Fifty-fourth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 27 - 30, 2016

Universal Covertness for Discrete Memoryless Sources

Rémi A. Chou, Matthieu R. Bloch, and Aylin Yener

***Abstract*—Consider a sequence S of length n emitted by a Discrete Memoryless Source (DMS) with unknown distribution p. We wish to construct a lossless source code that maps S to a sequence S' of minimal length m such that S' approximates in terms of Kullback-Leibler divergence a sequence emitted by another DMS with known distribution q. Our main result is the existence of a coding scheme that performs such a task with an asymptotically optimal compression rate, i.e., such that the limit of m/n is $H(p)/H(q)$ as n goes to infinity. Our coding scheme overcomes the challenges created by the lack of knowledge about p by relying on a sufficiently fine estimation of $H(p)$, followed by an appropriately designed type-based compression that jointly performs source resolvability and universal lossless source coding. Our result recovers several previous results that either assume p uniform, or q uniform, or p known. The price paid for these generalizations is the use of common randomness with vanishing rate. We further determine that the length of the latter roughly scales as the square root of n, by an analysis of second order asymptotics and error exponents.**

I. Introduction

Given n realizations X^n of a DMS $(\mathcal{X}, p_X)$, where $\mathcal{X}$ is a finite alphabet and p_X is an unknown distribution, we wish to form a sequence $\widehat{Y}^m$ of minimal lengh m, which, approximates a sequence emitted from the DMS $(\mathcal{Y}, p_Y)$ in terms of Kullback-Leibler divergence, where $\mathcal{Y}$ is a finite alphabet and p_Y is a given target distribution. Additionally, we require that X^n be losslessly reconstructed from Y^m. We refer to this problem as universal covertness for DMSs, since a warden observing $\widehat{Y}^m$ cannot distinguish it from m realizations of the DMS $(\mathcal{Y}, p_Y)$. The formal relation between the closeness of Y^m, in terms of Kullback-Leibler divergence, to the target distribution and the probability of detection by a warden follows by standard results on hypothesis testing [1], [2]. We implicitly assume that all parties share the same estimate of p_Y, obtained from a finite number of publicly available samples of $(\mathcal{Y}, p_Y)$. The problem is depicted in Figure 1.

Closely related settings have been studied in the literature. The closest is information-theoretic steganography [3]. Specifically, [3, Section 4] considers a similar model but does not address the problem of obtaining an optimal compression rate m/n. [3, Section 5] deals with a model in which p_Y is unknown but $H(Y)$ is known, and in which comparison of distributions is in terms of *normalized* Kullback-Leibler divergence.

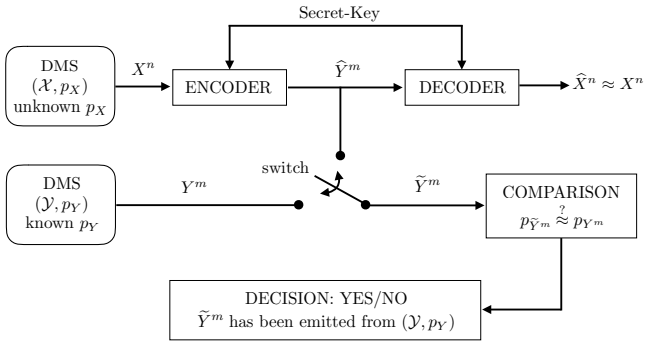

Fig. 1. Universal source covertness: problem description.

Our setting is also related to steganography as introduced in [4], yet with notable differences, including the sublinearity of the secret-key length used, the assumption that the source $(\mathcal{X}, p_X)$ is not necessarily uniform and has unknown statistics, our strict information-theoretic treatment without consideration of a distortion constraint between the encoder output $\widehat{Y}^m$ and the covertext Y^m, and the assumption that all parties, including the warden, only have an estimate of p_Y preventing the exact non-asymptotic requirement $p_{Y^m} = p_{\widehat{Y}^m}$, called perfect undetectability in [4].

Our model also recovers several notions including uniform lossless source coding [5] (by assuming p_X known and p_Y uniform), source resolvability [6] (by assuming p_X known and uniform, and by removing the reconstruction constraint), and random number conversion [7] (by assuming p_X known, and by removing the reconstruction constraint).

Finally, we stress that the special case of uniform lossless source coding for DMSs with *unknown distributions*, i.e., the case when p_Y is uniform, is of independent interest since uniformity of messages transmitted over a network is often a key assumption to establish secrecy results [8], [9].

The remainder of the paper is organized as follows. We formally describe the problem in Section II. We study the special case of uniform lossless source coding for DMSs with unknown distribution in Section III. Parts of the tech-

Rémi A. Chou and Aylin Yener are with the Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802 remi.chou@psu.edu,yener@ee.psu.edu

Matthieu R. Bloch is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0250 and with GT-CNRS UMI 2958, Metz, France matthieu.bloch@ece.gatech.edu

This work was supported in part by NSF under grants CIF-1319338, CNS-1314719, CCF-1320298, CIF-1527074, and by ANR with grant 13-BS03-0008.

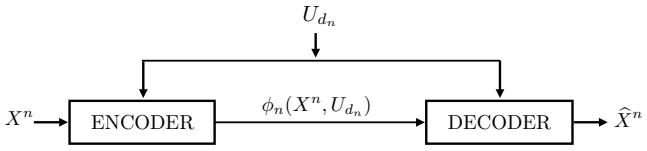

Fig. 2. Universal source covertness assisted with a seed.

nical proofs associated with Section III are relegated to the appendix to streamline presentation. Building on the results of Section III, we present our main result for universal source covertness in Section IV. Finally, we provide concluding remarks in Section V.

II. Problem statement

A. Notation

We define $[\![a,b]\!] \triangleq [\lfloor a \rfloor, \lceil b \rceil] \cap \mathbb{N}$. For two distributions p_X and $p_{X'}$ defined over a finite alphabet $\mathcal{X}$, we define $\mathbb{V}(p_X, p_{X'}) \triangleq \sum_{x\in\mathcal{X}} |p_X(x) - p_{X'}(x)|$ and refer to this quantity as the variational distance between p_X and $p_{X'}$. We denote the Kullback-Leibler divergence between two distributions by $\mathbb{D}(\cdot||\cdot)$. Unless specified otherwise, capital letters designate random variables, whereas lowercase letters designate realizations of associated random variables, e.g., x is a realization of the random variable X. We denote the set of all the distributions over $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$. We denote the indicator function by $\mathbb{1}\{\omega\}$, which is equal to 1 if the predicate ω is true and 0 otherwise. For any $x \in \mathbb{R}$, we define $[x]^+ \triangleq \max(0, x)$. Finally, we let p-lim denote convergence in probability.

B. Model for universal source covertness

Consider a discrete memoryless source $(\mathcal{X}, p_X)$. Let $n \in \mathbb{N}$, $d_n \in \mathbb{N}$, and let U_{d_n} be a uniform random variable over $\mathcal{U}_{d_n} \triangleq [\![1, 2^{d_n}]\!]$, independent of X^n. In the following we refer to U_{d_n} as the *seed* and d_n as its length. As illustrated in Figure 2, our objective is to design a source code to compress and reconstruct the source $(\mathcal{X}, p_X)$, whose distribution is unknown, with the assistance of a seed U_{d_n} and such that the encoder output approximates a target distribution with respect to the Kullback-Leibler divergence.

Definition 1. *An $(n, m, 2^{d_n})$ universal covert code with respect to the DMS $(\mathcal{Y}, p_Y)$ for n realizations of the DMS $(\mathcal{X}, p_X)$ with unknown distribution p_X consists of*

- *A seed set $\mathcal{U}_{d_n} \triangleq [\![1, 2^{d_n}]\!]$,*
- *An encoding function $\phi_n : \mathcal{X}^n \times \mathcal{U}_{d_n} \to \mathcal{Y}^m$,*
- *A decoding function $\psi_n : \mathcal{Y}^m \times \mathcal{U}_{d_n} \to \mathcal{X}^n$,*

where ϕ_n and ψ_n do not depend on prior knowledge about p_X.

The performance of the code is measured in terms of (i) reliability, i.e., average probability of error $\mathbb{P}[X^n \neq \psi_n(\phi_n(X^n, U_{d_n}), U_{d_n})]$, (ii) covertness, i.e., closeness of the encoder output to a target distribution $\mathbb{D}\left(p_{\phi_n(X^n, U_{d_n})} || p_{Y^m}\right)$, where $p_{Y^m} \triangleq \prod_{i=1}^m p_Y$ with p_Y a given distribution over $\mathcal{Y}$, (iii) its output length m, which should be minimal, i.e., asymptotically close to $nH(X)/H(Y)$, and (iv) the seed length d_n, which should be negligible compared to n.

III. Special case: Uniform lossless source coding for DMSs with unknown distribution

In this section, we first study a special case of the problem described in Section II-B, where p_Y is the uniform distribution over $\mathcal{Y}$. We refer to this special case as uniform lossless source coding for DMSs with *unknown distributions*. We will build upon the solution proposed for this special case to provide a solution for the general case in Section IV. The specific contributions of this section are the following.

- When the entropy of the source is known but its distribution is unknown, we show that uniform lossless source coding is possible if the shared seed has length on the order of $n^{1/2+\beta}$, $\beta > 0$, where n is the length of the sequence to compress.
- When lower and upper bounds on the entropy of the source are known but its distribution is unknown, and without allowing the encoder to refine the estimate of the entropy, we show that uniform lossless source coding requires a seed of length αn, where $\alpha > 0$ decreases with the gap between the lower and upper bounds.
- When the entropy and the distribution of the source are unknown and if one allows the encoder to estimate the former with the sequence to compress, e.g., with a plug-in estimate [10], uniform lossess source coding is possible with probability arbitrarily close to one as n goes to infinity, when the length of the seed is on the order of $n^{1/2+\beta}$, $\beta > 0$.

Our results generalize and complement an earlier result for DMSs with known distributions, which shows that uniform lossless source coding is possible if encoder and decoder share a *seed* [5], [8].

In the presence of sources with unknown distributions, the problem of uniform lossless source coding aims at *jointly* performing universal lossless source coding [11], [12] and universal randomness extraction [13]. The main technical challenges are (i) the design of an appropriate type-based source code that can support both reliability and uniformity constraints (see Section III-B and the proof of Theorem 1), (ii) the simplification of error exponents expressed as optimization problems (see Appendices A-C and A-D), (iii) the combination of entropy estimation and the coding scheme of Section III-B (see Theorem 3).

We describe our model, coding scheme, and results in Sections III-A, III-B, and III-C, respectively.

A. Model

As in Section II-B, consider a DMS $(\mathcal{X}, p_X)$, let $n \in \mathbb{N}$, $d_n \in \mathbb{N}$, and let U_{d_n} be a uniform random variable over $\mathcal{U}_{d_n} \triangleq [\![1, 2^{d_n}]\!]$, independent of X^n.

Definition 2. *A $(2^{nR_n}, n, 2^{d_n})$ uniform compression code is an $(n, nR_n, 2^{d_n})$ universal covert code with respect to the DMS $(\{0,1\}, (1/2, 1/2))$, for n realizations of the DMS $(\mathcal{X}, p_X)$ with unknown distribution p_X.*

B. Coding Scheme

We first recall some known facts about the method of types [11]. Let $n \in \mathbb{N}$. For any sequence $x^n \in \mathcal{X}^n$, the type of x^n is its empirical distribution

$$\left(\frac{\sum_{i=1}^n \mathbb{1}\{x_i = x\}}{n}\right)_{x\in\mathcal{X}}.$$

Let $\mathcal{P}_n(\mathcal{X})$ denote the set of all types over $\mathcal{X}$, and $T^n_{\bar{X}}$ denote the set of sequences x^n with type $p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})$. We will make use of the following lemma.

Lemma 1 ([11])**.** *The following properties hold.*

1) $|\mathcal{P}_n(\mathcal{X})| \leqslant (n+1)^{|\mathcal{X}|}$;
2) $(n+1)^{-|\mathcal{X}|} 2^{nH(\bar{X})} \leqslant |T^n_{\bar{X}}| \leqslant 2^{nH(\bar{X})}$;
3) For $x^n \in T^n_{\bar{X}}$, $p_{X^n}(x^n) = 2^{-n(H(\bar{X})+\mathbb{D}(p_{\bar{X}}||p_X))}$,

where $H(\bar{X})$ denotes the entropy of the type $p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})$.

Let $R_n > 0$ and define $\gamma(n) \triangleq |\mathcal{X}|\log(n+1)$. The encoder consists of a deterministic map

$$\begin{aligned}\phi_n : \mathcal{U}_{d_n} \times \mathcal{X}^n &\to [\![1, 2^{nR_n}]\!] \times [\![1, 2^{\gamma(n)}]\!] \\ (u, x^n) &\mapsto \left(\phi_n^{(1)}(u, x^n), \phi_n^{(2)}(u, x^n)\right),\end{aligned}$$

where for any seed u, $\phi_n^{(1)}(u, x^n)$ is an injective mapping whenever $H(\bar{X}) \leqslant R_n$, thus ensuring error-free compression in such a case, and $\phi_n^{(2)}(u, x^n)$ is an injective mapping that uniquely identifies the type $p_{\bar{X}}$ of x^n. The decoder consists of a deterministic map

$$\begin{aligned}\psi_n : \mathcal{U}_{d_n} \times [\![1, 2^{nR_n}]\!] \times [\![1, 2^{\gamma(n)}]\!] &\to \mathcal{X}^n \\ (u, i, j) &\mapsto \psi_n(u, i, j),\end{aligned}$$

where $\psi_n(u, i, j)$ is the unique x^n such that $\phi_n(u, x^n) = (i, j)$ when $H(\bar{X}) \leqslant R_n$ or an arbitrary sequence $\hat{x}^n$ otherwise.

We randomly generate ϕ_n described next. First, choose a mapping $\phi_n^{(1)}$ as follows.

- For all types $p_{\bar{X}}$ such that $H(\bar{X}) \leqslant R_n$ and for each $u \in \mathcal{U}_{d_n}$, choose $\phi_n^{(1)}(u, \cdot) : \mathcal{X}^n \to [\![1, 2^{nR_n}]\!]$ uniformly at random among the $\prod_{k=0}^{|T^n_{\bar{X}}|-1} \left(2^{nR_n} - k\right)$ possible injective mappings – this is possible by Lemma 1.
- For all types $p_{\bar{X}}$ such that $H(\bar{X}) > R_n$ and for each $u \in \mathcal{U}_{d_n}$, choose $\phi_n^{(1)}(u, \cdot) : \mathcal{X}^n \to [\![1, 2^{nR_n}]\!]$ uniformly at random among the $(2^{nR_n})^{|\mathcal{X}|^n}$ possible mappings.

Then, for each $u \in \mathcal{U}_{d_n}$, choose a mapping $\phi_n^{(2)}(u, \cdot) : \mathcal{X}^n \to [\![1, 2^{\gamma(n)}]\!]$ independently and uniformly at random among the $\prod_{k=0}^{|\mathcal{P}_n(\mathcal{X})|-1} \left(2^{\gamma(n)} - k\right)$ possible injective mappings.

Denote the random variables corresponding to these randomly generated mappings by Φ_n, $\Phi_n^{(1)}$, $\Phi_n^{(2)}$. The following lemma will prove useful later on.

Lemma 2. *For any $m \triangleq (i, j) \in [\![1, 2^{nR_n}]\!] \times [\![1, 2^{\gamma(n)}]\!]$,*

$$\mathbb{E}_{\Phi_n}\left[\mathbb{1}\{\phi_n(u, x^n) = m\}\right] = 2^{-(nR_n+\gamma(n))}. \tag{1}$$

The proof of Lemma 2 is omitted due to space constraints.

C. Results

Theorem 1. *For any $H > 0$, there exists a sequence of $(2^{nR_n}, n, 2^{d_n})$ uniform compression codes $\{\mathcal{C}_n\}_{n\geqslant 1}$ with $R_n \triangleq H + \frac{d_n}{2n}$ such that for any DMS with known entropy equal to H but unknown distribution, we have*

$$\lim_{n\to\infty} P_e(\phi_n, \psi_n) = 0, \quad \lim_{n\to\infty} U_e(\phi_n) = 0,$$
$$d_n = \Theta(n^{1/2+\beta}),$$

where $\beta > 0$ is arbitrary.

The proof of Theorem 1 is presented in Appendix A. Note that the entropy of the source to compress is only needed to specify the rates R_n of the codes. We next consider the case for which the entropy of the source to compress is not perfectly known.

Theorem 2. *For any $H > 0$, there exists a sequence of $(2^{nR_n}, n, 2^{d_n})$ uniform compression codes $\{\mathcal{C}_n\}_{n\geqslant 1}$ with $R_n \triangleq H + \epsilon_n$ and $\epsilon_n \triangleq \frac{\lceil n^{1/2+\beta}\rceil}{n}$, $\beta > 0$, such that for any discrete memoryless source with entropy known to belong to the interval $[H_0, H]$, where $H_0 < H$, but with unknown distribution, we have*

$$\lim_{n\to\infty} P_e(\phi_n, \psi_n) = 0, \quad \lim_{n\to\infty} U_e(\phi_n) = 0,$$
$$d_n = 2n\epsilon_n + n(H - H_0).$$

Theorem 2 is a consequence of Theorem 1, its proof is omitted due to space constraints. The bound in Theorem 2 is rather pessimistic, as the penalty paid, in terms of seed length, for not exactly knowing the entropy of the source is $\Theta(n)$. The following theorem shows how to mitigate this caveat.

Theorem 3. *For any $n \in \mathbb{N}$, there exists a set $\mathcal{S}_n$ of uniform compression codes, there exist two functions $f_n : (X^n, \mathcal{S}_n) \mapsto \phi_n$, $g_n : (M, \mathcal{S}_n) \mapsto \psi_n$, where $(\phi_n, \psi_n) \in \mathcal{S}_n$ and M is the compressed sequence observed by the decoder, such that for any DMS with unknown distribution, if the encoder chooses $f_n(X^n, \mathcal{S}_n) = \phi_n$ to encode X^n into M and the decoder chooses $g_n(M, \mathcal{S}_n) = \psi_n$ to recover X^n from M, then*

$$\lim_{n\to\infty} P_e(\phi_n, \psi_n) = 0, \ \operatorname*{p-lim}_{n\to\infty} U_e(\phi_n) = 0,$$
$$\operatorname*{p-lim}_{n\to\infty} R_n = H(X), \ d_n = \Theta(n^{1/2+\beta}),$$

where R_n is the rate of ϕ_n and $\beta > 0$. Note that R_n is estimated from X^n. More precisely, for any $r > 0$, the following uniformity condition is satisfied

$$\lim_{n\to\infty} \mathbb{P}\left[U_e(\phi_n) \leqslant n^{-r}\right] = 1.$$

Proof. We first describe the code construction. Let $n \in \mathbb{N}^*$. Let $t < 1/2$ and define

$$q \triangleq \lceil n^t \rceil, \quad \delta \triangleq \frac{\log|\mathcal{X}|}{n^t}.$$

We also define $a_i \triangleq i \times \delta, i \in [\![0, q-1]\!]$, $a_q \triangleq \log|\mathcal{X}|$, $a_{-1} \triangleq a_0$, and $a_{q+1} \triangleq a_q$ such that $\{[a_i, a_{i+1}]\}_{i\in[\![0,q-1]\!]}$ is a partition of $[0, \log|\mathcal{X}|]$. For every pair

$$(H_0, H) \in \mathcal{H} \triangleq \{(a_{i-2}, a_{i+1}) : i \in [\![1, q]\!]\},$$

we construct a uniform compression code for sources with entropy known to belong to the interval $[H_0, H]$ using Theorem 3. Let $\mathcal{S}_n$ denote the set of the q uniform compression codes constructed $\{(\phi_n^{(i)}, \psi_n^{(i)})\}_{i\in[\![1,q]\!]}$, $i \in [\![1, q]\!]$, and assume that encoder and decoder share the sequence $\{\mathcal{S}_n\}_{n\in\mathbb{N}^*}$. Note that this sequence is deterministic and shared before any observation of the sequence to compress.

We now describe the encoding/decoding process. Let x^n denote a realization of X^n and let $\widehat{H}(x^n)$ denote the plug-in estimate [10] of $H(X)$ using x^n. There exists $I_0 \in [\![1, q]\!]$ such that $\widehat{H}(x^n) \in [a_{I_0-1}, a_{I_0}]$. Define the mean and variance of the plug-in estimator

$$\mu \triangleq \mathbb{E}[\hat{H}(x^n)], \quad \sigma^2 \triangleq \mathbb{E}[(\hat{H}(x^n) - \mu)^2],$$

which are shown to be $\mu = H(X) + \delta_n$ and $\sigma^2 = O(n^{-1})$, with $\delta_n = O(n^{-1})$ in [10]. Also define the events

$$\mathcal{E} \triangleq \{(H(X) \geqslant a_{I_0+1}) \text{ or } (H(X) \leqslant a_{I_0-2})\},$$
$$\widetilde{\mathcal{E}} \triangleq \left\{\left(H(X) \geqslant H^{(n)}\right) \text{ or } \left(H(X) \leqslant H_0^{(n)}\right)\right\},$$

where

$$H_0^{(n)} \triangleq \hat{H}(x^n) - \delta_n - n^{-2t},$$
$$H^{(n)} \triangleq \hat{H}(x^n) - \delta_n + n^{-2t}.$$

We then have

$$\begin{aligned}
\mathbb{P}[\mathcal{E}] &\overset{(a)}{\leqslant} \mathbb{P}[\widetilde{\mathcal{E}}] \\
&= \mathbb{P}[|\hat{H}(x^n) - \delta_n - H(X)| \geqslant n^{-2t}] \\
&= \mathbb{P}[|\hat{H}(x^n) - \mu| \geqslant n^{-2t}] \\
&\overset{(b)}{\leqslant} \frac{\sigma^2}{n^{-4t^2}} \\
&= O(n^{-1+4t^2}), \qquad (2)
\end{aligned}$$

where (a) holds because for n large enough, $(\delta_n + n^{-2t}) = o(a_{I_0-1} - a_{I_0-2}) = o(n^{-t})$, $(-\delta_n + n^{-2t}) = o(a_{I_0+1} - a_{I_0}) = o(n^{-t})$, and $[H_0^{(n)}, H^{(n)}]$ is thus a subinterval of $[a_{I_0-1}, a_{I_0+2}]$, (b) holds by Chebyshev's inequality.

Since $[a_{I_0-2}, a_{I_0+1}] \in \mathcal{H}$, there exists a code $(\phi_n^{(I_0)}, \psi_n^{(I_0)})$ in $\mathcal{S}_n$ that has been designed for $(H_0, H) = (a_{I_0-2}, a_{I_0+1})$. The encoder uses $\phi_n^{(I_0)}$ to encode x^n. The decoder knows which code to choose in $\mathcal{S}_n$ via the length of the compressed sequence, which embeds the code rate and uniquely identifies one code in $\mathcal{S}_n$.

Finally, remembering that I_0 depends on x^n, define the set

$$\mathcal{E}_n \triangleq \{x^n \in \mathcal{X}^n : H(X) \in [a_{I_0-2}, a_{I_0+1}]\}.$$

We have

$$\begin{aligned}
&\mathbb{E}_{X^n}[U_e(\phi_n^{(I_0)})] \\
&= \sum_{x^n\in\mathcal{E}_n} p(x^n) U_e(\phi_n^{(I_0)}) + \sum_{x^n\notin\mathcal{E}_n} p(x^n) U_e(\phi_n^{(I_0)}) \\
&\leqslant U_e(\phi_n^{(I_0)}) + 2\mathbb{P}[X^n \notin \mathcal{E}_n] \xrightarrow{n\to\infty} 0,
\end{aligned}$$

where the limit holds by design of $\phi_n^{(I_0)}$, i.e., Theorem 2, and by (2). We thus have convergence in the mean, which implies convergence in probability. By the law of total probability, we also have $\lim_{n\to\infty} P_e(\phi_n^{(I_0)}, \psi_n^{(I_0)}) = 0$ by (2). □

Observe that $H(X)$ dictates the rate of the source code, and an underestimated $H(X)$ will prevent reliability, whereas an overestimated $H(X)$ will prevent a correct approximation of the target distribution p_Y by the encoder output. Consequently, a fine enough estimation of the entropy of the source is crucial, and makes our coding scheme variable-length, since this estimation depends on the sequence to compress.

IV. Covertness for DMSs with unknown distribution

Our coding scheme for universal covertness makes use of two building blocks, which are two special cases of the model described in Section II-B. (i) Uniform compression for DMS with unknown distribution, studied in Section III. (ii) Source resolvability [6] with a reliability constraint, which corresponds to the case where p_X is known to be the uniform distribution. We show in the following result how to combine (i) and (ii) to obtain universal source covertness.

Theorem 4. *For any $n \in \mathbb{N}$, for any sequence X^n emitted from a DMS with unknown distribution, there exists an encoding function ϕ_n and a decoding function ψ_n, such that if one defines $\widehat{Y}^m \triangleq \phi_n(X^n, U_{d_n})$, then*

$$\lim_{n\to\infty} \mathbb{P}[X^n \neq \psi_n(\widehat{Y}^m, U_{d_n})] = 0,$$
$$\operatorname*{p-lim}_{n\to\infty} \mathbb{D}\left(p_{\widehat{Y}^m} || p_{Y^m}\right) = 0,$$
$$\operatorname*{p-lim}_{n\to\infty} \frac{m}{n} = H(X)/H(Y),$$
$$d_n = \Theta(n^{1/2+\beta}),$$

where $\beta > 0$. Note that ϕ_n and thus m are determined given the realization of the sequence X^n. More precisely, for any

$r > 0$, the following covertness condition is satisfied

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{D}\left(p_{\widehat{Y}^m}||p_{Y^m}\right) \leqslant n^{-r}\right] = 1.$$

Proof overview. Define $R_Y \triangleq H(Y) - \epsilon$, $\epsilon > 0$, where $H(Y)$ is the entropy associated with the target distribution p_Y. Let $\beta > 0$, $n \in \mathbb{N}$. We consider a set $\mathcal{S}_n$ of uniform compression codes, and the functions $f_n : (X^n, \mathcal{S}_n) \mapsto \phi_n$, $g_n : (M, \mathcal{S}_n) \mapsto \psi_n$ provided by Theorem 3, where $(\phi_n, \psi_n) \in \mathcal{S}_n$ and M is the compressed sequence that the encoder outputs. We also consider q and $I_0 \in [\![1, q]\!]$ as in the proof of Theorem 3. We define I as the modulo-2 sum of the binary representation of I_0 and a uniformly distributed sequence $\widetilde{K}$ of $\log q$ bits. We assume that $\widetilde{K}$ is shared by the encoder and the decoder – note that, by definition of q, this additional amount of shared randomness is negligible compared to $n^{1/2+\beta}$. We then define the integer

$$m \triangleq \left\lceil \frac{|M|+|I|}{R_Y} \right\rceil,$$

where $|\cdot|$ denotes the length of a sequence, and the sequence

$$M' \triangleq [M||C||I],$$

where $||$ denotes concatenation and C is a uniformly distributed sequence of $|C|$ bits, with $|C| \in [\![0, R_Y]\!]$ such that

$$|M|+|C|+|I| = \lceil mR_Y \rceil.$$

Remark 1. *Note that if one chooses $m \triangleq \left\lceil \frac{|M|}{R_Y} \right\rceil$ with $I = \emptyset$, $C = \emptyset$, then only knowing m leads to an uncertainty on the length of M. We have thus added the extra information I, to allow the decoder to recover I_0 and thus the length of M, which in turn allows to select the right code in $\mathcal{S}_n$ for the reconstruction of X^n.*

Then, by definition of m, and Theorem 2,

$$\text{p-}\lim_{n\to\infty} \frac{m}{n} = \frac{H(X)}{H(Y) - \epsilon}.$$

Next, by means of random binning using, for instance, the method in [14], it is possible to construct a map h that performs source resolvability for $(\mathcal{Y}, p_Y)$ with lossless reconstruction, i.e., the output of the map has a distribution close to p_{Y^m} and the map input can be perfectly reconstructed from its output. Note that standard resolvability results [6] are insufficient for our purposes as we require perfect recoverability of the input from the output.

Finally, the encoder forms $\widehat{Y}^m \triangleq h(M')$, so that the decoder can determine from $\widehat{Y}^m$, in this order, M', then I, then I_0, then M, and finally approximate X^n using g_n. Note that we have obtained $\text{p-}\lim_{n\to\infty} \mathbb{V}\left(p_{\widehat{Y}^m}, p_{Y^m}\right) = 0$, but by the proof of [14, Theorem 1], which relies on strong typicality and by the proof of Theorem 3 we also have for any $r > 0$, $\lim_{n\to\infty} \mathbb{P}\left[\mathbb{V}\left(p_{\widehat{Y}^m}, p_{Y^m}\right) \leqslant n^{-r}\right] = 1$. We consequently get for any $r > 0$, $\lim_{n\to\infty} \mathbb{P}\left[\mathbb{D}\left(p_{\widehat{Y}^m}||p_{Y^m}\right) \leqslant n^{-r}\right] = 1$ by the following relation [15]

$$\mathbb{D}(p\|q) \leqslant \log\left(\frac{1}{\mu_q}\right) \mathbb{V}(p, q),$$

where p, q are two distributions over the finite alphabet $\mathcal{X}$ with supports equal to $\mathcal{X}$, $\mu_q \triangleq \min_{x\in\mathcal{X}} q(x)$.

V. Concluding Remarks

Our proposed construction consists of the combination of (i) a fine enough estimation of the source entropy via a plug-in estimator [10], and (ii) an appropriately designed type-based coding scheme able to simulatenously perform universal lossless source coding and source resolvablity. To simplify our analysis, we divide the problem into two simpler problems, source resolvability with lossless reconstruction and universal uniform lossless source coding. Our coding scheme makes use of a seed, i.e., a uniformly distributed sequence of bits, shared by the encoder and the decoder. Although the seed rate vanishes to zero as n grows and has a length in the order of $O(n^{1/2+\beta})$, $\beta > 0$, that favorably compares to the the seed length $\log(n!)$ required in [4], the question whether a seed is required at all for the convergence speed proposed remains open.

A future challenge would be to propose a low-complexity coding scheme for universal covertness. Note that in a non universal setting, i.e., when the distribution of the sequence to cover is known, it is possible to obtain a low-complexity coding scheme with polar codes by combining the uniform lossless source code of [9] and the source resolvability code in [16].

Appendix A
Proof of Theorem 1

We prove Theorem 1 for the uniformity constraint

$$U_e(\phi_n) \triangleq \mathbb{V}\left(p_{\phi_n(X^n, U_{d_n})}, p_{U_{M_n}}\right),$$

where $p_{U_{M_n}}$ is the uniform distribution over $\mathcal{M}_n$. To obtain the result for the uniformity constraint involving the Kullback-Leibler as in Definition 2, we can make use of [11, Lemma 2.7], which ensures that if $\lim_{n\to\infty} n\mathbb{V}\left(p_{\phi_n(X^n, U_{d_n})}, p_{U_{M_n}}\right) = 0$, then $\lim_{n\to\infty} \mathbb{D}\left(p_{\phi_n(X^n, U_{d_n})}||p_{U_{M_n}}\right) = 0$.

We provide upper bounds on average over the choice of Φ_n for the quantities P_e and U_e in Sections A-A and A-B. In Sections A-C and A-D, we further simplify the upper bounds proved in Sections A-A and A-B, which will allow us to study second order asymptotics. Finally, in Section A-E, we derive a sufficient condition on the seed length d_n to ensure a nearly uniform encoder output and near lossless reconstruction. To simplify notation, we drop the subscript n for Ψ_n and Φ_n.

A. Upper-bound on $\mathbb{E}_\Phi[P_e]$

For any $u \in \mathcal{U}_{d_n}$, for any $x^n \in \mathcal{X}^n$, we define

$$E(u, x^n) \triangleq \mathbb{1}\left\{\Psi(\Phi(u, x^n), u) \neq x^n\right\}.$$

By using Lemma 1 and standard techniques for universal compression via the method of types [11], one can show

$$\mathbb{E}_\Phi[P_e] \leqslant 2^{-n \min\limits_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) > R_n}} \mathbb{D}(p_{\bar{X}} || p_X) + \gamma(n)}. \tag{3}$$

B. Upper-bound on $\mathbb{E}_\Phi[U_e]$

We let M be the output of the encoder and define

$$A(\Phi) \triangleq \frac{1}{|T^n_{\bar{X}}|} \sum_{x^n \in T^n_{\bar{X}}} \sum_u p(u) \mathbb{1}\{\Phi(u, x^n) = m\}.$$

For $m \in [\![1, 2^{nR_n + \gamma(n)}]\!]$, it can be shown by (1),

$$\mathbb{E}_\Phi\left[A(\Phi)\right] = 2^{-(nR_n + \gamma(n))}. \tag{4}$$

Hence, by defining p_U as the uniform distribution over $[\![1, 2^{nR_n+\gamma(n)}]\!]$, it can be shown that,

$$\begin{aligned}
&\mathbb{E}_\Phi[U_e] \\
&= \mathbb{E}_\Phi\left[\sum_m \left| \sum_{p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})} \mathbb{P}[T^n_{\bar{X}}] \left(A(\Phi) - 2^{-(nR_n+\gamma(n))}\right)\right|\right].
\end{aligned} \tag{5}$$

Next, we have

$$\begin{aligned}
&\mathbb{E}_\Phi[U_e] \\
&\overset{(a)}{\leqslant} \mathbb{E}_\Phi\left[\sum_m \sum_{p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})} \mathbb{P}[T^n_{\bar{X}}] \left|A(\Phi) - 2^{-(nR_n+\gamma(n))}\right|\right] \\
&= \sum_{p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})} \mathbb{P}[T^n_{\bar{X}}] \sum_m \mathbb{E}_\Phi\left[\left|A(\Phi) - 2^{-(nR_n+\gamma(n))}\right|\right] \\
&\overset{(b)}{\leqslant} \sum_{p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})} \mathbb{P}[T^n_{\bar{X}}] \sum_m \sqrt{\mathrm{Var}_\Phi\left(A(\Phi)\right)},
\end{aligned} \tag{6}$$

where (a) holds by the triangle inequality and (5), (b) holds by Jensen's inequality and by Equation (4). We then have

$$\begin{aligned}
&\mathbb{E}_\Phi\left[\left(\sum_{x^n \in T^n_{\bar{X}}} \sum_u p(u) \mathbb{1}\{\Phi(u, x^n) = m\}\right)^2\right] \\
&= \sum_{x^n, \hat{x}^n} \sum_{u, \hat{u} \neq u} p(u) p(\hat{u}) \mathbb{E}_\Phi[\mathbb{1}\{\Phi(u, x^n) = m\} \mathbb{1}\{\Phi(\hat{u}, \hat{x}^n) = m\}] \\
&\quad + \sum_{x^n, \hat{x}^n \neq x^n} \sum_u p(u)^2 \mathbb{E}_\Phi[\mathbb{1}\{\Phi(u, x^n) = m\} \mathbb{1}\{\Phi(u, \hat{x}^n) = m\}] \\
&\quad + \sum_{x^n} \sum_u p(u)^2 \mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(u, x^n) = m\}^2\right] \\
&\overset{(a)}{\leqslant} \sum_{x^n, \hat{x}^n} \sum_{u, \hat{u}} p(u) p(\hat{u}) \frac{1}{2^{2(nR_n + \gamma(n))}} + \sum_{x^n, u} p(u)^2 \frac{1}{2^{(nR_n + \gamma(n))}} \\
&\overset{(b)}{=} \frac{|T^n_{\bar{X}}|^2}{2^{2(nR_n+\gamma(n))}} + |T^n_{\bar{X}}| 2^{-d_n} 2^{-(nR_n+\gamma(n))},
\end{aligned} \tag{7}$$

where in (a) the first expectation in the left-hand side (l.h.s.) of the inequality is, by (1) and since the choice of $x^n \mapsto \phi_n(u, x^n)$ is made independently for each $u \in \mathcal{U}_{d_n}$,

$$\begin{aligned}
&\mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(u, x^n) = m\} \mathbb{1}\{\Phi(\hat{u}, \hat{x}^n) = m\}\right] \\
&= \mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(u, x^n) = m\}\right] \mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(\hat{u}, \hat{x}^n) = m\}\right] \\
&= 2^{-2(nR_n+\gamma(n))}.
\end{aligned}$$

We differentiate two cases for the second expectation. If $H(\bar{X}) \leqslant R_n$, then by injectivity of Φ we have

$$\mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(u, x^n) = m\} \mathbb{1}\{\Phi(u, \hat{x}^n) = m\}\right] = 0,$$

and if $H(\bar{X}) > R_n$, we have

$$\begin{aligned}
&\mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(u, x^n) = m\} \mathbb{1}\{\Phi(u, \hat{x}^n) = m\}\right] \\
&= \mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(u, x^n) = m\}\right] \mathbb{E}_\Phi\left[\mathbb{1}\{\Phi(u, \hat{x}^n) = m\}\right] \\
&= 2^{-2(nR_n+\gamma(n))}.
\end{aligned}$$

Finally, the third expectation follows from (1). (b) holds by marginalization over u and $\hat{u}$, and because the sums over x^n have $|T^n_{\bar{X}}|$ terms. We thus obtain

$$\begin{aligned}
\mathrm{Var}_\Phi\left(A(\Phi)\right) &\overset{(a)}{=} \mathbb{E}[A(\Phi)^2] - \frac{1}{2^{2(nR_n+\gamma(n))}} \\
&\overset{(b)}{=} \frac{1}{|T^n_{\bar{X}}|} 2^{-d_n} 2^{-(nR_n+\gamma(n))} \\
&\overset{(c)}{\leqslant} 2^{-nH(\bar{X})+\gamma(n)} 2^{-d_n} 2^{-(nR_n+\gamma(n))},
\end{aligned} \tag{8}$$

where (a) holds by (4), (b) holds by (7), (c) holds by Lemma 1. Define

$$R(d_n) \triangleq R_n - d_n/n. \tag{9}$$

We have

$$\begin{aligned}
\sum_m \sqrt{\mathrm{Var}_\Phi\left(A(\Phi)\right)} &\overset{(a)}{=} \sum_m \sqrt{2^{-nH(\bar{X}) - d_n - nR_n}} \\
&= 2^{-\frac{n}{2}(H(\bar{X}) - R_n + d_n/n)} \\
&\overset{(b)}{\leqslant} 2 \cdot 2^{-\frac{n}{2}[H(\bar{X}) - R(d_n)]^+},
\end{aligned} \tag{10}$$

where (a) holds by (8), and (b) holds because in the l.h.s. of (6) $\sum_m \left|A(\Phi) - 2^{-(nR_n+\gamma(n))}\right|$ is a variational distance and is upper bounded by 2.

Finally, define

$$E(n) \triangleq \min_{p_{\bar{X}} \in \mathcal{P}(\mathcal{X})} [[H(\bar{X}) - R(d_n)]^+ + 2\mathbb{D}(p_{\bar{X}} || p_X)], \tag{11}$$

such that we obtain

$$\begin{aligned}
\mathbb{E}_\Phi[U_e] &\overset{(a)}{\leqslant} \sum_{p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})} \mathbb{P}[T^n_{\bar{X}}] \cdot 2 \cdot 2^{-\frac{n}{2}[H(\bar{X}) - R(d_n)]^+} \\
&\overset{(b)}{\leqslant} \sum_{p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})} 2^{-n\mathbb{D}(p_{\bar{X}} || p_X)} 2 \cdot 2^{-\frac{n}{2}[H(\bar{X}) - R(d_n)]^+}
\end{aligned}$$

$$\overset{(c)}{\leqslant} \sum_{p_{\bar{X}} \in \mathcal{P}_n(\mathcal{X})} 2 \cdot 2^{-\frac{n}{2}E(n)}$$
$$\overset{(d)}{\leqslant} 2 \cdot 2^{-\frac{n}{2}E(n)+\gamma(n)}, \tag{12}$$

where (a) holds by (6) and (10), (b) and (d) hold by Lemma 1, (c) holds by (11) and because $\mathcal{P}_n(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$.

C. Simpler upper-bound on $\mathbb{E}_\Phi[U_e]$

Let $(\epsilon_n)_{n\in\mathbb{N}}$ be a positive sequence decreasing and converging to zero. For $n \in \mathbb{N}$, let

$$R_n \triangleq d_n/n + H(X) - \epsilon_n.$$

We first show the following lemma.

Lemma 3. *Making use of the definitions* (9)*,* (11)*, we have*

$$E(n) \geqslant \min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X). \tag{13}$$

Proof. For any $p_{\bar{X}} \in \mathcal{P}(\mathcal{X})$, define

$$f(p_{\bar{X}}, n) \triangleq H(\bar{X}) - R(d_n) + \mathbb{D}(p_{\bar{X}}||p_X).$$

First, note that by positivity of the divergence we have

$$\begin{aligned} &E(n) \\ &\geqslant \min_{p_{\bar{X}} \in \mathcal{P}(\mathcal{X})} [H(\bar{X}) - R(d_n)]^+ + \mathbb{D}(p_{\bar{X}}||p_X) \\ &= \min \left(\min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X), \min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \geqslant R(d_n)}} f(p_{\bar{X}}, n) \right). \end{aligned} \tag{14}$$

Observe also that $f(p_{\bar{X}}, n)$ is a linear function of $p_{\bar{X}}$, since

$$H(\bar{X}) + \mathbb{D}(p_{\bar{X}}||p_X) = \sum_x p_{\bar{X}}(x) \log \frac{1}{p_X(x)}.$$

We then have,

$$\begin{aligned} \min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \geqslant R(d_n)}} f(p_{\bar{X}}, n) &\overset{(a)}{=} \max_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \geqslant R(d_n)}} [-f(p_{\bar{X}}, n)] \\ &\overset{(b)}{=} \max_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) = R(d_n)}} [-f(p_{\bar{X}}, n)] \\ &= \min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) = R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X)] \\ &\geqslant \min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X)], \end{aligned} \tag{15}$$

where (a) can be seen as a maximization of a continuous convex function over a convex compact set $\mathcal{C}$, since $(H(\bar{X}) + \mathbb{D}(p_{\bar{X}}||p_X))$ is a linear function of $p_{\bar{X}}$, and $-H(\bar{X})$ is a convex function of $p_{\bar{X}}$. It can be shown, as in the proof of [17, Proposition 5.2], that the maximum is attained at an extreme point of $\mathcal{C}$ using the maximum principle [18]. Moreover, as in [17, Proposition 5.2], one can show that the set of extreme points of $\mathcal{C}$ is a subset of the set of points that satisfy the constraint with equality, and consequently we obtain (b). Finally, the result follows by combining (14) and (15). □

The following lemma shows that the right hand side of Equation (13) converges to zero as n goes to infinity.

Lemma 4. *Consider the sequence* $\left(p_{\bar{X}}^{(n)}\right)_{n\in\mathbb{N}}$*, where for any* $n \in \mathbb{N}$ *we have defined*

$$p_{\bar{X}}^{(n)} \triangleq \operatorname*{arg\,min}_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X).$$

We have

$$\lim_{n\to\infty} \mathbb{D}\left(p_{\bar{X}}^{(n)}||p_X\right) = 0.$$

Proof overview. To show the lemma, it is sufficient to upper bound the sequence

$$\left(\min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X) \right)_{n\in\mathbb{N}}$$

with a sequence that goes to zero as n goes to infinity. By definition of R_n and $R(d_n)$, we have for $n \in \mathbb{N}$,

$$\min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X) = \min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant H(X) - \epsilon_n}} \mathbb{D}(p_{\bar{X}}||p_X). \tag{16}$$

Let x_0, x_1 be arbitrary elements of $\mathcal{X}$, define $p_0 \triangleq p_X(x_0)$ and $p_1 \triangleq p_X(x_1)$, and assume $p_0 > p_1$. We define a probability distribution $p_{\tilde{X}}$ as follows. Define for any $n \in \mathbb{N}$, for any $x \in \mathcal{X}\backslash\{x_0, x_1\}$, for any $\delta_n \in]0, \min(p_1, 1-p_0)[$,

$$p_{\tilde{X}}(x) \triangleq p_X(x),$$
$$p_{\tilde{X}}(x_0) \triangleq p_0 + \delta_n, \quad p_{\tilde{X}}(x_1) \triangleq p_1 - \delta_n.$$

It can be shown that

$$\mathbb{D}(p_{\tilde{X}}||p_X) \leqslant \frac{\delta_n}{\mu_X}, \tag{17}$$

where $\mu_X \triangleq \min_{x \in \text{Supp}(p_X)} p_X(x)$, and

$$H(X) - H(\tilde{X}) \geqslant \epsilon_n, \tag{18}$$

where we have chosen $\delta_n \triangleq \epsilon_n \left(\log \frac{p_0}{p_1}\right)^{-1}$.

Hence, we obtain

$$\begin{aligned} \min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}}||p_X) &\overset{(a)}{\leqslant} \mathbb{D}(p_{\tilde{X}}||p_X) \\ &\overset{(b)}{\leqslant} \frac{\delta_n}{\mu_X} \\ &\overset{(c)}{\leqslant} \frac{\epsilon_n}{\mu_X} \left(\log \frac{p_0}{p_1}\right)^{-1} \\ &\xrightarrow{n\to\infty} 0, \end{aligned}$$

where (a) holds by (16) and (18), (b) holds by (17), (c) holds by definition of δ_n. □

Using Lemma 4, we can now obtain the following lower bound for the right-hand side of Equation (13).

Lemma 5. *For any $\beta > 0$, we have for n large enough*

$$\min_{\substack{p_{\bar{X}} \in \mathcal{P}(\mathcal{X}) \\ H(\bar{X}) \leqslant R(d_n)}} \mathbb{D}(p_{\bar{X}} || p_X) \geqslant \epsilon_n^{2+\beta}.$$

Proof. We define for any $n \in \mathbb{N}$, $p_{\bar{X}}^{(n)}$ as in Lemma 4, and let $\bar{X}^{(n)}$ be a random variable distributed according to $p_{\bar{X}}^{(n)}$. We have by Lemma 4 and by Pinsker's inequality

$$\begin{aligned} \mathbb{V}\left(p_{\bar{X}}^{(n)}, p_X\right) &\leqslant \sqrt{2 \ln 2} \sqrt{\mathbb{D}(p_{\bar{X}}^{(n)} || p_X)} \\ &\xrightarrow{n \to \infty} 0. \end{aligned} \tag{19}$$

For any $\alpha > 0$, for any $C > 0$, for n large enough, we have

$$\begin{aligned} \epsilon_n &\overset{(a)}{\leqslant} H(X) - H(\bar{X}^{(n)}) \\ &\overset{(b)}{\leqslant} \mathbb{V}\left(p_{\bar{X}}^{(n)}, p_X\right) \log \frac{|\mathcal{X}|}{\mathbb{V}\left(p_{\bar{X}}^{(n)}, p_X\right)} \\ &\overset{(c)}{\leqslant} C \mathbb{V}\left(p_{\bar{X}}^{(n)}, p_X\right)^{1-\alpha}, \end{aligned} \tag{20}$$

where (a) holds by definition of $\bar{X}^{(n)}$ and R_n, (b) holds by [11][Lemma 2.7], then observe that for $\alpha > 0$, $\frac{x \log \frac{1}{x}}{x^{1-\alpha}} \xrightarrow{x \to 0^+} 0^+$ and thus for any $C > 0$, for n large enough, (c) holds by Equation (19).

Hence, for n large enough, we have

$$\begin{aligned} C^{-1/(1-\alpha)} \epsilon_n^{1/(1-\alpha)} &\overset{(a)}{\leqslant} \mathbb{V}\left(p_{\bar{X}}^{(n)}, p_X\right) \\ &\overset{(b)}{\leqslant} \sqrt{2 \ln 2} \sqrt{\mathbb{D}\left(p_{\bar{X}}^{(n)} || p_X\right)}, \end{aligned}$$

where (a) holds by Equation (20), (b) holds by Pinsker's inequality. We have thus obtained for n large enough

$$\frac{C^{-2/(1-\alpha)}}{2 \ln 2} \epsilon_n^{2/(1-\alpha)} \leqslant \mathbb{D}\left(p_{\bar{X}}^{(n)} || p_X\right).$$

We conclude by choosing $C \triangleq (2 \ln 2)^{(\alpha-1)/2}$ and $\alpha \triangleq (1 + 2/\beta)^{-1}$. □

Hence, combining Equation (12), Lemma 3, and Lemma 5, we have obtained for $\beta > 0$, for n large enough

$$\mathbb{E}_\Phi[U_e] \leqslant 2 \cdot 2^{-n \epsilon_n^{2+\beta} + \gamma(n)}. \tag{21}$$

D. Simpler upper-bound on $\mathbb{E}_\Phi[P_e]$

Let $(\epsilon_n)_{n \in \mathbb{N}}$ be a positive sequence decreasing and converging to zero. For $n \in \mathbb{N}$, let $R_n \triangleq H(X) + \epsilon_n$. Similar to Section A-C, it can be shown for $\beta > 0$,

$$\mathbb{E}_\Phi[P_e] \leqslant 2^{-n \epsilon_n^{2+\beta} + \gamma(n)}. \tag{22}$$

E. Sufficient condition for lossless compression with nearly uniform output

Let $\beta > 0$. For $n \in \mathbb{N}$, we define $\epsilon_n \triangleq \frac{\lceil n^{1/2+\beta} \rceil}{n}$.

We then have by Equation (22) for n large enough, for $R_n \triangleq H(X) + \epsilon_n$

$$\mathbb{E}_\Phi[P_e] \leqslant 2^{-n \epsilon_n^{2+\beta} + \gamma(n)} \leqslant 2^{-n^{3\beta/2} + \gamma(n)}.$$

We also have by Equation (21), for n large enough, for $R'_n \triangleq d_n/n + H(X) - \epsilon_n$

$$\mathbb{E}_\Phi[U_e] \leqslant 2 \cdot 2^{-n \epsilon_n^{2+\beta} + \gamma(n)} \leqslant 2 \cdot 2^{-n^{3\beta/2} + \gamma(n)}.$$

Hence, by choosing d_n such that $R_n = R'_n$, i.e., such that $d_n = 2n\epsilon_n = 2\lceil n^{1/2+\beta} \rceil$, we have

$$\mathbb{E}_\Phi[P_e + U_e] = \mathbb{E}_\Phi[P_e] + \mathbb{E}_\Phi[U_e] \xrightarrow{n \to \infty} 0,$$

and by Markov Lemma, there exists a choice of encoders/decoders (ϕ_n, ψ_n) for which $P_e + U_e \xrightarrow{n \to \infty} 0$.

REFERENCES

[1] U. Maurer, "Authentication theory and hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1350–1356, 2000.

[2] R. Blahut, *Principles and practice of information theory*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[3] C. Cachin, "An information-theoretic model for steganography," *Information and Computation*, vol. 192, no. 1, pp. 41–56, 2004.

[4] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2706–2722, 2008.

[5] R. Chou and M. Bloch, "Data compression with nearly uniform output," in *Proc. of IEEE Int. Symp. Inf. Theory*, 2013.

[6] T. