Codes for Tasks and Rényi Entropy Rate

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

Abstract—A task is randomly drawn from a finite set of tasks and is described using a fixed number of bits. All the tasks that share its description must be performed. Upper and lower bounds on the minimum ρ-th moment of the number of performed tasks are derived. The key is an analog of the Kraft Inequality for partitions of finite sets. When a sequence of tasks is produced by a source of a given Rényi entropy rate of order $1/(1+\rho)$ and n tasks are jointly described using nR bits, it is shown that for R larger than the Rényi entropy rate, the ρ-th moment of the ratio of performed tasks to n can be driven to one as n tends to infinity, and that for R less than the Rényi entropy rate it tends to infinity. This generalizes a recent result for IID sources by the same authors. A mismatched version of the direct part is also considered, where the code is designed according to the wrong law. The penalty incurred by the mismatch can be expressed in terms of a divergence measure that was shown by Sundaresan to play a similar role in the Massey-Arikan guessing problem.

I. INTRODUCTION

You are asked to complete a task X drawn according to a PMF P from a finite set of tasks $\mathcal{X}$. You do not get to see X but only its description $f(X)$, where

$$f\colon \mathcal{X} \to \{1, \ldots, M\}. \tag{1}$$

In other words, X is described to you using $\log M$ bits. You know the mapping f and you promise to complete X based on $f(X)$, which leaves you no choice but to complete every task in the set

$$f^{-1}(f(X)) = \{x \in \mathcal{X} : f(x) = f(X)\}. \tag{2}$$

In the interesting case where $M < |\mathcal{X}|$, you will sometimes have to perform multiple tasks, of which all but one are superfluous. (We use $|\cdot|$ to denote the cardinality of sets.)

Given M, the goal is to design f so as to minimize the ρ-th moment of the number of tasks you perform

$$\mathrm{E}\big[|f^{-1}(f(X))|^\rho\big] = \sum_{x \in \mathcal{X}} P(x)|f^{-1}(f(x))|^\rho, \tag{3}$$

where ρ is some given positive number. This minimum is at least one because X is in $f^{-1}(f(X))$; it decreases as M increases; and it is equal to one when $M \geq |\mathcal{X}|$.

Our first result is a pair of upper and lower bounds on this minimum as a function of M. The bounds are expressed in terms of the *Rényi entropy of X of order* $1/(1+\rho)$:

$$H_{\frac{1}{1+\rho}}(X) = \frac{1+\rho}{\rho} \log \sum_{x \in \mathcal{X}} P(x)^{\frac{1}{1+\rho}}. \tag{4}$$

Throughout $\log(\cdot)$ stands for $\log_2(\cdot)$, the logarithm to base 2. For typographic reasons we henceforth use the notation

$$\tilde{\rho} = \frac{1}{1+\rho}, \quad \rho > 0. \tag{5}$$

Theorem I.1. *Let $\rho > 0$.*

1) *For all positive integers M and every $f\colon \mathcal{X} \to \{1, \ldots, M\}$,*

$$\mathrm{E}\big[|f^{-1}(f(X))|^\rho\big] \geq 2^{\rho(H_{\tilde{\rho}}(X) - \log M)}. \tag{6}$$

2) *For every integer $M > \log|\mathcal{X}| + 2$ there exists $f\colon \mathcal{X} \to \{1, \ldots, M\}$ such that*

$$\mathrm{E}\big[|f^{-1}(f(X))|^\rho\big] < 1 + 2^{\rho(H_{\tilde{\rho}}(X) - \log \widetilde{M})}, \tag{7}$$

where $\widetilde{M} = (M - \log|\mathcal{X}| - 2)/4$.

A proof is provided in Section II. The lower bound is essentially [1, Lemma III.1].

Theorem I.1 is particularly useful when applied to the case where a sequence of tasks is produced by a source $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X}$ and the first n tasks $X^n = (X_1, \ldots, X_n)$ are jointly described using nR bits:

$$f\colon \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}. \tag{8}$$

We assume that the order in which the tasks are performed matters and that every n-tuple of tasks in the set $f^{-1}(f(X^n))$ must be performed. The total number of performed tasks is therefore $n|f^{-1}(f(X^n))|$, and the ratio of the number of performed tasks to the number of assigned tasks is $|f^{-1}(f(X^n))|$.

