

Universal Fixed-to-Variable Source Coding in the Finite Blocklength Regime

Oliver Kosut and Lalitha Sankar

Dept. of Electrical, Computer and Energy Engineering,

Arizona State University, Tempe, AZ 85287.

{okosut,lalithasankar}@asu.edu

Abstract—A universal source coding problem is considered in the finite blocklength regime for stationary memoryless sources. A new coding scheme is presented that encodes based on the type class size and the empirical support set of the sequence. It is shown that there is no loss in dispersion relative to the case when the source distribution is known. A new bound is obtained on the third order asymptotic coding rate. Numerical results are presented for finite blocklengths comparing the proposed coding scheme with a variety of coding schemes including Lempel-Ziv.

I. INTRODUCTION

We have entered an era in which large volumes of data are continually generated, accessed and stored in distributed servers. In contrast to the traditional data storage models in which large blocks of data are compressed and stored for lossless recovery, the evolving information access and storage model requires compressing (losslessly for many applications) arbitrarily small blocks of data over a range of time scales. Typical examples include online retailers and social network sites that are continuously collecting, storing, and analyzing user data for a variety of purposes.

Finite blocklength compression schemes could be well suited to these applications. There has been renewed interest in finite blocklength source and channel coding (see, for e.g., [1], [2], [3], [4], [5], [6]). The finite blocklength (near-) lossless source coding literature typically assumes knowledge of the underlying source distribution [3], [6]. In general, however, the distribution may neither be known *a priori* nor easy to estimate reliably in the small blocklength regime. There exist many well known and well studied lossless compression codes with asymptotically proven performance (e.g., [7]). The cost of universality in lossless source coding has been studied in [8], and more recently [9], the latter studying the problem in the finite length regime. These works differ from ours in that they assume prefix-free codes, and they use redundancy as a performance metric. Instead, we do not make the assumption that the code is prefix-free, and for our performance metric we use the probability of the code-length exceeding a given number of bits, more in-keeping with the finite blocklength literature (e.g., [3]). These appear to change the problem, as our bound on the third-order coding rate differs from the corresponding one from [8].

We consider a fixed-to-variable length coding scheme for a stationary memoryless, often referred to as independent and identically distributed (i.i.d.), source with unknown distribu-

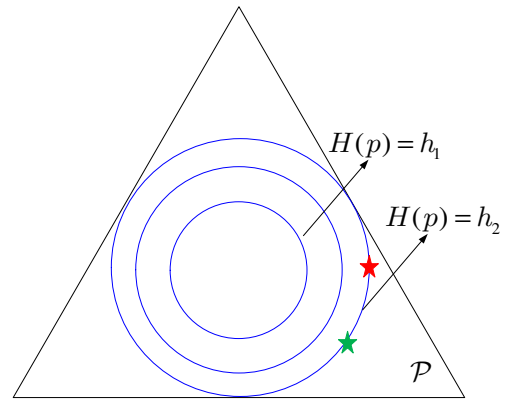


Fig. 1. A probability simplex \mathcal{P} and entropy contours.

tion P_X . For such a source, the minimal number of bits required to compress a length n sequence with probability $(1 - \epsilon)$ is at most¹

$$nH + \sqrt{nV}Q^{-1}(\epsilon) + c \log(n) + O(1), \quad (1)$$

where for the case when the source distribution is known [10]², we have

$$c = -\frac{1}{2}. \quad (2)$$

In this paper, we show that for the universal case, (1) holds with the same second-order dispersion term $\sqrt{nV}Q^{-1}(\epsilon)$ and we bound c as

$$c \leq \frac{1}{2} (|\mathcal{X}_{P_X}| - 3) \quad (3)$$

where \mathcal{X}_{P_X} is the support set of the distribution P_X . While it is common to develop the second order dispersion term as a more refined asymptotic analysis, the third order term in (1) is required to further capture the difference between the cases of known and unknown source distributions. Similar third-order analysis for channel coding over discrete memoryless channels is recently developed in [11].

