Explicit Renyi Entropy for Hidden Markov Models

Joachim Breitner*, Maciej Skorski†
* *University of Pennsylvania, USA*
joachim@cis.upenn.edu
† *University of Luxembourg, Luxembourg*
maciej.skorski@gmail.com

***Abstract*—Determining entropy rates of stochastic processes is a fundamental but difficult problem, with closed-form solutions known only for specific cases. This paper pushes the state-of-the-art by solving the problem for Hidden Markov Models (HMMs) and Renyi entropies. While computation of Renyi entropy for Markov chains reduces to studying the growth of a simple matrix product, computations for HMMs involve *products of random matrices*. As a result, this case is much harder and no explicit formulas have been known so far.**

In the finite-sample regime we circumvent this issue for Renyi entropy of integer orders, reducing the problem again to *single matrix products* where the matrix is built from transition and emission probabilities by means of *tensor products*. To obtain results in the asymptotic setting, we use a novel technique for determining the growth of non-negative matrix powers. The classical approach – Frobenius-Perron theory – requires positivity assumptions; we instead work directly with the spectral formula. As a consequence, our results do not suffer from limitations such as irreducibility and aperiodicity. This improves our understanding of the entropy rate even for standard (unhidden) chains.

A recently published side-channel attack against RSA was proven effective using our result.

I. Introduction

A. Renyi Entropy of Stochastic Processes

The notion of Renyi entropy, introduced by Renyi in [28] as an extension of Shannon entropy, finds a number of applications across many disciplines including coding theory [7], unsupervised learning [13, 32], anomaly detection [17], multiple source adaptation [21], image processing [19, 23, 29], password guessing [2, 8, 25], randomness extraction [5, 11], testing random number generators [16, 30], quantifying neural activity [24], or economy and finance [9, 14]. In the finite sample regime, one defines the α-th order Renyi entropy of a random variable Z over a finite alphabet $\mathcal{Z}$ as

$$H_\alpha(Z) = \frac{1}{1-\alpha} \log \sum_{z\in\mathcal{Z}} P_Z(z)^\alpha$$

whereas, for a stochastic source $Z = \{Z_i\}_{i=1}^{\infty}$ the quantity of interest is the entropy per output symbol in a *finite-length realization* $Z_1^n = Z_1, \ldots, Z_n$ and its limit

$$H_\alpha(Z) = \lim_{n\to\infty} \frac{H_\alpha(Z_1^n)}{n},$$

called the *entropy rate*. Finding explicit formulas for entropy rates or finite realizations for general sources is intractable, and it remains non-trivial even for restricted classes of sources. So far, the most general sources with known entropy formulas are *Markov chains*, with the asymptotic analysis given in [26] and the finite-length regimes studied recently in [15]. We focus on *Hidden Markov Models*, which are Markov chains observed through a noisy, memoryless channel. While the study of Renyi entropy is justified on its own right, investigating HMM is particularly important because they are very powerful models for sequential data, as seen in natural language processing or in bioinformatics.

B. Summary of Our Results and Related Works

Due to space contraints we were not able to accommodate all the proofs; they appear in the full version (available on arxiv).

We show how to *explicitly* compute the Renyi entropy of HMMs over finite alphabets in both *finite-length and asymptotic regimes*. This problem has been open so far; Wu et al. [31] discuss some convergence properties with no explicit formulas and under further restrictions.

For Shannon entropy the problem has been found hard and solvable only for specific cases, being related to the intractable task of finding top Lapunov exponents [12] in random matrix products. We find that calculating the Renyi entropy in finite-length regimes can be reduced to *powering of explicit substochastic matrices*; in the asymptotic regime powers can be approximated by *spectral analysis* which yields formulas on entropy rates.

To set up the right level of expectations, we note that the IID case and Markov chains are easy to handle due to factorization [26] but finite-sample probabilities of HMMs *do not factorize* into deterministic matrix products [1]. Nonetheless, we show that collision probabilities can be computed using *tensor products* of an explicit auxiliary matrix.

While this part assumes the the entropy order is an integer bigger than 1, it is a mild limitation. Firstly, most of applications of Renyi entropy use integer orders and, in fact, integer orders are preffered for algorithmic efficiency [1]. Secondly and more importantly, Renyi entropy obeys powerfull *interpolation properties* when its smoothed version is used [27]; by perturbing the distribution by a negligible mass of one makes all Renyi entropies of order bigger or equal than 2 close by a constant[2] (no matter the alphabet size), so that asymptotically all rates are close (by $o(1)$ with a large number of samples).

Interestingly, in the "spectral analysis" part, we *remove positivity assumptions* used before [12, 15, 26, 31] for computing entropy rates; the key ingredient is an elegant general lemma

[1] In HMMs factorization depends on randomness of emission probabilities.

[2] Perturbing ϵ fraction of the mass yields the gap of at most $\log(1/\epsilon)$ bits.

on the growth of matrix products. Table I gives a summary of our results compared to related literature.

TABLE I: Formulas on entropy of stochastic processes.