Theorem I.2. *Let $\{X_i\}_{i=1}^{\infty}$ be any source with finite alphabet $\mathcal{X}$.*

1) *If $R > \limsup_{n\to\infty} H_{\tilde{\rho}}(X^n)/n$, then there exist encoders $f_n\colon \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}$ such that[1]*

$$\lim_{n\to\infty} \mathrm{E}\big[|f_n^{-1}(f_n(X^n))|^\rho\big] = 1. \tag{9}$$

2) *If $R < \liminf_{n\to\infty} H_{\tilde{\rho}}(X^n)/n$, then for any choice of encoders $f_n\colon \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}$,*

$$\lim_{n\to\infty} \mathrm{E}\big[|f_n^{-1}(f_n(X^n))|^\rho\big] = \infty. \tag{10}$$

Proof: On account of Theorem I.1, for all n large enough so that $2^{nR} > n \log|\mathcal{X}| + 2$,

$$2^{n\rho\left(\frac{H_{\tilde{\rho}}(X^n)}{n} - R\right)} \leq \min_{f_n\colon \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}} \mathrm{E}\big[|f_n^{-1}(f_n(X^n))|^\rho\big]$$
$$< 1 + 2^{n\rho\left(\frac{H_{\tilde{\rho}}(X^n)}{n} - R + \delta_n\right)}, \tag{11}$$

where $\delta_n \to 0$ as $n \to \infty$. ∎

[1] Throughout 2^{nR} stands for $\lfloor 2^{nR} \rfloor$.

When it exists, the limit

$$\lim_{n \to \infty} \frac{H_\alpha(X^n)}{n} \qquad (12)$$

is called the *Rényi entropy rate of order* α. It exists for a large class of sources, including time-invariant Markov sources [2]–[4]. Theorem I.2 generalizes [1, Theorem IV.1] from IID sources to sources with memory and furnishes an operational characterization of the Rényi entropy rate for all orders in $(0, 1)$. Note that for IID sources the Rényi entropy rate reduces to the Rényi entropy because in this case $H_{\tilde{\rho}}(X^n) = n H_{\tilde{\rho}}(X_1)$.

The proof of the lower bound in Theorem I.1 hinges on the following simple observation.

Proposition I.3. *If* $\mathcal{L}_1, \dots, \mathcal{L}_M$ *is a partition of a finite set* $\mathcal{X}$ *into* M *nonempty subsets (i.e.,* $\bigcup_{m=1}^{M} \mathcal{L}_m = \mathcal{X}$ *and* $\mathcal{L}_m \cap \mathcal{L}_{m'} = \emptyset$ *if, and only if,* $m' \neq m$*), and* $L(x)$ *is the cardinality of the subset containing* x*, then*

$$\sum_{x \in \mathcal{X}} \frac{1}{L(x)} = M. \qquad (13)$$

Proof:

$$\sum_{x \in \mathcal{X}} \frac{1}{L(x)} = \sum_{m=1}^{M} \sum_{x \in \mathcal{L}_m} \frac{1}{L(x)} \qquad (14)$$

$$= \sum_{m=1}^{M} \sum_{x \in \mathcal{L}_m} \frac{1}{|\mathcal{L}_m|} \qquad (15)$$

$$= M. \qquad (16)$$

∎

Note that the reverse of Proposition I.3 is not true in the sense that if $\lambda \colon \mathcal{X} \to \mathbb{N} \triangleq \{1, 2, \dots\}$ satisfies

$$\sum_{x \in \mathcal{X}} \frac{1}{\lambda(x)} = \mu, \qquad (17)$$

then there need not exist a partition of $\mathcal{X}$ into $\lceil \mu \rceil$ subsets such that the cardinality of the subset containing x is at most $\lambda(x)$. A counterexample is $\mathcal{X} = \{a, b, c, d\}$ with $\lambda(a) = 1$, $\lambda(b) = 2$, and $\lambda(c) = \lambda(d) = 4$. In this example, $\mu = 2$, but we need 3 subsets to satisfy the cardinality constraints.