The challenge in designing a universal finite-length code is that in addition to the increased rate (relative to the asymptotic entropy rate) resulting from the finite length encoding, a single

¹ Q is the Gaussian cdf $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$, and Q^{-1} is its inverse function.

²Although there is a gap in Strassen's original proof; see discussion following (129) in [6].

code must be able to handle an arbitrary i.i.d. distribution. Thus, any finite length approximation should capture both effects. The difference in c in (2) and (3) when the source is known and unknown, respectively, captures the fact that with increasing larger alphabet size, the uncertainty about P_X increases thereby requiring more bits to represent a length n sequence.

A seemingly intuitive way to encode a universal sequence would be to encode its type and then encode the sequence within the type class [12, Chap. 13, pp. 433]. However, we show that this scheme can be improved to obtain the tighter bound on the third order term in (3). We do so by encoding based on the type class size and the empirical support set of the sequence. We illustrate our motivation in Fig. 1 in which the circular contours identify the distributions with the same empirical entropy over a simplex of distributions \mathcal{P} . Consider two type classes t_1 and t_2 with the same entropy h_2 , denoted by the two stars. While to first order the two classes have the same size (dependent on the entropy h_2), when one bounds the size of t_1 and t_2 using Stirling's approximation up to the third term it is seen that the class t_1 that is closer to the boundary of the simplex has a larger size than t_2 . We propose to use fewer bits to encode the types with smaller sizes and vice-versa. Exploiting these differences enables us to derive a better bound on the third order approximation term.

The paper is organized as follows. We introduce the finite-length lossless source coding problem and related definitions in Section II. In Section III we examine the case of a binary source in detail, and we show that for such a source the loss incurred due to the uncertainty in the distribution is only 1 bit. In Section IV we present our main theorems on the non-asymptotic and asymptotic bounds on coding rate. We also present a Lemma bounding the size of a type class up to the third term. We illustrate our results in Section V and conclude in Section VI.

II. PRELIMINARIES

First, a word on the nomenclature in the paper: we use \mathbb{P} to denote probability, \mathbb{E} to denote expectation, $x^+ = \max(x, 0)$, and $D(p||q)$ to denote the binary relative entropy. All logarithms and exponentials have base 2.

Let \mathcal{P} be the simplex of distributions over the finite alphabet \mathcal{X} . We consider a universal source coding problem in which a single code must compress a sequence X^n that is the output of an i.i.d. source with single letter distribution P_X , where P_X may be any element of \mathcal{P} . Any n -length sequence from the source is coded to a variable-length bit string via a coding function

$$\phi : \mathcal{X}^n \rightarrow \{0, 1\}^* = \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \dots\}. \quad (4)$$

We do not make the assumption that the code is prefix-free. Let $\ell(\phi(x^n))$ be the number of bits in the compressed binary string when x^n is the source sequence. The figure of merit is the probability that the length of the compressed string equals or exceeds an integer k , given by

$$\epsilon(k, P_X) = \mathbb{P}[\ell(\phi(X^n)) \geq k] \quad (5)$$

where the probability is taken with respect to $P_X \in \mathcal{P}$. The ϵ -coding rate at blocklength n is given by

$$R(n, \epsilon, P_X) = \max \left\{ \frac{k}{n} : \epsilon(k, P_X) \leq \epsilon \right\}. \quad (6)$$

Note that for a given code ϕ , the rate R depends on P_X .

The above definitions differ from [8], [9] in two ways. First, they assume prefix-free codes. Second, for the figure of merit they use *redundancy*, defined as the difference between the expected code length and the entropy of the true distribution:

$$\mathbb{E} \left[\ell(\phi(X^n)) - \log \frac{1}{P_X(X^n)} \right] \quad (7)$$

where the expectation is taken with respect to P_X . Using ϵ -coding rate rather than redundancy gives more refined information about the distribution of code lengths. In [8] it is proved that the optimal redundancy for a universal code is given by $\frac{d}{2} \log n + O(1)$ where d is the dimension of the set of possible source distributions (for i.i.d. sources, $d = |\mathcal{X}| - 1$). In Sec. IV, we show that with our model, there exists a universal code such that the gap to the optimal number of bits with known distribution is (using the d notation) $\frac{d-1}{2} \log n + O(1)$. Our lack of restriction to prefix-free codes seems to account for the difference seen between these two results. Interestingly, this difference is not found in the non-universal setting [3], [6].