Authors	Model	Entropy	Technique	Model Limitations
[15, 26]	Markov	Renyi	matrix powering	positivity assumptions
[12]	HMM (binary)	Shannon	random matrix products	positivity assumptions
[31]	HMM	Renyi	Markov-approximations	positivity assumptions no explicit formula
this paper	HMM	Renyi	tensoring + matrix powering new lemma on matrix products	none

C. *Our Result and Techniques*

Evaluation of Renyi entropy of a process $Z = \{Z_i\}_{i=1}^{\infty}$ over an alphabet $\mathcal{Z}$ reduces to the *collision probability*

$$\mathsf{CP}_{\alpha}(Z_1^n) = \sum_{z_1^n \in \mathcal{Z}^n} p(z_1^n)^{\alpha}. \quad (1)$$

If Z is a Markov chain, then we can factorize $p(z_1^n) = p(z_i) \cdot \prod_{i=2}^{n} p(z_i|z_{i-1})$ which can be computed as a *product of one matrix* because $p(z_i|z_{i-1}) = M(z_i; z_{i-1})$, where M is the state transition matrix, does not depend on i. Thus (1) depends on *matrix products* of the α-*entrywise* power of M, denoted by $M^{\diamond\alpha}$. Matrix powers, under extra positivity assumptions, can be approximated using Perron-Frobenius theory [20]. It follows that the asymptotic behavior is controlled by the biggest eigenvalue ρ of $M^{\diamond\alpha}$. In particular for large n we have [26]

$$\mathsf{CP}_{\alpha}(Z_1^n) = \Theta(1) \cdot \rho(M^{\diamond\alpha})^n \quad (2)$$

$$\frac{H_{\alpha}(Z_1^n)}{n} = \frac{1+o(1)}{1-\alpha} \log \rho(M^{\diamond\alpha}), \quad n \to \infty. \quad (3)$$

For Hidden Markov Models, which are observations of some Markov process $\{X_i\}_{i=1}^{\infty}$, this approach fails. This is because the factorization $p(z_1^n)^{\alpha} = p(z_1)^{\alpha} \cdot \prod_{i=2}^{n} p(z_i|z_{i-1})^{\alpha}$ boils down to *random matrix products*, as the transition from z_{i-1} to z_i depends the hidden states x_{i-1}, x_i which changes following a random process. The theory of asymptotic properties of random matrix products is not only fairly involved but so far insufficient for the problem at hand. We have explicit results for products of stochastic matrices [3]; however our matrices $p(z_i|z_{i-1})^{\alpha}$ are sub-stochastic because of the α-power. A tempting alternative might be the standard factorization conditioned on hidden states $p(z_1^n) = p(z_1, x_1) \cdot \sum_{x_1^n} \prod_{i=2}^{n} p(z_i|x_i)p(x_i|x_{i-1})$. However, while it can be computed recursively by dynamic programming, it does not reduce to matrix multiplication when raised to the power α (as opposed to the previous case).

From now, we assume that Z_i is a hidden Markov process with the underlying Markov chain X_i, both on finite alphabets.

1) Closed Formulas for Renyi Entropies of HMMs: To avoid random matrix products or recurrences with no explicit solution, we change the approach and observe that, in case of integer α, we can see (1) as the probability that α independent finite length realizations collide. Thus, we are interested in the event

$$E_n = \left\{ \forall i = 1, \ldots, n : \ Z_i^{(1)} = Z_i^{(2)} = \ldots = Z_i^{(\alpha)} \right\}$$

where $Z^{(1)}, \ldots, Z^{(\alpha)}$ are independent copies of Z. The probability of this event can be evaluated recursively by dynamic programming, conditioned on hidden states. More precisely, denote for shortness the tuples of random variables $X_i' = \left(X_i^{(1)}, \ldots, X_i^{(\alpha)}\right)$ and $Z_i' = \left(Z_i^{(1)}, \ldots, Z_i^{(\alpha)}\right)$. The probability of going from $Z_i^{(j)} = z_i^{(j)}$ to $Z_{i+1}^{(j)} = z_{i+1}^{(j)}$ is fully explained by the hidden states $x_i^{(j)}$ and $x_{i+1}^{(j)}$. Namely

$$\Pr[E_n] = \sum_{(x_i', z_i') \in \mathcal{C}} P_{X_i', Z_i' | X_{i-1}', Z_{i-1}'} \left(x_i', z_i' | x_{i-1}', z_{i-1}'\right) \quad (4)$$

summation over tuples x_i', z_i' satisfying *collision restrictions*

$$\mathcal{C} = \{x^1, z^1 \ldots x^{\alpha}, z^{\alpha} \in (\mathcal{X} \times \mathcal{Z})^{\alpha} : z^1 = \ldots = z^{\alpha}\}. \quad (5)$$

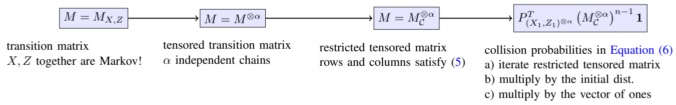

$M = M_{X,Z}$
$M = M^{\otimes\alpha}$
$M = M_{\mathcal{C}}^{\otimes\alpha}$
$P_{(X_1,Z_1)^{\otimes\alpha}}^T \left(M_{\mathcal{C}}^{\otimes\alpha}\right)^{n-1} \mathbf{1}$
transition matrix
X, Z together are Markov!
tensored transition matrix
α independent chains
restricted tensored matrix
rows and columns satisfy (5)
collision probabilities in Equation (6)
a) iterate restricted tensored matrix
b) multiply by the initial dist.
c) multiply by the vector of ones

Fig. 1: Our framework for computing integer-order Renyi entropies.

