Maximizing Rényi Entropy Rate

Christoph Bunte and Amos Lapidoth
ETH Zurich
Switzerland
Email: {bunte,lapidoth}@isi.ee.ethz.ch

***Abstract*—Of all univariate distributions on the nonnegative reals of a given mean, the distribution that maximizes the Rényi entropy is Lomax. But the memoryless Lomax stochastic process does not maximize the Rényi entropy rate: For Rényi orders smaller than one the supremum of the Rényi entropy rates is infinite, and for orders larger than one it is the differential Shannon entropy of the exponential distribution, which is the distribution that maximizes the differential Shannon entropy subject to these constraints. This is shown to be a special case of a much more general principle.**

I. Introduction

The order-α Rényi entropy of a probability density function (PDF) f is defined as

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

where α can be any positive number other than one. The integral on the RHS of (1) always exists, possibly taking on the value $+\infty$, in which case we define $h_\alpha(f) = +\infty$ if $0 < \alpha < 1$ and $h_\alpha(f) = -\infty$ if $\alpha > 1$. When a random variable (RV) X is of density f_X we sometimes write $h_\alpha(X)$ instead of $h_\alpha(f_X)$.

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

$$h_\alpha(\{X_k\}) = \lim_{n\to\infty} \frac{1}{n} h_\alpha\bigl(X_1^n\bigr),$$

whenever the limit exists. Here we use the notation X_i^j to denote the tuple $(X_i, \dots, X_j)$.

The Rényi entropy is closely related to the differential Shannon entropy:

$$h(f) = -\int_{-\infty}^{\infty} f(x) \log f(x) \, \mathrm{d}x. \tag{2}$$

(The integral on the RHS of (2) need not exist. If it does not, then we say that $h(f)$ does not exist.) Under some mild technical conditions [1],

$$h_\alpha(f) \le h(f), \quad \text{for } \alpha > 1; \tag{3}$$

$$h_\alpha(f) \ge h(f) \quad \text{for } 0 < \alpha < 1; \tag{4}$$

and

$$\lim_{\alpha\to 1} h_\alpha(f) = h(f). \tag{5}$$

The entropy of a pair of independent random variables is the sum of the individual entropies. This is true for both Rényi entropy and differential Shannon entropy. But the two entropies behave differently when the random variables are correlated: The differential Shannon entropy of a pair is always upper bounded by the sum of the individual differential Shannon entropies, but this inequality need not hold for Rényi entropy. Consequently, the SP that maximizes the Rényi entropy rate among all stochastic processes of a given marginal distribution need not be IID.

As we shall show, Rényi entropy rate is typically maximized by stochastic processes with memory. By introducing memory we can typically achieve Rényi rates that are infinite when α is smaller than one, and that—notwithstanding (3)—are equal to the Shannon rates when $\alpha > 1$.

II. Results

A. Maximizing Rényi Entropy Subject to Constraints

Consider the problem of maximizing $h_\alpha(f)$ subject to a set of constraints of the form

$$\left.\begin{aligned} & f(x) \ge 0, \quad \text{with equality if } x \notin \mathcal{S}, \\ & \int_{\mathcal{S}} f(x)\,\mathrm{d}x = 1, \\ & \int_{\mathcal{S}} f(x) r_i(x)\,\mathrm{d}x = \gamma_i, \quad i = 1,\dots,m. \end{aligned}\right\} \tag{6}$$

(The first two constraints make sure that f is a PDF that is zero outside the set $\mathcal{S} \subseteq \mathbb{R}^n$.) A classic result [2, Th. 12.1.1] is that if one can find constants $\lambda_0, \dots, \lambda_m$ such that

$$f^\star(x) = e^{\lambda_0 + \sum_{i=1}^m \lambda_i r_i(x)}, \quad x \in \mathcal{S}, \tag{7}$$

satisfies (6), then $f^\star$ is the unique maximizer of $h(f)$ with respect to all f satisfying (6). Our first result is the analogous result for Rényi entropy:

Theorem II.1.

