same way as Theorem 17 by replacing (138) and (139) by (201) and (202), respectively. ∎

F. Second Order

By applying the central limit theorem to Lemmas 26 and 30, and by using Theorem 2, we have the following.

Theorem 32 Suppose that a transition matrix W satisfies Assumption 1. For arbitrary $\varepsilon \in (0, 1)$, we have

$$
\lim_{n \to \infty} \frac{\log M(n, \varepsilon) - nH^W(X|Y)}{\sqrt{n}}
$$
$$
= \lim_{n \to \infty} \frac{\log \overline{M}(n, \varepsilon) - nH^W(X|Y)}{\sqrt{n}}
$$
$$
= \sqrt{\mathsf{V}^W(X|Y)} \Phi^{-1}(\varepsilon).
$$
(224)

Proof: The central limit theorem for Markovian process [41], [48], [49] [35, Corollary 6.2.] guarantees that the random variable $(\log P_{X^n|Y^n}(X^n|Y^n) - nH^W(X|Y))/\sqrt{n}$ asymptotically obeys the normal distribution with the average 0 and the variance $\mathsf{V}^W(X|Y)$. This theorem can be shown by the same way as Theorem 18 by replacing the roles of Lemmas 18 and 22 by those of Lemmas 26 and 30 with $Q_Y = P_Y$, respectively. ∎

G. Modified Mutual Information Rate (MMIR)

Taking the limit in Theorems 27 and 28, we have the following.

Theorem 33 Suppose that a transition matrix W satisfies Assumption 1. The modified mutual information rate (MMIR) is asymptotically calculated as

$$
\lim_{n \to \infty} \frac{1}{n} D(e^{nR}) = \lim_{n \to \infty} \frac{1}{n} \overline{D}(e^{nR}) = [R - H^W(X|Y)]_+.
$$
(225)

Proof: Theorem 33 can be shown as the same way as Theorem 19. ∎

V. DISCUSSION AND CONCLUSION

In this paper, we have derived the non-asymptotic bounds on the uniform random number generation with/without information leakage for the Markovian case. In these bounds, the difference between $\Delta(M)$ and $\overline{\Delta}(M)$ is asymptotically negligible at least in the moderate deviation regime and the second order regime. The same relation holds between $D(M)$ and $\overline{D}(M)$. Hence, we can conclude that it is enough to employ any two-universal hash function even for the Markovian case.

Here, to discuss the practical importance of non-asymptotic results, we shall remark a difference of the uniform random number generation from channel and source coding. When we construct a practical system, we need to consider two issues:

- How to *quantitatively* guarantee the performance,
- How to implement the system efficiently.

The uniform random number generation do not have to care about decoding complexity although the coding problems requires decoding, which requires huge amount of calculation complexity. Furthermore, it is also known that universal$_2$ hash functions can be constructed by combination of Toeplitz matrix and the identity matrix. This construction has small amount of complexity and was implemented in a real demonstration [9]. Hence, our non-asymptotic results can be directly used as a performance guarantee of a practical system even when the source distribution has a memory.

Recently, Tsurumaru et al [11] proposed a new class of hash functions, so called ε-almost dual universal hash functions. Then, the recent paper [10] invented more efficient hash functions with less random seeds, which belong to ε-almost dual universal hash functions. Hence, it is needed to extend our result to ε-almost dual universal hash functions. Fortunately, another recent paper [28] has already shown similar results with ε-almost dual universal hash functions in the i.i.d. case. So, it is not so difficult to extend the results in [28] to the Markovian case.

In this paper, we have assumed that the transition matrix describing the Markovian chain is irreducible. When the transition matrix has several irreducible components, we need to consider the mixture distribution among the possible irreducible components, which is defined by the initial distribution. As discussed in [54, Theorem 1], in the finite state space, the asymptotic behavior of the (conditional) Rényi entropy is characterized by the maximum (conditional) Rényi entropy among the possible irreducible components, which depend on the initial distribution. Hence, for large deviation and moderate deviation, the exponential decreasing rate of the leaked information can be evaluated by the minimum rate among the possible irreducible components. On the other hand, in the case of the mixture of the i.i.d. case, when we fix the first and second orders of the coding rate, the limit of the decoding error probability is given by the stochastic mixture of the Gaussian distributions corresponding to the i.i.d. sources [55]. So, for the second order analysis for the Markovian case, we can expect the similar characterization by using the stochastic mixture of the Gaussian distributions corresponding to the irreducible components. Such an analysis is remained for a future study.

APPENDIX A
TAIL PROBABILITY

In converse proofs, we use some techniques to bound tail probabilities in [34], [35]. For this purpose, we need to translate some terminologies in statistics into terminologies in information theory. In this appendix, we introduce some terminologies and bounds from [34], [35]. For proofs, see [34], [35].

A. Single-Shot Setting

Let Z be a real valued random variable with distribution P. Let

$$\phi(\rho) := \log \mathsf{E}\left[e^{\rho Z}\right] = \log \sum_z P(z)e^{\rho z} \tag{226}$$

be the cumulant generating function (CGF). Let us introduce an exponential family

$$P_\rho(z) := P(z)e^{\rho z - \phi(\rho)}. \tag{227}$$

By differentiating the CGF, we find that

$$\phi'(\rho) = \mathsf{E}_\rho[Z] := \sum_z P_\rho(z)z. \tag{228}$$

We also find that

$$\phi''(\rho) = \sum_z P_\rho(z)\left(z - \mathsf{E}_\rho[Z]\right)^2. \tag{229}$$

We assume that Z is not constant. Then, (229) implies that $\phi(\rho)$ is a strict convex function and $\phi'(\rho)$ is monotonically increasing. Thus, we can define the inverse function $\rho(a)$ of $\phi'(\rho)$ by

$$\phi'(\rho(a)) = a. \tag{230}$$

Let

$$D_{1+s}(P\|Q) := \frac{1}{s}\log \sum_z P(z)^{1+s}Q(z)^{-s} \tag{231}$$

be the Rényi divergence. Then, we have the following relation:

$$sD_{1+s}(P_{\tilde\rho}\|P_\rho) = \phi((1+s)\tilde\rho - s\rho) - (1+s)\phi(\tilde\rho) + s\phi(\rho). \tag{232}$$

The following bounds on tail probabilities will be used later.