Each distribution $P_{X_{i+1}', Z_{i+1}' | X_i', Z_i'}$ in Equation (4) is given by a fixed matrix, namely the transition matrix of the process $\{X_i', Z_i'\}_i$ which is Markov (once we have revealed hidden states, the proof is simple and appears in the full version). Denoting the transition matrix of $\{X_i, Z_i\}_i$ by M, we can find the matrix of $\{X_i', Z_i'\}_i$ as the α-fold *Kronecker tensor product* of M, denoted by $M^{\otimes\alpha}$, which is a matrix with rows and columns in the α-fold Cartesian product of $\mathcal{X} \times \mathcal{Z}$; we call it the *tensored matrix* of M. Let $M_{\mathcal{C}}^{\otimes\alpha}$ be its submatrix restricted to rows and columns satisfying the restriction $\mathcal{C}$ in Equation (5), referred to as the *restricted tensored matrix*, and let $\left(P_{X_1', Z_1'}\right)_{\mathcal{C}}$ be the restriction of the probability vector of X', Z' to indices from $\mathcal{C}$.

Equation (4) is then expressed in the following compact form (multiplications are understood as matrix/vector multiplications)

Theorem 1 (Renyi entropy in finite-length regimes)**.** *Renyi entropy of finite-length realizations of Z can be computed with powers of the restricted tensored product of M as follows*

$$H_{\alpha}(Z_1^n) = \frac{1}{1-\alpha} \log \left(\left(P_{X_1', Z_1'}\right)_{\mathcal{C}}^T \cdot \left(M_{\mathcal{C}}^{\otimes\alpha}\right)^{n-1} \cdot \mathbf{1} \right) \quad (6)$$

See Section III-A for the proof.

Remark 1 (Explicitly computing the base matrix)**.** *The matrix M can be computed from emission and transition probabilities, respectively $p(z_i|x_i)$ and $p(x_i|x_{i-1})$.*

Figure 1 illustrates how to compute explicit entropies.

2) Explicit Renyi Entropy Rates for HMMs Without Positivity: To approximate the iterated powers of non-negative matrices in (6) we can use classical Perron-Frobenius theory. A drawback is, however, that this requires *positivity assumptions* on the matrix – for example, that it has a strictly positive power. Phrased in terms of the stochastic process, this means that results would suffer from being only applicable to to matrices with *irreducible and aperiodic* supporting graphs. In principle, one can decompose the matrix into components obeying positivity assumptions (for example the canonical decomposition into irreducible parts), and apply the Perron-Frobenius theory separately. However, handling periodicity or even merging results from individual components is not immediate. One such counter-intuitive case is discussed in Section I-D2.

We give an elegant solution to a more general problem – the growth of certain *pseudonorms* of matrix powers. Namely, for any non-negative matrix A and a non-negative vector u we determine the rate of growth of $u^T \cdot A^n \cdot \mathbf{1}$ with n (which in particular fits Equation (6)). The mapping $B \to u^T \cdot |B| \cdot \mathbf{1}$, where $|B|$ is the element-wise application of the absolute value, is a weighted sum of absolute elements in B with non-negative coefficients and thus a pseudonorm. Our problem reduces to estimating how powers of A grow under this pseudonorm, which we handle by using the *spectral formula*. This result, stated later in Lemma 1 and of independent interest, allows us to compute the rate of any hidden Markov process.

Theorem 2 (Renyi entropy rates)**.** *Let ρ_i for $i \in I$ be the spectral radius of matrices corresponding to the irreducible components of $M_{\mathcal{C}}^{\otimes\alpha}$. Then the entropy rate is given by*

$$H_\alpha(Z) = \frac{1}{1-\alpha} \log\left(\max_{i \in I^+} \rho_i\right)$$

where the set I^+ of "reachable" components is defined as all components that can be reached in the associated graph of the matrix $M_{\mathcal{C}}^{\otimes\alpha}$ from tuples in $\mathcal{C}$ having positive probability under the initial distribution $P_{X_1',Z_1'}$.

See Section III-B for the proof.