1) *Let $0 < \alpha < 1$ and*

$$f^\star(x) = \begin{cases} \dfrac{c}{\bigl(1+\sum_{i=1}^m \lambda_i r_i(x)\bigr)^{\frac{1}{1-\alpha}}} & \text{if } x \in \mathcal{S}, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where the constants c and $\lambda_1, \dots, \lambda_m$ are such that $f^\star$ satisfies (6) and $\sum_{i=1}^m \lambda_i r_i(x) \ge -1$ for all $x \in \mathcal{S}$. Then $f^\star$ is the unique maximizer of $h_\alpha(f)$ with respect to all f satisfying (6).

2) *Let $\alpha > 1$ and*

$$f^\star(x) = \begin{cases} c\Bigl(1 + \sum_{i=1}^m \lambda_i r_i(x)\Bigr)^{\frac{1}{\alpha-1}} & \text{if } x \in \mathcal{S} \cap \mathcal{T}, \\ 0 & \text{otherwise,} \end{cases} \tag{9a}$$

where

$$\mathcal{T} = \Big\{ x : \sum_{i=1}^{m} \lambda_i r_i(x) \geq -1 \Big\}, \tag{9b}$$

and where the constants c *and* $\lambda_1, \dots, \lambda_m$ *are such that* $f^\star$ *satisfies* (6). *Then* $f^\star$ *is the unique maximizer of* $h_\alpha(f)$ *with respect to all* f *satisfying* (6).

The proof of Theorem II.1 is similar to the proof of [2, Thm. 12.1.1] with the KL-divergence replaced with Sundaresan's divergence [3]:

$$\Delta_\alpha(f||g) = \log\Bigg(\frac{\int_{-\infty}^{\infty} g(x)^\alpha \, \mathrm{d}x}{\big(\int_{-\infty}^{\infty} f(x)^\alpha \, \mathrm{d}x\big)^{\frac{1}{1-\alpha}}} \Big(\int_{-\infty}^{\infty} \frac{f(x)}{g(x)^{1-\alpha}} \, \mathrm{d}x \Big)^{\frac{\alpha}{1-\alpha}} \Bigg). \tag{10}$$

We omit the details. When the constraints are covariance constraints, the result can be found in [4].

B. The Lomax Distribution Maximizes the Rényi Entropy Over All Nonnegative RVs with a Given Mean

If we apply Theorem II.1 to the case where $\mathcal{S} = [0, \infty)$ and $\int_0^\infty x f(x) \, \mathrm{d}x = \mu$, we obtain that for $1/2 < \alpha < 1$ the maximizer is the Lomax density

$$f^\star(x) = \frac{1}{\mu} \frac{\frac{\alpha}{2\alpha - 1}}{\Big(1 + \frac{x}{\mu(\frac{1}{1-\alpha} - 2)}\Big)^{\frac{1}{1-\alpha}}}, \quad x \geq 0. \tag{11}$$

And for $\alpha > 1$ it is

$$f^\star(x) = \frac{\alpha}{\mu(2\alpha - 1)} \Big(1 - \frac{x}{\mu(\frac{\alpha}{\alpha-1} + 1)}\Big)^{\frac{1}{\alpha-1}}, \quad 0 \leq x \leq \mu\Big(\frac{\alpha}{\alpha - 1} + 1\Big). \tag{12}$$

It is instructive to also consider the vector case. If we wish to maximizes $h_\alpha(X_1, X_2)$ subject to the constraints that both X_1 and X_2 be nonnegative and of mean μ, then the joint densities we obtain from Theorem II.1 do not factorize. Under the maximizing distribution, X_1 and X_2 are thus not independent. This already indicates that the SP maximizing the Rényi rate will not be IID. But rather than studying this maximization for these particular constraints, we turn to the general case.

C. Maximizing Rényi Entropy Rate Subject to Constraints

Let $h^\star(\Gamma)$ denote the supremum of $h(f_X)$ over all densities f_X under which

$$\Pr(X \in \mathcal{S}) = 1 \quad \text{and} \quad \mathrm{E}[g(X)] \leq \Gamma. \tag{13}$$

Assume that for some Γ_0

$$h^\star(\Gamma_0) > -\infty, \tag{14a}$$

and

$$h^\star(\Gamma) < \infty \quad \text{for every } \Gamma \geq \Gamma_0. \tag{14b}$$

It then follows that $h^\star(\Gamma)$ is finite, nondecreasing, and concave in Γ for all $\Gamma \geq \Gamma_0$.

