Lossless and Lossy Source Compression with Near-Uniform Output: Is Common Randomness Always Required?

Badri N. Vellambi
New Jersey Institute of Technology
Newark, NJ 07102
Email: badri.vellambi@njit.edu

Matthieu Bloch, Rémi Chou
Georgia Institute of Technology
Atlanta, GA 30332
Email: {matthieu, remi.chou}@gatech.edu

Jörg Kliewer
New Jersey Institute of Technology
Newark, NJ 07102
Email: jkliewer@njit.edu

***Abstract*—It is known that a sub-linear rate of source-independent random seed (common randomness) can enable the construction of lossless compression codes whose output is nearly uniform under the variational distance (Chou-Bloch-ISIT'13). This work uses finite-blocklength techniques to present an alternate proof that for near-uniform lossless compression, the seed length has to grow strictly larger than $\sqrt{n}$, where n represents the blocklength of the lossless compression code. In the lossy setting, we show the surprising result that a seed is not required to make the encoder output nearly uniform.**

***Index Terms*—Source coding, Lossless coding, Rate-distortion, Finite-blocklength techniques.**

I. Introduction

The relationship between vanishing error probability probability of error and the uniformity of encoder outputs for lossless compression has received rigorous treatment [1], [2]. Specifically, Hayashi has shown that uniformity under variational distance and vanishing error probability (i.e., lossless compression) cannot be simultaneously met for discrete memoryless sources (DMSs) [2]. One way to guarantee both near-uniform outputs as well as lossless compression is to allow the encoder and decoder to share source-independent common randomness. With the aid of this randomness, we can, in effect, average of multiple codebooks. This setup was considered previously in [3], and it was shown that a random seed whose length that grows as the square root of the blocklength is both necessary and sufficient for uniform lossless compression.

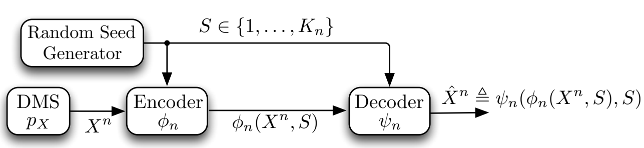

Fig. 1. The setup for common randomness-assisted compression.

In this work, we consider the same setup as in [3], where a DMS is compressed by means of a source-independent common randomness shared by both the encoder and the decoder (see Fig. 1). The motivation of this work is to understand the fundamental limits of the seed size for both lossless as well as lossy (rate-distortion) compression of DMSs by the use of recent developments in finite-blocklength and Gaussian approximation techniques [4]–[6]. Note that the tradeoff between uniformity under variational distance and operation at rates close to the rate-distortion boundary has not been studied before. Our contributions are as follows:

• For lossless compression, we provide an alternate intuitive proof that near-uniform outputs can be achieved only if the seed size grows faster than the square-root of the blocklength of the code. This is shown by arguing that for encoder output uniformity, the seed size has to exceed the standard deviation of n i.i.d. self-information random variables.

• For lossy compression, we show the interesting result that there exist codes that simultaneously : (a) operate close to the rate $R(D)$); (b) produce outputs whose distribution is near-uniform under the variation distance metric; and (c) require *no* random seed;. This is shown by partitioning each bin of an optimal R-D code into $\Theta(2^{\sqrt{n}\beta})$ bins for some $\beta > 0$. Here, the $\sqrt{n}$ dependence appears due to the standard deviation of the sum of n i.i.d. D-tilted information random variables.

The remainder of the paper is organized thus. Section II provides the notation, and Section III formally presents the problem. Section IV presents the allied technical results needed in this work. Lastly, Sections V and VI detail the results on uniform lossless and lossy compression codes.

II. Notation

For $n \in \mathbb{N}$, $[\![1, n]\!] \triangleq \{1, \dots, n\}$. Uppercase letters (e.g., X, Y) denote random variables (RVs), lowercase letters (e.g., x, y) denote their realizations, and the script versions (e.g., $\mathcal{X}$, $\mathcal{Y}$) denote their alphabets. In this work, all alphabets are assumed to be countable. Superscripts denote the vector lengths, and subscripts denote component indices. The variance of an RV X is given by $\mathsf{Var}(X)$. For a probability mass function (p.m.f.) p_X, the set of all ε-weakly typical sequences of length n is

$$\mathtt{W}_\varepsilon^n[p_X] \triangleq \left\{x^n \in \mathcal{X}^n : \left|\log_2 p_X(x^n) + nH(X)\right| < \varepsilon\right\}.$$

Given a random variable X with p.m.f. p_X, entropy $H(X)$ and $\mathsf{Var}(-\log_2 p_X(X)) = \sigma^2$, $a > b > 0$ and $n \in \mathbb{N}$,

$$\mathcal{T}_n(a, b) \triangleq \left\{x^n \in \mathcal{X}^n : -a < \frac{\log_2 p_X(x^n) + nH(X)}{\sigma\sqrt{n}} \le -b\right\}.$$

The probability of an event E occurring is given by $\mathbb{P}(E)$. Lastly, $\mathsf{len}(\mathbf{b})$ denotes the length of a binary string $\mathbf{b}$.

This work has been supported in part by the Australian Research Council Discovery Grant DP120102123, and NSF grants CCF-1440014, CCF-1439465, and CCF-1320298.