Remark 2 (Positivity assumptions removed)**.** *Note that the result depends on the initial distribution and the dominant eigenvalue, but the matrix can be arbitrary.*

Remark 3 (Note on computational aspects)**.** *For bounded α evaluating the rate is polynomial in the alphabet size; moreover the matrix size can be drastically reduced by constraints (e.g. for deterministic leakages or no hidden states). One might further use sparsity and entropy interpolation. These aspects are beyond the scope of this paper.*

3) Key Lemma: Growth of Non-negative Matrix Powers: Below we abstract our main technical ingredient: the lemma giving the growth rate of matrix powers under certain "pseudonorms". Since more limited results of this form have found applications in theory of random matrices [3] and previous works on entropy of HMMs [12, 26] we believe it to be of independent interest.

Lemma 1 (Weighted element sum of matrix powers)**.** *Let A be a non-negative matrix of size $m \times m$ and u be a non-negative vector of length m. Let I^+ contain all $i \in \{1, \ldots, m\}$ with*

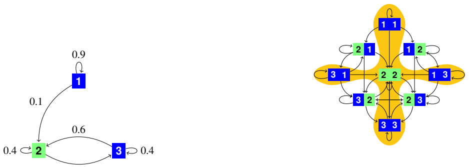
$$u^T \sum_{k=0}^{\infty} A^k \mathbf{e}_i > 0. \tag{7}$$

If A^+ is the submatrix of A with rows and columns I^+ then

$$\lim_{n\to\infty} \left(u^T A^n \mathbf{1}\right)^{\frac{1}{n}} = \rho(A^+). \tag{8}$$

Remark 4 (Simplification by the associated graph and irreducible decomposition)**.** *The description of A^+ can be simplified slightly by using the associated graph and the irreducible components. See the proof of Theorem 2 for details.*

The proof, which appears in the full version, combines the canonical decomposition, sandwiching argument and Gelfand's formula. A very special case when A is irreducible and u is positive follows from [22], and the case of any matrix A and positive u appears in [3]. Our result for non-negative weights is more general.



(a) The original markov chain.

(b) The second tensor. Colliding states $\mathcal{C}$ are highlighted.

Fig. 2: A three state hidden markov chain, the color is observable.

D. Examples and Applications

1) Example: Tensoring and Restricting Step by Step: Consider the Markov chain X in Figure 2. Its hidden states are $\{1, 2, 3\}$, of which the observer cannot distinguish state 1 and 3, as indicated by color. Let the starting distribution X_1 be $\frac{1}{3}$ on every state and the transition matrix

$$M = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0 & 0.6 & 0.4 \end{bmatrix}.$$

To calculate its collision entropy, we take the second tensor, and obtain the graph on the right of the picture. Intuitively, this models two independent copies of X. The transition matrix on the states $\{11, 12, 13, 21, 22, 23, 31, 32, 33\}$ is

$$M^{\otimes 2} = \begin{bmatrix} 0.81 & 0.09 & & 0.09 & 0.01 & & & & \\ & 0.36 & 0.54 & & 0.04 & 0.06 & & & \\ & 0.54 & 0.36 & & 0.06 & 0.04 & & & \\ & & & 0.36 & 0.04 & & 0.54 & 0.06 & \\ & & & & 0.16 & 0.24 & & 0.24 & 0.36 \\ & & & & 0.24 & 0.16 & & 0.36 & 0.24 \\ & & & 0.54 & 0.06 & & 0.36 & 0.04 & \\ & & & & 0.24 & 0.36 & & 0.16 & 0.24 \\ & & & & 0.36 & 0.24 & & 0.24 & 0.16 \end{bmatrix}.$$

But not all of these states are in $\mathcal{C} = \{11, 13, 31, 33, 22\}$, so we consider the restriction to $\mathcal{C}$

$$M_{\mathcal{C}}^{\otimes 2} = \begin{bmatrix} 0.81 & & 0.01 & & \\ & 0.36 & 0.06 & & \\ & & 0.16 & & 0.36 \\ & & 0.06 & 0.36 & \\ & & 0.36 & & 0.16 \end{bmatrix}.$$

The intuition here is that we only care about executions where the observed behaviors of the two independent copies of X are indistinguishable.

To investigate the asymptotic behavior, we find that largest eigenvalues of two irreducible components are $\rho_1 = 0.52$ and $\rho_2 = 0.81$. Intuitively, the eigenvectors $e_1 = (0, 0, 1, 0, 1)$ and $e_2 = (0.96666, 0, 0.02145, 0, 0.01188)$ describe a distribution that is stationary under the condition that we continue to observe the same output; the eigenvalue is the probability of continued collision. Since ρ_2 is the larger of the two, it is the asymptotically relevant, and according to Theorem 3 we have $H_2(X) = -\log(\rho_2) = 0.304$.

2) Example: (Lack of) Relation to Stationary Distribution and Recurrent States: Consider again the transition matrix M from the previous example, with all states being visible ($Z = X$). The Markov chain converges to a stationary distribution where state 1 has probability zero. This state is not recurrent: with probability 1, the chain hits it only a finite number of times because $\sum_k M_{11}^k < \infty$ (test for recurrent states).