Since $h(X_1, \dots, X_n) \leq \sum_{i=1}^n h(X_i)$, the SP that maximizes the Shannon rate $\lim_{n\to\infty} h(X^n)/n$ subject to the constraint that $\Pr(X_i \in \mathcal{S}) = 1$ and $\mathrm{E}[g(X_i)] \leq \Gamma$ for all i is IID. This is not the case for Rényi entropy rate:

Theorem II.2. *Let* Γ_0 *be such that* (14) *holds, and let* $\Gamma > \Gamma_0$.

1) *For every* $\alpha > 1$,

$$\lim_{n\to\infty} \sup \frac{h_\alpha(X^n)}{n} = h^\star(\Gamma), \tag{15}$$

where the supremum is over all joint distributions on $X_1, \dots, X_n$ *under which* $\Pr(X_i \in \mathcal{S}) = 1$ *and* $\mathrm{E}[g(X_i)] \leq \Gamma$.

2) *For every* $0 < \alpha < 1$,

$$\lim_{n\to\infty} \sup \frac{h_\alpha(X^n)}{n} = \lim_{\tilde{\Gamma}\to\infty} h^\star(\tilde{\Gamma}), \tag{16}$$

where the supremum is as in Part 1.

Before we present a proof, we point out that Theorem II.2 can be generalized in a straightforward fashion to account for multiple constraints: $\mathrm{E}[g_\ell(X_i)] \leq \Gamma_\ell$ for $\ell = 1, \dots, L$. However, for ease of presentation we focus on the case of a single constraint.

Proof. We begin with the proof of (15). The "converse part" that the LHS of (15) cannot exceed its RHS is an immediate consequence of the fact that the Rényi entropy of any order larger than 1 is upper-bounded by the differential Shannon entropy (3).

To prove the "direct part", let $f^\star$ be a density satisfying the constraints (13) and for which $h(f^\star) \geq h^\star(\Gamma) - \varepsilon$. Let $\mathcal{T}_n^\varepsilon(f^\star)$ denote the set of ε-weakly-typical sequences of length n with respect to $f^\star$:

$$\mathcal{T}_n^\varepsilon(f^\star) = \Big\{ x^n \in \mathcal{S}^n : 2^{-n(h(f^\star)+\varepsilon)} \leq \prod_{i=1}^n f^\star(x_i) \leq 2^{-n(h(f^\star)-\varepsilon)} \Big\}. \tag{17}$$

Further, let

$$\mathcal{G}_n^\varepsilon = \Big\{ x^n \in \mathcal{S}^n : \Gamma - \varepsilon \leq \frac{1}{n} \sum_{i=1}^n g(x_i) \leq \Gamma + \varepsilon \Big\}. \tag{18}$$

By the Law of Large Numbers, for all sufficiently large n,

$$1 - \varepsilon \leq \int_{\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon} \prod_{i=1}^n f^\star(x_i) \, \mathrm{d}x^n \tag{19}$$

$$\leq |\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon| 2^{-n(h(f^\star)-\varepsilon)}, \tag{20}$$

where the second line follows from the definition of $\mathcal{T}_n^\varepsilon(f^\star)$. (We use $|\cdot|$ to denote Lebesgue measure.) Rearranging gives

$$|\mathcal{T}_n^\varepsilon(f^\star) \cap \mathcal{G}_n^\varepsilon| \geq (1 - \varepsilon) 2^{n(h(f^\star)-\varepsilon)}. \tag{21}$$

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

$$h_\alpha(f_n) = \log|\mathcal{T}_n^\varepsilon(f^\star)\cap\mathcal{G}_n^\varepsilon| \tag{22}$$
$$\geq n\big(h(f^\star)-\varepsilon\big)+\log(1-\varepsilon) \tag{23}$$
$$\geq n\big(h^\star(\Gamma)-2\varepsilon\big)+\log(1-\varepsilon). \tag{24}$$

Consequently, if X^n has density f_n, then

$$\varliminf_{n\to\infty}\frac{h_\alpha(X^n)}{n}\geq h^\star(\Gamma)-2\varepsilon. \tag{25}$$