Proposition 3 ([35, Theorem A.2]) For any $a > \mathsf{E}[Z]$, we have

$$-\log P\{Z \geq a\}$$
$$\leq \inf_{\substack{s>0 \\ \tilde\rho \in \mathbb{R}, \sigma \geq 0}} \frac{1}{s}\left[\phi((1+s)\tilde\rho) - (1+s)\phi(\tilde\rho)\right.$$
$$\left. - (1+s)\log\left(1 - e^{-[\sigma a - \phi(\tilde\rho+\sigma)+\phi(\tilde\rho)]}\right)\right] \tag{233}$$

$$\leq \inf_{\substack{s>0 \\ \tilde\rho > \rho(a)}} \frac{1}{s}\left[\phi((1+s)\tilde\rho) - (1+s)\phi(\tilde\rho)\right.$$
$$\left. - (1+s)\log\left(1 - e^{-[(\tilde\rho-\rho(a))a - \phi(\tilde\rho+\sigma)+\phi(\tilde\rho)]}\right)\right]. \tag{234}$$

Similarly, for any $a < \mathsf{E}[Z]$, we have

$$-\log P\{Z \leq a\}$$
$$\leq \inf_{\substack{s>0 \\ \tilde\rho \in \mathbb{R}, \sigma \geq 0}} \frac{1}{s}\left[\phi((1+s)\tilde\rho) - (1+s)\phi(\tilde\rho)\right.$$
$$\left. - (1+s)\log\left(1 - e^{-[\sigma a - \phi(\tilde\rho+\sigma)+\phi(\tilde\rho)]}\right)\right] \tag{235}$$

$$\leq \inf_{\substack{s>0 \\ \tilde\rho < \rho(a)}} \frac{1}{s}\left[\phi((1+s)\tilde\rho) - (1+s)\phi(\tilde\rho)\right.$$
$$\left. - (1+s)\log\left(1 - e^{-[(\rho(a)-\tilde\rho)a - \phi(\tilde\rho+\sigma)+\phi(\tilde\rho)]}\right)\right]. \tag{236}$$

B. Transition Matrix

The discussion in this and the next subsections is a generalization of that for the lower conditional Rényi entropy $H_{1+\theta}^{\downarrow,W}(X|Y)$ in the following sense. In these subsections, the set $\mathcal{Z}$, and the functions g, $\tilde{g}$, and $\phi(\rho)$ are addressed. The set $\mathcal{Z}$ is the generalization of $\mathcal{X} \times \mathcal{Y}$, and the functions g, $\tilde{g}$, and $\phi(\rho)$ are the generalizations of $\log W - \log W_Y$, $\log P_{X_1 Y_1} - \log P_{Y_1}$, and $-\theta H_{1+\theta}^{\downarrow,W}(X|Y)$, respectively. Under this generalization, the same notation has the same meaning as for the lower conditional Rényi entropy $H_{1+\theta}^{\downarrow,W}(X|Y)$.

Let $\{W(z|z')\}_{(z,z')\in\mathcal{Z}^2}$ be an ergodic and irreducible transition matrix, and let $\tilde{P}$ be its stationary distribution. For a function $g : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$, let

$$\mathsf{E}[g] := \sum_{z,z'} \tilde{P}(z')W(z|z')g(z,z'). \tag{237}$$

We also introduce the following tilted matrix:

$$\tilde{W}_\rho(z|z') := W(z|z')e^{\rho g(z,z')}. \tag{238}$$

Let λ_ρ be the Perron-Frobenius eigenvalue of W_ρ. Then, the CGF for W with generator g is defined by

$$\phi(\rho) := \log \lambda_\rho. \tag{239}$$

Lemma 33 The function $\phi(\rho)$ is a convex function of ρ, and it is strict convex iff. $\phi''(0) > 0$.

From Lemma 33, $\phi'(\rho)$ is monotone increasing function. Thus, we can define the inverse function $\rho(a)$ of $\phi'(\rho)$ by

$$\phi'(\rho(a)) = a. \tag{240}$$

C. Markov Chain

Let $\mathbf{Z} = \{Z^n\}_{n=1}^\infty$ be the Markov chain induced by $W(z|z')$ and an initial distribution P_{Z_1}. For functions $g : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ and $\tilde{g} : \mathcal{Z} \to \mathbb{R}$, let $S_n := \sum_{i=2}^n g(Z_i, Z_{i-1}) + \tilde{g}(Z_1)$. Then, the CGF for S_n is given by

$$\phi_n(\rho) := \log \mathsf{E}\left[e^{\rho S_n}\right]. \tag{241}$$

We will use the following finite evaluation for $\phi_n(\rho)$.

