Non-Asymptotic Bounds for Optimal Fixed-to-Variable Lossless Compression without Prefix Constraints

Igal Sason
Andrew and Erna Viterbi Faculty of Electrical Engineering
Technion-Israel Institute of Technology
Haifa 32000, Israel
E-mail: sason@ee.technion.ac.il

Sergio Verdú
Department of Electrical Engineering
Princeton University
New Jersey 08544, USA
E-mail: verdu@princeton.edu

***Abstract*—Bounds on optimal guessing moments serve to improve non-asymptotic bounds on the cumulant generating function of the codeword lengths for fixed-to-variable optimal lossless source coding without prefix constraints. Non-asymptotic bounds on the reliability function of discrete memoryless sources are presented as well. Lower bounds on the cumulant generating function of the codeword lengths are given, by means of the smooth Rényi entropy, for source codes that allow decoding errors.**

Index Terms – Cumulant generating function, lossless source coding, Rényi information measures.

I. Introduction

One of the major achievements in information theory is the development of lossless data compression algorithms, and the derivation of the fundamental limits that establish the highest achievable compression efficiency that is compatible with perfect data recovery. For uniquely-decodable lossless source coding, Campbell ([3], [4]) proposed the normalized cumulant generating function of the codeword lengths as a generalization to the frequently used design criterion of normalized average code length. Campbell's motivation in [3] was to control the contribution of the longer codewords via a free parameter in the cumulant generating function: if the value of this parameter tends to zero, then the resulting design criterion becomes the normalized average code length; on the other hand, by increasing the value of the free parameter, the penalty for longer codewords becomes more severe, and the resulting code optimization yields a reduction in the fluctuations of the codeword lengths. In [3], asymptotically tight upper and lower bounds on the minimum normalized cumulant generating function were obtained for discrete memoryless stationary sources with finite alphabet. These bounds, expressed in terms of the Rényi entropy, imply that for sufficiently long source sequences, it is possible to make the normalized cumulant generating function of the codeword lengths approach the Rényi entropy as closely as desired by a proper fixed-to-variable uniquely-decodable source code; moreover, a converse result in [3] shows that there is no uniquely-decodable source code for which the normalized cumulant generating function of its codeword lengths lies below the Rényi entropy. In addition, this type of bounds was studied in the context of other various coding problems, including guessing.

Kontoyiannis and Verdú [8] studied the behavior of the best achievable rate and other fundamental limits in variable-rate lossless source compression without prefix constraints. In the non-asymptotic regime, the fundamental limits of fixed-to-variable lossless compression with and without prefix constraints were shown to be tightly coupled. Reference [8] obtains non-asymptotic upper and lower bounds on the distribution of codeword lengths, along with a rigorous proof of the Gaussian approximation put forward in 1962 by Strassen [17] for memoryless sources. An alternative approach was followed by Courtade and Verdú in [5], where they derived non-asymptotic bounds for the normalized cumulant generating function of the codeword lengths for optimal variable-length lossless codes without prefix constraints; these bounds are used in [5] to obtain simple proofs of the asymptotic normality and the reliability function of memoryless sources allowing countably infinite alphabets.

In [10], Kostina *et al.* studied the fundamental limits of the minimum average length of lossless and lossy variable-length compression, allowing a nonzero error probability $\varepsilon \in [0, 1)$ for almost lossless compression. The bounds in [10] were used to obtain a Gaussian approximation on the speed of convergence of the minimum average length, which was shown to be quite accurate for all but small blocklengths. In [7], Koga and Yamamoto followed an information-spectrum approach to obtain asymptotic properties of the codeword lengths for prefix fixed-to-variable source codes, allowing decoding errors. This work was refined in the non-asymptotic setting by Kuzuoka [9], which bounds the cumulant generating function of the codeword lengths via the smooth Rényi entropy.

Section II defines the Rényi measures used in this paper. Section III provides improved bounds on the normalized cumulant generating function of the codeword lengths for fixed-to-variable optimal codes, and on the non-asymptotic reliability function of discrete memoryless sources, tightening the bounds by Courtade and Verdú [5]. Due to space limitations, proofs are omitted here and appear in [15, Section 4].