Define the self information under a distribution P_X as

$$\iota_{P_X}(x) = \log \frac{1}{P_X(x)}. \quad (8)$$

For a sequence $x^n \in \mathcal{X}^n$, let t_{x^n} be the type of x^n , so that

$$t_{x^n}(x) = \frac{|\{i : x_i = x\}|}{n}. \quad (9)$$

For a type t , let T_t be the type class of t , i.e.

$$T_t = \{x^n \in \mathcal{X}^n : t_{x^n} = t\}. \quad (10)$$

For any distribution P_X on \mathcal{X} , let

$$\mathcal{Z}(P_X) = \{x \in \mathcal{X} : P_X(x) > 0\}. \quad (11)$$

From (8), the source entropy and the varentropy are given as

$$H = \mathbb{E}[\iota_{P_X}(X)] \text{ and } V = \text{Var}[\iota_{P_X}(X)] \quad (12)$$

where the expectation and variance are over P_X .

III. BINARY SOURCES

We begin by examining universal codes for binary i.i.d. sources. Consider first the optimal code when the distribution is known. These codes were studied in detail in [3], [6]. It is easy to see that the optimal code simply sorts all sequences in decreasing order of probability, and then assigns sequences to $\{0, 1\}^*$ in this order. Thus the more likely sequences will be assigned fewer bits. For example, consider an i.i.d. source with $\mathcal{X} = \{A, B\}$ where $P_X(A) = \delta$ and $\delta > 0.5$. The probability of a sequence is strictly increasing with the number of As, so the optimal code will assign sequences to $\{0, 1\}^*$ in an order where sequences with more As precede those with fewer. For

example, for $n = 3$, one optimal order is (sequences with the same type can always be exchanged)

$$\text{AAA, AAB, ABA, BAA, ABB, BAB, BBA, BBB.} \quad (13)$$

Interestingly, this is an optimal code for any binary source with $\delta \geq 0.5$. If $\delta < 0.5$, the optimal code assigns sequences to $\{0, 1\}^*$ in the reverse order. That is, there are only two optimal codes.³ To design a universal code, we can simply interleave the beginnings of each of these codes, so for $n = 3$ the sequences would be in the following order:

$$\text{AAA, BBB, AAB, BBA, ABA, BAB, BAA, ABB.} \quad (14)$$

In this order, any given sequence appears in a position at most twice as deep as in the two optimal codes. Hence, this code requires at most one additional bit as compared to the optimal code when the distribution is known. This holds for any n , as stated in the following theorem.

Theorem 1: Let $R^*(n, \epsilon, P_X)$ be the optimal fixed-to-variable rate when the distribution P_X is known. If $|\mathcal{X}| = 2$, there exists a universal code achieving

$$nR(n, \epsilon, P_X) \leq nR^*(n, \epsilon, P_X) + 1. \quad (15)$$

IV. ACHIEVABLE SCHEME

A typical approach to encode sequences from an unknown i.i.d. distribution is to use a two-stage descriptor to encode the type t of the sequence x^n first followed by its index within the type class T_t [12, Chap. 13, pp. 433]. We refer to such a coding scheme as a *Type-Sequence Code*. While bounds on c can be developed for this scheme, in the interest of space, we focus on the main achievable scheme based on the size of the type class, which we henceforth refer to as *Type Size Code*.

To enable comparisons with other schemes, the following Theorem bounds the achievable rate of the Type-Sequence Code at finite blocklength.