III. Problem Definition

The problem setup is identical to that in [3]. We consider the compression of a discrete memoryless source (DMS) p_X over a countable alphabet $\mathcal{X}$ with the aid of a source-independent random seed such that the following two conditions are met:

- The output of the decoding/reconstruction function meets the lossless/lossy reconstruction constraint; and
- The output of the encoder is near-uniform under the variational distance metric.

For the sake of completeness, we define uniform lossless and uniform lossy compression codes as follows.

Definition 1 (Uniform Lossless Compression Code): Given DMS p_X over an alphabet $\mathcal{X}$ and source-independent random seed $S \in [\![1, K_n]\!]$, $n \in \mathbb{N}$, an $(M_n, n, K_n, \varepsilon)$-uniform lossless compression code C of blocklength n comprises of an encoder $\phi_n : \mathcal{X}^n \times [\![1, K_n]\!] \to [\![1, M_n]\!]$ and a decoder $\psi_n : [\![1, M_n]\!] \times [\![1, K_n]\!] \to \mathcal{X}^n$ such that

$$P_e(\phi_n, \psi_n) \triangleq \mathbb{P}\left[X^n \neq \psi_n\left(\phi_n(X^n, S), S\right)\right] \le \varepsilon$$
$$U_e(\phi_n) \triangleq \sum_{i=1}^{M_n} \left|\mathbb{P}\big[\phi_n(X^n, S) = i\big] - \tfrac{1}{M_n}\right| \le \varepsilon$$

Definition 2 (Uniform Lossy Compression Code): Given DMS p_X over an alphabet $\mathcal{X}$, finite (reconstruction) alphabet $\hat{\mathcal{X}}$, distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, d_{\max}]$, source-independent random seed $S \in [\![1, K_n]\!]$, $n \in \mathbb{N}$, and distortion $D \in (0, D_{\max})$, an $(M_n, n, K_n, D, \varepsilon)$-uniform lossy compression code C comprises of an encoder $\phi_n : \mathcal{X}^n \times [\![1, K_n]\!] \to [\![1, M_n]\!]$ and a reconstruction function $\psi_n : [\![1, M_n]\!] \times [\![1, K_n]\!] \to \hat{\mathcal{X}}^n$ such that

$$\mathbb{P}\left[\sum_{j=1}^{n} \frac{d\big(X_i, (\psi_n\left(\phi_n(X^n, S), S\right))_i\big)}{n} > D\right] \le \varepsilon$$
$$U_e(\phi_n) \triangleq \sum_{i=1}^{M_n} \left|\mathbb{P}\big[\phi_n(X^n, S) = i\big] - \tfrac{1}{M_n}\right| \le \varepsilon$$

The remainder presents the following two main results:

1. For vanishing block error probability and near-uniform encoder output, the seed length has to grow faster than $\sqrt{n}$, where n is the blocklength of the compression code.
2. For lossy compression, there is no need for a common random seed to achieve near-uniform encoder output.

However, before we present them, we present some preliminary results that we require in our proofs.

IV. Some Preliminary Results

Lemma 1: Let $\{X_i\}_{i \in \mathbb{N}}$ be emitted by a DMS p_X over a countable set $\mathcal{X}$. Furthermore, suppose that p_X is such that $H(X) < \infty$, $\sigma^2 \triangleq \mathsf{Var}\big(-\log_2 p_X(X_1)\big) > 0$, and $\rho \triangleq E\big[|\log_2 p_X(X_1) + H(X)|^3\big] < \infty$. Then, there exists an $\alpha > 0$ such that for any $a > b > 0$,

$$\eta_{a,b} \triangleq \Big|\mathbb{P}[X^n \in \mathcal{T}_n(a, b)] - \big(\Phi(-b) - \Phi(-a)\big)\Big| \le \frac{2\alpha\rho}{\sigma^3\sqrt{n}},$$

where Φ is the cumulative distribution function of the standard normal distribution.

Proof: Let $S_n \triangleq \frac{nH(X) + \sum_{j=1}^{n} \log_2 p_X(X_j)}{\sigma\sqrt{n}}$. Then,

$$\begin{aligned}\mathbb{P}[X^n \in \mathcal{T}_n(a, b)] &= \mathbb{P}\left[-a < S_n \le -b\right] \\ &= \mathbb{P}\left[S_n \le -b\right] - \mathbb{P}\left[S_n \le -a\right].\end{aligned}$$

Hence, by triangle inequality, we have

$$\eta_{a,b} \le \sum_{\lambda \in \{a,b\}} \big|\mathbb{P}\left[S_n \le -\lambda\right] - \Phi(-\lambda)\big| \le \frac{2\alpha\rho}{\sigma^3\sqrt{n}}, \quad (1)$$

where (1) follows from the Berry-Esséen Theorem [7, Theorem 1.5] and α depends only on the source p.m.f. p_X. ∎

Remark 1: Similarly, for $b > 0$,

$$\eta_{\infty,b} \triangleq \Big|\mathbb{P}[X^n \in \mathcal{T}_n(\infty, b)] - \Phi(-b)\Big| \le \frac{\alpha\rho}{\sigma^3\sqrt{n}}. \quad (2)$$

Lemma 2: Let Z be a random variable over a countable alphabet $\mathcal{Z}$ and let $\mathcal{Z}' \subseteq \mathcal{Z}$ be given such that $\mathbb{P}[Z \notin \mathcal{Z}'] \le \delta_1$ for some $\delta_1 > 0$. Let $\phi : \mathcal{Z} \to \mathcal{A}$ and $\mathcal{A}' \subseteq \mathcal{A}$ be given such that $\mathbb{P}[\phi(Z) \in \mathcal{A}'] \ge 1 - \delta_2$ for some $\delta_2 > 0$. Let

$$\mathcal{B} \triangleq \Big\{a \in \mathcal{A}' : \mathbb{P}\big[Z \in \mathcal{Z}' | \phi(Z) = a\big] \ge 1 - \sqrt{\delta_1}\Big\}.$$

Then, $\mathbb{P}[\phi(Z) \in \mathcal{B}] \ge 1 - \sqrt{\delta_1} - \delta_2$.