However, as our next result shows, allowing a slightly larger number of subsets suffices:

Proposition I.4. *If* $\mathcal{X}$ *is a finite set,* $\lambda \colon \mathcal{X} \to \mathbb{N} \cup \{+\infty\}$ *and*

$$\sum_{x \in \mathcal{X}} \frac{1}{\lambda(x)} = \mu \qquad (18)$$

(with the convention $1/\infty = 0$*), then there exists a partition of* $\mathcal{X}$ *into at most*

$$\min_{\alpha > 1} \lfloor \alpha \mu + \log_\alpha |\mathcal{X}| + 2 \rfloor \qquad (19)$$

subsets such that

$$L(x) \leq \min\{\lambda(x), |\mathcal{X}|\}, \quad \text{for all } x \in \mathcal{X}, \qquad (20)$$

where $L(x)$ *is the cardinality of the subset containing* x*.*

Proposition I.4 is the key to the upper bound in Theorem I.1. Combined with Proposition I.3 it can be considered an analog of the Kraft Inequality [5, Theorem 5.5.1] for partitions of finite sets. A proof is given in Section III.

The construction of the encoder in the derivation of the upper bound in Theorem I.1 requires knowledge of the distribution P of X (see Section II-B). In Section IV we consider a mismatched version of this direct part where the construction is carried out based on the law Q instead of P. We show that the penalty incurred by the mismatch between P and Q can be expressed in terms of the divergence measures

$$\Delta_\alpha(P||Q) \triangleq \log \frac{\sum_{x \in \mathcal{X}} Q(x)^\alpha}{\left(\sum_{x \in \mathcal{X}} P(x)^\alpha\right)^{\frac{1}{1-\alpha}}} \left(\sum_{x \in \mathcal{X}} \frac{P(x)}{Q(x)^{1-\alpha}} \right)^{\frac{\alpha}{1-\alpha}}, \qquad (21)$$

where α can be any positive number not equal to one. (We use the convention $0/0 = 0$ and $a/0 = +\infty$ if $a > 0$.) This family of divergence measures was proposed by Sundaresan [6], who showed that it plays a similar role in the Massey-Arikan guessing problem [7], [8].

II. Proof of Theorem I.1

A. The Lower Bound (Converse)

The proof of the lower bound is inspired by the proof of [8, Theorem 1]. Fix an encoder $f \colon \mathcal{X} \to \{1, \dots, M\}$, and note that it gives rise to a partition of $\mathcal{X}$ into the M subsets

$$\{x \in \mathcal{X} : f(x) = m\}, \quad m \in \{1, \dots, M\}. \qquad (22)$$

Let N denote the number of nonempty subsets in this partition. Also note that for this partition the cardinality of the subset containing x is

$$L(x) = |f^{-1}(f(x))|, \quad \text{for all } x \in \mathcal{X}. \qquad (23)$$

Recall Hölder's Inequality: If $a, b \colon \mathcal{X} \to [0, \infty)$, $p, q > 1$ and $1/p + 1/q = 1$, then

$$\sum_{x \in \mathcal{X}} a(x)b(x) \leq \left(\sum_{x \in \mathcal{X}} a(x)^p \right)^{1/p} \left(\sum_{x \in \mathcal{X}} b(x)^q \right)^{1/q}. \qquad (24)$$

Rearranging (24) gives

$$\sum_{x \in \mathcal{X}} a(x)^p \geq \left(\sum_{x \in \mathcal{X}} b(x)^q \right)^{-p/q} \left(\sum_{x \in \mathcal{X}} a(x)b(x) \right)^p. \qquad (25)$$

Substituting $p = 1 + \rho$, $q = (1 + \rho)/\rho$, $a(x) = P(x)^{\frac{1}{1+\rho}} |f^{-1}(f(x))|^{\frac{\rho}{1+\rho}}$ and $b(x) = |f^{-1}(f(x))|^{-\frac{\rho}{1+\rho}}$ in (25), we obtain

$$\sum_{x \in \mathcal{X}} P(x) |f^{-1}(f(x))|^\rho \qquad (26)$$

$$\geq \left(\sum_{x \in \mathcal{X}} \frac{1}{|f^{-1}(f(x))|} \right)^{-\rho} \left(\sum_{x \in \mathcal{X}} P(x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \qquad (27)$$

$$= 2^{\rho(H_{\tilde{\rho}}(X) - \log N)} \qquad (28)$$

$$\geq 2^{\rho(H_{\tilde{\rho}}(X) - \log M)}, \qquad (29)$$

where (28) follows from (4), (23), and Proposition I.3; and where (29) follows because $N \leq M$.