Theorem 2: The ϵ -rate achieved by the Type-Sequence Code for any $P_X \in \mathcal{P}$ is bounded by

$$R(n, \epsilon, P_X) \leq \frac{1}{n} [(|\mathcal{X}| - 1) \log(n + 1)] + k^*(\epsilon) / n \quad (16)$$

where

$$k^*(\epsilon) = \min \{ k : \sum_t \mathbb{P}(T_t) |1 - 2^k / |T_t||^+ < \epsilon \}. \quad (17)$$

Proof: The first term in (16) follows from the fact that there are at most $(n + 1)^{|\mathcal{X}| - 1}$ types on the alphabet \mathcal{X} . The second term follows directly from the code construction and (6). ■

Type Size Code: Recall that $\mathcal{Z}(t_{X^n})$ is the support set under t_{X^n} . The encoding function ϕ outputs two strings:

- 1) a string of $|\mathcal{X}|$ bits recording $\mathcal{Z}(t_{X^n})$, i.e. the elements of \mathcal{X} that appear in the observed sequence, and
- 2) a string that assigns sequences to $\{0, 1\}^*$ in order based on the size of the type class of the type of

³Here our assumption that the code may not be prefix-free becomes relevant, since it is not the case that there are only two optimal prefix-free codes for binary sources.

X^n , among all types t with $\mathcal{Z}(t) = \mathcal{Z}(t_{X^n})$. That is, if $\mathcal{Z}(t_{x^n}) = \mathcal{Z}(t_{x'^n})$ and $|T_{t_{x^n}}| < |T_{t_{x'^n}}|$, then $\ell(\phi(x^n)) \leq \ell(\phi(x'^n))$.

Note that the code described in Section III for binary sources is very similar to the Type Size Code. The support set string is omitted, and type classes with the same size are interleaved rather than ordered one after the other, but in essential aspects the codes are the same.

Theorem 3: For the Type Size Code

$$R(n, \epsilon, P_X) \leq \frac{1}{n} |\mathcal{X}| + \frac{1}{n} \lceil \log M(\epsilon) \rceil \quad (18)$$

where

$$M(\epsilon) = \inf_{\tau: \mathbb{P}[\frac{1}{n} \log |T_{t_{X^n}}| > \tau] \leq \epsilon} \sum_{\substack{t: \frac{1}{n} \log |T_t| \leq \tau \\ \mathcal{Z}(t) = \mathcal{Z}(t_{X^n})}} |T_t|. \quad (19)$$

Proof: The bound in (18) follows directly from the code construction and (6). ■

Obtaining an asymptotic bound on $R(n, \epsilon, P_X)$ in (18) requires bounding the size of the type classes. The size of a type class is closely related to the empirical entropy of the type, but importantly one is not strictly increasing with the other. The following Lemma, from an exercise in [13] makes this precise.

Lemma 1 (Exercise 1.2.2 in [13]): The size of the class of type t is bounded as

$$nf(t) + C^- \leq \log |T_t| \leq nf(t) \quad (20)$$

where $C^- = \frac{1 - |\mathcal{X}|}{2} \log(2\pi) - \frac{|\mathcal{X}|}{12 \ln 2}$ and

$$f(t) = H(t) + \frac{1 - |\mathcal{X}|}{2n} \log n + \frac{1}{2n} \sum_{x \in \mathcal{X}} \min \{ \log n, -\log t(x) \}.$$

The following theorem bounds the asymptotic rate achieved by the Type Size Code.

Theorem 4: For the Type Size Code and any $P_X \in \mathcal{P}$,

$$R(n, \epsilon, P_X) \leq H + \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \frac{1}{2} (|\mathcal{Z}(P_X)| - 3) \frac{\log n}{n} + O\left(\frac{1}{n}\right) \quad (21)$$

where H and V are defined in (12) and $\mathcal{Z}(\cdot)$ in (11).