Intuitively, such a state should be negligible in the asymptotic entropy analysis. However the opposite happens: for Renyi entropy of order $\alpha = 2$ and any finite length n the first state contributes most to the matrix powering (0.81^n as opposed to λ^n where $|\lambda| < 0.81$ contributed by the component formed by the second and third state). Thus in the asymptotic setting the entropy depends only on the first state.

Corollary 1. *Renyi entropy rate* is not a function of the stationary distribution. *It depends even on non-recurrent states.*

3) Noiseless Observations: We consider the hidden Markov model where the state chain is observed through *noiseless measurements*. More precisely, for a deterministic mapping $T : \mathcal{X} \rightarrow \mathcal{Z}$ the observed (hidden) chain Z_i is given by mapping the base Markov chain: $Z_i = T(X_i)$. While this is less general than our result in Theorem 2, this particular case leads to a *very sparse tensored matrix* and simpler formula for the entropy rate (independent on the dimension of Z).

Theorem 3 (More compact formula for noisyless case)**.** *Let X_i be as above with the transition matrix M. Let $M_{\mathcal{C}}^{\otimes\alpha}$ be the α-fold Kronecker tensor product of M restricted to the tuples of indices $s = (s_1, \ldots, s_\alpha)$ such that $T(s_1) = T(s_2) = \ldots = T(s_\alpha)$. Then for any integer $\alpha > 1$:*

- *the entropy rate of Z_i under Renyi entropy of order α is given by Theorem 2 applied to $M_{\mathcal{C}}^{\otimes\alpha}$ as above and the initial distribution being $P_{X_1^{\otimes\alpha}}$ (α-fold product of X_1).*
- *if M is irreducible and aperiodic, the entropy rate is*

$$H_\alpha(\{Z_i\}_i) = \frac{1}{\alpha - 1} \cdot \rho(M_{\mathcal{C}}^{\otimes\alpha})$$

where $\rho(\cdot)$ denotes the spectral radius.

The proof appears in the full version.

4) Modeling Side Channel Leakage: The motivation and first application for this work was the theoretical analysis of a side-channel attack against RSA encryption [6]. By observing memory access timing, the attacker gains knowledge about the instructions the victim's encryption program is executing. In this particular case, while the victim performs the modular exponentiation necessary for encryption using a sliding-window square-and-multiply algorithm, the attacker learns the sequence of squares and multiplies performed.

The attacker tries to recover the secret key using an established *search-and-prune* technique [10], which is practical if the size of this search tree is linear in the size of the secret key. The contribution of [6] was a more aggressive pruning strategy which, empirically, made attacks against 2048 bit RSA feasible. But why was this attack so effective?

It turns out that the formula that bounds the size of the search tree from above depends directly on the collision entropy of the stochastic process that models the leaked observations (assuming a random and uniformly distributed key). This make intuitive sense, as the tree is pruned if two independent copies of this process (the real one and the guessed one) no longer collide. Concretely, the tree size is linear in the key size if the entropy rate is $H > 0.5$. By modelling the states of the square-and-multiply algorithm as a Markov chain, and modeling the observation as a HMM, we can apply Theorem 3 of the present paper, obtain $H = 0.545$ and thus gain a rigorous understanding of why this attacks works so well.

5) Entropy Rates for Finite Markov Chains: If T in Theorem 3 is a one-to-one mapping, and X_i is aperiodic and irreducible, then the rate equals $\frac{1}{1-\alpha}\log\rho(M^{\diamond\alpha})$ where $M^{\diamond\alpha}$ is the α-fold Hadamard product. This reproves the formula for Markov chains with no hidden states [26]. Illustrative example and details appear in the full version.

In fact, the more general part of Theorem 3 holds with no positivity (aperiodicity and irreducibility) assumptions. Moreover, although it uses the assumption that α is integer, for this case we can use directly Lemma 1 in the analysis [26], instead of Perron-Frobenius theory. We thus extend the classical result to possibly periodic and reducible Markov chains

Corollary 2 (Renyi entropy for any Markov chain)**.** *Let $\alpha \neq 1$ be a positive real number. Let Z_i be a finite-alphabet Markov chain and M its transition matrix. Let ρ_i for $i \in I$ be the spectral radius of all irreducible components of $M^{\diamond\alpha}$. Then the entropy rate is given by*

$$H_\alpha(Z) = \frac{1}{1-\alpha}\log\left(\max_{i\in I^+}\rho_i\right)$$

where the maximum is over all 'positive' components I^+ that are assigned positive mass under the initial distribution X_1.

6) Binary Markov chains and Bernoulli Noise: If the chain outcomes are flipped with probability ϵ by a noisy channel, the rate (for order $\alpha > 1$) changes by an $O(\epsilon)$ term. The exact expression up to $O(\epsilon^2)$ has been studied in [12]. We provide an alternative characterization of the entropy rate, as the root of an *explicit polynomial of degree 8*. In particular, for any ϵ we can compute the exact value numerically, without asymptotic expressions. Here we assume that the transition matrix is positive, to apply perturbation theory. The detailed discussion appears in the full version

E. Algebraic Equations for the Entropy Rate

Our results imply that the Renyi entropy rate of integer order $\alpha > 1$ is characterized by an algebraic equation.