∎

B. The Upper Bound (Direct Part)

Since Hölder's Inequality (24) holds with equality if, and only if, (iff) $a(x)^p$ is proportional to $b(x)^q$, it follows that the lower bound in Theorem I.1 holds with equality iff $|f^{-1}(f(x))|$ is proportional to $P(x)^{-1/(1+\rho)}$. We derive the upper bound in Theorem I.1 by constructing a partition that approximately satisfies this relationship. To this end, we use Proposition I.4 with $\alpha = 2$ in (19) and

$$\lambda(x) = \begin{cases} \left\lceil \beta P(x)^{-\frac{1}{1+\rho}} \right\rceil & \text{if } P(x) > 0, \\ +\infty & \text{if } P(x) = 0, \end{cases} \quad (30)$$

where we choose β just large enough to guarantee the existence of a partition of $\mathcal{X}$ into at most M subsets satisfying (20). This is accomplished by the choice

$$\beta = \frac{2 \sum_{x \in \mathcal{X}} P(x)^{\frac{1}{1+\rho}}}{M - \log|\mathcal{X}| - 2}. \quad (31)$$

(This is where we need $M > \log|\mathcal{X}| + 2$.) Indeed,

$$\mu = \sum_{x \in \mathcal{X}} \frac{1}{\lambda(x)} \quad (32)$$

$$\leq \sum_{x \in \mathcal{X}} \frac{P(x)^{\frac{1}{1+\rho}}}{\beta} \quad (33)$$

$$= \frac{M - \log|\mathcal{X}| - 2}{2}, \quad (34)$$

and hence

$$2\mu + \log|\mathcal{X}| + 2 \leq M. \quad (35)$$

Let then the partition $\mathcal{L}_1, \ldots, \mathcal{L}_N$ with $N \leq M$ be as promised by Proposition I.4, and construct $f : \mathcal{X} \to \{1, \ldots, M\}$ by setting $f(x) = m$ if $x \in \mathcal{L}_m$. For this encoder,

$$\sum_{x \in \mathcal{X}} P(x) |f^{-1}(f(x))|^\rho = \sum_{x : P(x) > 0} P(x) L(x)^\rho \quad (36)$$

$$\leq \sum_{x : P(x) > 0} P(x) \lambda(x)^\rho \quad (37)$$

$$< 1 + 2^{\rho(H_{\bar{\rho}}(X) - \log \widetilde{M})}, \quad (38)$$

where the strict inequality follows from (30) and the inequality

$$\lceil \xi \rceil^\rho < 1 + 2^\rho \xi^\rho, \quad \text{for all } \xi \geq 0, \quad (39)$$

which is easily checked by considering separately the cases $0 \leq \xi \leq 1$ and $\xi > 1$. ∎

III. PROOF OF PROPOSITION I.4

We describe a procedure for constructing a partition of $\mathcal{X}$ with the desired properties. Since the labels do not matter, we may assume for convenience of notation that $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$ and

$$\lambda(1) \leq \lambda(2) \leq \cdots \leq \lambda(|\mathcal{X}|). \quad (40)$$

The first subset in the partition we construct is

$$\mathcal{L}_0 = \{x \in \mathcal{X} : \lambda(x) \geq |\mathcal{X}|\}. \quad (41)$$

If $\mathcal{X} = \mathcal{L}_0$, then the construction is complete and (19) and (20) are clearly satisfied. Otherwise we follow the steps below to construct additional subsets $\mathcal{L}_1, \ldots, \mathcal{L}_M$.

Step 1: If

$$|\mathcal{X} \setminus \mathcal{L}_0| \leq \lambda(1), \quad (42)$$

then we complete the construction by setting $\mathcal{L}_1 = \mathcal{X} \setminus \mathcal{L}_0$ and $M = 1$. Otherwise we set

$$\mathcal{L}_1 = \{1, \ldots, \lambda(1)\} \quad (43)$$

and go to Step 2.