II. Unconditional and Smooth Rényi Entropy

The information measures used in this paper apply to discrete random variables. All the definitions in this section extend in a natural way to random vectors.

Definition 1: [12] Let X be a discrete random variable taking values on a finite or countably infinite set $\mathcal{X}$, and let P_X be its probability mass function. The Rényi entropy of order $\alpha \in (0,1) \cup (1,\infty)$ is given by[1]

$$H_\alpha(X) = H_\alpha(P_X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X^\alpha(x). \tag{1}$$

By its continuous extension,

$$H_0(X) = \log \big|\{x \in \mathcal{X} \colon P_X(x) > 0\}\big|, \tag{2}$$

$$H_1(X) = H(X), \tag{3}$$

$$H_\infty(X) = \log \frac{1}{\max\limits_{x \in \mathcal{X}} P_X(x)}. \tag{4}$$

Another Rényi information measure used in this paper is the smooth Rényi entropy, introduced by Renner and Wolf [14] (after a different definition in [13]).

Definition 2: [14] Let X be a discrete random variable taking values on $\mathcal{X}$, and let P_X denote the probability mass function of X. Let $\alpha \in (0,1) \cup (1,\infty)$ and $\varepsilon \in [0,1)$. The *ε-smooth Rényi entropy of order α* is given by

$$H_\alpha^{(\varepsilon)}(X) = \frac{1}{1-\alpha} \min_{\mu \in B^{(\varepsilon)}(P_X)} \log \sum_{x \in \mathcal{X}} \mu^\alpha(x) \tag{5}$$

$$B^{(\varepsilon)}(P_X) \triangleq \Big\{\mu \colon \mathcal{X} \to [0,1] : \sum_{x \in \mathcal{X}} \mu(x) \geq 1-\varepsilon, \quad \mu(x) \leq P_X(x),\ \forall x \in \mathcal{X}\Big\}. \tag{6}$$

The ε-smooth Rényi entropy becomes the Rényi entropy when $\varepsilon = 0$, i.e.,

$$H_\alpha^{(0)}(X) = H_\alpha(X), \quad \alpha \in (0,1) \cup (1,\infty). \tag{7}$$

III. Non-Asymptotic Bounds for Optimal Fixed-to-Variable Lossless Compression

This section applies the improved bounds on guessing moments in [15, Section 3] to analyze non-prefix one-to-one binary optimal codes, which do not satisfy Kraft's inequality. These codes are one-shot codes that assign distinct codewords to source symbols; their average length, which is smaller than the Shannon entropy of the source, is analyzed in [18].

A. Basic setup, notation and preliminaries

Definition 3: A variable-length lossless compression binary code for a discrete set $\mathcal{X}$ is an injective mapping:

$$f \colon \mathcal{X} \to \{0,1\}^* = \{\varnothing, 0, 1, 00, 01, 10, 11, 000, \ldots\} \tag{8}$$

where $f(x)$ is the codeword which is assigned to $x \in \mathcal{X}$; the length of this codeword is denoted by $\ell(f(x))$ where $\ell \colon \{0,1\}^* \to \{0,1,2,\ldots\}$ with the convention that $\ell(\varnothing) = 0$.

Definition 4: [19] A variable-length lossless source code is *compact* whenever it contains a codeword only if all shorter codewords also belong to the code.

[1]Unless explicitly stated, the logarithm base can be chosen by the reader, with exp indicating the inverse function of log.

Definition 5: [19] Given a probability mass function P_X on $\mathcal{X}$, a variable-length lossless source code is *P_X-efficient* if for all $(a,b) \in \mathcal{X}^2$,

$$\ell(f(a)) < \ell(f(b)) \Longrightarrow P_X(a) \geq P_X(b). \tag{9}$$

Definition 6: [19] Given a probability mass function P_X on $\mathcal{X}$, a variable-length lossless source code is *P_X-optimal* if it is both compact and P_X-efficient.