Corollary 3. *The Renyi entropy rate of integer order $\alpha > 1$ of a hidden Markov process (the base chain over a finite alphabet) is the absolute value of a root of an explicit polynomial.*

This is interesting when compared to the Shannon entropy rate, which can be characterized by a more complicated functional equation [18]. We can thus derive exact expansions, approximations or study perturbations (e.g. due to noise).

F. Evaluating Security of True Random Number Generators

As per modern paradigms in the design of harwadre random number generators [4], the security is evaluated by developing a model for the stochastic source and evaluating the entropy in the outcome. An appealing model for the raw source is a Markov chai [4]. however as the output gets post-processed (condensers or extractors that increase the entropy rate of a raw source) we technically work with hidden chains. In this context our result may be used to explicitly determine the entropy rate.

II. Preliminaries and notational conventions

For a process $Z = Z_1, Z_2, \ldots$ we define the finite realization of length n as $Z_1^n = Z_1, \ldots, Z_n$. To simplify the notation, we use the standard convention that probabilities involving events of the form $A_i = a_i, B_i = b_i$ are written with capital symbols omitted, that is $P(A_i = a_i) = p(a_i)$, $P(A_i = a_i | B_i = b_i) = p(a_i | b_i)$ and so on. We identify the probability distribution of a random variable S with values in (finite) $\mathcal{S}$ with the vector with coordinates indexed by $\mathcal{S}$. For any vector μ or matrix A indexed by $\mathcal{S}$, by $\mu_{\mathcal{S}'}$ respectively $A_{\mathcal{S}'}$ we understand restrictions to indices from $\mathcal{S}'$ where $\mathcal{S}' \subseteq \mathcal{S}$. Single vectors are understood as columns; y^T denotes the transposition of a vector or matrix y. All logarithms are taken at base 2.

Definition 1 (Associated graph of a matrix). *For a non-negative matrix A its* associated graph *is the directed graph with all matrix indices $1, \ldots, m$ as nodes, and edges $i \to j$ iff $A_{i,j} > 0$.*

Definition 2 (Renyi Entropy [28]). *The Renyi entropy of order α of a discrete random variable Z is defined as*

$$H_\alpha(Z) = \frac{1}{1-\alpha} \log \sum_z P_Z(z)^\alpha$$

The limit $\alpha = 1$ is Shannon entropy $H_1(Z) = -\sum_z P_Z(z) \log P_Z(z)$ and min-entropy $H_\infty(Z) = \min_z \log(1/P_Z(z))$.

Definition 3 (Entropy Rate). *The Renyi entropy rate of order α of a discrete process $Z = \{Z_i\}_{i \geqslant 1}$ is defined as*

$$H_\alpha(Z) = \lim_{n \to \infty} \frac{1}{n} H_\alpha(Z_1, \ldots, Z_n)$$

Definition 4 (Kronnecker Tensor Product). *The Kronecker product of two square matrices A, B over $\mathcal{X} \times \mathcal{X}$, is a matrix $A \otimes B = C$ with entries $C((i,j),(i',j')) = A(i,j) \cdot B(i',j')$.*

The α-fold tensor product of a matrix A is denoted by A^α. Sometimes for shortness we will also denote by $P_{Y^{\otimes\alpha}}$ the joint distribution of α-independent copies of a distribution P_Y.

Definition 5 (Hidden Markov Model). *The hidden Markov model consists of the base (hidden) chain X_i and observations Z_i, for $i = 1, 2, \ldots$ such that*

1) $P_{X_i|X_{i-1},\ldots,X_1} = P_{X_i|X_{i-1}}$ *(Markov assumption)*
2) $P_{Z_i|X_1,Z_1\ldots X_T,Z_T} = P_{Z_i|X_i}$ *(Output independence)*

and the transition $P_{X_i|X_{i-1}}$ and emission $P_{Z_i|X_i}$ probabilities do not change with time i.

III. Main Results

A. Proof of Theorem 1

Proof. For convenience, we assume that the processes X and Z are indexed starting from $i = 0$. For every $z \in \mathcal{Z}$ we have

$$\begin{aligned} P_{Z^n}(z)^\alpha &= p(z_0)^\alpha \prod_{i=1}^n p(z_i|z_{i-1})^\alpha \\ &= \left(\sum_{x_0,\ldots,x_n} p(z_0, x_0) \prod_{i=1}^n p(z_i, x_i | z_{i-1}, x_{i-1}) \right)^\alpha. \end{aligned}$$