Lemma 34 ([35, Lemma 5.1]) Let v_ρ be the eigenvector of $\tilde{W}_\rho^T$ with respect to the Perron-Frobenius eigenvalue λ_ρ such

that $\min_z v_\rho(z) = 1$. Let $w_\rho(z) := P_{Z_1}(z)e^{\rho\tilde{g}(z)}$. Then, we have

$$(n-1)\phi(\rho) + \underline{\delta}_\phi(\rho) \leq \phi_n(\rho) \leq (n-1)\phi(\rho) + \overline{\delta}_\phi(\rho), \quad (242)$$

where

$$\begin{aligned}
\overline{\delta}_\phi(\rho) &:= \log\langle v_\rho|w_\rho\rangle, & (243)\\
\underline{\delta}_\phi(\rho) &:= \log\langle v_\rho|w_\rho\rangle - \log\max_z v_\rho(z). & (244)
\end{aligned}$$

From this lemma, we have the following.

Corollary 1 For any initial distribution and $\rho \in \mathbb{R}$, we have

$$\lim_{n\to\infty}\phi_n(\rho) = \phi(\rho). \quad (245)$$

The relation

$$\lim_{n\to\infty}\frac{1}{n}\mathsf{E}[S_n] = \phi'(0) = \mathsf{E}[g] \quad (246)$$

is well known. Furthermore, we also have the following.

Lemma 35 For any initial distribution, we have

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{Var}[S_n] = \phi''(0). \quad (247)$$

Finally, we also use the following bound on tail probabilities.

Proposition 4 ([35, Theorem 7.2]) For any $a > \mathsf{E}[g]$, we have

$$\begin{aligned}
&-\log P\{S_n \geq an\}\\
&\leq \inf_{\substack{s>0\\ \tilde{\rho}>\rho(a)}} \frac{1}{s}\Bigg[(n-1)\big(\phi((1+s)\tilde{\rho}) - (1+s)\phi(\tilde{\rho})\big) + \delta_1\\
&\quad - (1+s)\log\left(1 - e^{(n-1)[(\tilde{\rho}-\rho(a))a + \phi(\rho(a)) - \phi(\tilde{\rho})] + \delta_2}\right)\Bigg],
\end{aligned}$$
$$(248)$$

where

$$\begin{aligned}
\delta_1 &:= \overline{\delta}_\phi((1+s)\tilde{\rho}) - (1+s)\underline{\delta}_\phi(\tilde{\rho}), & (249)\\
\delta_2 &:= (\tilde{\rho}-\rho(a))a + \overline{\delta}_\phi(\rho(a)) - \underline{\delta}_\phi(\tilde{\rho}). & (250)
\end{aligned}$$

Similarly, for any $a < \mathsf{E}[g]$, we have

$$\begin{aligned}
&-\log P\{S_n \leq an\}\\
&\leq \inf_{\substack{s>0\\ \tilde{\rho}<\rho(a)}} \frac{1}{s}\Bigg[(n-1)\big(\phi((1+s)\tilde{\rho}) - (1+s)\phi(\tilde{\rho})\big) + \delta_1\\
&\quad - (1+s)\log\left(1 - e^{(n-1)[(\tilde{\rho}-\rho(a))a + \phi(\rho(a)) - \phi(\tilde{\rho})] + \delta_2}\right)\Bigg].
\end{aligned}$$
$$(251)$$

APPENDIX B
PROOF OF LEMMA 12

We first prove the following lemma.

Lemma 36 Suppose that $x_1 = x_n$. Then, we have

$$\prod_{i=2}^{n} W(x_i|x_{i-1}) \leq e^{-(n-1)H_\infty^W(X)}. \quad (252)$$

Proof: When cycle $c = \{(x_1, x_2), \ldots, (x_{n-1}, x_n)\}$ is a Hamilton cycle, the statement is obvious from the definition of $H_\infty^W(X)$. Otherwise, there exists a Hamilton cycle $c' = \{(x_j, x_{j+1}), \ldots, (x_{k-1}, x_k)\}$ in c. Then, we have

$$\begin{aligned}
&\prod_{i=2}^{n} W(x_i|x_{i-1})\\
&= \prod_{(x',x)\in c\backslash c'} W(x|x') \prod_{(x',x)\in c'} W(x|x')\\
&\leq \prod_{(x',x)\in c\backslash c'} W(x|x')e^{-(k-j)H_\infty^W(X)}. \quad (253)
\end{aligned}$$

Since $c\backslash c'$ is also a cycle, by repeating this procedure, we have the statement of the lemma. ∎

We now go back to the proof of Lemma 12. To prove the left hand side inequality of (88), we need to upper bound $\max_{x^n} P_{X^n}(x^n)$.

For a given x^n satisfying the relation $x_1 \neq x_n$, we chose an extension $x^m = (x_1, \ldots, x_m)$ of x^n as follows. (1) x_m is chosen to be x_1. (2) The path $c = \{(x_n, x_{n+1}), \ldots, (x_{m-1}, x_m)\}$ from x_n to x_m is chosen as the Hamilton path $\operatorname{argmax}_{c\in\mathcal{C}_{x_n,x_1}} \prod_{(x_a,x_b)\in\hat{c}} W(x_b|x_a)$. Then, we have

$$\begin{aligned}
AP_{X^n}(x^n) &\leq P_{X^m}(x^m) \overset{(a)}{\leq} \max_x P_{X_1}(x)e^{-(m-1)H_\infty^W(X)}\\
&\leq \max_x P_{X_1}(x)e^{-(n-1)H_\infty^W(X)}, \quad (254)
\end{aligned}$$

where (a) follows from Lemma 36. For a given x^n satisfying the relation $x_1 = x_n$, Lemma 36 implies that

$$P_{X^n}(x^n) \leq \max_x P_{X_1}(x)e^{-(n-1)H_\infty^W(X)}. \quad (255)$$

Since $A \leq 1$, we have the left hand side inequality of (88) in the both case.

