Maximum Rényi Entropy Rate

Christoph Bunte and Amos Lapidoth, *Fellow, IEEE*

***Abstract*— The supremum of the Rényi entropy rate over the class of discrete-time stationary stochastic processes, whose marginals are supported by some given set and satisfy some given cost constraint, is computed. Unlike the Shannon entropy, the Rényi entropy of a random vector can exceed the sum of the Rényi entropies of its components, and the supremum is, therefore, typically not achieved by memoryless processes. It is nonetheless related to Shannon's entropy: when the Rényi parameter exceeds one, the supremum is equal to the corresponding supremum of Shannon's entropy, and when it is smaller than one, the supremum equals the logarithm of the volume of the support set. A Burg-like supremum of the Rényi entropy rate over the class of stochastic processes, whose autocovariance function begins with some given values, is also solved. It is not achieved by Gauss–Markov processes, but it is nonetheless related to Burg's supremum: the two are equal when the Rényi parameter exceeds one, and the former is infinite otherwise.**

***Index Terms*— Burg's Theorem, entropy rate, maximization, Rényi entropy, Rényi entropy rate, spectrum estimation.**

I. Introduction

The object of our study is the supremum of the Rényi entropy rate over the class of all stationary stochastic processes $\{Z_k\}_{k\in\mathbb{Z}}$ whose marginals Z_k satisfy

$$\Pr[Z_k \in \mathcal{S}] = 1, \quad \mathrm{E}[r(Z_k)] \leq \Gamma, \quad k \in \mathbb{Z}. \tag{1}$$

Here $\mathcal{S}$ is the "support set," the mapping $r : \mathcal{S} \to \mathbb{R}$ is the "cost function," and $\Gamma \in \mathbb{R}$ is the "maximal-allowed average cost." The sets $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ (some of which will appear only later) denote the reals, the integers, and the positive integers. Our interest is in two cases: the "discrete setting" where the support set $\mathcal{S}$ is finite, and the "continuous setting" where $\mathcal{S}$ is a Borel measurable subset of the reals; $r(\cdot)$ is Borel measurable; and we restrict attention to stochastic processes whose finite-dimensional distributions (FDDs) have densities with respect to the corresponding Lebesgue measures.

In the discrete setting the supremum yields worst-case results for the "guessing problem" [1]–[5], for the task-encoding problem [6], and for Campbell's measure of typical code length [7]. In the continuous setting it yields worst-case results on the quantization of densities [8].

We shall focus on the continuous setting, because the results and proofs for this setting can be easily translated to the discrete setting: Replace integrals with sums; probability density functions with probability mass functions, and interpret the "volume" of a set as its cardinality.

If instead of *Rényi* rate we had maximized the *Shannon* rate, we could have limited ourselves to memoryless processes, because the Shannon entropy of a random vector is upper-bounded by the sum of the Shannon entropies of its components, and this upper bound is tight when the components are independent.[1] But this bound does not hold for Rényi entropy: the Rényi entropy of a vector with dependent components can exceed the sum of the Rényi entropies of its components. Consequently, the supremum of the Rényi rate subject to (1) is typically not achieved by memoryless processes. This supremum and the structure of the stochastic processes that approach it is the subject of this paper.

We emphasize that our focus here is on the maximization of Rényi *rate* and not entropy. The latter is studied in [10]–[13]. The Rényi entropy of some specific multivariate densities are computed in [14]. To the best of our knowledge, the maximization of Rényi rate has not been studied before. But the Rényi rate has been computed for some specific stochastic processes: It was computed for finite-state Markov chains by Rached *et al.* [15] with extensions to countable state space in [16]. It was computed for stationary Gaussian processes by Golshani and Pasha in [17]. Extensions are explored in [18].

Another class of stochastic processes that we shall consider is related to Burg's work on spectral estimation [19], [9, Th. 12.6.1]. It comprises all one-sided stochastic processes $\{X_i\}_{i\in\mathbb{N}}$ that, for some given $\alpha_0, \dots, \alpha_p \in \mathbb{R}$, satisfy

$$\mathrm{E}[X_i X_{i+k}] = \alpha_k, \quad \Big(i \in \mathbb{N}, k \in \{0, \dots, p\}\Big). \tag{2}$$

While Burg studied the maximum over this class of the Shannon rate, we will study the maximum of the Rényi rate.

The rest of the paper is organized as follows. Section II contains the statements of our main results along with a discussion and the required definitions. We discuss the constraints (1) on the marginals and the Burg-like constraints (2) separately. The proofs pertaining to the former are in Section IV and to the latter in Section V. Section III derives and collects some of the results we shall need to prove the main results.

II. Main Results and Discussion

To describe our results we need some definitions. Those are presented next, along with the basic bounds that form their context.

Manuscript received January 3, 2015; revised December 31, 2015; accepted January 9, 2016. Date of publication January 25, 2016; date of current version February 12, 2016. This paper was presented at the 2014 Seventh Joint Workshop on Coding and Communications and the 2014 IEEE 28-th Convention of Electrical and Electronics Engineers in Israel.

C. Bunte is with UBS AG, Europastrasse 1, 8152 Opfikon, Switzerland (e-mail: christoph.bunte@ubs.com).

A. Lapidoth is with the Signal and Information Processing Laboratory, ETH Zurich, Zurich CH-8092, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

Communicated by R. Sundaresan, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2016.2521364

[1] Except in the discrete setting, in this paper "Shannon entropy" refers to differential Shannon entropy [9, Ch. 8].

A. Definitions and Basic Bounds

The order-α Rényi entropy $h_\alpha(f)$ of a probability density function (PDF) f is

$$h_\alpha(f) = \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f^\alpha(x)\, \mathrm{d}x, \tag{3}$$

where α—the "Rényi parameter"—can be any positive number other than one. The integrand is nonnegative, so the integral on the RHS of (3) always exists, possibly taking on the value $+\infty$, in which case we define $h_\alpha(f)$ as $+\infty$ if $0 < \alpha < 1$ and as $-\infty$ if $\alpha > 1$. With this convention the Rényi entropy always exists and

$$h_\alpha(f) > -\infty, \quad 0 < \alpha < 1, \tag{4}$$
$$h_\alpha(f) < +\infty, \quad \alpha > 1. \tag{5}$$

When a random variable (RV) X is of density f_X we sometimes write $h_\alpha(X)$ instead of $h_\alpha(f_X)$.

If the support of f is contained in $\mathcal{S}$, then

$$h_\alpha(f) \leq \log|\mathcal{S}|, \quad \Big(\alpha > 0, \alpha \neq 1\Big), \tag{6}$$

where $|\mathcal{A}|$ denotes the Lebesgue measure of the set $\mathcal{A}$, and where we interpret $\log|\mathcal{S}|$ as $+\infty$ when $|\mathcal{S}|$ is infinite.[2] (Throughout this paper we define $\log \infty = \infty$ and $\log 0 = -\infty$.)

The Rényi entropy is closely related to the Shannon entropy:

$$h(f) = -\int_{-\infty}^{\infty} f(x) \log f(x)\, \mathrm{d}x. \tag{7}$$

(The integral on the RHS of (7) need not exist. If it does not, then we say that $h(f)$ does not exist.) Depending on whether α is smaller or larger than one, the Rényi entropy can be larger or smaller than the Shannon entropy. Indeed, if f is of Shannon entropy $h(f)$ (possibly $+\infty$), then by [20, Lemma 5.1 (iv)]:

$$h_\alpha(f) \leq h(f), \quad \text{for } \alpha > 1; \tag{8}$$
$$h_\alpha(f) \geq h(f), \quad \text{for } 0 < \alpha < 1. \tag{9}$$

Moreover, under some mild technical conditions [20, Lemma 5.1 (ii)]:

$$\lim_{\alpha \to 1} h_\alpha(f) = h(f). \tag{10}$$

The order-α Rényi rate $h_\alpha(\{X_k\})$ of a stochastic process (SP) $\{X_k\}$ is defined as

$$h_\alpha(\{X_k\}) = \lim_{n\to\infty} \frac{1}{n} h_\alpha\big(X_1^n\big) \tag{11}$$

whenever the limit exists.[3] Here X_i^j denotes the tuple $(X_i, \ldots, X_j)$.

Notice that if each X_k takes value in $\mathcal{S}$, then X_1^n takes value in $\mathcal{S}^n$, and it then follows from (6) that $h_\alpha(X_1^n) \leq \log|\mathcal{S}|^n$ and thus

$$h_\alpha(\{X_k\}) \leq \log|\mathcal{S}|. \tag{12}$$

[2] In the discrete setting $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$.

[3] We say that the limit exists and is equal to $+\infty$ if for every $\mathsf{M} > 0$ there exists some n_0 such that for all $n > n_0$ the Rényi entropy $h_\alpha(X_1, \ldots, X_n)$ exceeds $n\mathsf{M}$, possibly by being $+\infty$.

Another upper bound on $h_\alpha(\{X_k\})$, one that is valid for $\alpha > 1$, can be obtained by noting that when $\alpha > 1$ we can use (8) to obtain

$$h_\alpha(X_1^n) \leq h(X_1^n) \tag{13}$$
$$\leq \sum_{i=1}^n h(X_i), \tag{14}$$

and thus, by (13),

$$h_\alpha(\{X_k\}) \leq h(\{X_k\}), \quad \alpha > 1, \tag{15}$$

whenever both $h_\alpha(\{X_k\})$ and the Shannon rate $h(\{X_k\})$ exist.

B. Constrained Marginals

To describe our results on the maximum Rényi rate subject to the constraint (1) we need one more definition. We define $h^\star(\Gamma)$ to be the supremum of $h(f_X)$ over all densities f_X under which

$$\Pr[X \in \mathcal{S}] = 1 \quad \text{and} \quad \mathrm{E}[r(X)] \leq \Gamma. \tag{16}$$

Here and throughout the supremum should be interpreted as $-\infty$ if (16) does not hold under any probability distribution that is absolutely continuous with respect to the Lebesgue measure. Thus, if no density satisfies (16), then $h^\star(\Gamma)$ is $-\infty$. We shall assume that $\mathcal{S}$ and $r(\cdot)$ are such that for some $\Gamma_0 \in \mathbb{R}$

$$h^\star(\Gamma_0) > -\infty, \tag{17a}$$

and

$$h^\star(\Gamma) < \infty \quad \text{for every } \Gamma_0 \leq \Gamma < \infty. \tag{17b}$$

For example, if $r(x)$ is x^2 and $\mathcal{S}$ is the reals, then this conditions holds whenever Γ_0 is positive; $h^\star(\Gamma)$ equals

$$\frac{1}{2}\log(2\pi e\Gamma), \quad \Gamma > 0; \tag{18}$$

and $h^\star(\Gamma)$ is achieved by a variance-Γ centered Gaussian $f^\star \sim \mathcal{N}(0, \Gamma)$ [9, Th. 8.6.5].