Defining, as in Equation (5), the set of colliding states

$\mathcal{C} = \{(x^1, z^1, \ldots, x^{(\alpha)}, z^{(\alpha)}) \in (\mathcal{X} \times \mathcal{Z})^\alpha : z^{(1)} = \ldots = z^{(\alpha)}\}$

we can write $p = \sum_z P_{Z^n}(z)^\alpha$ as

$$p = \sum_{(x_i,z_i)\in\mathcal{C}} \prod_{i=1}^n \prod_{j=1}^\alpha p(z_i^{(j)}, x_i^{(j)} | z_{i-1}^{(j)}, x_{i-1}^{(j)}) \prod_{j=1}^\alpha p(z_0^{(j)}, x_0^{(j)})$$

where $x_i = (x_i^{(1)}, \ldots, x_i^{(\alpha)})$, $z_i = (z_i^{(1)}, \ldots, z_i^{(\alpha)})$, and for brevity we use the isomorphism $(\mathcal{X} \times \mathcal{Z})^\alpha \cong \mathcal{X}^\alpha \times \mathcal{Z}^\alpha$

$$(x_i^{(1)}, z_i^{(1)}, \ldots, x_i^{(\alpha)}, z_i^{(\alpha)}) \cong ((x_i^{(1)}, \ldots, x_i^{(\alpha)}), (z_i^{(1)}, \ldots, z_i^{(\alpha)})) = (x_i, z_i).$$

This can be further simplified: let $M_{\mathcal{C}}^{\otimes\alpha}$ be the α-fold tensor product of the matrix $M = p((z,x)|(z',x'))$ restricted to $\mathcal{C}$, μ be the vector with elements $\prod_{j=1}^\alpha p(x_0^j, z_0^j)$ over all choices $x_0, z_0 \in \mathcal{C}$, and $\mathbf{1}$ be the vector of ones indexed by $\mathcal{C}$. Then

$$\sum_z P_{Z^n}(z)^\alpha = \mu^T \cdot \left(M_{\mathcal{C}}^{\otimes\alpha}\right)^n \cdot \mathbf{1}$$

which, together with Definition 2, implies Equation (6) (there we go back to the original numeration starting from $i = 1$). □

B. Proof of Theorem 2

Proof. By Theorem 1 we obtain

$$\lim_{n\to\infty} \frac{H_\alpha(Z_1^n)}{n} = \frac{1}{1-\alpha} \log \left(\lim_{n\to\infty} \left(\left(P_{X_1',Z_1'}\right)_{\mathcal{C}}^T \cdot \left(M_{\mathcal{C}}^{\otimes\alpha}\right)^{n-1} \cdot \mathbf{1} \right)^{1/n} \right)$$

The result will follow now from Lemma 1 applied with $A = M_{\mathcal{C}}^{\otimes\alpha}$, once we prove the following about "positive indices".

Claim 1. *Let A be a non-negative matrix and u be a non-negative vector. Let $I_1, \ldots, I_d$ be the subsets of indices corresponding to d irreducible classes in the canonical decomposition of A. Then $\sum_k uA^k e_i > 0$ if and only if the associated graph of A connects i with some j such that $u_j > 0$ (in a finite number of steps).*

Proof of Claim. This follows immediately from the properties of the adjacency matrix, namely $A_{i,j}^k > 0$ if there is a path from i to j of length k. □

Translated to the setting of Theorem 2, the claim implies that we consider only irreducible components reachable from points with positive measure under $(P_{X_1',Z_1'})_{\mathcal{C}}$ (these points are tuples in $\mathcal{C}$ with positive initial probaiblity $P_{X_1',Z_1'}$). □

IV. Conclusion

We have analytically omputed Renyi entropies of hidden Markov processes, when the entropy order is an integer bigger than 1. The main technical contribution is a result on *pseudonorms of iterated matrices*, that allow us (for the first time) to *get rid of positivity assumptions*. Some problems we leave for future work are:

- Analytical tractability of rates for non-integer α?
- Rates of smooth Renyi entropy?
- Seed of the convergence towards the entropy rate.
- Use of perturbation theory to handle small leakages.

V. Acknowledgements

This work was supported in part by the Luxembourg National Research Fund under grant no. C17/IS/11613923.

References

[1] J. Acharya, A. Orlitsky, A. T. Suresh, and H. Tyagi. "The Complexity of Estimating Rényi Entropy". In: *SODA 2015*. 2015, pp. 1855–1869.

[2] E. Arikan. "An inequality on guessing and its application to sequential decoding". In: *IEEE Trans. Information Theory* 42.1 (1996), pp. 99–105.

[3] D. Bajović, J. Xavier, and B. Sinopoli. "Products of stochastic matrices: large deviation rate for Markov chain temporal dependencies". In: *Annual Allerton Conference on Communication, Control, and Computing*. 2012, pp. 724–729.

[4] E. Barker and J. Kelsey. "Recommendation for the entropy sources used for random bit generation". In: *Draft NIST Special Publication* (2012), pp. 800–900.

[5] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer. "Generalized privacy amplification". In: *IEEE Trans. Information Theory* 41.6 (1995), pp. 1915–1923.

[6] D. J. Bernstein, J. Breitner, D. Genkin, L. G. Bruinderink, N. Heninger, T. Lange, C. van Vredendaal, and Y. Yarom. "Sliding right into disaster: Left-to-right sliding windows leak". In: *CHES*. Springer. 2017, pp. 555–576.