Proof: Define

$$\mathcal{B}_0 \triangleq \Big\{a \in \mathcal{A} : \mathbb{P}\big[Z \in \mathcal{Z}' | \phi(Z) = a\big] \ge 1 - \sqrt{\delta_1}\Big\}.$$

Then,

$$\begin{aligned}\delta_1 &\ge \mathbb{P}[Z \notin \mathcal{Z}'] \ge \mathbb{P}[Z \notin \mathcal{Z}', \phi(Z) \notin \mathcal{B}_0] \\ &\ge \sqrt{\delta_1}\, \mathbb{P}[\phi(Z) \notin \mathcal{B}_0]. \end{aligned} \quad (3)$$

Finally, the claim follows from the following argument.

$$\mathbb{P}\big[\phi(Z) \notin \mathcal{B}\big] \le \mathbb{P}\big[\phi(Z) \notin \mathcal{B}_0\big] + \mathbb{P}\big[\phi(Z) \notin \mathcal{A}'\big],$$

which by (3) and the hypothesis is no more than $\sqrt{\delta_1} + \delta_2$. ∎

Lemma 3: Let $a, b \in \mathbb{N}$ with $a > b$. Let $\frac{1}{a} \le x \le \frac{1}{b}$. Then,

$$\sup_{a \le n \le b} \sum_{j=1}^{n} \left|\tfrac{1}{n} - x\right| = \sup_{a \le n \le b} |1 - nx| \le \frac{a - b}{b}. \quad (4)$$

Proof: The term $|nx - 1|$ is the largest it can be when nx is either the largest or smallest value it can be. Hence, $\sup_{a \le n \le b} |1 - nx| \le \max\left\{1 - \frac{b}{a}, \frac{a}{b} - 1\right\} = \frac{a-b}{b}$. ∎

Lemma 4: Let $n \in \mathbb{N}$, $M_n \in \mathbb{N}$, $\gamma_n \in [0, M_n]$, and function $\phi_n : \mathcal{X}^n \times [\![1, K_n]\!] \to [\![1, M_n]\!]$ be given. Then,

$$U_e(\phi_n) \ge 2\,\mathbb{P}\left[p_X(X^n) > \tfrac{K_n}{\gamma_n}\right] - \tfrac{2\gamma_n}{\mathcal{M}_n}. \quad (5)$$

Proof: This follows directly from Lemma 2.1.2 of [8]. ∎

V. Uniform Lossless Compression Codes

For the lossless setting, we only present a new proof of the converse, since complete details of the achievability for the optimal seed size is given in [3]. In the achievable scheme, the encoder is an instance of random mapping from source realization and seed pair to bin indices. With this construction, a seed length of $\Theta(n^{\frac{1}{2}+\delta})$ is shown to be sufficient for any $\delta > 0$. A new intuitive proof of the necessity is as follows.

Theorem 1 (Converse): Let a non-uniform DMS p_X meeting the conditions of Lemma 1 be given. For $i \in \mathbb{N}$, let $\mathcal{C}_i$ be an $(M_{n_i}, n_i, K_{n_i}, \varepsilon_i)$-uniform lossless compression code with blocklength n_i, encoding function ϕ_{n_i}, and decoding function ψ_{n_i} such that

$$\lim_{i\to\infty} P_e(\phi_{n_i}, \psi_{n_i}) = \lim_{i\to\infty} U_e(\phi_{n_i}) = \lim_{i\to\infty} \varepsilon_i = 0.$$

Then, $\liminf_{i\to\infty} n_i^{-1} \log_2 M_{n_i} \geq H(X)$. Furthermore,

$$\lim_{i\to\infty} n_i^{-\frac{1}{2}} (\log_2 M_{n_i} - n_i H(X)) = \infty \tag{6}$$

$$\lim_{i\to\infty} n_i^{-\frac{1}{2}} \log_2 K_{n_i} = \infty. \tag{7}$$

Proof: Since p_X is not uniform, $P_e(\phi_{n_i}, \psi_{n_i}) \to 0$ and $U_e(\phi_{n_i}) \to 0$ can be jointly met only if $n_i \to \infty$ as $i \to \infty$. The encoding function ϕ_{n_i} utilizes a source-independent seed, say, S_i taking values in $[\![1, K_{n_i}]\!]$. Hence, for each $i \in \mathbb{N}$, we can find $s_i^* \in [\![1, K_{n_i}]\!]$ such that

$$\mathbb{P}\big[X^{n_i} \neq \psi_{n_i}(\phi_{n_i}(X_i^n, s_i^*), s_i^*)\big] \leq P_e(\phi_{n_i}, \psi_{n_i}). \tag{8}$$