The optimality in Definition 6 is justified in Proposition 1. Let $f_X^* \colon \mathcal{X} \to \{0,1\}^*$ be a P_X-optimal variable-length lossless source code. If $|\mathcal{X}| < \infty$, then

a) $\varnothing$ is assigned to the most likely element in $\mathcal{X}$.
b) All the 2^ℓ binary strings of length ℓ are assigned to the 2^ℓ-th through $(2^{\ell+1}-1)$-th most likely elements with $\ell \in \{1, \ldots, \lfloor \log_2(1+|\mathcal{X}|) \rfloor - 1\}$. For example, 0 and 1 (or 1 and 0) are assigned, respectively, to the second and third most likely elements in $\mathcal{X}$.
c) If $\log_2(1+|\mathcal{X}|)$ is not an integer, then codewords of length $\lfloor \log_2(1+|\mathcal{X}|) \rfloor$ are assigned to each of the remaining $1 + |\mathcal{X}| - 2^{\lfloor \log_2(1+|\mathcal{X}|) \rfloor}$ elements in $\mathcal{X}$.

As long as $|\mathcal{X}| > 1$, there is more than one P_X-optimal code since compactness and P_X-efficiency are preserved by swapping codewords of the same length (and, if $|\mathcal{X}| = 2$, then the second most likely element can be either assigned 0 or 1). In the presence of ties among probabilities, the value of $\ell(f_X^*(x))$ for some $x \in \mathcal{X}$ may depend on the choice of f_X^*. The following result provides several relevant properties of optimal codes.

Proposition 1: ([8], [19]) Fix a probability mass function P_X on a finite set $\mathcal{X}$. The following results hold for P_X-optimal codes $f_X^* \colon \mathcal{X} \to \{0,1\}^*$:

a) The distribution of $\ell(f_X^*(X))$ is invariant to the actual choice of f_X^*, and it only depends on P_X.
b) For every lossless data compression code f, and for all $r \geq 0$,
$$\mathbb{P}\big[\ell\big(f(X)\big) \leq r\big] \leq \mathbb{P}\big[\ell\big(f_X^*(X)\big) \leq r\big]. \tag{10}$$
Furthermore, the inequality in (10) is strict for some $r \geq 0$ if f is not P_X-optimal.
c)
$$\sum_{x \in \mathcal{X}} 2^{-\ell(f_X^*(x))} \leq \log_2(1+|\mathcal{X}|) \tag{11}$$
with equality if and only if $|\mathcal{X}|+1$ is a positive integral power of 2. Furthermore, all compact codes for $\mathcal{X}$ achieve the same value of $\sum\limits_{x \in \mathcal{X}} 2^{-\ell(f(x))}$, which is larger than that achieved by a non-compact code.

Definition 7: The *cumulant generating function* of the codeword lengths of P_X-optimal binary codes is given by

$$\Lambda^*(\rho) \triangleq \log \mathbb{E}\big[2^{\rho\, \ell(f_X^*(X))}\big], \quad \rho \in \mathbb{R}. \tag{12}$$

Remark 1: (12) is actually a scaled cumulant generating function. The cumulant generating function of a random variable X is given by

$$\Lambda_X(\rho) = \log_{\mathrm{e}} \mathbb{E}\big[e^{\rho X}\big], \quad \rho \in \mathbb{R} \tag{13}$$

whereas, following Campbell [3], it is more natural to study the function given by

$$\widetilde{\Lambda}_X(\rho) = \log \mathbb{E}\big[2^{\rho X}\big]. \tag{14}$$

Note, however, that (13) and (14) satisfy

$$\widetilde{\Lambda}_X(\rho) = \Lambda_X(\rho \log_{\mathrm{e}} 2) \, \log \mathrm{e}, \tag{15}$$

which implies that they can be obtained from each other by proper linear scalings of the axes.