1) The Case of $\alpha > 1$: For $\alpha > 1$ we note that (11), (14), and the definition of $h^\star(\Gamma)$ imply that for every SP $\{Z_k\}$ satisfying (1)

$$h_\alpha(\{Z_k\}) \leq h^\star(\Gamma), \quad \alpha > 1, \tag{19}$$

and consequently,

$$\sup h_\alpha(\{Z_k\}) \leq h^\star(\Gamma), \quad \alpha > 1, \tag{20}$$

where the supremum is over all SPs satisfying (1). Perhaps surprisingly, this bound is tight:

Theorem 1 (Max Rényi Rate for $\alpha > 1$): Suppose that $\alpha > 1$, and that $\Gamma > \Gamma_0$, where Γ_0 satisfies (17). *Then for every $\tilde{\varepsilon} > 0$ there exists a stationary SP $\{Z_k\}$ satisfying* (1) *whose Rényi rate exists and exceeds $h^\star(\Gamma) - \tilde{\varepsilon}$.*

Proof: See Section IV. □

As the following heuristic argument demonstrates, one has to walk a fine line in order to achieve the supremum promised in Theorem 1. To see why, let us focus on the case where $h^\star(\cdot)$ is strictly increasing and where there exist real constants $\lambda_0, \lambda_1 \in \mathbb{R}$ for which the function $f^\star(x) = \exp\big(\lambda_0 + \lambda_1 r(x)\big)\, \mathrm{I}\{x \in \mathcal{S}\}$ is a density achieving $h^\star(\Gamma)$. For any other density g supported on $\mathcal{S}$ and satisfying

$$\int_{\mathcal{S}} g(x)\, r(x)\, \mathrm{d}x = \Gamma \tag{21}$$

we then have (as in the proof of [9, Th. 12.1.1])

$$h(g) = h(f^\star) - D(g \| f^\star) \tag{22}$$
$$= h^\star(\Gamma) - D(g \| f^\star), \tag{23}$$

where $D(g \| f^\star)$ denotes relative entropy [9, Sec. 8.5].

Using this and (14) we thus obtain that if $\{Z_k\}$ is a stationary SP and if f_Z is the density of Z_1 and

$$\int_{\mathcal{S}} f_Z(x)\, r(x)\, \mathrm{d}x = \Gamma, \tag{24}$$

then

$$h_\alpha(\{Z_k\}) \leq h^\star(\Gamma) - D(f_Z \| f^\star), \quad \alpha > 1. \tag{25}$$

Thus, for $h_\alpha(\{Z_k\})$ to be close to $h^\star(\Gamma)$, the density of Z_1 must be "close" (in relative-entropy) to $f^\star$.[4] We can repeat this argument for the joint density of Z_1, Z_2 to infer that Z_1 and Z_2 must be "nearly independent" with each being of density "nearly" $f^\star$. More generally, for every fixed $m \in \mathbb{N}$ the joint density of $Z_1, \ldots, Z_m$ must be nearly of a product form. But, of course choosing $\{Z_k\}$ IID will not work, because this choice would lead to a Rényi rate equal to $h_\alpha(f_{Z_1})$, which is typically smaller than $h(Z_1)$ (unless Z_1 is uniform); see (8).

2) The Case of $0 < \alpha < 1$: For $0 < \alpha < 1$ we can use (12) to obtain for the same supremum

$$\sup h_\alpha(\{Z_k\}) \leq \log|\mathcal{S}|, \quad 0 < \alpha < 1. \tag{26}$$

This seemingly crude bound is tight:

Theorem 2 (Max Rényi Rate for $0 < \alpha < 1$): Suppose that $0 < \alpha < 1$ and that $\Gamma > \Gamma_0$, where Γ_0 satisfies (17).

- *If $|\mathcal{S}| = \infty$, then for every $\mathsf{M} \in \mathbb{R}$ there exists a stationary SP $\{Z_k\}$ satisfying* (1) *whose Rényi rate exists and exceeds M.*
- *If $|\mathcal{S}| < \infty$, then for every $\tilde{\varepsilon} > 0$ there exists a stationary SP $\{Z_k\}$ satisfying* (1) *whose Rényi rate exists and exceeds $\log|\mathcal{S}| - \tilde{\varepsilon}$.*

Proof: See Section IV. □

Remark 3: Theorems 1 and 2 can be generalized in a straightforward fashion to account for multiple constraints:

$$\mathrm{E}[r_i(Z_k)] \leq \Gamma_i, \quad i = 1, \ldots, m. \tag{27}$$

The proofs require only slight modifications, but to ease the presentation we focus on the case of a single constraint.

3) A Second-Moment Constraint: A special case of Theorems 1 and 2 is when the cost is quadratic, i.e., $r(x) = x^2$ and where there are no restrictions on the support, i.e., $\mathcal{S} = \mathbb{R}$. In this case we can slightly strengthen the results of the above theorems: When we consider the proofs of these theorems for this case, we see that the proposed distributions are radially-symmetric.[5] We can thus establish that the constructed SP is centered and uncorrelated:

Proposition 4 (A Second-Moment Constraint):

1) *For every $\alpha > 1$, every $\sigma > 0$, and every $\tilde{\varepsilon} > 0$ there exists a centered stationary SP $\{Y_k\}$ whose Rényi rate exists and exceeds $\frac{1}{2}\log(2\pi e\sigma^2) - \tilde{\varepsilon}$ and for which*

$$\mathrm{E}[Y_k Y_{k'}] = \sigma^2\, \mathrm{I}\{k = k'\}. \tag{28}$$

(Here $\mathrm{I}\{$*statement*$\}$ *is one when the statement is true and zero otherwise.)*

2) *For every $0 < \alpha < 1$, every $\sigma > 0$, and every $\mathsf{M} \in \mathbb{R}$ there exists a centered stationary SP $\{Y_k\}$ whose Rényi rate exists and exceeds M and for which* (28) *holds.*

This proposition will be the key to the proof of Theorem 5 ahead.

C. A Burg-Like Constraint

We next present our result for the constraint (2) on the values of the autocovariance function at the lags $0, \ldots, p$. Given $\alpha_0, \ldots, \alpha_p \in \mathbb{R}$, we consider the family of all one-sided stochastic processes $X_1, X_2, \ldots$ satisfying (2). We assume that the $(p+1) \times (p+1)$ matrix whose Row-ℓ Column-m element is $\alpha_{|\ell - m|}$ is positive definite.

Theorem 5: Let p be a nonnegative integer, and let the $p+1$ constants $\alpha_0, \ldots, \alpha_p \in \mathbb{R}$ be such that the $(p+1) \times (p+1)$ matrix whose Row-ℓ Column-m element is $\alpha_{|\ell - m|}$ is positive definite. The supremum of the order-α Rényi rate over all one-sided stochastic processes satisfying (2) *is $+\infty$ for $0 < \alpha < 1$ and is equal to the Shannon rate of the p-th order Gauss-Markov process for $\alpha > 1$.*

Proof: See Section V. □

Theorem 5 has bearing on the spectral estimation problem, i.e., the problem of extrapolating the values of the autocovariance sequence from its first $p + 1$ values. One approach is to choose the extrapolated sequence to be the autocovariance sequence of the stochastic process that—among all stochastic processes that have an autocovariance sequence that starts with these $p + 1$ values—maximizes the Shannon rate, namely the p-th order Gauss-Markov process (Burg's theorem).

A different approach might be to choose some $\alpha > 1$ and to replace the maximization of the Shannon rate with that of the order-α Rényi rate. As we next argue, Theorem 5 shows that this would result in the same extrapolated sequence. Indeed, inspecting the proof of the theorem we see that the stochastic process $\{X_i\}$ that we construct, while not a Gauss-Markov process, has the same autocovariance sequence as the p-th order Gauss-Markov process that satisfies the constraints. Moreover, for $\alpha > 1$ the supremum can only be achieved by a stochastic process of this autocovariance sequence: for any other autocovariance function the Rényi rate is upper bounded by the Shannon rate (because $\alpha > 1$), and the latter is upper bounded by the Shannon rate of the Gaussian process,

[4] We are ignoring here the fact that one might consider approaching the supremum with (24) only being an inequality.

[5] For the case at hand $f^\star$ is a centered Gaussian, $\mathcal{S}$ is $\mathbb{R}$, and $r\colon x \mapsto x^2$, so by (37) and (40) $\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)$ is the intersection of two rings and is thus also a ring. Consequently, by (75), f_n is uniform over a ring and hence radially symmetric.

which, unless the autocovariance sequence is that of the p-th order Gauss-Markov process, is strictly smaller than the supremum (Burg's theorem).

As in the proofs of Theorems 1 and 2, the lion's share of the proof of Theorem 5 is dedicated to the construction of stochastic processes that approach the promised suprema. We construct these processes by filtering the stochastic process $\{Y_k\}$ of Corollary 4 with carefully-chosen initial conditions, which allow us to relate the Rényi rate of the filter's output to that of its input. A different approach might have been to further generalize the constraints in (27) so as to allow for constraints such as those of (2), which cannot be expressed as constraints on the marginals. But this would have complicated the weak-typicality arguments and would have made the stationarization more difficult.

III. Preliminaries

A. On the Max Shannon Entropy $h^\star(\Gamma)$

We collect here some of the results on $h^\star(\Gamma)$ that will be needed to prove Theorem 1. These hardly require proof in the discrete setting.

Proposition 6: Let Γ_0 satisfy (17)*. Then over the interval $[\Gamma_0, \infty)$ the function $h^\star(\cdot)$ is finite, nondecreasing, and concave. It is continuous over (Γ_0, ∞), and*

$$\lim_{\Gamma\to\infty} h^\star(\Gamma) = \log|\mathcal{S}|. \tag{29}$$

Proof: Monotonicity is immediate from the definition because increasing Γ enlarges the set of densities that satisfy (16). Concavity follows from the concavity of Shannon entropy, and continuity follows from concavity. It remains to establish (29). To this end we first argue that for every Γ,

$$h^\star(\Gamma) \le \log|\mathcal{S}|. \tag{30}$$

When $|\mathcal{S}|$ is infinite this is trivial, and when $|\mathcal{S}|$ is finite this follows by noting that $h^\star(\Gamma)$ cannot exceed the maximum of the Shannon entropy in the absence of cost constraints, and the latter is achieved by a uniform distribution on $\mathcal{S}$ and is equal to $\log|\mathcal{S}|$. In view of (30), our claim (29) will follow once we establish that

$$\varliminf_{\Gamma\to\infty} h^\star(\Gamma) \ge \log|\mathcal{S}|, \tag{31}$$

which is what we set out to prove next.