Fix $a > b > 0$, and define $\mathcal{L}_i(a,b)$ as

$$\mathcal{L}_i(a,b) \triangleq \big\{x^{n_i} \in \mathcal{T}^{n_i}(a,b) : x^{n_i} = \psi_{n_i}(\phi_{n_i}(x^{n_i}, s_i^*), s_i^*)\big\}.$$

Note that

$$\begin{aligned}\mathbb{P}[X^{n_i} \in \mathcal{L}_i(a,b)] &\geq \begin{bmatrix} \mathbb{P}\big[X^{n_i} \in \mathcal{T}_{n_i}(a,b)\big] \\ -\mathbb{P}\big[X^{n_i} \neq \psi_{n_i}(\phi_{n_i}(X_i^n, s_i^*), s_i^*)\big] \end{bmatrix} \\ &\overset{(8)}{\geq} \mathbb{P}[X^{n_i} \in \mathcal{T}_{n_i}(a,b)] - P_e(\phi_{n_i}, \psi_{n_i}) \\ &\geq \underbrace{\Phi(-b) - \Phi(-a) - \frac{2\alpha\rho}{\sigma^3\sqrt{n_i}} - P_e(\phi_{n_i}, \psi_{n_i})}_{\triangleq \eta_i(a,b)},\end{aligned}$$

where the last inequality follows from Lemma 1. Note that for any $x^{n_i} \in \mathcal{L}_i(a,b)$, $p_X(x^{n_i}) \leq 2^{-n_i H(X) - b\sigma\sqrt{n_i}}$, where $\sigma^2 \triangleq \mathsf{Var}(-\log_2 p_X(X_1))$. Hence,

$$\frac{|\mathcal{L}_i(a,b)|}{2^{n_i H(X) + b\sigma\sqrt{n_i}}} \geq \mathbb{P}[X^{n_i} \in \mathcal{L}_i(a,b)] \geq \eta_i(a,b). \tag{9}$$

Since $\mathcal{L}_i(a,b)$ is a subset of source realizations for which the code offers perfect reconstruction (when $S_i = s_i^*$), we have

$$M_{n_i} \geq |\mathcal{L}_i(a,b)| \overset{(9)}{\geq} \eta_i(a,b)\, 2^{n_i H(X) + b\sigma\sqrt{n_i}}. \tag{10}$$

Note that since $\lim_{i\to\infty} \eta_i(a,b) = \Phi(-b) - \Phi(-a)$, we have

$$\lim_{i\to\infty} n_i^{-\frac{1}{2}} \log_2 \eta_i(a,b) = 0$$
$$\liminf_{i\to\infty} n_i^{-1} \log_2 M_{n_i} \geq H(X).$$

Rearranging (10), we have

$$\liminf_{i\to\infty} \frac{\log M_{n_i} - n_i H(X)}{\sqrt{n_i}\sigma} \geq b + \liminf_{i\to\infty} \frac{\log_2 \eta_i(a,b)}{\sqrt{n_i}\sigma} = b.$$

Since b is any arbitrary positive number, (6) follows by letting $b \to \infty$. To prove (7), we use Lemma 4 with

$$\gamma_{n_i} \triangleq \eta_i(a,b)\, 2^{n_i H(X)}$$

$$M_{n_i} \geq \eta_i(a,b)\, 2^{n_i H(X) + b\sigma\sqrt{n_i}},$$

which yields

$$\begin{aligned}\mathbb{P}\left[p_X(X^{n_i}) > \tfrac{K_{n_i}}{\gamma_{n_i}}\right] &\leq \frac{1}{2} U_e(\phi_{n_i}) + \frac{\gamma_{n_i}}{M_{n_i}} \\ &\leq \frac{1}{2} U_e(\phi_{n_i}) + 2^{-b\sigma\sqrt{n_i}}.\end{aligned} \tag{11}$$

From Remark 1 of Section IV, it follows that

$$\begin{aligned}\mathbb{P}\left[p_X(X^{n_i}) \leq \tfrac{K_{n_i}}{\gamma_{n_i}}\right] &= \mathbb{P}\left[p_X(X^{n_i}) \leq \frac{K_{n_i}}{\eta_i(a,b)\, 2^{n_i H(X)}}\right] \\ &\leq \Phi\left(\frac{\log_2 \frac{K_{n_i}}{\eta_i(a,b)}}{\sigma\sqrt{n_i}}\right) + \frac{\alpha\rho}{\sigma^3\sqrt{n_i}}.\end{aligned} \tag{12}$$

Combining (11) and (12), we obtain

$$\Phi\left(\frac{\log_2 \frac{K_{n_i}}{\eta_i(a,b)}}{\sigma\sqrt{n_i}}\right) \geq \beta_i \triangleq 1 - \frac{U_e(\phi_{n_i})}{2} - 2^{-b\sigma\sqrt{n_i}} - \frac{\alpha\rho}{\sigma^3\sqrt{n_i}}.$$

Rearranging terms and applying the appropriate limit, we get

$$\begin{aligned}\lim_{i\to\infty} \frac{\log_2 K_{n_i}}{\sigma\sqrt{n_i}} &= \Phi^{-1}\left(\Phi\left(\lim_{i\to\infty} \frac{\log_2 K_{n_i}}{\sigma\sqrt{n_i}}\right)\right) \\ &= \Phi^{-1}\left(\lim_{i\to\infty} \Phi\left(\frac{\log_2 K_{n_i}}{\sigma\sqrt{n_i}}\right)\right) \\ &\geq \Phi^{-1}\left(\lim_{i\to\infty} \beta_i\right) = \Phi^{-1}(1) = \infty,\end{aligned}$$

where in the above arguments, we have used the fact that Φ is invertible, continuous and increasing. ■