As mentioned in the introduction, the cumulant generating function of the codeword lengths provides an important design criterion. In particular, it yields the average length via the equality

$$\lim_{\rho \to 0} \frac{\Lambda^*(\rho)}{\rho} = \mathbb{E}[\ell(f_X^*(X))]. \tag{16}$$

Theorem 1: [5, Theorem 1] If $\rho \in (-\infty, -1]$, then

$$H_\infty(X) - \log \log_2(1 + |\mathcal{X}|) \le -\Lambda^*(\rho) \le H_\infty(X), \tag{17}$$

and, if $\rho \in (-1, 0) \cup (0, \infty)$, then

$$H_{\frac{1}{1+\rho}}(X) - \log \log_2(1 + |\mathcal{X}|) \le \frac{\Lambda^*(\rho)}{\rho} \le H_{\frac{1}{1+\rho}}(X). \tag{18}$$

By invoking the Chernoff bound and using Theorem 1, the following result holds.

Theorem 2: [5, Theorem 2] For all $H(X) < R < \log |\mathcal{X}|$

$$\log \frac{1}{\mathbb{P}[\ell(f_X^*(X)) \ge R]} \ge \sup_{\rho > 0} \Big\{ \rho R - \rho H_{\frac{1}{1+\rho}}(X) \Big\} \tag{19}$$

$$= D(X_\alpha \| X) \tag{20}$$

where $\alpha \in (0, 1)$ is a function of R chosen so that $R = H(X_\alpha)$, and X_α has the scaled probability mass function

$$P_{X_\alpha}(x) = \frac{P_X^\alpha(x)}{\sum\limits_{a \in \mathcal{X}} P_X^\alpha(a)}, \quad x \in \mathcal{X}. \tag{21}$$

B. Improved bounds on the distribution of the optimal code lengths

We provide bounds on the cumulant generating function and the complementary cumulative distribution of optimal lengths for lossless compression of a random variable X which takes values on a finite set $\mathcal{X}$. These bounds improve those in Theorems 1 and 2, and in Section III-C we use them to derive non-asymptotic bounds for optimal fixed-to-variable lossless codes. Due to space limitations, proofs are omitted here and they are available in [15, Section 4].

We start by generalizing [5, Lemma 1] from $\beta = 1$ to arbitrary $\beta \in \mathbb{R}$.

Lemma 1: For an optimal binary code, and for all $\beta \in \mathbb{R}$

$$\sum_{x \in \mathcal{X}} 2^{-\beta \, \ell(f_X^*(x))} = \begin{cases} (2^\Delta - 1) s_\beta^m + \dfrac{1 - s_\beta^m}{1 - s_\beta}, & \beta \ne 1 \\ m + 2^\Delta - 1, & \beta = 1 \end{cases} \tag{22}$$

$$\triangleq t(\beta, |\mathcal{X}|) \tag{23}$$

where

$$s_\beta = 2^{1-\beta}, \tag{24}$$

$$m = \big\lfloor \log_2(1 + |\mathcal{X}|) \big\rfloor, \tag{25}$$

$$\Delta = \log_2(1 + |\mathcal{X}|) - m \in [0, 1). \tag{26}$$

Lemma 2: Let X be a random variable taking values on a finite set $\mathcal{X}$, and let $\rho \ne 0$. Then, for an optimal binary code,

$$\frac{1}{\rho} \log \mathbb{E}\big[2^{\rho \, \ell(f_X^*(X))}\big] \ge \sup_{\beta \in (-\rho, \infty) \setminus \{0\}} \frac{1}{\beta} \Big[H_{\frac{\beta}{\beta+\rho}}(X) - \log \, t(\beta, |\mathcal{X}|) \Big], \tag{27}$$

where $t(\cdot)$ is defined in (23).

The problem of guessing discrete random variables has been extensively studied in the information theory literature (see, e.g., [1], [11], [15] and references therein). The central object of interest is the distribution (or the moments) of the number of guesses required to identify a realization of a random variable X, taking values on a finite or countably infinite set $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$, by repeatedly asking questions of the form "Is X equal to x?" until the value of X is guessed correctly. A *guessing function* is a one-to-one function $g \colon \mathcal{X} \to \mathcal{X}$, which can be viewed as a permutation of the elements of $\mathcal{X}$ in the order in which they are guessed. The average number of guesses is minimized by taking the guessing function to be the *ranking function* g_X, for which $g_X(x) = k$ if $P_X(x)$ is the k-th largest mass [11]. Although the tie breaking affects the choice of g_X, the distribution of $g_X(X)$ does not depend on how ties are resolved. Not only does this strategy minimize the average number of guesses, but it also minimizes the ρ-th moment of the number of guesses for every $\rho > 0$.

The following result tightens [1, Proposition 4] (see [15, Section 3]):

Lemma 3: Let X be a discrete random variable taking values on a set $\mathcal{X}$, and let g_X be the ranking function according to P_X. Then, for all $\rho \ge 0$,

$$\mathbb{E}[g_X^\rho(X)] \le \frac{1}{1+\rho} \Big[\exp\Big(\rho H_{\frac{1}{1+\rho}}(X) \Big) - 1 \Big] + \exp\Big((\rho - 1)^+ H_{\frac{1}{\rho}}(X) \Big) \tag{28}$$

where $(x)^+ \triangleq \max\{x, 0\}$ for $x \in \mathbb{R}$.

Lemma 4: Let X be a random variable taking values on a finite set $\mathcal{X}$, and let g_X be a ranking function of X. Then, for every optimal binary code and for all $\rho > 0$,

$$2^{-\rho} \, \mathbb{E}[g_X^\rho(X)] < \mathbb{E}\big[2^{\rho \, \ell(f_X^*(X))}\big] \le \mathbb{E}[g_X^\rho(X)]. \tag{29}$$

This leads us to the following result:

Theorem 3: Let X be a random variable taking values on a finite set $\mathcal{X}$. Then, for every optimal binary code, the cumulant

generating function in (12) satisfies

$$\sup_{\beta\in(-\rho,\infty)\setminus\{0\}} \frac{1}{\beta}\left[H_{\frac{\beta}{\beta+\rho}}(X) - \log\, t(\beta,|\mathcal{X}|)\right] \leq \frac{\Lambda^*(\rho)}{\rho} \tag{30}$$

$$\leq H_{\frac{1}{1+\rho}}(X) + \frac{1}{\rho}\,\log\!\Bigg(\frac{1}{1+\rho}\left[1-\exp\left(-\rho H_{\frac{1}{1+\rho}}(X)\right)\right] + \exp\left((\rho-1)^+ H_{\frac{1}{\rho}}(X) - \rho H_{\frac{1}{1+\rho}}(X)\right)\Bigg), \tag{31}$$

for all $\rho > 0$, where $t(\cdot)$ is given in (23). Moreover, (30) also holds for $\rho < 0$.

Proof: The lower bound in the left side of (30) is Lemma 2, and the upper bound in the right side of (31) follows from Lemmas 3 and 4. ∎

Remark 2: For $\rho \in (-1,0)\cup(0,\infty)$, loosening the bound in the left side of (30) by the sub-optimal choice of $\beta = 1$ and invoking $t(1,|\mathcal{X}|) \leq \log_2(1+|\mathcal{X}|)$ (in view of Lemma 1, and since $2^x - 1 \leq x$ for $x \in [0,1]$) recovers the lower bound in (18).

Remark 3: The second term in the right side of (31) is non-positive for all $\rho \geq 1$ [15]; due to the non-negativity of the Rényi entropy, this also holds for $\rho \in (0,1)$. Hence, for $\rho > 0$, the upper bound in (31) improves the bound in the right side of (18).

The Chernoff bound and (31) readily yield the following lower bound.

Theorem 4: In the setting of Theorem 3, for $R < \log|\mathcal{X}|$,

$$\log\left(\frac{1}{\mathbb{P}\big[\ell(f_X^*(X)) > R\big]}\right) \geq \sup_{\rho>0}\Bigg\{\rho R - \rho H_{\frac{1}{1+\rho}}(X) - \log\!\Bigg(\frac{1}{1+\rho}\left[1-\exp\left(-\rho H_{\frac{1}{1+\rho}}(X)\right)\right] + \exp\left((\rho-1)^+ H_{\frac{1}{\rho}}(X) - \rho H_{\frac{1}{1+\rho}}(X)\right)\Bigg)\Bigg\}. \tag{32}$$

C. Non-asymptotic bounds for fixed-to-variable lossless source codes

We consider the fixed-to-variable-length lossless compression in Definition 6 where the object to be compressed $x^n = (x_1,\ldots,x_n) \in \mathcal{A}^n$ is a string of length n (n is known to both encoder and decoder), whose letters are drawn from a finite alphabet $\mathcal{X}$ according to the probability mass function $P_{X^n}(x^n) = \prod_{i=1}^{n} P_X(x_i)$ for all $x^n \in \mathcal{A}^n$. We consider the following non-asymptotic measures for optimal fixed-to-variable lossless compression:

- The cumulant generating function of the codeword lengths is given by

$$\Lambda_n(\rho) := \frac{1}{n}\,\log\mathbb{E}\left[2^{\rho\,\ell(f_{X^n}^*(X^n))}\right], \quad \rho\in\mathbb{R}. \tag{33}$$

- The non-asymptotic version of the source reliability function is given by

$$E_n(R) = \frac{1}{n}\,\log\left(\frac{1}{\mathbb{P}\big[\frac{1}{n}\,\ell(f_{X^n}^*(X^n)) \geq R\big]}\right). \tag{34}$$

In view of Theorems 3 and 4, the following result is obtained (see [15]):

Theorem 5: Consider a memoryless and stationary source of finite alphabet $\mathcal{X}$, and let $f_{X^n}^*: \mathcal{A}^n \to \{0,1\}^*$ be an optimal compression code. Then, the following bounds hold:

a) For all $\rho > 0$

$$\sup_{\beta\in(-\rho,\infty)\setminus\{0\}} \frac{\rho}{\beta}\left[H_{\frac{\beta}{\beta+\rho}}(X) - \frac{1}{n}\,\log t(\beta,|\mathcal{A}|^n)\right] \leq \Lambda_n(\rho) \tag{35}$$

$$\leq \rho H_{\frac{1}{1+\rho}}(X) + \frac{1}{n}\,\log\!\Bigg(\frac{1}{1+\rho}\left[1-\exp\left(-n\rho H_{\frac{1}{1+\rho}}(X)\right)\right] + \exp\left(n\left[(\rho-1)^+ H_{\frac{1}{\rho}}(X) - \rho H_{\frac{1}{1+\rho}}(X)\right]\right)\Bigg)$$

where $t(\cdot)$ is as defined in (23).

b) For $R < \log|\mathcal{X}|$

$$E_n(R) \geq \sup_{\rho>0}\Bigg\{\rho R - \rho H_{\frac{1}{1+\rho}}(X) - \frac{1}{n}\log\!\Bigg(\frac{1}{1+\rho}\left[1-\exp\left(-n\rho H_{\frac{1}{1+\rho}}(X)\right)\right] + \exp\left(n\left[(\rho-1)^+ H_{\frac{1}{\rho}}(X) - \rho H_{\frac{1}{1+\rho}}(X)\right]\right)\Bigg)\Bigg\}. \tag{36}$$

Remark 4: The non-asymptotic bounds on the cumulant generating function in (35) recover the known asymptotic result in [5, (29)] where for all $\rho > 0$

$$\Lambda(\rho) := \lim_{n\to\infty}\Lambda_n(\rho) = \rho H_{\frac{1}{1+\rho}}(X), \tag{37}$$

which, incidentally, coincides with Arikan's asymptotic fundamental limit for $\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[g_{X^n}^{\rho}(X^n)]$ when X^n is i.i.d. [1]. To this end, note that $\lim_{n\to\infty}\frac{1}{n}\log t(1,|\mathcal{A}|^n) = 0$, and selecting $\beta = 1$ in the left side of (35) yields

$$\underline{\lim_{n\to\infty}}\,\Lambda_n(\rho) \geq \rho H_{\frac{1}{1+\rho}}(X). \tag{38}$$