To show the opposite inequality, let $\tilde{x} = \operatorname{argmax}_x P_{X_1}(x)$. Assume that $\tilde{x} \neq x^*$. Then, let x^m be the sequence such that it start with $\tilde{x}$, the first part constitutes a Hamilton path $c_o = \operatorname{argmax}_{c\in\mathcal{C}_{\tilde{x},x^*}} \prod_{(x_a,x_b)\in\hat{c}} W(x_b|x_a)$ and then the sequence corresponding to the cycle c^* is repeated $\lceil (n-|c_o|)/|c^*|\rceil$ times. Then, we have

$$\begin{aligned}
\max_{x^n} P_{X^n}(x^n) &\geq \max_{x^{m'}} P_{X^m}(x^{m'}) \geq P_{X^m}(x^m)\\
&\geq P_{X_1}(\tilde{x})Ae^{-\lceil (n-|c_o|)/|c^*|\rceil|c^*|H_\infty^W(X)}\\
&\geq P_{X_1}(\tilde{x})Ae^{-\{(n-|c_o|)+|c^*|\}H_\infty^W(X)}\\
&\geq P_{X_1}(\tilde{x})Ae^{-\{(n-1)+|c^*|\}H_\infty^W(X)}. \quad (256)
\end{aligned}$$

Assume that $\tilde{x} = x^*$. Then, we construct x^m in the same way with omitting the first part. So, we have

$$
\begin{aligned}
\max_{x^n} P_{X^n}(x^n) &\geq \max_{x^{m\prime}} P_{X^m}(x^{m\prime}) \geq P_{X^m}(x^m) \\
&\geq P_{X_1}(\tilde{x}) e^{-\lceil n/|c^*|\rceil |c^*| H_\infty^W(X)} \\
&= P_{X_1}(\tilde{x}) e^{-\{n + |c^*|\} H_\infty^W(X)}
\end{aligned}
\tag{257}
$$

Combining (256) and (257), we have the right hand side inequality of (88). ∎

APPENDIX C
PROOF OF LEMMA 11

To prove (86), we use the limiting results (68) and (91). More precisely, we have

$$
\lim_{\theta\to\infty} H_{1+\theta}^W(X) = \lim_{\theta\to\infty} \lim_{n\to\infty} \frac{1}{n} H_{1+\theta}(X^n)
$$
$$
= \lim_{n\to\infty} \lim_{\theta\to\infty} \frac{1}{n} H_{1+\theta}(X^n) = \lim_{n\to\infty} \frac{1}{n} H_\infty(X^n) = H_\infty^W(X).
\tag{258}
$$

To complete the proof, we need to show that the order of the limits can be changed, which is justified if $\overline{\delta}(\theta)/\theta$ and $\underline{\delta}(\theta)/\theta$ are bounded. For this purpose, it suffices to show $w_\theta(x) \leq M^{1+\theta}$ and $v_\theta(x) \leq \tilde{M}^{1+\theta}$ for some constants $M, \tilde{M}$ because these relations imply that

$$
-\frac{1}{\theta} \log |\mathcal{X}| (M\tilde{M})^{1+\theta} \leq \frac{\underline{\delta}(\theta)}{\theta} \leq \frac{\overline{\delta}(\theta)}{\theta}
$$
$$
\leq \frac{\underline{\delta}(\theta)}{\theta} + \frac{1}{\theta} \log \tilde{M}^{1+\theta} \leq \frac{1}{\theta} \log \tilde{M}^{1+\theta}.
$$

The former is obvious. To prove the latter, without loss of generality, we can assume that $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$ and that $v_\theta(1) \geq \cdots \geq v_\theta(|\mathcal{X}|) = 1$. Since $\tilde{W}_\theta^T$ is irreducible, we can fix an integer m such that $(\tilde{W}_\theta^T)^m(|\mathcal{X}||1) > 0$. Since v_θ is an eigenvector, we have

$$
\sum_{x'} (\tilde{W}_\theta^T)^m(x|x') v_\theta(x') = (\lambda_\theta)^m v_\theta(x).
\tag{259}
$$

On the other hand, we have

$$
\begin{aligned}
&(\tilde{W}_\theta^T)^m(1|x') \\
&= \sum_{x_1, x_2, \ldots, x_{m-1}} \tilde{W}_\theta^T(1|x_{m-1}) \cdots \tilde{W}_\theta^T(x_2|x_1) \tilde{W}_\theta^T(x_1|x') \\
&\leq |\mathcal{X}|^{m-1} \left(\max_{x,\bar{x}} \tilde{W}_\theta^T(x|\bar{x}) \right)^m = |\mathcal{X}|^{m-1} \left(\max_{x,\bar{x}} W(\bar{x}|x)^{1+\theta} \right)^m \\
&= |\mathcal{X}|^{m-1} \left(\max_{x,\bar{x}} W(\bar{x}|x) \right)^{m(1+\theta)}.
\end{aligned}
\tag{260}
$$

Since there exists, at least, one sequence $x_1, x_2, \ldots, x_{m-1}$ such that $\tilde{W}_\theta^T(|\mathcal{X}||x_{m-1}) \cdots \tilde{W}_\theta^T(x_2|x_1) \tilde{W}_\theta^T(x_1|1) > 0$, we have

$$
\begin{aligned}
&(\tilde{W}_\theta^T)^m(|\mathcal{X}||1) \\
&= \sum_{x_1, x_2, \ldots, x_{m-1}} \tilde{W}_\theta^T(|\mathcal{X}||x_{m-1}) \cdots \tilde{W}_\theta^T(x_2|x_1) \tilde{W}_\theta^T(x_1|1) \\
&\geq \left(\min_{\substack{x,\bar{x} \\ W(\bar{x}|x)>0}} \tilde{W}_\theta^T(x|\bar{x}) \right)^m = \left(\min_{\substack{x,\bar{x} \\ W(\bar{x}|x)>0}} W(\bar{x}|x) \right)^{m(1+\theta)}.
\end{aligned}
\tag{261}
$$

Thus, we have

$$
\begin{aligned}
v_\theta(1) &= \frac{(\lambda_\theta)^m v_\theta(1)}{(\lambda_\theta)^m v_\theta(|\mathcal{X}|)} \overset{(a)}{=} \frac{\sum_{x'} (\tilde{W}_\theta^T)^m(1|x') v_\theta(x')}{\sum_{x'} (\tilde{W}_\theta^T)^m(|\mathcal{X}||x') v_\theta(x')} \\
&\leq \frac{\sum_{x'} (\tilde{W}_\theta^T)^m(1|x') v_\theta(x')}{(\tilde{W}_\theta^T)^m(|\mathcal{X}||1) v_\theta(1)} \leq \sum_{x'} \frac{(\tilde{W}_\theta^T)^m(1|x')}{(\tilde{W}_\theta^T)^m(|\mathcal{X}||1)} \\
&\overset{(b)}{\leq} \sum_{x'} \frac{|\mathcal{X}|^{m-1} (\max_{x,\bar{x}} W(\bar{x}|x))^m}{\left(\min_{\substack{x,\bar{x} \\ W(\bar{x}|x)>0}} W(\bar{x}|x) \right)^{m(1+\theta)}} \\
&= |\mathcal{X}|^m \left(\frac{(\max_{x,\bar{x}} W(\bar{x}|x))^m}{\left(\min_{\substack{x,\bar{x} \\ W(\bar{x}|x)>0}} W(\bar{x}|x) \right)^m} \right)^{1+\theta} \\
&\leq \left(\frac{|\mathcal{X}|^m (\max_{x,\bar{x}} W(\bar{x}|x))^{m(1+\theta)}}{\left(\min_{\substack{x,\bar{x} \\ W(\bar{x}|x)>0}} W(\bar{x}|x) \right)^m} \right)^{1+\theta},
\end{aligned}
\tag{262}
$$

where (a) and (b) follow from (259) and the pair of (260) and (261), respectively. Hence, we have the desired bound. ∎

APPENDIX D
PROOF OF LEMMA 15

Since $1 \leq \sum_x \left(\frac{W_{X|X',Y',Y}(x|x',y',y)}{T(y|y')} \right)^{1+\theta} \leq |\mathcal{X}|$, we have

$$
\begin{aligned}
&K_\theta(y|y') \\
&= W_Y(y|y') T(y|y') \left(\sum_x \left(\frac{W_{X|X',Y',Y}(x|x',y',y)}{T(y|y')} \right)^{1+\theta} \right)^{\frac{1}{1+\theta}} \\
&\to W_Y(y|y') T(y|y')
\end{aligned}
\tag{263}
$$

as $\theta \to \infty$. Thus, by the continuity of eigenvalues with respect to the matrix, we have $\kappa_\theta \to \kappa_\infty$, which implies (100). ∎

APPENDIX E
PROOF OF THEOREM 6

To prove (101), we note that $P_{X^n|Y^n}$ can be written as

$$
\begin{aligned}
&P_{X^n|Y^n}(x^n|y^n) \\
&= P_{X_1|Y_1}(x_1|y_1) \prod_{i=2}^n W_{X|X',Y',Y}(x_i|x_{i-1}, y_{i-1}, y_i).
\end{aligned}
\tag{264}
$$

Thus, in a similar manner as the proof of Lemma 12, we can derive an upper bound and a lower bound on $H_\infty^\downarrow(X^n|Y^n)$, from which we can derive (101).

On the other hand, to show (102), we have

$$
\begin{aligned}
&e^{-H_\infty^\uparrow(X^n|Y^n)} \\
&= \sum_{y^n} P_{Y^n}(y^n) \max_{x^n} P_{X^n|Y^n}(x^n|y^n) \\
&= P_{Y_1}(y_1) \max_{x_1} P_{X_1|Y_1}(x_1|y_1) \prod_{i=2}^n W_Y(y_i|y_{i-1}) T(y_i|y_{i-1}).
\end{aligned}
\tag{265}
$$

Thus, in a similar manner as the proof of Lemma 9 shown in [30, Lemma 10], we can derive an upper bound and a lower bound on $H_\infty^\uparrow(X^n|Y^n)$, from which we can derive (102).

APPENDIX F
PROOF OF LEMMA 24

Let

$$\Omega = \left\{ x : \log \frac{1}{P_X(x)} \le a \right\}. \tag{266}$$

Then, for $\rho \le 1$, we have

$$|\Omega| \le \sum_{x \in \Omega} e^{(1-\rho)\left(a - \log \frac{1}{P_X(x)}\right)}$$
$$\le \sum_x P_X(x)^{1-\rho} e^{(1-\rho)a} = e^{(1-\rho)a + \phi(\rho;P)}, \tag{267}$$

where $\phi(\rho; P)$ is defined in (226). Here, we set $\rho = \rho(a)$ and $a = a(R)$. Then, by noting (50), we have

$$|\Omega| \le e^R = M\nu. \tag{268}$$

Thus, by using Lemma 23, we have (131). ∎

APPENDIX G
PROOF OF THEOREM 12

The proof proceed almost in a similar manner as the proof of Lemma 24. Let

$$\Omega = \left\{ x^n : \log \frac{1}{P_{X^n}(x^n)} \le an \right\}. \tag{269}$$

Then, for any $\rho \le 1$, we have

$$\begin{aligned} |\Omega| &\le e^{(1-\rho)an + \phi(\rho; P_{X^n})} \\ &= e^{(1+\theta)an - \theta H_{1+\theta}(X^n)} \\ &\le e^{(1+\theta)an - (n-1)\theta H_{1+\theta}^W(X) - \underline{\delta}(\theta)}, \end{aligned} \tag{270}$$

where we changed variable as $\rho = -\theta$ and used Lemma 8. Here, we set $\theta = \theta(a)$ and $a = a(R)$. Then, by noting (50), we have

$$|\Omega| \le e^{(n-1)R + \left\{(1+\theta(a(R)))a(R) - \underline{\delta}(\theta(a(R)))\right\}} = \frac{M_n}{2}. \tag{271}$$

Thus, by using Lemma 23, we have

$$\overline{\Delta}(M_n) \ge \frac{1}{4} P_{X^n} \left\{ \log \frac{1}{P_{X^n}(x^n)} \le a(R)n \right\}. \tag{272}$$

Finally, by using Proposition 4, and changing the variable as $\tilde\rho = -\tilde\theta$, we have the assertion of the theorem. ∎

APPENDIX H
PROOF OF LEMMA 32

Let

$$\Omega_y = \left\{ x : \log \frac{P_Y^{(1+\theta)}(y)}{P_{XY}(x,y)} \le a \right\}. \tag{273}$$

Then, for any $\theta \ge -1$, we have

$$\begin{aligned} |\Omega_y| &\le \sum_{x \in \Omega_y} e^{(1+\theta)\left(a - \log \frac{P_Y^{(1+\theta)}(y)}{P_{XY}(x,y)}\right)} \\ &\le e^{(1+\theta)a} \sum_x \frac{P_{XY}(x,y)^{1+\theta}}{P_Y^{(1+\theta)}(y)^{1+\theta}} \\ &\overset{(a)}{=} e^{(1+\theta)a} \sum_x \left[P_{XY}(x,y)^{1+\theta} \right. \\ &\qquad \left. \cdot \frac{\left[\sum_y \left(\sum_{x'} P_{XY}(x',y)^{1+\theta}\right)^{\frac{1}{1+\theta}}\right]^{1+\theta}}{\sum_{x''} P_{XY}(x'',y)^{1+\theta}} \right] \\ &\overset{(b)}{=} e^{(1+\theta)a - \theta H_{1+\theta}^\uparrow(X|Y)}, \end{aligned} \tag{274}$$

where (a) and (b) follow from (11) and (10), respectively. Thus, by setting $\theta = \theta(a)$ and $a = a(R)$, and by noting (21), we have

$$|\Omega_y| \le e^R = M\nu. \tag{275}$$

Thus, from Lemma 31, we have (195). ∎

APPENDIX I
PROOF OF THEOREM 26

The proof proceed in a similar manner as the proof of Lemma 32. Let

$$\Omega_{y^n} = \left\{ x^n : \log \frac{P_{Y^n}^{(1+\theta)}(y^n)}{P_{X^n Y^n}(x^n, y^n)} \le an \right\}. \tag{276}$$

Then, for any $\theta \ge -1$, we have (cf. the proof of Lemma 32)

$$\begin{aligned} |\Omega_{y^n}| &\le e^{(1+\theta)an - \theta H_{1+\theta}^\uparrow(X^n|Y^n)} \\ &\le e^{(1+\theta)an - (n-1)\theta H_{1+\theta}^{\uparrow,W}(X|Y) - (1+\theta)\underline{\xi}(\theta)}, \end{aligned} \tag{277}$$

where we used Lemma 9 in the inequality. Here, we set $\theta = \theta(a)$ and $a = a(R)$. Then, by noting (58), we have

$$\begin{aligned} |\Omega_{y^n}| &\le e^{(n-1)R + \left\{(1+\theta(a(R)))(a(R) - \underline{\xi}(\theta(a(R))))\right\}} \\ &= \frac{M_n}{2}. \end{aligned} \tag{278}$$

Thus, by using Lemma 31, we have

$$\overline{\Delta}(M_n) \ge \frac{1}{4} P_{X^n Y^n} \left\{ \log \frac{P_{Y^n}^{(1+\theta(a(R)))}(y^n)}{P_{X^n Y^n}(x^n, y^n)} \le a(R)n \right\} \tag{279}$$

Here, we denote the CGF with $Z = \log \frac{Q_Y(Y)}{P_{XY}(X,Y)}$ by $\phi(\theta; P_{XY}|Q_Y)$. Then, we have

$$\theta H_{1+\theta}^\uparrow(P_{XY}|Q_Y) = -\phi(-\theta; P_{XY}|P_Y^{(1+\theta(a(R)))}). \tag{280}$$

Applying (235) of Proposition 3 to the random variable $Z = \log \frac{P_Y^{(1+\theta(a(R)))}(Y)}{P_{XY}(X,Y)}$, we have

$$- \log P_{X^n Y^n} \left\{ \log \frac{P_{Y^n}^{(1+\theta(a(R)))}(y^n)}{P_{X^n Y^n}(x^n, y^n)} \le a(R)n \right\}$$

$$\le \inf_{\substack{s > 0 \\ \tilde\rho \in \mathbb{R}, \sigma \ge 0}} \frac{1}{s} \left[\phi((1+s)\tilde\rho; P_{X^n Y^n}|P_{Y^n}^{(1+\theta(a(R)))}) \right.$$

$$\left. - (1+s)\phi(\tilde\rho; P_{X^n Y^n}|P_{Y^n}^{(1+\theta(a(R)))}) - (1+s)\log\left(1 - e^{C_5}\right) \right],$$

where

$$C_5 := -\Big[\sigma a - \phi(\tilde{\rho} + \sigma; P_{X^n Y^n} | P_{Y^n}^{(1+\theta(a(R)))})$$
$$+ \phi(\tilde{\rho}; P_{X^n Y^n} | P_{Y^n}^{(1+\theta(a(R)))})\Big].$$