Remark 2: Note that there is no requirement for the seed to be uniform. Further, the result in Theorem 1 holds provided

1. For each n, $(X_1, \ldots, X_n)$ is conditionally i.i.d. for any $s_i \in [\![1, K_{n_i}]\!]$, and
2. There exists $0 < \sigma_0 < \infty$ such that

$$0 < \sup_{s_i \in [\![1, K_{n_i}]\!]} \mathsf{Var}(X_1 | S = s_i) < \sigma_0.$$

VI. Uniform Lossy Compression Codes

For the main result, we require the following two results (Theorems 2 and 3) proven in [4]. Note that throughout this section, $\jmath_X(x, D)$ denotes the D-tilted information [6, Def. 6].

Theorem 2 (Achievability): Let $\{X_i\}_{i\in\mathbb{N}}$ generated by DMS p_X over a alphabet countable $\mathcal{X}$, and distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to (0, D_{\max})$ be given. Then, for $D \in (0, D_{\max})$, there exist: (a) a sequence of codes $\{C_n^*\}_{n\in\mathbb{N}}$ (where C_n^* is a code over n symbols with encoder ϕ_n^*, and reconstruction function ψ_n^*) *operating*[1] at distortion level D; (b) a sequence of variable-length binary prefix-free encoders $\{\varphi_n^*\}_{n\in\mathbb{N}}$, where φ_n^* maps codewords of C_n^* to binary strings; and (c) $k, \nu \geq 0$ such that the following holds asymptotically almost surely.

$$\begin{aligned}l_n^*(X^n) &\triangleq \mathsf{len}\big(\varphi_n^*(\phi_n^*(X^n))\big) \\ &\leq \sum_{i=1}^n \jmath_X(X_i, D) + k \log_2 n + \nu.\end{aligned} \tag{13}$$

[1]Both [4] and Theorem 2 employ the stringent *zero* excess distortion criterion, i.e, the reconstruction $\hat{X}^n$ satisfies $\mathbb{P}\big[\sum_{i=1}^n d(X_i, \hat{X}_i) > nD\big] = 0$.

Outline of Code Construction of [4]: The code is a modification of the standard construction for average per-symbol distortion constraint. Let $p_{\hat{X}|X}$ be a test channel that meets both the required distortion level D and the condition $I(X;\hat{X}) = R(D)$, and let $p_{\hat{X}}$ denote the corresponding marginal. To construct a code of length n, generate *sufficiently many* codewords with components of every codeword drawn i.i.d. according to $p_{\hat{X}}$. For a realization $X^n = x^n$, if there is no such codeword meeting the distortion constraint, pick a sequence $\hat{x}^n \in \hat{\mathcal{X}}^n$ that meets the distortion D or less and convey that in $\lceil n \log_2 |\hat{\mathcal{X}}| \rceil$ bits. An additional flag describes which of the two events (whether or not a suitable codeword was found) was realized is also conveyed to the decoder. Analysis of the required size of the codebook yields the result.

Theorem 3 (Converse): Let $\{X_i\}_{i\in\mathbb{N}}$ be emitted by a DMS p_X over alphabet $\mathcal{X}$. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to (0, D_{\max})$ be a distortion measure. Let $D \in (0, D_{\max})$ and $\{b_n\}_{n\in\mathbb{N}} \in {\mathbb{R}^+}^{\mathbb{N}}$ be such that $\sum_{j=1}^{\infty} 2^{-b_j} < \infty$. Then, for any sequence of codes $\{C_n\}_{n\in\mathbb{N}}$ operating at distortion level D (where C_n is a code over n symbols with encoder ϕ_n, and reconstruction function ψ_n), and any sequence of variable-length binary prefix-free encoders $\{\varphi_n\}_{n\in\mathbb{N}}$ where φ_n maps codewords of C_n to binary strings, the following holds asymptotically almost surely.

$$l_n(X^n) \triangleq \mathsf{len}(\varphi_n(\phi_n(X^n))) \geq \sum_{i=1}^{n} \jmath_X(X_i, D) - b_n. \quad (14)$$

We are now equipped to present the main result of this section.

Theorem 4 (Achievability): Let a DMS p_X over a countable alphabet $\mathcal{X}$, and distortion measure $d : \mathcal{X} \times \widehat{\mathcal{X}} \to [0, D_{\max}]$ be given. Let $R(D)$ denote the R-D function for the given source under the measure d. Let $D \in (0, D_{\max})$ be given such that $R(D) < H(X)$ and $V(D) \triangleq \mathsf{Var}(\jmath(X, D)) > 0$. Then, for each $i \in \mathbb{N}$, we can construct a R-D code $\mathcal{C}_i$ of sufficiently large blocklength n_i, encoder $\phi_{n_i} : \mathcal{X}^{n_i} \to [\![1, M_{n_i}]\!]$ and a reconstruction function $\psi_{n_i} : [\![1, M_{n_i}]\!] \to \hat{\mathcal{X}}^{n_i}$ such that

$$\begin{aligned}
\lim_{i\to\infty} \textstyle\sum_{i=1}^{M_{n_i}} \left|\mathbb{P}[\phi_{n_i}(X^{n_i}) = i] - \frac{1}{M_{n_i}}\right| &= 0 \\
\lim_{i\to\infty} \left|n_i^{-1} \log_2 M_{n_i} - R(D)\right| &= 0 \\
\lim_{i\to\infty} \mathbb{P}\Big[\textstyle\sum_{j=1}^{n_i} d\big(X_j, \big(\psi_{n_i}\big(\phi_{n_i}(X^{n_i})\big)\big)_j\big) > n_i D\Big] &= 0.
\end{aligned}$$