Proof: Let $\bar{\mathcal{X}} = \mathcal{Z}(P_X)$ and $p_\tau = \mathbb{P}[\log |T_{t_{X^n}}| / n > \tau]$. For specific constants B and D defined later, let

$$\tau = H(P_X) + \sqrt{\frac{V}{n}} Q^{-1}\left(\epsilon - \frac{B}{\sqrt{n}} - |\bar{\mathcal{X}}| \exp(-nD)\right) + \frac{1 - |\bar{\mathcal{X}}|}{2} \frac{\log n}{n} - \frac{1}{2n} \sum_{x \in \bar{\mathcal{X}}} \log \frac{P_X(x)}{2} \quad (22)$$

We now show that for this choice of τ , $p_\tau \leq \epsilon$ and use it to bound $M(\epsilon)$ in (19). Using Chernoff bounds, for $D \equiv \min_{x \in \bar{\mathcal{X}}} D(P_X(x)/2 \| P_X(x)) > 0$, we have

$$\sum_{x \in \bar{\mathcal{X}}} \mathbb{P}\left[t_{X^n}(x) < \frac{P_X(x)}{2}\right] \leq |\bar{\mathcal{X}}| \exp(-nD). \quad (23)$$

We can write

$$p_\tau \leq \mathbb{P} \left[\frac{1}{n} \log |T_{t(X^n)}| > \tau, t_{X^n}(x) \geq \frac{P_X(x)}{2}, \forall x \in \bar{\mathcal{X}} \right] \\ + \sum_{x \in \bar{\mathcal{X}}} \mathbb{P} \left[t_{X^n}(x) < \frac{P_X(x)}{2} \right] \\ \leq \mathbb{P} \left[f(t_{X^n}) > \tau, t_{X^n}(x) \geq \frac{P_X(x)}{2} \right] \quad (24)$$

$$+ |\bar{\mathcal{X}}| \exp(-nD) \\ \leq \mathbb{P}[f'(t_{X^n}) > \tau] + |\bar{\mathcal{X}}| \exp(-nD) \quad (25)$$

$$\leq \mathbb{P} \left[\sum_{x \in \bar{\mathcal{X}}} -t_{X^n}(x) \log P_X(x) > g(\tau, n) \right] \quad (26)$$

$$+ |\bar{\mathcal{X}}| \exp(-nD) \\ = \mathbb{P} \left[\frac{1}{n} \sum_{i=1}^n -\log P_X(X_i) > g(\tau, n) \right] + |\bar{\mathcal{X}}| \exp(-nD) \\ \leq Q \left(\sqrt{\frac{n}{V}} [g(\tau, n) - H(P_X)] \right) + \frac{B}{\sqrt{n}} + |\bar{\mathcal{X}}| \exp(-nD) \quad (27)$$

$$= \epsilon \quad (28)$$

where (24) follows from Lemma 1, we have defined

$$f'(t_{X^n}) = H(t_{X^n}) + \frac{1-|\bar{\mathcal{X}}|}{2} \frac{\log n}{n} + \frac{1}{2n} \sum_{x \in \bar{\mathcal{X}}} -\log \frac{P_X(x)}{2}, \\ g(\tau, n) = \tau - \frac{1-|\bar{\mathcal{X}}|}{2} \frac{\log n}{n} + \frac{1}{2n} \sum_{x \in \bar{\mathcal{X}}} \log \frac{P_X(x)}{2},$$

(26) holds because $H(t_{X^n}) \leq H(t_{X^n}) + D(t_{X^n} \| P_X) = \sum_{x \in \bar{\mathcal{X}}} -t_{X^n}(x) \log P_X(x)$, (27) holds by the Berry-Esseen theorem, and (28) holds by the choice of τ in (22). The constant B in (27) results from the Berry-Esseen theorem and is related to the third moment of $\iota_{P_X}(X)$, which is finite for distributions with finite alphabets.

We now bound $M(\epsilon)$ using (19). As seen above, the probability that $\mathcal{Z}(t_{X^n}) \neq \bar{\mathcal{X}}$ is exponentially small, so we may assume they are equal. Fixing $\Delta > 0$ we may write

$$M(\epsilon) = \sum_{\substack{t: \frac{1}{n} \log |T_t| \leq \tau \\ \mathcal{Z}(t) = \bar{\mathcal{X}}}} |T_t| \leq \sum_{\substack{t: f(t) + \frac{C^-}{n} \leq \tau \\ \mathcal{Z}(t) = \bar{\mathcal{X}}}} \exp\{nf(t)\} \quad (29)$$

$$= \sum_{i=0}^{\infty} \sum_{\substack{t: f(t) \in \mathcal{A}_i \\ \mathcal{Z}(t) = \bar{\mathcal{X}}}} \exp\{nf(t)\} \quad (30)$$

$$\leq \sum_{i=0}^{\infty} |\{t: f(t) \in \mathcal{A}_i, \mathcal{Z}(t) = \bar{\mathcal{X}}\}| 2^{\{n\tau - C^- - ni\Delta\}} \quad (31)$$

where $\mathcal{A}_i = (\tau - C^-/n - (i+1)\Delta, \tau - C^-/n - i\Delta]$.