The permutation invariance of the sets $\mathcal{T}_n^\varepsilon(f^\star)$ and $\mathcal{G}_n^\varepsilon$ imply that the X_i's have identical marginals. And since $n^{-1}\sum_{i=1}^n \mathrm{E}[g(X_i)]\leq\Gamma+\varepsilon$ by the definition of $\mathcal{G}_n^\varepsilon$ and f_n, it follows that $\mathrm{E}[g(X_i)]\leq\Gamma+\varepsilon$ for every i. The proof of (15) now follows from the continuity of $h^\star(\Gamma)$ in Γ.

To establish the direct part of (16), we will show that for any fixed $\Gamma_1>\Gamma$ we can achieve $h^\star(\Gamma_1)$. The result will then follow by letting Γ_1 tend to infinity. Fix then some $\Gamma_1>\Gamma$. Let f_0 achieve $h^\star(\Gamma_0)$ and let f_1 achieve $h^\star(\Gamma_1)$.[1] Since $\Gamma_1>\Gamma_0$, it follows that $h^\star(\Gamma_1)\geq h^\star(\Gamma_0)$.

Define

$$\mathcal{S}_i=\mathcal{T}_n^\varepsilon(f_i)\cap\mathcal{G}_n^\varepsilon(f_i),\quad i\in\{0,1\}, \tag{26}$$

and fix some $\delta>0$ small enough so that

$$(1-\delta)\Gamma_0+\delta\Gamma_1\leq\Gamma. \tag{27}$$

Consider now the "mixture" density

$$f_n(x^n)=(1-\delta)\frac{1}{|\mathcal{S}_0|}1\{x^n\in\mathcal{S}_0\}+\delta\frac{1}{|\mathcal{S}_1|}1\{x^n\in\mathcal{S}_1\}, \tag{28}$$

where $1\{\cdot\}$ denotes the indicator function. Let X^n be of density f_n. As in the proof of (15) it follows that the X_i's have identical marginals and $\mathrm{E}[g(X_i)]\leq\Gamma$. Moreover, since $\mathcal{S}_0$ and $\mathcal{S}_1$ are disjoint (for small enough $\varepsilon>0$ that guarantees that $\Gamma_1-\varepsilon>\Gamma_0+\varepsilon$),

$$\frac{h_\alpha(X^n)}{n} = \frac{1}{n(1-\alpha)}\log\Big((1-\delta)^\alpha|\mathcal{S}_0|^{1-\alpha}+\delta^\alpha|\mathcal{S}_1|^{1-\alpha}\Big). \tag{29}$$

On account of (21), this implies

$$\varliminf_{n\to\infty}\frac{h_\alpha(X^n)}{n}\geq h^\star(\Gamma_1)-\varepsilon \tag{30}$$

and thus completes the direct part of (16).

We next turn to the converse. If $\lim_{\tilde\Gamma\to\infty}h^\star(\tilde\Gamma)=\infty$, then the converse is trivial: no achievable Rényi rate can exceed $+\infty$. To conclude the converse we therefore now consider the case where

$$\lim_{\tilde\Gamma\to\infty}h^\star(\tilde\Gamma)<\infty. \tag{31}$$

[1]For simplicity we assume that the suprema are achieved. Otherwise use $h(f_0)\geq h^\star(\Gamma_0)-\varepsilon$ and $h(f_1)\geq h^\star(\Gamma_1)-\varepsilon$.

We will show that in this case the Lebesgue measure of the support set $\mathcal{S}$ must be finite

$$|\mathcal{S}|<\infty \tag{32}$$

and

$$\lim_{\tilde\Gamma\to\infty}h^\star(\tilde\Gamma)=\log|\mathcal{S}|. \tag{33}$$

To this end, observe that

$$h^\star(k)\geq\log|\{x\in\mathcal{S}:g(x)\leq k\}| \tag{34}$$

because the RHS can be achieved by a uniform distribution on the set $\{x\in\mathcal{S}:g(x)\leq k\}$. Moreover, by the monotone convergence theorem,

$$\lim_{k\to\infty}|\{x\in\mathcal{S}:g(x)\leq k\}|=|\mathcal{S}|. \tag{35}$$

Combining (34) and (35) we obtain

$$\lim_{\tilde\Gamma\to\infty}h^\star(\tilde\Gamma)\geq\log|\mathcal{S}|, \tag{36}$$

which establishes that (31) indeed implies (32). And since the uniform distribution on $\mathcal{S}$ maximizes the differential Shannon entropy,

$$h^\star(\tilde\Gamma)\leq\log|\mathcal{S}|, \tag{37}$$

which combines with (36) to establish (33).