Proof: We begin with a sequence of codes $\{(\phi_n^*, \psi_n^*, \varphi_n^*)\}_{n\in\mathbb{N}}$ constructed using Theorem 2. These codes in fact meet the more-stringent **zero** excess distortion constraint [4]. Let $I_n \triangleq \phi_n^*(X^n)$ denote the output of the encoder ϕ_n^*. Let the description length of a prefix-free Shannon-Elias-Fano code for the source $\phi_n^*(X^n)$ be $l_n(X^n) = 1 + \lceil -\log_2 p_{I_n}(\phi_n^*(X^n)) \rceil$ [9], and let $l^*(X^n) = \mathsf{len}(\varphi_n(\phi_n^*(X^n)))$ denote the description length of the variable-length code φ_n^* for the source $\phi_n^*(X^n)$. By the competitive optimality of the Shannon-Elias-Fano code [9, Theorem 5.10.1], we have

$$\mathbb{P}\big[l_n(X^n) < l_n^*(X^n) + \log_2 n\big] \geq 1 - \tfrac{2}{n}. \quad (15)$$

Since it is true that

$$\log_2 \frac{1}{p_{I_n}\big(\phi_n^*(X^n)\big)} \leq l_n(X^n) \leq \log_2 \frac{4}{p_{I_n}\big(\phi_n^*(X^n)\big)},$$

we can combine (13), (14), and (15) with $b_n = 2\log_2 n$, to conclude that for some $\kappa > 0$

$$\mathbb{P}\left[\left|\log_2 \frac{1}{p_{I_n}\big(\phi_n^*(X^n)\big)} - \sum_{i=1}^{n} \jmath_X(X_i, D)\right| > \kappa \log_2 n\right] \underset{n\to\infty}{\longrightarrow} 0.$$

Fix $\varepsilon < H(X) - R(D)$ and let $N_0 \in \mathbb{N}$ be an integer such that for $n > N_0$, the above probability is no greater than $\frac{\varepsilon}{4}$.

Also note that by the Central Limit Theorem, we have

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} \frac{\jmath_X(X_i, D) - R(D)}{\sqrt{nV(D)}}\right| > Q^{-1}\left(\tfrac{\varepsilon}{10}\right)\right] \underset{n\to\infty}{\longrightarrow} \frac{\varepsilon}{5}.$$

Let $N_1 \in \mathbb{N}$ be chosen such that for $n > N_1$, the above probability is no more than $\frac{\varepsilon}{4}$.

By the AEP, we have $\mathbb{P}[X^n \notin \mathbb{W}_\varepsilon^n[p_X]] \to 0$ as $n \to \infty$. Let $N_2 \in \mathbb{N}$ be chosen such that the probability of realizing an atypical sequence is no more than $\frac{\varepsilon^2}{4}$.

We see that for $n > \max\{N_0, N_1, N_2\}$ such that

$$\kappa \log_2 n < \sqrt{nV(D)} Q^{-1}\big(\tfrac{\varepsilon}{10}\big), \quad (16)$$

$$2^{-n(H(X)-\varepsilon)} < \frac{\varepsilon}{3} 2^{-nR(d) - 3\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)}, \quad (17)$$

$$2^{-\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} < \varepsilon, \quad (18)$$

we are guaranteed that $\mathbb{P}\big[X^n \notin \mathbb{W}_\varepsilon^n[p_X]\big] < \frac{\varepsilon^2}{4}$ as well as

$$\mathbb{P}\left[\left|\log_2 \frac{2^{-nR(D)}}{p_{I_n}\big(\phi_n^*(X^n)\big)}\right| > 2\sqrt{nV(D)} Q^{-1}\big(\tfrac{\varepsilon}{10}\big)\right] < \frac{\varepsilon}{2}. \quad (19)$$

To arrive at (19), we required the description length characterizations of Theorems 2 and 3, and the variable-length code sequence $\{\varphi_n^*\}_{n\in\mathbb{N}}$. Moving forward, we only require the sequence of rate-distortion codes $\{(\phi_n^*, \psi_n^*)\}_{n\in\mathbb{N}}$. Now, define

$$\mathcal{I}_n \triangleq \left\{i : \left|\log_2 \frac{2^{-nR(D)}}{p_{I_n}(i)}\right| \leq 2\sqrt{nV(D)} Q^{-1}\big(\tfrac{\varepsilon}{10}\big)\right\}. \quad (20)$$

Now, filter the indices in $\mathcal{I}_n$ to create $\mathscr{I}_n$ by defining

$$\mathscr{I}_n \triangleq \left\{i \in \mathcal{I}_n : \mathbb{P}\big[X^n \in \mathbb{W}_\varepsilon^n[p_X] \big| \phi_n^*(X^n) = i\big] \geq 1 - \frac{\varepsilon}{2}\right\}.$$

These encoder outputs are precisely those which *nearly* have the same probability of occurrence, and have been generated predominantly by typical source realizations. A straightforward application of Lemma 2 yields $\mathbb{P}[I_n \in \mathscr{I}_n] \geq 1 - \varepsilon$.