We choose the variable $\tilde{\rho}$ to be $-\tilde{\theta}$ and restrict the variable σ to be $\tilde{\theta} - \theta(a(R))$ with the condition $\tilde{\theta} > \theta(a(R))$. Then, we use (280) and Lemma 10. Hence, we have the assertion of theorem. ∎

ACKNOWLEDGMENT

REFERENCES

[1] M. Hayashi and S. Watanabe, "Non-asymptotic and asymptotic analyses on Markov chains in several problems," in *Proceedings of 2014 Information Theory and Applications Workshop*, Catamaran Resort, San Diego, USA, February 2014, pp. 1–14.

[2] P. Elias, "The efficient construction of an unbiased random sequence," *Ann. Math. Statist.*, vol. 43, pp. 865–870, 1972.

[3] S. Vembu and S. Verdú, "Generating random bits from arbitrary source:fundamental limits," *IEEE Trans. Inform. Theory*, vol. 41, no. 5, pp. 1322–1332, September 1995.

[4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[5] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 4947–4966, November 2009.

[6] ——, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4619–4637, October 2008, arXiv:cs/0503089.

[7] R. Renner and R. König, "Universally composable privacy amplification against quantum adversaries," in *Second Theory of Cryptography Conference TCC*, ser. Lecture Notes in Computer Science, J. Killian, Ed., vol. 3378. Cambridge, MA, USA: Springer-Verlag, 2005, pp. 407–425, arXiv:quant-ph/0403133.

[8] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inform. Theory*, vol. 57, no. 6, pp. 3989–4001, June 2011.

[9] T. Asai and T. Tsurumaru, "Efficient privacy amplification algorithms for quantum key distribution (in japanese)," in *IEICE Technical Report (ISEC2010-121)*, 2011.

[10] M. Hayashi and T. Tsurumaru, "More efficient privacy amplification with less random seeds via dual universal hash function," *Accepted for publication in IEEE Trans. Inform. Theory*, (arXiv:1311.5322).

[11] T. Tsurumaru and M. Hayashi, "Dual universality of hash functions and its applications to quantum cryptography," *IEEE Trans. Inform. Theory*, vol. 59, pp. 4700–4717, 2013, arXiv:1101.0064v2.

[12] M. Hayashi, "Tight exponential analysis of universally composable privacy amplification and its applications," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7728–7746, November 2013.

[13] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer, 2003.

[14] R. G. Gallager, *Low Density Parity Check Codes*. M.I.T. Press, 1963.

[15] Y. Altug and A. B. Wagner, "Moderate deviation analysis of channel coding: Discrete memoryless case," in *Proceedings of IEEE International Symposium on Information Theory*, Austin, Texas, USA, June 2010, pp. 265–269.

[16] D. He, L. A. Lastras-Montano, E. Yang, A. Jagmohan, and J. Chen, "On the redundancy of slepian-wolf coding," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5607–5627, December 2009.

[17] S. Kuzuoka, "A simple technique for bounding the redundancy of source coding with side information," in *Proc. IEEE Int. Symp. Inf. Theory 2012*, Cambridge, MA, 2012, pp. 915–919.

[18] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Trans. Third. Prague Conf. Inf. Th.*, 1962, pp. 689–723.

[19] V. Y. F. Tan, "Moderate-deviations of lossy source coding for discrete and gaussian sources," in *Proc. IEEE Int. Symp. Inf. Theory 2012*, Cambridge, MA, 2012, pp. 920 – 924.

[20] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[21] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography–part I: Secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[22] C. H. Bennett, G. Brassard, and J. M. Robert, "Privacy amplification by public discussion," *SIAM Journal on Computing*, vol. 17, no. 2, pp. 210–229, Apr. 1988.

[23] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer, "Generalized privacy amplification," *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 1915–1923, Nov. 1995.

[24] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Ludy, "A pseudorandom generator from any one-way function," *SIAM Journal on Computing*, vol. 28, no. 4, pp. 1364–1396, 1999.

[25] R. Renner, "Security of quantum key distribution," Ph.D. dissertation, Dipl. Phys. ETH, Switzerland, February 2005.

[26] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7693–7710, November 2013.

[27] S. Watanabe and M. Hayashi, "Non-asymptotic analysis of privacy amplification via Rényi entropy and inf-spectral entropy," in *Proc. IEEE Int. Symp. Inf. Theory 2013*, Istanbul, Turkey, 2013, pp. 2715–2719, arXiv:1211.5252.

[28] M. Hayashi, "Security analysis of ε-almost dual universal$_2$ hash functions: smoothing of min entropy vs. smoothing of rényi entropy of order 2," 2013, arXiv:1309.1596.

[29] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3047–3061, December 2004.

[30] M. Hayashi and S. Watanabe, "Non-asymptotic and asymptotic analyses of information processing on markov chains," 2013, arXiv:1309.7528.

[31] A. Teixeira, A. Matos, and L. Antunes, "Conditional Rényi entropies," *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4273–4277, July 2012.

[32] M. Iwamoto and J. Shikata, "Information theoretic security for encryption based on conditional Rényi entropies," 2013, http://eprint.iacr.org/2013/440.pdf.

[33] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," *Colloquia Mathematica Societatis Janos Bolyai, 16. Topics in Information Theory*, pp. 41–52, 1975.