[7] I. Csiszár. "Generalized cutoff rates and Renyi's information measures". In: *IEEE Trans. Information Theory* 41.1 (1995), pp. 26–34.

[8] M. K. Hanawal and R. Sundaresan. "Guessing Revisited: A Large Deviations Approach". In: *IEEE Trans. Information Theory* 57.1 (2011), pp. 70–78.

[9] P. E. Hart. "Moment Distributions in Economics: An Exposition". In: *Journal of the Royal Statistical Society. Series A (General)* 138.3 (1975), pp. 423–434.

[10] N. Heninger and H. Shacham. "Reconstructing RSA Private Keys from Random Key Bits". In: *CRYPTO 2009*. Springer, 2009, pp. 1–17.

[11] R. Impagliazzo and D. Zuckerman. "How to recycle random bits". In: *30th Annual Symposium on Foundations of Computer Science*. IEEE. 1989, pp. 248–253.

[12] P. Jacquet, G. Seroussi, and W. Szpankowski. "On the entropy of a hidden Markov process". In: *Theor. Comput. Sci.* 395.2-3 (2008), pp. 203–219.

[13] R. Jenssen, K. E. Hild, D. Erdogmus, J. C. Principe, and T. Eltoft. "Clustering using Renyi's entropy". In: *International Joint Conference on Neural Networks*. Vol. 1. 2003, 523–528 vol.1.

[14] P. Jizba, H. Kleinert, and M. Shefaat. "Rényi's information transfer between financial time series". In: *Physica A: Statistical Mechanics and its Applications* 391.10 (2012), pp. 2971–2989.

[15] S. Kamath and S. Verdú. "Estimation of entropy rate and Rényi entropy rate for Markov chains". In: *ISIT*. 2016, pp. 685–689.

[16] D. E. Knuth. *The Art of Computer Programming: Sorting and Searching*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc., 1998.

[17] K. Li, W. Zhou, S. Yu, and B. Dai. "Effective DDoS attacks detection using generalized entropy metric". In: *International Conference on Algorithms and Architectures for Parallel Processing*. Springer. 2009, pp. 266–280.

[18] J. Luo and D. Guo. "On the Entropy Rate of Hidden Markov Processes Observed Through Arbitrary Memoryless Channels". In: *IEEE Trans. Information Theory* 55.4 (2009), pp. 1460–1467.

[19] B. Ma, A. Hero, J. Gorman, and O. Michel. "Image registration with minimum spanning tree algorithm". In: *International Conference on Image Processing*. Vol. 1. 2000, pp. 481–484.

[20] C. R. MacCluer. "The Many Proofs and Applications of Perron's Theorem." In: *SIAM Review* 42.3 (2000), pp. 487–498.

[21] Y. Mansour, M. Mohri, and A. Rostamizadeh. "Multiple source adaptation and the Rényi divergence". In: *Conference on Uncertainty in Artificial Intelligence*. 2009, pp. 367–374.

[22] M. Marcus and M. Newman. "The sum of the elements of the powers of a matrix." In: *Pacific J. Math.* 12.2 (1962), pp. 627–635.

[23] H. Neemuchwala, A. O. H. III, S. Zabuawala, and P. L. Carson. "Image registration methods in high-dimensional space". In: *Int. J. Imaging Systems and Technology* 16.5 (2006), pp. 130–145.

[24] L. Paninski. "Estimation of Entropy and Mutual Information". In: *Neural Comput.* 15.6 (June 2003), pp. 1191–1253.

[25] C. E. Pfister and W. G. Sullivan. "Rényi Entropy, Guesswork Moments, and Large Deviations". In: *IEEE Trans. Information Theory* 50.11 (2004), pp. 2794–2800.

[26] Z. Rached, F. Alajaji, and L. L. Campbell. "Rényi's divergence and entropy rates for finite alphabet Markov sources". In: *IEEE Trans. Information Theory* 47.4 (2001), pp. 1553–1561.

[27] R. Renner and S. Wolf. "Smooth Renyi entropy and applications". In: *International Symposium onInformation Theory*. June 2004, pp. 233–.

[28] A. Rényi. "On measures of information and entropy". In: *Berkeley Symposium on Mathematics, Statistics and Probability*. 1960, pp. 547–561.

[29] P. K. Sahoo and G. Arora. "A thresholding method based on two-dimensional Renyi's entropy". In: *Pattern Recognition* 37.6 (2004), pp. 1149–1161.

[30] P. C. van Oorschot and M. J. Wiener. "Parallel Collision Search with Cryptanalytic Applications". In: *J. Cryptology* 12.1 (1999), pp. 1–28.

[31] C. Wu, E. L. Xu, and G. Han. "Renyi entropy rate of hidden Markov processes". In: *IEEE International Symposium on Information Theory (ISIT)*. 2017, pp. 2970–2974.

[32] D. Xu. "Energy, Entropy and Information Potential for Neural Computation". AAI9935317. PhD thesis. Gainesville, FL, USA, 1998.