Using this together with (20), we conclude that

$$|\mathscr{I}_n| \leq 2^{nR(D) + 2\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} \quad (21)$$

$$|\mathscr{I}_n| \geq (1 - \varepsilon) 2^{nR(D) - 2\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)}. \quad (22)$$

Note that even though the indices in $\mathscr{I}_n$ occur with nearly the same probability, their distribution is not close to uniform. The next step therefore is to further subdivide the pre-images ${\phi_n^*}^{-1}(i)$ for $i \in \mathscr{I}_n$ so that the resultant bins occur with near-equal probabilities. To do so, define $\omega_n : \mathscr{I}_n \to \mathbb{N}$ by

$$\omega_n(i) \triangleq \Big\lfloor p_{I_n}(i) 2^{nR(D) + 3\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} \Big\rfloor.$$

For $i \in \mathscr{I}_n$, $\omega_n(i)$ denotes the number of bins the typical sequences in the pre-image ${\phi_n^*}^{-1}(i) \cap \mathbb{W}_\varepsilon^n[p_X]$ will be subdivided to effect near-uniformity. Now, let $M_n \triangleq 1 + \sum_{i \in \mathscr{I}_n} \omega_n(i)$. This quantity represents the total number of bins generated after the subdivision process. We can bound M_n as follows.

$$M_n \le 1 + \mathbb{P}[I_n \in \mathscr{I}_n] 2^{nR(D)+3\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} \le \underbrace{1 + 2^{nR(D)+3\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)}}_{\triangleq \overline{M}_n}. \tag{23}$$

Similarly,

$$\begin{aligned} M_n &\ge 1 + \sum_{i \in \mathscr{I}_n} \left[p_{I_n}(i) 2^{nR(D)+3\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} - 1 \right] \\ &\overset{(21)}{\ge} \underbrace{\left[\begin{array}{c} 1 + (1-\varepsilon) 2^{nR(D)+3\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} \\ -2^{nR(D)+2\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} \end{array} \right]}_{\triangleq \underline{M}_n} \end{aligned} \tag{24}$$

Let $\zeta_n \triangleq (1-\varepsilon) 2^{-nR(D)-3\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)}$ denote the target probability for each of the bins generated by subdividing the pre-images $\phi_n^{*\,-1}(i) \cap \mathtt{W}_\varepsilon^n[p_X]$ for $i \in \mathscr{I}_n$. For each $i \in \mathscr{I}_n$, partition $\phi_n^{*\,-1}(i) \cap \mathtt{W}_\varepsilon^n[p_X]$ into sets $\mathcal{S}(i,j)$, $j = 1, \ldots, \Omega_n(i)$, such that for $i \in \mathscr{I}_n$ and $1 \le j < \Omega_n(i)$, we have

$$\mathbb{P}[X^n \in \mathcal{S}(i,j)] \in (\zeta_n, \zeta_n + 2^{-n(H(X)-\varepsilon)}), \tag{25}$$

$$\mathbb{P}[X^n \in \mathcal{S}(i, \Omega_n(i))] \in (0, \zeta_n + 2^{-n(H(X)-\varepsilon)}). \tag{26}$$

It is to be remarked here that the above partitioning can be arbitrary as long as the conditions (25) and (26) are met. Then,

$$\begin{aligned} \omega_n(i)(\zeta_n + 2^{-n(H(X)-\varepsilon)}) &\overset{(17)}{\le} (1 - \tfrac{2\varepsilon}{3}) \mathbb{P}[X^n \in \phi_n^{*\,-1}(i)] \\ &< \mathbb{P}\big[X^n \in \phi_n^{*\,-1}(i) \cap \mathtt{W}_\varepsilon^n[p_X]\big]. \\ &< \Omega_n(i)(\zeta_n + 2^{-n(H(X)-\varepsilon)}), \end{aligned}$$

where the last inequality follows from (25) and (26). Hence, for each $i \in \mathscr{I}_n$, $\Omega_n(i) > \omega_n(i)$. We therefore have

$$\begin{aligned} \kappa_n &\triangleq \sum_{\substack{i \in \mathscr{I}_n \\ 1 \le j \le \omega_n(i)}} \mathbb{P}\big[X^n \in \mathcal{S}(i,j)\big] \ge \sum_{i \in \mathscr{I}_n} \zeta_n \omega_n(i) \\ &\ge (1-\varepsilon) \mathbb{P}\big[\phi_n^*(X^n) \in \mathscr{I}_n\big] - \zeta_n |\mathscr{I}_n| \\ &\ge (1-\varepsilon)^2 - (1-\varepsilon) 2^{-\sqrt{nV(D)}Q^{-1}\left(\frac{\varepsilon}{10}\right)} \overset{(18)}{>} 1 - 3\varepsilon \end{aligned} \tag{27}$$

Now, let $\mathbb{I}_n = \big\{(i,j) : i \in \mathscr{I}_n, 1 \le j \le \omega_n(i)\big\} \cup \big\{(0,0)\big\}$ denote the set of 'bin indices' for the uniform lossy compression code to be constructed. Then, by definition, $|\mathbb{I}_n| = M_n$. Now, define the uniform lossy encoder map $\hat{\phi}_n : \mathcal{X}^n \to \mathbb{I}_n$ by

$$\hat{\phi}_n(x^n) \triangleq \begin{cases} (i,j) & x^n \in S(i,j), i \in \mathscr{I}_n \text{ and } j \le \omega_n(i) \\ (0,0) & \text{otherwise} \end{cases}.$$