Moreover, since $H_\alpha(X)$ is monotonically non-increasing in α,

$$\rho H_{\frac{1}{1+\rho}}(X) - (\rho-1)^+ H_{\frac{1}{\rho}}(X) \geq \min\{\rho,1\}\, H_{\frac{1}{\rho}}(X) \tag{39}$$

and, if X is non-deterministic, then the rightmost inequality in (35) and (39) yield

$$\overline{\lim_{n\to\infty}}\,\Lambda_n(\rho) \leq \rho H_{\frac{1}{1+\rho}}(X), \tag{40}$$

recovering (37) from (38) and (40). Furthermore, (36) and (39) imply that

$$E(R) := \underline{\lim_{n\to\infty}}\,E_n(R) \geq \sup_{\rho>0}\Big\{\rho R - \rho H_{\frac{1}{1+\rho}}(X)\Big\}. \tag{41}$$

Although, as noted in Remark 4, the improvement in the bounds afforded by Theorem 5 washes out asymptotically, the following example illustrates the improvement in the non-asymptotic regime. The following example illustrates the non-asymptotic bounds in this work, and comparing them with the bounds in [5]. Due to space limitations, the reader is referred to plots in the full paper version [15].

Example 1: Consider a discrete memoryless source which emits n letters from the ternary alphabet $\mathcal{A} = \{a, b, c\}$ with $P_X(a) = \frac{4}{7}$, $P_X(b) = \frac{2}{7}$ and $P_X(c) = \frac{1}{7}$. The bounds on the cumulant generating function in [5, Theorem 1] (see (18)) are given by

$$\rho\, H_{\frac{1}{1+\rho}}(X) - \tfrac{\rho}{n}\, \log\log_2(1 + |\mathcal{A}|^n) \leq \Lambda_n(\rho) \leq \rho H_{\frac{1}{1+\rho}}(X)$$

for $\rho > 0$. For numerical results, see [15, Figures 3 and 4]. The match between the upper and lower bounds in Theorem 5 improves by increasing n, and the tightening obtained by the lower bound in Theorem 5 can be significant for small n. Furthermore, numerical results show the improvement of the lower bound on the non-asymptotic source reliability function $E_n(R)$ in Theorem 2 over the bound in (36) for small to moderate values of n.

D. Variable-Length Source Coding Allowing Errors

Following a recent study by Kuzuoka [9], the analysis in [15, Section 5] leads to a derivation of improved lower bounds on the cumulant generating function of the codeword lengths for variable-length source coding allowing errors (which, in contrast to the conventional fixed-to-fixed paradigm, are not necessarily detectable by the decoder) by means of the smooth Rényi entropy in Definition 2. In contrast to [9], the bounds in [15, Section 5] are derived for source codes without the prefix condition when either the maximal or average decoding error probabilities are limited not to exceed a given value $\varepsilon \in [0, 1)$.

Theorem 6: Let X take values on a finite set $\mathcal{X}$, and let $f\colon \mathcal{X} \to \mathcal{C}$ be an encoder (possibly stochastic) with a finite codebook $\mathcal{C} \subseteq \{0,1\}^*$. Let $\ell\colon \mathcal{C} \to \{0, 1, \ldots, \}$ be the length function of the codewords in $\mathcal{C}$. Fix $\varepsilon \in [0, 1)$ and $\rho > 0$.

1) If the *average* decoding error probability cannot be larger than ε, then

$$\frac{1}{\rho}\, \log \mathbb{E}\left[2^{\rho\, \ell(f(X))}\right] \geq \sup_{\beta > 0} \frac{1}{\beta} \left[H^{(\varepsilon)}_{\frac{\beta}{\beta+\rho}}(X) - \log t(\beta, |\mathcal{X}|)\right]$$

where $H^{(\varepsilon)}_{\alpha}(X)$ is the ε-smooth Rényi entropy of order α, and $t(\cdot)$ is given in (23).

2) If the *maximal* decoding error probability cannot be larger than ε, then also

$$\frac{1}{\rho}\, \log \mathbb{E}\left[2^{\rho\, \ell(f(X))}\right] \geq \sup_{\beta \in (-\rho, 0)} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log t(\beta, |\mathcal{X}|)\right] - \frac{1}{\rho}\, \log \frac{1}{1-\varepsilon}.$$

Proof: See [15, Section 5]. ∎

Remark 5: If the maximal decoding error probability cannot be larger than $\varepsilon \in (0, 1)$, then neither of the bounds in Theorem 6 is superseded by the other, as can be verified by numerical experimentation.

The reader is referred to [15, Section 5] for a discussion on Theorem 6, including numerical results of the new improved lower bounds in Theorem 6 with a comparison to [9].

REFERENCES

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

[2] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory - 2nd Colloquium*, Keszthely, Hungary, 1975, Colloquia Mathematica Societatis Janós Bolyai (I. Csiszár and P. Elias editors), Amsterdam, the Netherlands: North Holland, vol. 16, pp. 41-52, 1977.

[3] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, no. 4, pp. 423–429, August 1965.

[4] L. L. Campbell, "Definition of entropy by means of a coding problem," *Probability Theory and Related Fields*, vol. 6, no. 2, pp. 113–118, June 1966.

[5] T. Courtade and S. Verdú, "Cumulant generating function of codeword lengths in optimal lossless compression," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2494–2498, Honolulu, Hawaii, USA, July 2014.

[6] T. Courtade and S. Verdú, "Variable-length lossy compression and channel coding: Non-asymptotic converses via cumulant generating functions," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 2499–2503, Honolulu, Hawaii, USA, July 2014.

[7] H. Koga and H. Yamamoto, "Asymptotic properties on codeword lengths of an optimal FV code for general sources," *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1546–1555, April 2005.

[8] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: non-asymptotics and asymptotics," *IEEE Trans. on Information Theory*, vol. 60, no. 2, pp. 777–795, February 2014.

[9] S. Kuzuoka, "On the smooth Rényi entropy and variable-length source coding allowing errors," *Proceedings of the 2016 IEEE International Symposium on Info. Theory*, pp. 745–749, Barcelona, Spain, July 2016.

[10] V. Kostina, Y. Polyanskiy and S. Verdú, "Variable-length compression allowing errors," *IEEE Trans. on Information Theory*, vol. 61, no. 8, pp. 4316–4330, August 2015.

[11] J. L. Massey, "Guessing and entropy," *Proceedings of the 1994 IEEE International Symposium on Information Theory*, p. 204, Trondheim, Norway, June 1994.

[12] A. Rényi, "On measures of entropy and information," *Proceedings of the 4th Berkeley Symposium on Probability Theory and Mathematical Statistics*, pp. 547–561, Berkeley, California, USA, 1961.

[13] R. Renner and S. Wolf, "Smooth Rényi entropy and applications," *Proceedings of the 2004 IEEE International Symposium on Information Theory*, p. 232, Chicago, Illinois, USA, July 2004.

[14] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," *Advances in Cryptology - ASIACRYPT 2005*, Lecture Notes in Computer Science, vol. 3788, pp. 199–216, Springer, 2005.

[15] I. Sason and S. Verdú, "Improved bounds on lossless source coding and guessing moments via Rényi measures," *IEEE Trans. on Information Theory*, vol. 64, no. 6, pp. 4323–4346, June 2018.

[16] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July-October 1948.

[17] V. Strassen, "Asymptotische abschäzungen in Shannon's informationstheorie," *Transactions of the Third Prague Conference on Information Theory, Statistics, Decision Functions, Random Processes (Liblice, 1962)*. Prague: Publ. House Czech. Acad. Sci., pp. 689–723, 1964.

[18] W. Szpankowski and S. Verdú, "Minimum expected length of fixed-to-variable lossless compression without prefix constraints," *IEEE Trans. on Information Theory*, vol. 57, no. 7, pp. 4017–4025, July 2011.

[19] S. Verdú, *Information Theory*, in preparation.