[34] M. Hayashi and S. Watanabe, "Information geometry approach to parameter estimation in Markov chains," *Accepted for publication in Annals of Statistics*, (arXiv:1401.3814).

[35] S. Watanabe and M. Hayashi, "Finite-length analysis on tail probability and simple hypothesis testing for Markov chain," 2014, arXiv:1401.3801.

[36] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[37] M. Tomamichel and V. Y. F. Tan, "ε-capacities and second-order coding rates for channels with general state," 2013, arXiv:1305.6789.

[38] J. G. Kemeny and J. Snell, *Finite Markov Chains*. Springer, 1976.

[39] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 2*. Wiley, 1971.

[40] A. N. Tikhomirov, "On the convergence rate in the central limit theorem for weakly dependent random variables," *Theory of Probability & Its Applications*, vol. 25, no. 4, pp. 5591–508, 1980.

[41] I. Kontoyiannis and P. Meyn, "Spectral theory and limit theorems for geometrically ergodic Markov processes," *The Annals of Applied Probability*, vol. 13, no. 1, pp. 304–362, 2003.

[42] L. Hervé, J. Ledoux, and V. Patilea, "A uniform Berry-Esseen theorem on m-estimators for geometrically ergodic Markov chains," *Bernoulli*, vol. 18, no. 2, pp. 703–734, 2012.

[43] M. Tomamichel, M. Berta, and M. Hayashi, "A duality relation connecting different quantum generalizations of the conditional Rényi entropy," *J. Math. Phys.*, vol. 55, p. 082206, 2014.

[44] M. Hayashi, "Large deviation analysis for quantum security via smoothing of renyi entropy of order 2," *IEEE Trans. Inform. Theory*, vol. 60, no. 10, pp. 6702–6732, October 2014.

[45] O. S. S. F. M. Muller-Lennert, F. Dupuis and M. Tomamichel, "On quantum rényi entropies: a new definition and some properties," *J. Math. Phys.*, vol. 54, no. 12, p. 122203, 2013.

[46] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Express, 1985.

[47] R. Matsumoto and M. Hayashi, "Universal strongly secure network coding with dependent and non-uniform messages," arXiv:1111.4174.

[48] G. L. Jones, "On the Markov chain central limit theorem," *Probability Surveys*, vol. 1, pp. 299–320, 2004.

[49] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.

[50] R. Canetti, "Universally composable security: a new paradigm for cryptographic protocols," in *Proc. 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, Oct. 2001, pp. 136–145.

[51] B. Pfitzmann and M. Waidner, "Composition and integrity preservation of secure reactive systems," in *7th ACM Conference on Computer and Communications Security*. ACM press, 2000, pp. 245–254.

[52] J. Shikata, "Formalization of information-theoretic security for key agreement, revisited," in *Proc. IEEE Int. Symp. Inf. Theory 2013*, Istanbul, Turkey, 2013, pp. 2720–2724.

[53] M. Koashi, "Simple security proof of quantum key distribution based on complementarity," *New Journal of Physics*, vol. 11, no. 4, p. 045018, April 2009.

[54] F. A. Z. Rached and L. L. Campbell, "Rényi's divergence and entropy rates for finite alphabetmarkov sources," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1553 – 1561, April 2001.

[55] R. Nomura and T. S. Han, "Second-order resolvability, intrinsic randomness, and fixed-length source coding for mixed sources: Information spectrum approach," *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 1–16, January 2013.

Shun Watanabe (M'09) received the B.E., M.E., and Ph.D. degrees from the Tokyo Institute of Technology in 2005, 2007, and 2009, respectively. During April 2009 to February 2015, he was an Assistant Professor in the Department of Information Science and Intelligent Systems at the University of Tokushima. During April 2013 to March 2015, he was a visiting Assistant Professor in the Institute for Systems Research at the University of Maryland, College Park. Since February 2015, he has been an Associate Professor in the Department of Computer and Information Sciences at Tokyo University of Agriculture and Technology. His current research interests are in the areas of information theory, quantum information theory, cryptography, and computer science.

Masahito Hayashi (M'06–SM'13) was born in Japan in 1971. He received the B.S. degree from the Faculty of Sciences in Kyoto University, Japan, in 1994 and the M.S. and Ph.D. degrees in Mathematics from Kyoto University, Japan, in 1996 and 1999, respectively.

He worked in Kyoto University as a Research Fellow of the Japan Society of the Promotion of Science (JSPS) from 1998 to 2000, and worked in the Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN from 2000 to 2003, and worked in ERATO Quantum Computation and Information Project, Japan Science and Technology Agency (JST) as the Research Head from 2000 to 2006. He also worked in the Superrobust Computation Project Information Science and Technology Strategic Core (21st Century COE by MEXT) Graduate School of Information Science and Technology, The University of Tokyo as Adjunct Associate Professor from 2004 to 2007. In 2006, he published the book "Quantum Information: An Introduction" from Springer. He worked in the Graduate School of Information Sciences, Tohoku University as Associate Professor from 2007 to 2012. In 2012, he joined the Graduate School of Mathematics, Nagoya University as Professor. He also worked in Centre for Quantum Technologies, National University of Singapore as Visiting Research Associate Professor from 2009 to 2012 and as Visiting Research Professor from 2012 to now. In 2011, he received Information Theory Society Paper Award (2011) for Information-Spectrum Approach to Second-Order Coding Rate in Channel Coding. In 2016, he received the Japan Academy Medal from the Japan Academy and the JSPS Prize from Japan Society for the Promotion of Science.

He is on the Editorial Board of *International Journal of Quantum Information* and *International Journal On Advances in Security*. His research interests include classical and quantum information theory and classical and quantum statistical inference.