The reconstruction function $\hat{\psi}_n : \mathbb{I}_n \to \hat{\mathcal{X}}^n$ for $\hat{\phi}_n$ is given by

$$\hat{\psi}_n(i,j) \triangleq \begin{cases} \psi_n^*(i) & (i,j) \ne (0,0) \\ \hat{x}^n & \text{otherwise} \end{cases},$$

where $\hat{x}^n$ is a particular element of $\hat{\mathcal{X}}^n$. By construction, the distortion constraint is not met only when $\hat{\phi}_n(X^n) = (0,0)$. The probability of this event occurring can be bounded by

$$\mathbb{P}[\hat{\phi}_n(X^n) = (0,0)] = 1 - \kappa_n \overset{(27)}{<} 3\varepsilon. \tag{28}$$

Lastly, the L_1-norm between the actual distribution of $\hat{\phi}_n(X^n)$ and the uniform distribution on $\mathbb{I}_n$ is bounded by

$$\begin{aligned} U_e(\hat{\phi}_n) &\triangleq \sum_{\iota \in \mathbb{I}_n} \left| \mathbb{P}[\hat{\phi}_n(X^n) = \iota] - \tfrac{1}{M_n} \right| \\ &\le \left(\begin{array}{l} \sum_{\iota \in \mathbb{I}_n} \left| \mathbb{P}[\hat{\phi}_n(X^n) = \iota] - \zeta_n \right| \\ \quad + \sum_{\iota \in \mathbb{I}_n} \left| \zeta_n - \tfrac{\zeta_n}{1-\varepsilon} \right| + \sum_{\iota \in \mathbb{I}_n} \left| \tfrac{\zeta_n}{1-\varepsilon} - \tfrac{1}{M_n} \right| \end{array} \right) \quad (29) \\ &\overset{(4)}{<} M_n \, 2^{-nH(X)+n\varepsilon} + 3\varepsilon + \zeta_n + \frac{\varepsilon M_n \zeta_n}{1-\varepsilon} + \frac{\overline{M}_n - \underline{M}_n}{\underline{M}_n} \\ &< \frac{6\varepsilon}{1-\varepsilon} + o(1), \end{aligned} \tag{30}$$

where we have used Lemma 3 for the last sum in (29), since $\underline{M}_n \le \frac{1-\varepsilon}{\zeta_n} \le \overline{M}_n$. Hence, for a sufficiently large $n \in \mathbb{N}$, there exists an $(M_n, n, 1, D, 7\varepsilon)$-uniform lossy compression code for the DMS p_X under the distortion measure d.

Now, pick $\{\varepsilon_i\}_{i \in \mathbb{N}}$ such that $\varepsilon_i \downarrow 0$ as $i \to \infty$. For each $i \in \mathbb{N}$, we can construct an $(M_{n_i}, n_i, 1, D, 7\varepsilon_i)$-uniform lossy compression code with encoding function $\hat{\phi}_{n_i}$ and reconstruction function $\hat{\psi}_{n_i}$ using the above technique. Thus, for this sequence of codes,

$$\begin{aligned} &\lim_{i \to \infty} U_e(\hat{\phi}_{n_i}) &&= 0, \\ &\lim_{i \to \infty} \left| n_i^{-1} \log_2 M_{n_i} - R(D) \right| &&= 0, \\ &\lim_{i \to \infty} \mathbb{P}\Big[n_i^{-1} \textstyle\sum_{j=1}^{n_i} d(X_j, \big(\hat{\psi}_{n_i}\big(\hat{\phi}_{n_i}(X^{n_i})\big)\big)_j > D \Big] &&= 0, \end{aligned}$$

which follow from (23), (24), (28), and (30). ■

Remark 3: Since the explicit rates of convergence in (13) and (14) are absent, the above proof does not guarantee the existence of sequences of R-D codes satisfying both $\frac{1}{n_i} \log_2 M_{n_i} \to R(D)$ and $D_{\mathrm{KL}}(p_{I_{n_i}} || \mathsf{unif}([\![1, M_{n_i}]\!])) \to 0$, where D_{KL} is the Kullback-Leibler divergence functional, and $\mathsf{unif}([\![1, M_{n_i}]\!])$ is the uniform distribution over $[\![1, M_{n_i}]\!]$.

REFERENCES

[1] T. S. Han, "Folklore in source coding: Information-spectrum approach," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 747–753, February 2005.

[2] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4619–4637, October 2008.

[3] R. Chou and M. Bloch, "Data compression with nearly uniform output," *2013 IEEE International Symposium on Information Theory*, pp. 1979–1983, 2013.

[4] I. Kontoyiannis, "Pointwise redundancy in lossy data compression and universal lossy data compression," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 136–152, January 2000.

[5] E.-H. Yang and Z. Zhang, "On the redundancy of lossy source coding with abstract alphabets," *IEEE Transactions on Information Theory, IEEE Transactions on*, vol. 45, no. 4, pp. 1092–1110, May 1999.

[6] V. Kostina and S. Verdu, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.

[7] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 1-2, pp. 1–184, 2014.

[8] T. S. Han, *Information-Spectrum Methods in Information Theory*, 1st ed. Springer, 2003.

[9] T. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.