We bound the number of types with $f(t) \in \mathcal{A}_i$ in the Appendix. Applying the results from (47) in the Appendix to (31), we obtain

$$M(\epsilon) \leq \sum_{i=0}^{\infty} \frac{Kn^{|\bar{\mathcal{X}}|-1}}{d} \left[\Delta + \frac{C}{n} \right] \exp\{n\tau - C^- - ni\Delta\} \\ = \frac{Kn^{|\bar{\mathcal{X}}|-1}}{d} \left[\Delta + \frac{C}{n} \right] \frac{\exp\{n\tau - C^-\}}{1 - \exp\{-n\Delta\}}.$$

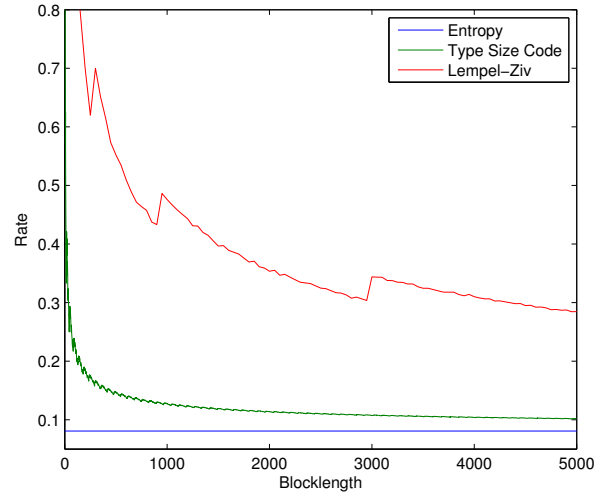


Fig. 2. Numerical results showing the rate $R(n, \epsilon, P_X)$ for a binary source with bias 0.01, $\epsilon = 10^{-2}$, comparing Lempel-Ziv to the Type Size Code. We omit other bounds, as they would be nearly indistinguishable from the rate of the Type Size Code. The Lempel-Ziv curve is computed using 2000 Monte-Carlo runs.

The above holds for any $\Delta > 0$, so we may take $\Delta = \frac{C}{n}$ to write

$$\log M(\epsilon) \leq n\tau - C^- \quad (32)$$

$$+ \log \left[\frac{Kn^{|\bar{\mathcal{X}}|-1}}{d} \frac{2C}{n} \frac{1}{1 - \exp\{-C\}} \right] \\ = n\tau + (|\bar{\mathcal{X}}| - 2) \log n - C^- \quad (33)$$

$$+ \log \left[\frac{2KC}{d(1 - \exp\{-C\})} \right] \\ = nH(P_X) + \sqrt{nV}Q^{-1}(\epsilon) \quad (34) \\ + \frac{1}{2}(|\bar{\mathcal{X}}| - 3) \log n + O(1)$$

where we have used the definition of τ from (22). ■

V. NUMERICAL RESULTS

We present several numerical results comparing different universal coding schemes. Fig. 2 compares the performance of the Type Size Code to Lempel-Ziv for a binary source. The figure illustrates that while Lempel-Ziv asymptotically achieves the entropy rate for ergodic sources, it can perform quite poorly at low blocklengths. On the other hand, since Lempel-Ziv is by no means limited to i.i.d. sources, it is not surprising that it performs worse than a code optimized for this purpose. Fig. 3 examines a ternary source, comparing the Type Size Code to the Type-Sequence Code (computed using Theorems 3 and 2 respectively), as well as the optimal code when the underlying distribution is known exactly. The latter is obviously a lower bound for the universal problem, and it is computed using [6, Theorem 6]. Also shown is the three-term approximation in (21). For the specific example shown in the figure, the approximation gives the rate achieved by the Type Size Code extremely accurately.

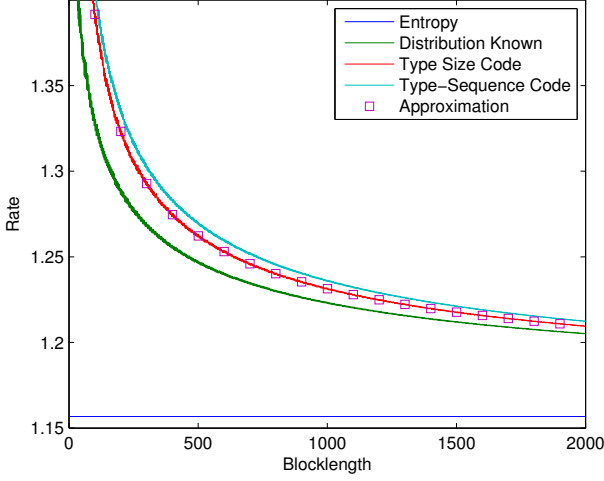


Fig. 3. Numerical results showing the rate $R(n, \epsilon, P_X)$ for a ternary source with distribution $[0.1 \ 0.2 \ 0.7]$, $\epsilon = 10^{-2}$. The Type Size Code is compared with the Type-Sequence Code, the optimal code when the distribution is known, and the three-term approximation.

VI. CONCLUDING REMARKS

We have presented an achievable scheme for universal source coding in the finite blocklength regime. Through a novel encoding based on the type class size of the sequence, we have demonstrated that finer distinction can be made in encoding sequences with the same empirical entropy for non-binary sources. Generalizations to sources with memory is an important problem that needs addressing.

APPENDIX: BOUND ON $|\mathcal{A}_i|$ IN (31)

Define a 2-norm ball of radius $1/2n$ around a distribution P as $B(P) = \{Q : \|P - Q\|_2 < \frac{1}{2n}\}$. Note that for any two different types t_1, t_2 , $\|t_1 - t_2\|_2 \geq \frac{1}{n}$ so $B(t_1)$ and $B(t_2)$ are always disjoint. Let $\mathcal{P}_{\bar{\mathcal{X}}}$ be the set of distributions P with $\mathcal{Z}(P) = \bar{\mathcal{X}}$. Since $\mathcal{P}_{\bar{\mathcal{X}}}$ is a $|\bar{\mathcal{X}}| - 1$ -dimensional space, we define volumes on $\mathcal{P}_{\bar{\mathcal{X}}}$ via the $|\bar{\mathcal{X}}| - 1$ -dimensional Lebesgue measure. For any type $t \in \mathcal{P}_{\bar{\mathcal{X}}}$, $t(x) \geq 1/n$ for all $x \in \bar{\mathcal{X}}$, so

$$\text{Vol}(B(t) \cap \mathcal{P}_{\bar{\mathcal{X}}}) = d/n^{|\bar{\mathcal{X}}|-1} \quad (35)$$

for a constant d that depends only on $|\bar{\mathcal{X}}|$. Let \mathcal{P}_n be the set of types of length n . For any set of distributions \mathcal{A} , we may bound the number of types in \mathcal{A} with $\mathcal{Z}(t) = \bar{\mathcal{X}}$ by

$$|\mathcal{A} \cap \mathcal{P}_n \cap \mathcal{P}_{\bar{\mathcal{X}}}| = \sum_{t \in \mathcal{A} \cap \mathcal{P}_n \cap \mathcal{P}_{\bar{\mathcal{X}}}} \frac{n^{|\bar{\mathcal{X}}|-1}}{d} \text{Vol}(B(t) \cap \mathcal{P}_{\bar{\mathcal{X}}}) \quad (36)$$

$$= \frac{n^{|\bar{\mathcal{X}}|-1}}{d} \text{Vol}\left(\bigcup_{t \in \mathcal{A} \cap \mathcal{P}_n \cap \mathcal{P}_{\bar{\mathcal{X}}}} B(t) \cap \mathcal{P}_{\bar{\mathcal{X}}}\right) \quad (37)$$

$$\leq \frac{n^{|\bar{\mathcal{X}}|-1}}{d} \text{Vol}\left(\bigcup_{P \in \mathcal{A}} B(P) \cap \mathcal{P}_{\bar{\mathcal{X}}}\right) \quad (38)$$

where (37) holds because the balls are disjoint. There exists a constant C so that for any distributions P and Q ,

$$|f(P) - f(Q)| \leq C\|P - Q\|_2. \quad (39)$$

In particular, for any $Q \in B(P)$,

$$|f(Q) - f(P)| \leq C/2n. \quad (40)$$

For any $\lambda \geq 0$ let

$$h(\lambda) = \text{Vol}(\{P \in \mathcal{P}_{\bar{\mathcal{X}}} : f(P) \leq \lambda\}). \quad (41)$$

Let K be the constant so that for all a, b ,

$$|h(a) - h(b)| \leq K|a - b|. \quad (42)$$

Note that K depends only on $|\bar{\mathcal{X}}|$. For any real a ,

$$|\{t \in \mathcal{P}_n \cap \mathcal{P}_{\bar{\mathcal{X}}} : a < f(t) \leq a + \Delta\}| \quad (43)$$

$$\leq \frac{n^{|\bar{\mathcal{X}}|-1}}{d} \text{Vol}\left(\bigcup_{P: a < f(P) \leq a + \Delta} B(P) \cap \mathcal{P}_{\bar{\mathcal{X}}}\right) \quad (44)$$

$$\leq \frac{n^{|\bar{\mathcal{X}}|-1}}{d} \text{Vol}(\{P \in \mathcal{P}_{\bar{\mathcal{X}}} : f(P) \in (a - \frac{C}{2n}, a + \frac{C}{2n} + \Delta]\}) \quad (45)$$

$$= \frac{n^{|\bar{\mathcal{X}}|-1}}{d} \left[h\left(a + \Delta + \frac{C}{2n}\right) - h\left(a - \frac{C}{2n}\right) \right] \quad (46)$$

$$\leq \frac{Kn^{|\bar{\mathcal{X}}|-1}}{d} \left[\Delta + \frac{C}{n} \right] \quad (47)$$

where (44) holds by (38), (45) holds by (40), (46) holds by the definition of h in (41), and (47) holds by (42).

REFERENCES

- [1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, pp. 2307–2359, 2010.
- [2] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, Jun. 2012.
- [3] S. Verdú and I. Kontoyiannis, "Lossless data compression rate: Asymptotics and non-asymptotics," in *Proc. 46th Annual Conf. Inform. Sciences and Systems (CISS)*, Mar. 2012, pp. 1–6.
- [4] V. Y. F. Tan and O. Kosut, "The dispersion of Slepian-Wolf coding," in *Proc. 2012 IEEE Intl. Symp. Inform. Theory (ISIT)*, Jul. 2012, pp. 915–919.
- [5] Y. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, July 2012, pp. 831–835.
- [6] I. Kontoyiannis and S. Verdú, "Lossless data compression rate at finite blocklengths," arxiv.org:1212.2668.
- [7] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. 24, no. 5, pp. 530–536, Sep 1978.
- [8] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inform. Theory*, vol. 36, no. 3, pp. 453–471, May 1990.
- [9] A. Beirami and F. Fekri, "Results on the redundancy of universal compression for finite-length sequences," in *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, Aug. 2011.
- [10] V. Strassen, "Asymptotic approximations in Shannon's information theory," <http://www.math.cornell.edu/~pmlut/strassen.pdf>, Aug. 2009, english translation of original Russian article in Trans. Third Prague Conf. on Inform. Th., Statistics, Decision Functions, Random Processes (Liblice, 1962), Prague, 1964.
- [11] M. Tomanich and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics of discrete memoryless channels," arxiv.org:1212.3689.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Ed. New York: Wiley, 2006.
- [13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, 1981.