We first note that for every $\Gamma \in \mathbb{R}$

$$h^\star(\Gamma) \ge \log|\{x \in \mathcal{S} : r(x) \le \Gamma\}| \tag{32}$$

because when the RHS is finite it can be achieved by a uniform distribution on the set $\{x \in \mathcal{S} : r(x) \le \Gamma\}$, a distribution under which (16) clearly holds, and when it is infinite, it can be approached by uniform distributions on ever-increasing compact subsets of this set. We next note that, by the Monotone Convergence Theorem (MCT),

$$\lim_{\Gamma\to\infty} |\{x \in \mathcal{S} : r(x) \le \Gamma\}| = |\mathcal{S}|. \tag{33}$$

Combining (32) and (33) establishes (31) and hence completes the proof of (29). □

The following proposition demonstrates that $h^\star$ can be approached by bounded densities.

Proposition 7: Suppose that $\Gamma \in (\Gamma_0, \infty)$, where Γ_0 satisfies (17)*. Then for every $\delta > 0$ there exists some bounded density $f^\star$ supported by $\mathcal{S}$ such that*

$$\int f^\star(x) r(x)\, \mathrm{d}x < \Gamma + \delta, \tag{34a}$$

$$h(f^\star) > h^\star(\Gamma) - \delta. \tag{34b}$$

Proof: See the appendix. □

B. Weak Typicality

Given a density f on $\mathcal{S}$ of finite Shannon entropy

$$-\infty < h(f) < \infty, \tag{35}$$

a positive integer n, and some $\varepsilon > 0$, we follow [9, Sec. 8.2] and denote by $\mathcal{T}_n^\varepsilon(f)$ the set of ε-weakly-typical sequences of length n with respect to f:

$$\mathcal{T}_n^\varepsilon(f) = \left\{x_1^n \in \mathcal{S}^n : 2^{-n(h(f)+\varepsilon)} \le \prod_{k=1}^n f(x_k) \le 2^{-n(h(f)-\varepsilon)}\right\}. \tag{36}$$

For example, if f is a centered Gaussian and $\mathcal{S}$ is $\mathbb{R}$, then $\mathcal{T}_n^\varepsilon(f)$ is a "ring"

$$\mathcal{T}_n^\varepsilon(f) = \left\{\mathbf{x} \in \mathbb{R}^n : a \le n^{-1}\|\mathbf{x}\|^2 \le b\right\}, \quad f \sim \mathcal{N}\left(0, \sigma^2\right), \tag{37}$$

where $\mathbf{x}$ stands for x_1^n, and where a and b are determined by the variance of f and by ε.

By the AEP, if $X_1, \ldots, X_n$ are drawn IID according to some such f, then the probability of $(X_1, \ldots, X_n)$ being in $\mathcal{T}_n^\varepsilon(f)$ tends to 1 as $n \to \infty$ (with ε held fixed) [9, Th. 8.2.2].

Given some measurable function $r : \mathcal{S} \to \mathbb{R}$, some density f that is supported on $\mathcal{S}$ and that satisfies

$$\int_{\mathcal{S}} f(x)|r(x)|\, \mathrm{d}x < \infty, \tag{38}$$

and given some $n \in \mathbb{N}$ and $\varepsilon > 0$, we define

$$\mathcal{G}_n^\varepsilon(f) = \left\{x_1^n \in \mathcal{S}^n : \left|\frac{1}{n}\sum_{k=1}^n r(x_k) - \int_{\mathcal{S}} f(x) r(x)\, \mathrm{d}x\right| < \varepsilon\right\}. \tag{39}$$

For example, if $r(x)$ is x^2 and $\mathcal{S}$ is $\mathbb{R}$, then $\mathcal{G}_n^\varepsilon(f)$ is the ring

$$\mathcal{G}_n^\varepsilon(f) = \left\{\mathbf{x} \in \mathbb{R}^n : \left|\frac{1}{n}\|\mathbf{x}\|^2 - \int f(x) x^2\, \mathrm{d}x\right| < \varepsilon\right\}, \quad r : x \mapsto x^2. \tag{40}$$

By the Law of Large Numbers (LLN), if $X_1, \ldots, X_n$ are drawn IID according to some density f that satisfies the above conditions, then the probability of $(X_1, \ldots, X_n)$ being in $\mathcal{G}_n^\varepsilon(f)$ tends to 1 as $n \to \infty$ (with ε held fixed).

From the above observations on $\mathcal{T}_n^\varepsilon(f)$ and $\mathcal{G}_n^\varepsilon(f)$ we conclude that if $X_1, \ldots, X_n$ are drawn IID according to some density f that is supported by $\mathcal{S}$ and that satisfies (35) and (38), then the probability of $(X_1, \ldots, X_n)$ being in the intersection $\mathcal{T}_n^\varepsilon(f) \cap \mathcal{G}_n^\varepsilon(f)$ tends to 1 as $n \to \infty$. Thus, for all sufficiently large n,

$$1-\varepsilon \le \int_{\mathcal{T}_n^\varepsilon(f)\cap\mathcal{G}_n^\varepsilon(f)} \prod_{k=1}^n f(x_k)\,\mathrm{d}x^n \le |\mathcal{T}_n^\varepsilon(f)\cap\mathcal{G}_n^\varepsilon(f)| 2^{-n(h(f)-\varepsilon)},$$

where the second inequality holds by (36). We thus conclude that if the support of f is contained in $\mathcal{S}$, the expectation of $|r(X)|$ under f is finite, and $h(f)$ is defined and is finite, then

$$|\mathcal{T}_n^\varepsilon(f)\cap\mathcal{G}_n^\varepsilon(f)| \ge (1-\varepsilon)2^{n(h(f)-\varepsilon)}, \quad n \text{ large.} \tag{41}$$

Note that if f is a centered Gaussian, $\mathcal{S}$ is $\mathbb{R}$, and $r\colon x \mapsto x^2$, then (37) and (40) imply that $\mathcal{T}_n^\varepsilon(f)\cap\mathcal{G}_n^\varepsilon(f)$ is the intersection of two rings and is thus also a ring.

C. On the Rényi Entropy of Mixtures

To construct stationary processes from random vectors we shall concatenate independent replicas of the vectors and then introduce a random jitter to stationarize the result. To control the behavior of the Rényi entropy under this jitter, we need some results on the Rényi entropy of mixtures. Those are presented here.

The following lemma provides a lower bound on the Rényi entropy of a mixture of densities in terms of the Rényi entropy of the individual densities.

Lemma 8: Let $f_1,\dots,f_p$ be probability density functions on $\mathbb{R}^n$ and let the nonnegative numbers $q_1,\dots,q_p \ge 0$ sum to one. Let f be the mixture density

$$f(\mathbf{x}) = \sum_{\ell=1}^p q_\ell f_\ell(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^n.$$

Then

$$h_\alpha(f) \ge \min_{1\le\ell\le p} h_\alpha(f_\ell).$$

Proof: For $0<\alpha<1$ this follows by the concavity of Rényi entropy. Consider now $\alpha>1$:

$$\begin{aligned}\log\int f^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x} &= \log\int\Bigl(\sum_{\ell=1}^p q_\ell f_\ell(\mathbf{x})\Bigr)^\alpha \mathrm{d}\mathbf{x}\\ &\le \log\int\sum_{\ell=1}^p q_\ell f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &= \log\Bigl(\sum_{\ell=1}^p q_\ell\int f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\Bigr)\\ &\le \log\max_{1\le\ell\le p}\int f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &= \max_{1\le\ell\le p}\log\int f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x},\end{aligned}$$

from which the claim follows because $1/(1-\alpha)$ is negative. Here the first inequality follows from the convexity of the mapping $\xi\mapsto\xi^\alpha$ (for $\alpha>1$), and the second inequality follows by upper-bounding the average by the maximum. □

We next turn to upper bounds.

Lemma 9: Consider the setup of Lemma 8.

1) If $\alpha>1$ then

$$h_\alpha(f) \le \min_{1\le\ell\le p}\Bigl\{\frac{\alpha}{1-\alpha}\log q_\ell + h_\alpha(f_\ell)\Bigr\}. \tag{42}$$

2) If $0<\alpha<1$ then

$$h_\alpha(f) \le \frac{1}{1-\alpha}\log p + \max_{1\le\ell\le p} h_\alpha(f_\ell). \tag{43}$$

Proof: We begin with the case where $\alpha>1$. Since the densities and weights are nonnegative,

$$\Bigl(\sum_{\ell=1}^p q_\ell f_\ell(\mathbf{x})\Bigr)^\alpha \ge \bigl(q_{\ell'} f_{\ell'}(\mathbf{x})\bigr)^\alpha, \quad \ell'\in\{1,\dots,p\}. \tag{44}$$

Integrating this inequality; taking logarithms, and dividing by $1-\alpha$ (which is negative) we obtain

$$h_\alpha(f) \le \frac{\alpha}{1-\alpha}\log q_{\ell'} + h_\alpha(f_{\ell'}), \quad \ell'\in\{1,\dots,p\}. \tag{45}$$

Since this holds for every $\ell'\in\{1,\dots,p\}$, we can minimize over ℓ' to obtain (42).

We next turn to the case where $0<\alpha<1$.

$$\begin{aligned}\log\int\Bigl(\sum_{\ell=1}^p q_\ell f_\ell(\mathbf{x})\Bigr)^\alpha\mathrm{d}\mathbf{x} &\le \log\int\max_{1\le\ell\le p} f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &\le \log\int\sum_{\ell=1}^p f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &= \log\sum_{\ell=1}^p\int f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &\le \log\Bigl(p\max_{1\le\ell\le p}\int f_\ell^\alpha(\mathbf{x})\Bigr)\mathrm{d}\mathbf{x}\\ &= \log p + \log\max_{1\le\ell\le p}\int f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}\\ &= \log p + \max_{1\le\ell\le p}\log\int f_\ell^\alpha(\mathbf{x})\,\mathrm{d}\mathbf{x}.\end{aligned}$$

Dividing this inequality by $1-\alpha$ (positive) yields (43). □

D. Bounded Densities

Proposition 10: If a density f is bounded, and if $\alpha>1$, then $h_\alpha(f)>-\infty$.

Proof: Let f be a density that is upper-bounded by the constant M (which must therefore be positive), and suppose that $\alpha>1$. In this case

$$f^\alpha(x) = f^{\alpha-1}(x)f(x) \le \mathsf{M}^{\alpha-1}f(x),$$

because $\xi\mapsto\xi^{\alpha-1}$ is monotonically increasing when $\alpha>1$. Integrating over x we obtain

$$\int f^\alpha(x)\,\mathrm{d}x \le \mathsf{M}^{\alpha-1} < \infty.$$

Since $\alpha>1$, this implies that

$$\frac{1}{1-\alpha}\log\int_{-\infty}^{\infty} f^\alpha(x)\,\mathrm{d}x > -\infty. \qquad □$$

E. The Marginals of the Uniform Density on $\mathcal{T}_n^\varepsilon(f) \cap \mathcal{G}_n^\varepsilon(f)$

Lemma 11: Let $f^\star$ be a density on $\mathcal{S}$ having finite order-α Rényi entropy

$$h_\alpha(f^\star) > -\infty \tag{46}$$

for some

$$\alpha > 1 \tag{47}$$

and satisfying (35) *and* (38) *(with $f^\star$ substituted for f). For every $n \in \mathbb{N}$, let $(X_1, \ldots, X_n)$ be drawn uniformly from the set $\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)$, where ε is some fixed positive number. Then for every sufficiently large n the following holds: for any $\rho \in \{1, \ldots, n\}$ the ρ-tuple $(X_1, \ldots, X_\rho)$ has finite order-α Rényi entropy*

$$h_\alpha(X_1, \ldots, X_\rho) > -\infty, \quad \Big(\rho \in \{1, \ldots, n\}, \alpha > 1\Big). \tag{48}$$

Proof: Denote the uniform density over $\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)$ by f_n, and let q_n be the product density

$$q_n(\mathbf{x}) = \prod_{k=1}^{n} f^\star(x_k), \quad \mathbf{x} \in \mathcal{S}^n. \tag{49}$$

Henceforth let n be sufficiently large for (41) to hold. Consequently,

$$f_n(\mathbf{x}) \le \frac{1}{1-\varepsilon} 2^{-n(h(f^\star)-\varepsilon)}, \quad \mathbf{x} \in \mathcal{S}^n. \tag{50}$$

Using this inequality and the definition in (36) of $\mathcal{T}_n^\varepsilon(f^\star)$, we can upper-bound f_n in terms of q_n for tuples in $\mathcal{T}_n^\varepsilon(f^\star)$:

$$f_n(\mathbf{x}) \le \frac{1}{1-\varepsilon} 2^{2n\varepsilon} q_n(\mathbf{x}), \quad \mathbf{x} \in \mathcal{T}_n^\varepsilon(f^\star). \tag{51}$$

For every $\rho \in \{1, \ldots, n\}$ we can obtain the density $f_n(x_1, \ldots, x_\rho)$ of $(X_1, \ldots, X_\rho)$ by integrating $f_n(x_1, \ldots, x_n)$ over $x_{\rho+1}, \ldots, x_n$:

$$\begin{aligned}
&f_n(x_1, \ldots, x_\rho)\\
&= \int f_n(\mathbf{x}) \,\mathrm{d}x_{\rho+1} \cdots \mathrm{d}x_n\\
&= \int f_n(\mathbf{x})\, \mathrm{I}\big\{\mathbf{x} \in \mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)\big\} \,\mathrm{d}x_{\rho+1} \cdots \mathrm{d}x_n\\
&\le \frac{1}{1-\varepsilon} 2^{2n\varepsilon} \int q_n(\mathbf{x})\, \mathrm{I}\big\{\mathbf{x} \in \mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)\big\} \,\mathrm{d}x_{\rho+1} \cdots \mathrm{d}x_n\\
&\le \frac{1}{1-\varepsilon} 2^{2n\varepsilon} \int q_n(\mathbf{x}) \,\mathrm{d}x_{\rho+1} \cdots \mathrm{d}x_n\\
&= \frac{1}{1-\varepsilon} 2^{2n\varepsilon} f^\star(x_1) \cdots f^\star(x_\rho), \quad x_1, \ldots, x_\rho \in \mathcal{S},
\end{aligned} \tag{52}$$

where $\mathrm{I}\{\cdot\}$ denotes the indicator function, and the first inequality follows from (51); the second by increasing the range of integration; and the final equality follows from (49).

Using (52) we can now lower-bound $h_\alpha(X_1, \ldots, X_\rho)$ as follows. If a density f is upper-bounded by $\mathsf{K}g$, where g is some other density and K is some positive constant, and if $\alpha > 1$, then

$$\begin{aligned}
h_\alpha(f) &= \frac{1}{1-\alpha} \log \int f^\alpha(\mathbf{x}) \,\mathrm{d}\mathbf{x}\\
&\ge \frac{1}{1-\alpha} \log \int \mathsf{K}^\alpha g^\alpha(\mathbf{x}) \,\mathrm{d}\mathbf{x}\\
&= \frac{\alpha}{1-\alpha} \log \mathsf{K} + h_\alpha(g),
\end{aligned} \tag{53}$$

where the inequality holds because $\alpha > 1$ so the pre-log is negative. Using this and (52) we obtain

$$h_\alpha(X_1, \ldots, X_\rho) \ge \frac{\alpha}{1-\alpha} \log\left(\frac{1}{1-\varepsilon} 2^{2n\varepsilon}\right) + \rho h_\alpha(f^\star) > -\infty.$$

□

F. Stationarization and Rényi Entropy

The following proposition is useful for the construction of a stationary process from a distribution on $\mathbb{R}^n$.

Proposition 12: Let f_n be some density on $\mathcal{S}^n$ having order-α Rényi entropy $h_\alpha(f_n)$ and satisfying

$$\sum_{k=1}^{n} \mathrm{E}[r(X_k)] \le n\Gamma, \quad (X_1, \ldots, X_n) \sim f_n. \tag{54}$$

Then there exists a stationary SP $\{Z_k\}$ satisfying (1) *for which the following holds:*

- *If*
$$h_\alpha(X_1, \ldots, X_\rho), h_\alpha(X_{n-\rho'+1}, \ldots, X_n) > -\infty, \quad \rho, \rho' \in \{1, \ldots, n-1\}, \tag{55}$$
whenever $(X_1, \ldots, X_n) \sim f_n$ and $\rho, \rho' \in \{1, \ldots, n-1\}$, then
$$\varliminf_{m \to \infty} \frac{1}{m} h_\alpha(Z_1, \ldots, Z_m) \ge \frac{1}{n} h_\alpha(f_n). \tag{56}$$
- *If*
$$h_\alpha(X_1, \ldots, X_\rho), h_\alpha(X_{n-\rho'+1}, \ldots, X_n) < +\infty, \quad \rho, \rho' \in \{1, \ldots, n-1\}, \tag{57}$$
whenever $(X_1, \ldots, X_n) \sim f_n$ and $\rho, \rho' \in \{1, \ldots, n-1\}$, then
$$\varlimsup_{m \to \infty} \frac{1}{m} h_\alpha(Z_1, \ldots, Z_m) \le \frac{1}{n} h_\alpha(f_n). \tag{58}$$
- *And if both* (55) *and* (57) *hold, then*
$$\lim_{m \to \infty} \frac{1}{m} h_\alpha(Z_1, \ldots, Z_m) = \frac{1}{n} h_\alpha(f_n). \tag{59}$$

Proof: Consider first the (nonstationary) SP $\{Y_k\}$ that we construct by drawing

$$\ldots, Y_{-n+1}^{0}, Y_1^n, Y_{n+1}^{2n}, \ldots \sim \text{IID } f_n.$$

To stationarize it, let T be drawn uniformly over $\{0, \ldots, n-1\}$ independently of $\{Y_k\}$, and define the stationary SP

$$Z_k = Y_{k+T}, \quad k \in \mathbb{Z}. \tag{60}$$

It satisfies (1). Consider now any m larger than $2n$, and express Z_1^m in one of two different way depending on whether T is zero or not. For $T = 0$

$$Z_1^m = \underbrace{Y_1^n, \ldots, Y_{\tilde{\nu}n-n+1}^{\tilde{\nu}n}}_{\tilde{\nu} = \lfloor m/n \rfloor \text{ } n\text{-tuples}}, \quad \underbrace{Y_{\tilde{\nu}n+1}, \ldots, Y_m}_{\tilde{\rho} = m - n\lfloor m/n \rfloor \text{ terms}} \tag{61}$$

where

$$\tilde{\nu} = \left\lfloor \frac{m}{n} \right\rfloor, \tag{62a}$$

$$\tilde{\rho} = m - n\left\lfloor \frac{m}{n} \right\rfloor \in \{0, \ldots, n-1\}. \tag{62b}$$

And for $T \in \{1, \dots, n-1\}$

$$Z_1^m = \underbrace{Y_{T+1}, \dots, Y_n}_{\rho' = n-T \text{ terms}}, \underbrace{Y_{n+1}^{2n}, \dots, Y_{\nu n+1}^{(\nu+1)n}}_{\nu\ n\text{-tuples}}, \underbrace{Y_{(\nu+1)n+1}, \dots, Y_{m+T}}_{\rho \text{ terms}} \tag{63}$$

where

$$\rho' = n - T \in \{1, \dots, n-1\}, \tag{64a}$$
$$\nu = \left\lfloor \frac{m-n+T}{n} \right\rfloor, \tag{64b}$$
$$\rho = m - n + T - n \left\lfloor \frac{m-n+T}{n} \right\rfloor \in \{0, \dots, n-1\}. \tag{64c}$$

Denote the density of Z_1^m by $f_{\mathbf{Z}}$ and its conditional density given $T = t$ by $f_{\mathbf{Z}|T=t}$.

To establish (56) we use Lemma 8, which implies that

$$h_\alpha\big(f_{\mathbf{Z}}\big) \geq \min_{0 \leq t \leq n-1} h_\alpha\big(f_{\mathbf{Z}|T=t}\big). \tag{65}$$

To compute $h_\alpha\big(f_{\mathbf{Z}|T=0}\big)$ we use (61) to obtain

$$h_\alpha\big(f_{\mathbf{Z}|T=0}\big) = \left\lfloor \frac{m}{n} \right\rfloor h_\alpha(f_n) + h_\alpha(X_1, \dots, X_{\tilde{\rho}}) \tag{66}$$
$$\geq \left\lfloor \frac{m}{n} \right\rfloor h_\alpha(f_n) + 0 \wedge \min_{1 \leq \rho \leq n-1} \big\{h_\alpha(X_1, \dots, X_\rho)\big\}, \tag{67}$$

where the second term on the RHS of (66) should be interpreted as zero when $\tilde{\rho}$ is zero, and where $a \wedge b$ denotes the minimum of a and b.

And to compute $h_\alpha\big(f_{\mathbf{Z}|T=t}\big)$ for $t \in \{1, \dots, n-1\}$ we use (63) to obtain

$$h_\alpha\big(f_{\mathbf{Z}|T=t}\big) = h_\alpha(X_{n-\rho'+1}, \dots, X_n) + \left\lfloor \frac{m-n+t}{n} \right\rfloor h_\alpha(f_n) + h_\alpha(X_1, \dots, X_\rho), \tag{68}$$

where ρ, ρ' are obtained from (64) by substituting t for T, and the last term on the RHS should be interpreted as zero when ρ is zero.

It thus follows from (65), (67), (68), and the above interpretation that

$$h_\alpha\big(f_{\mathbf{Z}}\big) \geq 0 \wedge \min_{1 \leq \rho' \leq n-1} \Big\{h_\alpha(X_{n-\rho'+1}, \dots, X_n)\Big\} + 0 \wedge \min_{1 \leq \rho \leq n-1} \Big\{h_\alpha(X_1, \dots, X_\rho)\Big\} + \min_{1 \leq t \leq n} \left\{ \left\lfloor \frac{m-n+t}{n} \right\rfloor h_\alpha(f_n) \right\}. \tag{69}$$

The first two terms do not depend on m and are greater than $-\infty$ whenever (55) holds. Dividing (69) by m and letting m tend to infinity (with n held fixed), establishes (56).

To establish (58) we need an upper bound on $h_\alpha\big(f_{\mathbf{Z}}\big)$. Such a bound can be obtained from Lemma 9. The exact form of the bound depends on whether α exceeds 1 or not. But either form leads to (58) upon dividing by m and letting it tend to infinity.

To conclude the proof we note that (59) follows from (58) and (56). □

IV. Proofs of Theorems 1 and 2

Proof of Theorem 1: Since $h^\star(\cdot)$ is continuous on the ray (Γ_0, ∞), and since $\Gamma > \Gamma_0$ by the theorem's hypotheses, $h^\star(\cdot)$ is continuous at Γ. Consequently, we can find some Γ' for which

$$\Gamma' < \Gamma \tag{70a}$$
$$h^\star(\Gamma') > h^\star(\Gamma) - \tilde{\varepsilon}. \tag{70b}$$

These inequalities imply that we can find some $\delta > 0$ small enough so that

$$\Gamma' + \delta < \Gamma \tag{71a}$$
$$h^\star(\Gamma') - \delta > h^\star(\Gamma) - \tilde{\varepsilon}. \tag{71b}$$

By Proposition 7, there exists some bounded density $f^\star$ supported by $\mathcal{S}$ such that

$$\int f^\star(x) r(x) \, \mathrm{d}x < \Gamma' + \delta, \tag{72a}$$
$$h(f^\star) > h^\star(\Gamma') - \delta. \tag{72b}$$

Moreover, the boundedness of $f^\star$, the hypothesis that $\alpha > 1$, and Proposition 10 imply that

$$h_\alpha(f^\star) > -\infty. \tag{72c}$$

These inequalities combine with (71) to imply

$$\int f^\star(x) r(x) \, \mathrm{d}x < \Gamma \tag{73a}$$
$$h(f^\star) > h^\star(\Gamma) - \tilde{\varepsilon}. \tag{73b}$$

We can hence choose $\varepsilon > 0$ small enough so that

$$\int f^\star(x) r(x) \, \mathrm{d}x < \Gamma - \varepsilon \tag{74a}$$
$$h(f^\star) > h^\star(\Gamma) - \tilde{\varepsilon} + \varepsilon. \tag{74b}$$

Let f_n be the uniform density over $\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)$

$$f_n \sim \mathrm{Unif}\Big(\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)\Big). \tag{75}$$

The cost of f_n can be bounded by noting that its support is contained in $\mathcal{G}_n^\varepsilon(f^\star)$, and

$$x_1^n \in \mathcal{G}_n^\varepsilon(f^\star) \implies \frac{1}{n} \sum_{k=1}^n r(x_k) < \int f^\star(x) r(x) \, \mathrm{d}x + \varepsilon$$
$$\implies \frac{1}{n} \sum_{k=1}^n r(x_k) < \Gamma,$$

where the second implication follows from (74a). Thus,

$$\int_{\mathcal{S}^n} f_n(\mathbf{x}) \sum_{i=1}^n r(x_i) \, \mathrm{d}\mathbf{x} \leq n\Gamma. \tag{76}$$

To lower-bound its Rényi entropy, we note that by the LLN (in combination with (74a)) and the AEP (see Section III-B)

$$|\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon(f^\star)| \geq (1-\varepsilon) 2^{n(h(f^\star)-\varepsilon)}, \quad n \text{ large}. \tag{77}$$

Consequently,

$$h_\alpha(f_n) \geq n\big(h(f^\star) - \varepsilon\big) + \log(1-\varepsilon) \quad n \text{ large},$$

or, upon dividing by n,

$$\frac{1}{n}h_\alpha(f_n) \geq h(f^\star) - \varepsilon + \frac{1}{n}\log(1-\varepsilon) \tag{78}$$

for all sufficiently large n. We now choose n large enough so that not only will (78) hold but also its RHS satisfy

$$h(f^\star) - \varepsilon + \frac{1}{n}\log(1-\varepsilon) > h^\star(\Gamma) - \tilde{\varepsilon}.$$

(This is possible by (74b).) For this n we thus have

$$\frac{1}{n}h_\alpha(f_n) > h^\star(\Gamma) - \tilde{\varepsilon}. \tag{79}$$

The inequalities (79) and (76) indicate that f_n is a good candidate for the application of Proposition 12. We hence proceed to check its hypotheses.

By Lemma 11 and (72c), if $X_1, \ldots, X_n \sim f_n$ then

$$h_\alpha(X_1, \ldots, X_\rho) > -\infty, \quad \rho \in \{1, \ldots, n-1\}, \tag{80}$$

and, since f_n is permutation invariant, we also infer

$$h_\alpha(X_{n-\rho'+1}, \ldots, X_n) > -\infty, \quad \rho' \in \{1, \ldots, n-1\} \tag{81}$$

so (55) holds. And, since $\alpha > 1$, it follows from (5) that (57) also holds. We can thus apply Proposition 12 to conclude the proof. □

Proof of Theorem 2: We first prove the theorem when $|\mathcal{S}| = \infty$. We distinguish between two cases. The first case, which is the case with which we begin, is when there exists some $n \in \mathbb{N}$ and a density $f_n^\star$ on $X_1, \ldots, X_n$ such that

$$\Pr[X_i \in \mathcal{S}] = 1, \quad \mathrm{E}[r(X_i)] \leq \Gamma, \quad i \in \{1, \ldots, n\} \tag{82}$$

and

$$h_\alpha(X_1, \ldots, X_n) = +\infty. \tag{83}$$

To apply Proposition 12 to this density, we note that, since $0 < \alpha < 1$, Inequality (4) implies (55), and the proposition thus guarantees the existence of a stationary SP $\{Z_k\}$ satisfying (1) and (56) so

$$\lim_{m\to\infty} \frac{1}{m}h_\alpha(Z_1, \ldots, Z_m) = +\infty. \tag{84}$$

This concludes the proof for the case at hand.

We next turn to the second case where $|\mathcal{S}|$ is still infinite, but any tuple whose components satisfy the constraints has Rényi entropy smaller than ∞:

$$\Big(\Pr[X_i \in \mathcal{S}] = 1, \mathrm{E}[r(X_i)] \leq \Gamma, i \in \{\nu_1, \ldots, \nu_2\}\Big) \implies \Big(h_\alpha(X_{\nu_1}, \ldots, X_{\nu_2}) < \infty\Big). \tag{85}$$

Since $|\mathcal{S}|$ is infinite, it follows from Proposition 6 that $h^\star(\Gamma) \to \infty$ as $\Gamma \to \infty$. Consequently, there exists some Γ_1 such that

$$h^\star(\Gamma_1) > \mathsf{M}. \tag{86}$$

Since $h^\star$ is monotonic, there is no loss in generality in assuming, as we shall, that

$$\Gamma_1 > \Gamma. \tag{87}$$

Let $\varepsilon \in (0, 1)$ be small enough so that

$$h^\star(\Gamma_1) > \mathsf{M} + 3\varepsilon \tag{88}$$

$$\Gamma_0 + \varepsilon < \Gamma < \Gamma_1 - \varepsilon. \tag{89}$$

Let the densities $f^{(0)}$ and $f^{(1)}$ be within ε of achieving $h^\star(\Gamma_0)$ and $h^\star(\Gamma_1)$ in the sense that their support is contained in $\mathcal{S}$ and

$$\left(\int_{\mathcal{S}} f^{(\ell)}(x)r(x)\,\mathrm{d}x \leq \Gamma_\ell, h\big(f^{(\ell)}\big) > h^\star(\Gamma_\ell) - \varepsilon\right), \quad \ell \in \{0, 1\}. \tag{90}$$

For every $n \in \mathbb{N}$, define

$$\mathcal{S}_\ell = \mathcal{T}_n^\varepsilon\big(f^{(\ell)}\big) \cap \mathcal{G}_n^\varepsilon(f^{(\ell)}), \quad \ell \in \{0, 1\}. \tag{91}$$

It follows from the LLN and AEP that, for all sufficiently large n,

$$|\mathcal{S}_\ell| \geq (1-\varepsilon)2^{n(h(f^{(\ell)})-\varepsilon)}, \quad \ell \in \{0, 1\}. \tag{92}$$

Assume now that n is large enough for this to hold. Let $\delta > 0$ be small enough so that

$$(1-\delta)(\Gamma_0 + \varepsilon) + \delta(\Gamma_1 + \varepsilon) \leq \Gamma. \tag{93}$$

(Such a δ can be found in view of (89).)

Consider now the mixture density

$$f_n(x_1^n) = (1-\delta)\frac{1}{|\mathcal{S}_0|}\mathrm{I}\{x_1^n \in \mathcal{S}_0\} + \delta\frac{1}{|\mathcal{S}_1|}\mathrm{I}\{x_1^n \in \mathcal{S}_1\}. \tag{94}$$

Let X_1^n be of density f_n. Using (93) and an argument similar to the one leading to (76) we obtain

$$\sum_{k=1}^{n} \mathrm{E}[r(X_k)] \leq n\Gamma. \tag{95}$$

In fact, the permutation invariance of f_n implies the stronger statement

$$\mathrm{E}[r(X_k)] \leq \Gamma, \quad k = 1, \ldots, n. \tag{96}$$

We next lower-bound $h_\alpha(X_1^n)$. To this end, we first argue that the sets $\mathcal{S}_0$ and $\mathcal{S}_1$ are disjoint. To see this, note that by the definition of the sets $\mathcal{G}_n^\varepsilon(f^{(0)})$, $\mathcal{G}_n^\varepsilon(f^{(1)})$ and by (90)

$$x_1^n \in \mathcal{G}_n^\varepsilon(f^{(0)}) \implies \frac{1}{n}\sum_{k=1}^{n} r(x_k) < \int f^{(0)}(x)r(x)\,\mathrm{d}x + \varepsilon \implies \frac{1}{n}\sum_{k=1}^{n} r(x_k) < \Gamma_0 + \varepsilon, \tag{97}$$

and

$$x_1^n \in \mathcal{G}_n^\varepsilon(f^{(1)}) \implies \frac{1}{n}\sum_{k=1}^{n} r(x_k) > \int f^{(1)}(x)r(x)\,\mathrm{d}x - \varepsilon \implies \frac{1}{n}\sum_{k=1}^{n} r(x_k) > \Gamma_1 - \varepsilon. \tag{98}$$

From (89), (97), and (98) we now conclude that $\mathcal{G}_n^\varepsilon(f^{(0)})$ and $\mathcal{G}_n^\varepsilon(f^{(1)})$ are disjoint and hence also $\mathcal{S}_0$ and $\mathcal{S}_1$.

Having established that $\mathcal{S}_0$ and $\mathcal{S}_1$ are disjoint, we can now compute $h_\alpha(f_n)$ directly to obtain:

$$\frac{h_\alpha(X_1^n)}{n} = \frac{1}{n(1-\alpha)}\log\Big((1-\delta)^\alpha|\mathcal{S}_0|^{1-\alpha} + \delta^\alpha|\mathcal{S}_1|^{1-\alpha}\Big) \geq \frac{1}{n(1-\alpha)}\log\Big(\delta^\alpha|\mathcal{S}_1|^{1-\alpha}\Big). \tag{99}$$

From this, (92), (90), and (88) it now follows that we can find some sufficiently large n for which

$$\frac{h_\alpha(X_1^n)}{n} > \mathsf{M}. \tag{100}$$

To apply Proposition 12 we note that (96) and (85) imply that (57) holds. And the fact that $\alpha \in (0,1)$ implies by (4) that (55) holds. Hence, by the proposition, there exists a stationary SP satisfying the constraints and whose Rényi rate is $n^{-1}h_\alpha(X_1^n)$ and thus exceeds M. This concludes the proof when $|\mathcal{S}| = \infty$.

The proof when $|\mathcal{S}| < \infty$ is very similar. In fact, it is a bit simpler because $|\mathcal{S}| < \infty$ implies (85). We begin the proof by noting that, since $|\mathcal{S}| < \infty$, Proposition 6 implies that $h^\star(\Gamma) \to \log|\mathcal{S}|$ as $\Gamma \to \infty$. Consequently, there exists some Γ_1 such that

$$h^\star(\Gamma_1) > \log|\mathcal{S}| - \tilde{\varepsilon}. \tag{101}$$

Replacing M with $\log|\mathcal{S}| - \tilde{\varepsilon}$ in the derivation that leads from (86) to (100), we obtain a density f_n for which

$$\frac{h_\alpha(X_1^n)}{n} > \log|\mathcal{S}| - \tilde{\varepsilon}. \tag{102}$$

The result then follows from Proposition 12 by noting that the LHS of (57) is upper bounded by $n\log|\mathcal{S}|$ and by noting that (55) holds by (4) because $0 < \alpha < 1$. □

V. Proof of Theorem 5

Proof of Theorem 5: Recall the assumption that the $(p+1)\times(p+1)$ matrix whose Row-ℓ Column-m element is $\alpha_{|\ell-m|}$ is positive definite. This implies [21] that there exist constants $a_1,\ldots,a_p,\sigma^2$ and a $p\times p$ positive definite matrix K_p such that the following holds[6]: if the random p-vector $(W_{1-p},\ldots,W_0)$ is of second-moment matrix K_p (not necessarily centered) and if $\{Z_i\}_{i=1}^\infty$ are independent of $(W_{1-p},\ldots,W_0)$ with

$$\mathrm{E}[Z_i] = 0, \quad i \in \mathbb{N}, \tag{103a}$$

$$\mathrm{E}[Z_i Z_j] = \sigma^2\,\mathrm{I}\{i=j\}, \quad i,j \in \mathbb{N}, \tag{103b}$$

then the stochastic process defined inductively via

$$X_i = \sum_{k=1}^p a_i X_{i-k} + Z_i, \quad i \in \mathbb{N} \tag{104}$$

with the initialization

$$(X_{1-p},\ldots,X_0) = (W_{1-p},\ldots,W_0) \tag{105}$$

satisfies the constraints (2).

[6]The Row-ℓ Column-m element of the matrix K_p is $\alpha_{|\ell-m|}$. This matrix is thus the result of deleting the last column and last row of the $(p+1)\times(p+1)$ matrix that we assumed was positive definite.

(By Burg's maximum entropy theorem [9, Th. 12.6.1], of all stochastic processes satisfying (2) the one of highest Shannon rate is the p-th order Gauss-Markov process. It is obtained when $(W_{1-p},\ldots,W_0)$ is a centered Gaussian and $\{Z_i\}$ are IID $\sim \mathcal{N}(0,\sigma^2)$. Its Shannon entropy rate is $(1/2)\log(2\pi e\sigma^2)$.)

We first consider the case where $\alpha > 1$. Let $a_1,\ldots,a_p$, σ^2 and K_p be as above, and let $\varepsilon > 0$ be arbitrarily small. By Proposition 4 there exists a SP $\{Z_i\}$ such that (103a) holds and such that

$$\lim_{n\to\infty}\frac{1}{n}h_\alpha(Z_1,\ldots,Z_n) \geq \frac{1}{2}\log(2\pi e\sigma^2) - \varepsilon. \tag{106}$$

The matrix K_p is positive definite, so by the spectral representation theorem we can find vectors $\mathbf{w}_1,\ldots,\mathbf{w}_p \in \mathbb{R}^p$ and constants $q_1,\ldots,q_p > 0$ with $q_1+\cdots+q_p = 1$ such that

$$\mathsf{K}_p = \sum_{\ell=1}^p q_\ell \mathbf{w}_\ell \mathbf{w}_\ell^\mathsf{T}. \tag{107}$$

(The vectors are eigenvectors of K_p, and the constants $q_1,\ldots,q_p$ are the scaled eigenvalues of K_p.) Draw the random vector $\mathbf{W}$ independently of $\{Z_i\}$ with

$$\Pr[\mathbf{W} = \mathbf{w}_\ell] = q_\ell, \quad \ell = 1,\ldots,p$$

so that, by (107),

$$\mathrm{E}[\mathbf{W}\mathbf{W}^\mathsf{T}] = \mathsf{K}_p. \tag{108}$$

Construct now the stochastic process $\{X_i\}$ using (104) initialized with $(X_{1-p},\ldots,X_0)^\mathsf{T}$ being set to $\mathbf{W}$.

By (108), the resulting SP satisfies (2). We next study its Rényi rate. To that end, we study the Rényi entropy of the vector X_1^n for $n \in \mathbb{N}$. Let $f_\mathbf{X}$ denote its density, and let $f_{\mathbf{X}|\mathbf{w}_\ell}$ denote its conditional density given $\mathbf{W} = \mathbf{w}_\ell$, so

$$f_\mathbf{X}(\mathbf{x}) = \sum_{\ell=1}^p q_\ell f_{\mathbf{X}|\mathbf{w}_\ell}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n.$$

Consequently, by Lemma 8,

$$h_\alpha(f_\mathbf{X}) \geq \min_{1\leq\ell\leq p} h_\alpha(f_{\mathbf{X}|\mathbf{w}_\ell}), \tag{109}$$

and by Lemma 9

$$h_\alpha(f_\mathbf{X}) \leq \min_{1\leq\ell\leq p}\Big\{\frac{\alpha}{1-\alpha}\log q_\ell + h_\alpha(f_{\mathbf{X}|\mathbf{w}_\ell})\Big\}. \tag{110}$$

We next study $h_\alpha(f_{\mathbf{X}|\mathbf{w}_\ell})$ for any given $\ell \in \{1,\ldots,p\}$. Recalling that $\mathbf{W}$ and $\{Z_i\}$ are independent, we conclude that, conditional on $\mathbf{W} = \mathbf{w}_\ell$, the random variables $X_1,\ldots,X_n$ are generated inductively via (104) with the initialization

$$(X_{1-p},\ldots,X_0)^\mathsf{T} = \mathbf{w}_\ell.$$

Conditionally on $\mathbf{W} = \mathbf{w}_\ell$, the random variables $X_1,\ldots,X_n$ are thus an affine transformation of $Z_1,\ldots,Z_n$. The transformation is of unit Jacobian (because the partial-derivatives matrix has 1's on the diagonal and 0's on the upper triangle), and thus

$$h_\alpha(f_{\mathbf{X}|\mathbf{w}_\ell}) = h_\alpha(Z_1,\ldots,Z_n), \quad \ell \in \{1,\ldots,p\}. \tag{111}$$

From this, (109), and (110) it follows that

$$h_\alpha(Z_1^n) \le h_\alpha(f_{\mathbf{X}}) \le \min_{1\le \ell \le p}\left\{\frac{\alpha}{1-\alpha}\log q_\ell\right\} + h_\alpha(Z_1^n).$$

Dividing by n and using (106) establishes the result.

We next turn to the case $0 < \alpha < 1$. For every $\mathsf{M} > 0$ arbitrarily large, we use Proposition 4 to construct $\{Z_i\}$ as above but with

$$\lim_{n\to\infty} \frac{1}{n} h_\alpha(Z_1,\ldots,Z_n) \ge \mathsf{M}.$$

The proof continues as for the case where α exceeds one. □

APPENDIX

In this appendix we present two lemmas, which we then use to prove Proposition 7 on approaching $h^\star(\Gamma)$ using bounded densities.

Lemma 13: Let f be a density supported by $\mathcal{S}$ for which $h(f)$ is defined;

$$\int f(x)|r(x)|\,\mathrm{d}x < \infty; \tag{112}$$

and for which

$$\int f(x)r(x)\,\mathrm{d}x \le \Gamma \tag{113}$$

for some $\Gamma \in \mathbb{R}$. Then for every $\delta > 0$ there exists a density $\tilde{f}$ that is bounded, supported by $\mathcal{S}$, and that satisfies

$$\int \tilde{f}(x)r(x)\,\mathrm{d}x \le \Gamma + \delta \tag{114}$$

and

$$h(\tilde{f}) \ge h(f) - \delta. \tag{115}$$

Proof: Let $0 < \varepsilon < 1$ be fixed (small), with its choice specified later. It follows from (112) and the MCT that there exists some M_1 sufficiently large so that

$$\int \Big(f(x) - \big(f(x)\wedge \mathsf{M}_1\big)\Big)|r(x)|\,\mathrm{d}x < \varepsilon,$$

where we recall that $a \wedge b$ stands for $\min\{a, b\}$. Since the density f integrates to 1, we can find some M_2 sufficiently large so that

$$\int \big(f(x)\wedge \mathsf{M}_2\big)\,\mathrm{d}x > 1-\varepsilon.$$

Define now

$$\mathsf{M} = \max\{1, \mathsf{M}_1, \mathsf{M}_2\}. \tag{116}$$

For this M we have:

$$\int \big(f(x)\wedge \mathsf{M}\big)\,\mathrm{d}x > 1-\varepsilon, \tag{117a}$$

$$\int \Big(f(x) - \big(f(x)\wedge \mathsf{M}\big)\Big)|r(x)|\,\mathrm{d}x < \varepsilon, \tag{117b}$$

$$\Big(f(x) \ge 1\Big) \implies \Big(f(x)\wedge \mathsf{M} \ge 1\Big). \tag{117c}$$

Consider now the bounded density

$$\tilde{f}(x) = \frac{1}{\beta}\big(f(x)\wedge \mathsf{M}\big), \tag{118a}$$

where

$$\beta = \int \big(f(\tilde{x})\wedge \mathsf{M}\big)\,\mathrm{d}\tilde{x}. \tag{118b}$$

Note that because $f(x)\wedge\mathsf{M}$ is upper-bounded by $f(x)$, which integrates to one, and because of (117a)

$$1-\varepsilon \le \beta \le 1, \tag{119}$$

so

$$\big(f(x)\wedge \mathsf{M}\big) \le \tilde{f}(x) \le \frac{1}{1-\varepsilon}\big(f(x)\wedge\mathsf{M}\big). \tag{120}$$

Moreover, $\tilde{f}$ is supported by $\mathcal{S}$.

Given $\delta > 0$ we next show that by choosing ε sufficiently small we can guarantee that both (114) and (115) hold. We begin with the former. Starting with (118a) we have

$$\begin{aligned}
&\int \tilde{f}(x)r(x)\,\mathrm{d}x \\
&\quad= \frac{1}{\beta}\int \big(f(x)\wedge\mathsf{M}\big)r(x)\,\mathrm{d}x \\
&\quad= \frac{1}{\beta}\int \Big(f(x) - \big(f(x) - f(x)\wedge\mathsf{M}\big)\Big)r(x)\,\mathrm{d}x \\
&\quad= \frac{1}{\beta}\int f(x)r(x)\,\mathrm{d}x \\
&\qquad+\frac{1}{\beta}\int \Big(f(x) - \big(f(x)\wedge\mathsf{M}\big)\Big)\big(-r(x)\big)\,\mathrm{d}x \\
&\quad\le \frac{1}{\beta}\Gamma + \frac{1}{\beta}\int \Big(f(x) - \big(f(x)\wedge\mathsf{M}\big)\Big)|r(x)|\,\mathrm{d}x \\
&\quad\le \frac{1}{\beta}\Gamma + \frac{1}{\beta}\varepsilon \\
&\quad\le \Gamma + \frac{\varepsilon}{1-\varepsilon}|\Gamma| + \frac{\varepsilon}{1-\varepsilon},
\end{aligned} \tag{121}$$

where the first inequality follows from (113); the second from (117b); and the last from (119).

We next study $h(\tilde{f})$. Starting with the definition of $\tilde{f}$,

$$\begin{aligned}
h(\tilde{f}) &= \int \frac{1}{\beta}\big(f(x)\wedge\mathsf{M}\big)\log\frac{\beta}{f(x)\wedge\mathsf{M}}\,\mathrm{d}x \\
&= \log\beta + \frac{1}{\beta}\int \big(f(x)\wedge\mathsf{M}\big)\log\frac{1}{f(x)\wedge\mathsf{M}}\,\mathrm{d}x \\
&= \log\beta + \frac{1}{\beta}\int_{x:\, f(x)\le 1} \big(f(x)\wedge\mathsf{M}\big)\log\frac{1}{f(x)\wedge\mathsf{M}}\,\mathrm{d}x \\
&\quad+\frac{1}{\beta}\int_{x:\, f(x)> 1} \big(f(x)\wedge\mathsf{M}\big)\log\frac{1}{f(x)\wedge\mathsf{M}}\,\mathrm{d}x.
\end{aligned} \tag{122}$$

By (116), $f(x)\wedge\mathsf{M} = f(x)$ whenever $f(x) \le 1$, so

$$\begin{aligned}
&\int_{x:\, f(x)\le 1} \big(f(x)\wedge\mathsf{M}\big)\log\frac{1}{f(x)\wedge\mathsf{M}}\,\mathrm{d}x \\
&\qquad= \int_{x:\, f(x)\le 1} f(x)\log\frac{1}{f(x)}\,\mathrm{d}x.
\end{aligned} \tag{123}$$

Since $\xi\log\xi^{-1}$ is decreasing for $\xi > 1$, and since $f(x) > 1$ implies $f(x)\wedge\mathsf{M} > 1$ (by (117c)),

$$\big(f(x)\wedge\mathsf{M}\big)\log\frac{1}{f(x)\wedge\mathsf{M}} \ge f(x)\log\frac{1}{f(x)}, \quad \Big(f(x) > 1\Big)$$

and hence

$$\int_{x:\, f(x)>1} \bigl(f(x) \wedge \mathsf{M}\bigr) \log \frac{1}{f(x) \wedge \mathsf{M}} \,\mathrm{d}x \geq \int_{x:\, f(x)>1} f(x) \log \frac{1}{f(x)} \,\mathrm{d}x. \tag{124}$$

Summing (123) and (124) we obtain

$$\int \bigl(f(x) \wedge \mathsf{M}\bigr) \log \frac{1}{f(x) \wedge \mathsf{M}} \,\mathrm{d}x \geq h(f). \tag{125}$$

Using this, (122), and (119) we conclude that

$$h(\tilde{f}) = h(f), \quad \text{whenever } h(f) = \infty$$

and

$$h(\tilde{f}) \geq \log(1-\varepsilon) + h(f) - \frac{\varepsilon}{1-\varepsilon}|h(f)|, \quad \text{whenever} \quad |h(f)| < \infty. \tag{126}$$

And obviously $h(\tilde{f}) \geq h(f)$ whenever $h(f) = -\infty$.

The result now follows by choosing ε small enough to guarantee that the RHS of (121) does not exceed $\Gamma + \delta$ and—if $h(f)$ is finite—that the RHS of (126) exceeds $h(f) - \delta$. □

The following lemma addresses the case where (112) does not hold.

Lemma 14: Let the density f supported by $\mathcal{S}$ be such that

$$\int f(x) r(x) \,\mathrm{d}x = -\infty \tag{127}$$

and $h(f)$ is defined and exceeds $-\infty$

$$h(f) > -\infty. \tag{128}$$

Then there exists a sequence of densities $\{\tilde{f}_k\}$ supported by $\mathcal{S}$ for which

$$\int \tilde{f}_k(x) |r(x)| \,\mathrm{d}x < \infty,$$

$$\lim_{k\to\infty} h(\tilde{f}_k) = h(f),$$

and

$$\lim_{k\to\infty} \int \tilde{f}_k(x) r(x) \,\mathrm{d}x = -\infty.$$

Proof: Define $r^+ \triangleq \max\{r, 0\}$ and $r^- \triangleq \max\{-r, 0\}$, so $r = r^+ - r^-$ with $r^+(x), r^-(x) \geq 0$. By (127),

$$\int f(x) r^-(x) \,\mathrm{d}x = \infty, \tag{129a}$$

$$\int f(x) r^+(x) \,\mathrm{d}x < \infty. \tag{129b}$$

Define for every $k \in \mathbb{N}$

$$\mathcal{D}_k \triangleq \bigl\{x : r^-(x) \leq k\bigr\}. \tag{130}$$

By the MCT

$$\lim_{k\to\infty} \int_{\mathcal{D}_k} f(x) r^+(x) \,\mathrm{d}x = \int f(x) r^+(x) \,\mathrm{d}x < \infty \tag{131a}$$

and

$$\lim_{k\to\infty} \int f(x) r^-(x) \,\mathrm{I}\{x \in \mathcal{D}_k\} \,\mathrm{d}x = \infty. \tag{131b}$$

Consequently,

$$\lim_{k\to\infty} \int_{\mathcal{D}_k} f(x) r(x) \,\mathrm{d}x = -\infty. \tag{132}$$

The lemma's hypotheses guarantee that $h(f)$ is defined and exceeds $-\infty$. Consequently,

$$h(f) = h^+(f) - h^-(f),$$

with

$$h^-(f) < \infty, \quad h^+(f) \leq \infty, \tag{133}$$

where,

$$h^+(f) \triangleq \int f(x) \log \frac{1}{f(x)} \,\mathrm{I}\{f(x) \leq 1\} \,\mathrm{d}x,$$

$$h^-(f) \triangleq \int f(x) \log f(x) \,\mathrm{I}\{f(x) > 1\} \,\mathrm{d}x.$$

By the MCT

$$\int_{\mathcal{D}_k} f(x) \log \frac{1}{f(x)} \,\mathrm{I}\{f(x) \leq 1\} \,\mathrm{d}x \uparrow h^+(f)$$

and

$$\int_{\mathcal{D}_k} f(x) \log f(x) \,\mathrm{I}\{f(x) > 1\} \,\mathrm{d}x \uparrow h^-(f)$$

so, upon subtracting (and recalling $h^-(f) < \infty$)

$$\lim_{k\to\infty} \int_{\mathcal{D}_k} f(x) \log \frac{1}{f(x)} \,\mathrm{d}x = h(f). \tag{134}$$

Define

$$\beta_k \triangleq \int_{\mathcal{D}_k} f(x) \,\mathrm{d}x.$$

Note that since f is a density,

$$\beta_k \leq 1$$

and (by the MCT)

$$\beta_k \uparrow 1. \tag{135}$$

Consequently,

$$0 < \beta_k \leq 1, \quad k \text{ large}. \tag{136}$$

For every such sufficiently large k, define the density

$$\tilde{f}_k(x) \triangleq \beta_k^{-1} f(x) \,\mathrm{I}\{x \in \mathcal{D}_k\}.$$

It is supported by $\mathcal{S}$, and its entropy $h(\tilde{f}_k)$ can be expressed as

$$\begin{aligned} h(\tilde{f}_k) &= \int \tilde{f}_k(x) \log \frac{1}{\tilde{f}_k(x)} \,\mathrm{d}x \\ &= \int_{\mathcal{D}_k} \tilde{f}_k(x) \log \frac{1}{\tilde{f}_k(x)} \,\mathrm{d}x \\ &= \int_{\mathcal{D}_k} \frac{1}{\beta_k} f(x) \log \frac{\beta_k}{f(x)} \,\mathrm{d}x \\ &= \log \beta_k + \frac{1}{\beta_k} \int_{\mathcal{D}_k} f(x) \log \frac{1}{f(x)} \,\mathrm{d}x. \end{aligned}$$

From this, (134), and (135) we obtain

$$\lim_{k\to\infty} h(\tilde{f}_k) = h(f). \tag{137}$$

And as to the expectation of $r(x)$ under $\tilde{f}_k$:

$$\begin{aligned}\int \tilde{f}_k(x) r(x)\,\mathrm{d}x &= \frac{1}{\beta_k}\int_{\mathcal{D}_k} f(x)r(x)\,\mathrm{d}x \\ &= \frac{1}{\beta_k}\int_{\mathcal{D}_k} f(x)r^+(x)\,\mathrm{d}x - \frac{1}{\beta_k}\int_{\mathcal{D}_k} f(x)r^-(x)\,\mathrm{d}x.\end{aligned}$$

The first term on the RHS is finite by (136) and (129b). Moreover, its limsup as $k \to \infty$ is finite by (135) and (131a). The second term tends to $-\infty$ by (135) and (132). Hence,

$$\lim_{k\to\infty}\int \tilde{f}_k(x)r(x)\,\mathrm{d}x = -\infty. \tag{138}$$

Moreover,

$$\begin{aligned}\int \tilde{f}_k(x)|r(x)|\,\mathrm{d}x &= \frac{1}{\beta_k}\int_{\mathcal{D}_k} f(x)r^+(x)\,\mathrm{d}x + \frac{1}{\beta_k}\int_{\mathcal{D}_k} f(x)r^-(x)\,\mathrm{d}x \\ &\le \frac{1}{\beta_k}\int f(x)r^+(x)\,\mathrm{d}x + k \\ &< \infty,\end{aligned} \tag{139}$$

where the first inequality follows from the nonnegativity of r^+ and from the definition of the set $\mathcal{D}_k$ (130), and the second inequality follows from (129b) and (136).

The lemma now follows from (139), (137), and (138). □

Proof of Proposition 7: Since Γ exceeds Γ_0, it follows from (17) that

$$-\infty < h^\star(\Gamma) < \infty. \tag{140}$$

Let the density f nearly achieve $h^\star(\Gamma)$ in the sense that it is supported by $\mathcal{S}$ and that

$$\int f(x)\,r(x)\,\mathrm{d}x \le \Gamma, \quad \text{and} \quad h(f) > h^\star(\Gamma) - \frac{\delta}{2}. \tag{141}$$

By (140), (141), and the definition of $h^\star(\Gamma)$,

$$-\infty < h(f) < \infty. \tag{142}$$

If $\int f(x)|r(x)|\,\mathrm{d}x$ is finite, then the result follows directly from Lemma 13. It remains to prove the result when this integral is infinite. In this case $\int f(x)r(x)\,\mathrm{d}x = -\infty$ by (141) (because $\Gamma < \infty$). Using this, the finiteness of $h(f)$ (142), and Lemma 14, we infer the existence of a density $\tilde{f}$ that supported by $\mathcal{S}$ and for which

$$\int \tilde{f}(x)|r(x)|\,\mathrm{d}x < \infty, \tag{143a}$$

$$h(\tilde{f}) > h(f) - \frac{\delta}{2}, \tag{143b}$$

$$\int \tilde{f}(x)r(x)\,\mathrm{d}x < \Gamma. \tag{143c}$$

Applying Lemma 13 to the density $\tilde{f}$, we conclude that there exists a bounded density $f^\star$ that is supported by $\mathcal{S}$ and that satisfies

$$h(f^\star) > h(\tilde{f}) - \frac{\delta}{2} \quad \text{and} \quad \int f^\star(x)r(x)\,\mathrm{d}x \le \Gamma + \delta \tag{144}$$

and hence, in view of (143) and (141),

$$h(f^\star) > h^\star(\Gamma) - \delta \quad \text{and} \quad \int f^\star(x)r(x)\,\mathrm{d}x \le \Gamma + \delta. \tag{145}$$

The existence of $f^\star$ concludes the proof of the proposition for the case where $\int f(x)|r(x)|\,\mathrm{d}x$ is infinite. □

Acknowledgment

Discussions with Stefan M. Moser and Igal Sason are gratefully acknowledged. The authors also thank the anonymous reviewers and the Associate Editor for their helpful tips on how to better organize the paper.

References

[1] E. Arikan, “An inequality on guessing and its application to sequential decoding,” *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, Jan. 1996.
[2] D. Malone and W. G. Sullivan, “Guesswork and entropy,” *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 525–526, Mar. 2004.
[3] C. E. Pfister and W. G. Sullivan, “Rényi entropy, guesswork moments, and large deviations,” *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2794–2800, Nov. 2004.
[4] M. K. Hanawal and R. Sundaresan, “Guessing revisited: A large deviations approach,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 70–78, Jan. 2011.
[5] M. M. Christiansen and K. R. Duffy, “Guesswork, large deviations, and Shannon entropy,” *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 796–802, Feb. 2013.
[6] C. Bunte and A. Lapidoth, “Encoding tasks and Rényi entropy,” *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5065–5076, Sep. 2014.
[7] L. L. Campbell, “A coding theorem and Rényi's entropy,” *Inf. Control*, vol. 8, no. 4, pp. 423–429, Aug. 1965.
[8] C. Bunte and A. Lapidoth, “Rényi entropy and quantization for densities,” in *Proc. Inf. Theory Workshop*, Nov. 2014, pp. 258–262.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[10] C. Bunte and A. Lapidoth, “Maximizing Rényi entropy rate,” in *Proc. IEEE 28th Conv. Electr. Electron. Eng.*, Eilat, Israel, Dec. 2014, pp. 1–4.
[11] M. A. Kumar and R. Sundaresan, “Minimization problems based on relative α-entropy I: Forward projection,” *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 5063–5080, Sep. 2015.
[12] E. Lutwak, D. Yang, and G. Zhang, “Moment-entropy inequalities,” *Ann. Probab.*, vol. 32, no. 1B, pp. 757–774, 2004.
[13] J. Costa, A. Hero, and C. Vignat, “On solutions to multivariate maximum α-entropy problems,” in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Berlin, Germany: Springer Verlag, 2003, pp. 211–226.
[14] K. Zografos and S. Nadarajah, “Expressions for Rényi and Shannon entropies for multivariate distributions,” *Statist. Probab. Lett.*, vol. 71, no. 1, pp. 71–84, 2005.
[15] Z. Rached, F. Alajaji, and L. L. Campbell, “Rényi's divergence and entropy rates for finite alphabet Markov sources,” *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1553–1561, May 2001.
[16] L. Golshani, E. Pasha, and G. Yari, “Some properties of Rényi entropy and Rényi entropy rate,” *Inf. Sci.*, vol. 179, no. 14, pp. 2426–2433, 2009.
[17] L. Golshani and E. Pasha, “Rényi entropy rate for Gaussian processes,” *Inf. Sci.*, vol. 180, no. 8, pp. 1486–1491, 2010.
[18] M. Khodabin, “ADK entropy and ADK entropy rate in irreducible-aperiodic Markov chain and Gaussian processes,” *J. Iranian Statist. Soc.*, vol. 9, no. 2, pp. 115–126, 2010.
[19] J. P. Burg, “Maximum entropy spectral analysis,” in *Modern Spectrum Analysis*, D. G. Childers, Ed. New York, NY, USA: IEEE Press, 1978, pp. 34–41.
[20] L. Wang and M. Madiman, “Beyond the entropy power inequality, via rearrangements,” *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5116–5137, Sep. 2014.
[21] M. Pourahmadi, *Foundations of Time Series Analysis and Prediction Theory* (Wiley Series in Probability and Statistics). New York, NY, USA: Wiley, 2001.

Christoph Bunte received the B.Sc., M.Sc., and Ph.D. degrees in Electrical Engineering and Information Technology in 2010, 2011, and 2014, all from ETH Zurich in Switzerland. He is a recipient of the 2014 IEEE Jack Keil Wolf ISIT Student Paper Award.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree in Mathematics (summa cum laude, 1986), the B.Sc. degree in Electrical Engineering (summa cum laude, 1986), and the M.Sc. degree in Electrical Engineering (1990) all from the Technion–Israel Institute of Technology. He received the Ph.D. degree in Electrical Engineering from Stanford University in 1995. In the years 1995–1999 he was an Assistant and Associate Professor at the department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, and was the KDD Career Development Associate Professor in Communications and Technology. He is now Professor of Information Theory at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. He served in the years 2003-2004 and 2009 as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY. Dr. Lapidoth's research interests are in Digital Communications and Information Theory. He is the author of the textbook *A Foundation in Digital Communication*, published by Cambridge University Press in 2009.