Having established that (31) implies (32) and (33), we are now ready to conclude the converse.

Since the uniform distribution maximizes both the differential entropy and the Rényi entropy when the support is finite, it follows that for every tuple $X_1,\ldots,X_n$ for which $\Pr(X_i\in\mathcal{S})=1$ for all $1\leq i\leq n$, we have

$$\frac{h_\alpha(X^n)}{n}\leq\log|\mathcal{S}| \tag{38}$$
$$=\lim_{\tilde\Gamma\to\infty}h^\star(\tilde\Gamma). \tag{39}$$

This completes the converse part of (16). □

D. *Consistency and Stationarity*

Let Γ_0 and Γ be as in Theorem II.2. We now make the additional assumptions that the cost function g is nonnegative

$$g(x)\geq 0,\quad x\in\mathcal{S}, \tag{40a}$$

and that its sublevel sets are of finite Lebesgue measure

$$|\{x\in\mathcal{S}:g(x)\leq\tilde\Gamma\}|<\infty,\quad\tilde\Gamma\geq\Gamma_0. \tag{40b}$$

Consider first the case where $\alpha>1$. Inspecting the proof of Theorem II.2 we see that we can find some sufficiently large n such that the n-vector whose density f_n is uniform over $\mathcal{T}_n^\varepsilon(f^\star)\cap\mathcal{G}_n^\varepsilon$ satisfies (24). Being uniform, the density f_n is bounded. And since its support is contained in $\mathcal{G}_n^\varepsilon$—which by (40b) and the nonnegativity of g is a subset of $\mathcal{S}^n$ of finite Lebesgue measure—all the marginals of f_n have bounded density. Consequently, since $\alpha>1$,

$$h_\alpha(X_1,\ldots,X_r)>-\infty,\quad r\in\{1,\ldots,n-1\}. \tag{41}$$

Consider now the SP $\{Y_k\}$ that we construct by drawing

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

As we next show, for this SP the Rényi rate exists and is lower bounded by $h^\star(\Gamma) - 2\varepsilon$. To see this consider a (large) positive integer m, and express it as $m = qn + r$ where $q = \lfloor m/n \rfloor$ and the remainder r is in $\{0, \ldots, n-1\}$. For such m we have by independence

$$\frac{1}{m} h_\alpha\bigl(Y_1^m\bigr) = \frac{1}{m} \Bigl\lfloor \frac{m}{n} \Bigr\rfloor h_\alpha(f_n) + \frac{1}{m} h_\alpha(X_1, \ldots, X_r),$$

from which the result follows by taking m to infinity and using (41) and (24).

In fact, we can even construct a *stationary* SP of this Rényi rate by introducing a random shift

$$Z_k = Y_{k+P},$$

where P is uniform over $\{0, \ldots, n-1\}$. The density f_Z of Z_1^m is now the mixture of $f_{\mathbf{Z}|p}$, where $f_{\mathbf{Z}|p}$ is the density of Z_1^m conditional on the shift being p. While h_α need not be concave for $\alpha > 1$, we can establish that $\{Z_k\}$ has the desired Rényi entropy rate using the inequality

$$\log \int f_{\mathbf{Z}}^{\alpha}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \log \int \left(\frac{1}{n} \sum_{p=0}^{n-1} f_{\mathbf{Z}|P=p}(\mathbf{z}) \right)^{\alpha} \mathrm{d}\mathbf{z} \tag{42}$$

$$\leq \log \int n^{-1} \sum_{p=0}^{n-1} f_{\mathbf{Z}|P=p}^{\alpha}(\mathbf{z})\,\mathrm{d}\mathbf{z} \tag{43}$$

$$= \log \left(n^{-1} \sum_{p=0}^{n-1} \int f_{\mathbf{Z}|P=p}^{\alpha}(\mathbf{z})\,\mathrm{d}\mathbf{z} \right) \tag{44}$$

$$\leq \log \max_{0 \leq p < n} \int f_{\mathbf{Z}|P=p}^{\alpha}(\mathbf{z})\,\mathrm{d}\mathbf{z}, \tag{45}$$

$$= \max_{0 \leq p < n} \log \int f_{\mathbf{Z}|P=p}^{\alpha}(\mathbf{z})\,\mathrm{d}\mathbf{z}, \tag{46}$$

which holds because $\alpha > 1$. For such α, the term $1/(1-\alpha)$ that multiplies the log in (1) is negative, so this upper bound on the log leads to a lower bound on the Rényi entropy.

We next turn to the case where $0 < \alpha < 1$. Here too we can construct a stationary process whose Rényi entropy rate exists and is arbitrarily large or infinite. Once again we choose n large enough so that the density f_n of (28) will have sufficiently large Rényi entropy. We next consider two cases. In the first no subset of $X_1, \ldots, X_n$ has infinite Rényi entropy. In this case we proceed as for the case where $\alpha > 1$ except that we replace (46) with the lower bound on the Rényi entropy that results from its concavity (for $0 < \alpha < 1$).

In the second case where some subset of $X_1, \ldots, X_n$ has infinite Rényi entropy, we define n' to be the cardinality of the smallest such subset. Since the distribution of $X_1, \ldots, X_n$ is permuation invariant, such a subset is $X_1, \ldots, X_{n'}$. Thus, $h_\alpha(X_1, \ldots, X_{n'}) = +\infty$ and for no $1 \leq r < n'$ can the Rényi entropy of $(X_1, \ldots, X_r)$ be infinite. We now construct the SP $\{\tilde{Y}_k\}$ by drawing

$$\ldots, \tilde{Y}_{-n'+1}^{0}, \tilde{Y}_1^{n'}, \tilde{Y}_{n'+1}^{2n'}, \ldots, \sim \text{IID } f_{n'}.$$

where f' is the joint density of $X_1, \ldots, X_{n'}$. We then proceed as in the first case but with a random delay that is uniform on $\{0, \ldots, n'-1\}$.

As special cases we now obtain:

Corollary II.3 (Rényi Rate under a Mean Constraint)**.**

1) *For every $\alpha > 1$, every $\mu > 0$, and every $\varepsilon > 0$ there exists a stationary SP $\{Y_k\}$ whose Rényi entropy rate exceeds $\log(\mu e) - \varepsilon$ and which satisfies*

$$Y_k \geq 0 \quad \mathrm{E}[Y_k] = \mu. \tag{47}$$

2) *For every $0 < \alpha < 1$, every $\mu > 0$, and every $M > 0$ there exists a stationary SP $\{Y_k\}$ whose Rényi entropy rate exceeds M and which satisfies* (47).

We can similarly treat a second moment constraint. Here we note that the densities we have proposed are isotropic, and we can thus establish that the SP is centered and uncorrelated.

Corollary II.4 (Rényi Rate under a Variance Constraint)**.**

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

$$\mathrm{E}[Y_k Y_{k'}] = \sigma^2 1\{k = k'\}. \tag{48}$$

2) *For every $0 < \alpha < 1$, every $\sigma > 0$, and every $M > 0$ there exists a centered stationary SP $\{Y_k\}$ whose Rényi entropy rate exceeds M and which satisfies* (48).

Some of the implications of Corollary II.4 are explored in [5].

References

[1] L. Wang and M. Madiman, "Beyond the entropy power inequality, via rearrangements," *arXiv preprint arXiv:1307.6018*, 2013.

[2] T. Cover and J. Thomas, *Elements of information theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006.

[3] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 269–287, 2007.

[4] 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*. Springer, 2003, pp. 211–226.

[5] C. Bunte and A. Lapidoth, "On maximum Rényi entropy rate," in *Joint Workshop on Coding and Communications (JWCC) November 14-15, 2014, Barcelona, Catalonia*, 2014.