Step $m \geq 2$: If

$$\left| \mathcal{X} \setminus \bigcup_{i=0}^{m-1} \mathcal{L}_i \right| \leq \lambda(|\mathcal{L}_1| + \ldots + |\mathcal{L}_{m-1}| + 1), \quad (44)$$

then we complete the construction by setting $\mathcal{L}_m = \mathcal{X} \setminus \bigcup_{i=0}^{m-1} \mathcal{L}_i$ and $M = m$. Otherwise we let $\mathcal{L}_m$ contain the $\lambda(|\mathcal{L}_1| + \ldots + |\mathcal{L}_{m-1}| + 1)$ smallest elements of $\mathcal{X} \setminus \bigcup_{i=0}^{m-1} \mathcal{L}_i$, i.e., we set

$$\mathcal{L}_m = \{ |\mathcal{L}_1| + \ldots + |\mathcal{L}_{m-1}| + 1, \ldots,$$
$$|\mathcal{L}_1| + \ldots + |\mathcal{L}_{m-1}| + \lambda(|\mathcal{L}_1| + \ldots + |\mathcal{L}_{m-1}| + 1) \} \quad (45)$$

and go to Step $m + 1$.

We next verify that (20) is satisfied and that the total number of subsets $M + 1$ does not exceed (19). Clearly, $L(x) \leq |\mathcal{X}|$ for every $x \in \mathcal{X}$, so to prove (20) we check that $L(x) \leq \lambda(x)$ for every $x \in \mathcal{X}$. It is clear that $L(x) \leq \lambda(x)$ for all $x \in \mathcal{L}_0$. Let $k(x)$ denote the smallest element in the subset containing x. Then $L(x) \leq \lambda(k(x))$ for all $x \in \bigcup_{m=1}^{M} \mathcal{L}_m$ by construction, and since $k(x) \leq x$, we have $\lambda(k(x)) \leq \lambda(x)$ by the assumption (40), and hence $L(x) \leq \lambda(x)$ for all $x \in \mathcal{X}$.

It remains to check that $M + 1$ does not exceed (19). This is clearly true when $M = 1$, so we assume that $M \geq 2$. Since $L(x) = \lambda(k(x))$ for all $x \in \bigcup_{m=1}^{M-1} \mathcal{L}_m$, we have on account of Proposition I.3

$$M = \sum_{x \in \bigcup_{m=1}^{M} \mathcal{L}_m} \frac{1}{L(x)} \quad (46)$$

$$= 1 + \sum_{x \in \bigcup_{m=1}^{M-1} \mathcal{L}_m} \frac{1}{L(x)} \quad (47)$$

$$= 1 + \sum_{x \in \bigcup_{m=1}^{M-1} \mathcal{L}_m} \frac{1}{\lambda(k(x))}. \quad (48)$$

Fix an arbitrary $\alpha > 1$ and let $\mathcal{M}$ be the set of indices $m \in \{1, \ldots, M-1\}$ such that there is an $x \in \mathcal{L}_m$ with $\lambda(x) > \alpha \lambda(k(x))$. We next argue that $|\mathcal{M}| < \log_\alpha |\mathcal{X}|$. To this end, enumerate the indices in $\mathcal{M}$ as $m_1 < m_2 < \cdots < m_{|\mathcal{M}|}$. For each $i \in \{1, \ldots, |\mathcal{M}|\}$ select $x_i \in \mathcal{L}_{m_i}$ such that $\lambda(x_i) > \alpha \lambda(k(x_i))$. Then

$$\lambda(x_1) > \alpha \lambda(k(x_1)) \quad (49)$$

$$\geq \alpha. \quad (50)$$

Note that if $m < m'$ and $x \in \mathcal{L}_m$ and $x' \in \mathcal{L}_{m'}$, then $x < x'$. Thus, $x_1 < k(x_2)$ because $x_1 \in \mathcal{L}_{m_1}$ and $k(x_2) \in \mathcal{L}_{m_2}$, and $m_1 < m_2$. Consequently,

$$\lambda(x_2) > \alpha\lambda(k(x_2)) \tag{51}$$
$$\geq \alpha\lambda(x_1) \tag{52}$$
$$> \alpha^2. \tag{53}$$

Iterating this argument shows that

$$\lambda(x_{|\mathcal{M}|}) > \alpha^{|\mathcal{M}|}. \tag{54}$$

And since $\lambda(x) \leq |\mathcal{X}|$ for $x \in \bigcup_{m=1}^{M} \mathcal{L}_m$ by (41), it follows that $|\mathcal{M}| < \log_\alpha |\mathcal{X}|$. Continuing from (48) with $\mathcal{M}^c \triangleq \{1, \ldots, M-1\} \setminus \mathcal{M}$,

$$M = 1 + |\mathcal{M}| + \sum_{x \in \bigcup_{m \in \mathcal{M}^c} \mathcal{L}_m} \frac{1}{\lambda(k(x))} \tag{55}$$

$$< 1 + \log_\alpha |\mathcal{X}| + \alpha \sum_{x \in \bigcup_{m \in \mathcal{M}^c} \mathcal{L}_m} \frac{1}{\lambda(x)} \tag{56}$$

$$\leq 1 + \log_\alpha |\mathcal{X}| + \alpha\mu, \tag{57}$$

where the first inequality follows because $\lambda(x) \leq \alpha\lambda(k(x))$ for $x \in \bigcup_{m \in \mathcal{M}^c} \mathcal{L}_m$, and where the second inequality follows from the hypothesis of the proposition. Since $M+1$ is an integer and $\alpha > 1$ is arbitrary, it follows from (57) that $M+1$ is upper-bounded by (19). ∎

IV. MISMATCH

The key to the upper bound in Theorem I.1 was to use Proposition I.4 with λ as in (30) and (31) to obtain a partition of $\mathcal{X}$ for which the cardinality of the subset containing x is approximately proportional to $P(x)^{-1/(1+\rho)}$. Evidently, this construction requires knowledge of the distribution P of X. In this section, we derive the penalty when P is replaced with Q in (30) and (31). Since it is then still true that

$$\mu \leq \frac{M - \log|\mathcal{X}| - 2}{2}, \tag{58}$$

Proposition I.4 guarantees the existence of a partition of $\mathcal{X}$ into at most M subsets satisfying (20). Constructing f from this partition as in Section II-B and proceeding similarly as in (36) to (38), we obtain

$$\sum_{x \in \mathcal{X}} P(x)|f^{-1}(f(x))|^\rho < 1 + 2^{\rho(H_{\tilde{\rho}}(X) + \Delta_{\tilde{\rho}}(P||Q) - \log \widetilde{M})}, \tag{59}$$

where $\Delta_{\tilde{\rho}}(P||Q)$ is as in (21) and $\widetilde{M}$ is as in Theorem I.1. (Note that $\Delta_{\tilde{\rho}}(P||Q) < \infty$ only if the support of P is contained in the support of Q.) The penalty in the exponent when compared to the upper bound in Theorem I.1 is thus given by $\Delta_{\tilde{\rho}}(P||Q)$. To reinforce this, further note that

$$\Delta_\alpha(P^n||Q^n) = n\Delta_\alpha(P||Q), \tag{60}$$

where P^n and Q^n are the n-fold products of P and Q. Consequently, if the source $\{X_i\}_{i=1}^\infty$ is IID P and we construct

$f_n \colon \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}$ similarly as above based on Q^n instead of P^n, we obtain the bound

$$\mathrm{E}\big[|f_n^{-1}(f_n(X^n))|^\rho\big] < 1 + 2^{n\rho(H_{\tilde{\rho}}(X_1) + \Delta_{\tilde{\rho}}(P||Q) - R + \delta_n)}, \tag{61}$$

where $\delta_n \to 0$ as $n \to \infty$. The RHS of (61) tends to one provided that $R > H_{\tilde{\rho}}(X_1) + \Delta_{\tilde{\rho}}(P||Q)$. Thus, in the IID case $\Delta_{\tilde{\rho}}(P||Q)$ is the rate penalty incurred by the mismatch between P and Q.

We conclude this section with some properties of $\Delta_\alpha(P||Q)$. Properties 1–3 (see below) were given in [6]; we repeat them here for completeness. Note that Rényi's divergence (see, e.g., [9])

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha}, \tag{62}$$

satisfies Properties 1 and 3 but none of the others in general.

Proposition IV.1. *Let* $\mathrm{supp}(P)$ *and* $\mathrm{supp}(Q)$ *denote the support sets of* P *and* Q. *The functional* $\Delta_\alpha(P||Q)$ *has the following properties.*

1) $\Delta_\alpha(P||Q) \geq 0$ *with equality iff* $P = Q$.
2) $\Delta_\alpha(P||Q) = \infty$ *iff* $(0 < \alpha < 1$ *and* $\mathrm{supp}(P) \nsubseteq \mathrm{supp}(Q))$ *or* $(\alpha > 1$ *and* $\mathrm{supp}(P) \cap \mathrm{supp}(Q) = \emptyset.)$
3) $\lim_{\alpha \to 1} \Delta_\alpha(P||Q) = D(P||Q)$.
4) $\lim_{\alpha \to 0} \Delta_\alpha(P||Q) = \log \frac{|\mathrm{supp}(Q)|}{|\mathrm{supp}(P)|}$ *if* $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$.
5) $\lim_{\alpha \to \infty} \Delta_\alpha(P||Q) = \log \frac{\max_{x \in \mathcal{X}} P(x)}{\frac{1}{|\mathcal{Q}|} \sum_{x' \in \mathcal{Q}} P(x')}$, *where*

$$\mathcal{Q} = \big\{x \in \mathcal{X} : Q(x) = \max_{x' \in \mathcal{X}} Q(x')\big\}.$$

Proof: Property 2 follows by inspection of (21). Properties 3–5 follow by simple calculus. As to Property 1, consider first the case where $0 < \alpha < 1$. In view of Property 2, we may assume that $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$. Hölder's Inequality (24) with $p = 1/\alpha$ and $q = 1/(1-\alpha)$ gives

$$\sum_{x \in \mathcal{X}} P(x)^\alpha = \sum_{x \in \mathrm{supp}(P)} \frac{P(x)^\alpha}{Q(x)^{\alpha(1-\alpha)}} Q(x)^{\alpha(1-\alpha)} \tag{63}$$

$$\leq \left(\sum_{x \in \mathrm{supp}(P)} \frac{P(x)}{Q(x)^{1-\alpha}} \right)^\alpha \left(\sum_{x \in \mathrm{supp}(P)} Q(x)^\alpha \right)^{1-\alpha}$$

$$\leq \left(\sum_{x \in \mathcal{X}} \frac{P(x)}{Q(x)^{1-\alpha}} \right)^\alpha \left(\sum_{x \in \mathcal{X}} Q(x)^\alpha \right)^{1-\alpha}. \tag{64}$$

The conditions for equality in Hölder's Inequality imply that equality holds iff $P = Q$. Consider next the case where $\alpha > 1$. By Hölder's Inequality with $p = \alpha$ and $q = \alpha/(\alpha - 1)$,

$$\sum_{x \in \mathcal{X}} \frac{P(x)}{Q(x)^{1-\alpha}} = \sum_{x \in \mathcal{X}} P(x)Q(x)^{\alpha-1} \tag{65}$$

$$\leq \left(\sum_{x \in \mathcal{X}} P(x)^\alpha \right)^{\frac{1}{\alpha}} \left(\sum_{x \in \mathcal{X}} Q(x)^\alpha \right)^{\frac{\alpha-1}{\alpha}}, \tag{66}$$

with equality iff $P = Q$. ∎

REFERENCES

[1] C. Bunte and A. Lapidoth, "Source coding, lists, and Rényi entropy," in *Information Theory Workshop (ITW), 2013 IEEE*, 2013, pp. 350–354.

[2] Z. Rached, F. Alajaji, and L. Campbell, "Rényi's divergence and entropy rates for finite alphabet Markov sources," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1553–1561, 2001.

[3] C.-E. Pfister and W. Sullivan, "Rényi entropy, guesswork moments, and large deviations," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2794–2800, 2004.

[4] D. Malone and W. G. Sullivan, "Guesswork and entropy," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 525–526, 2004.

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

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

[7] J. Massey, "Guessing and entropy," in *Information Theory Proceedings (ISIT), 1994 IEEE International Symposium on*, 1994, p. 204.

[8] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[9] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 26–34, 1995.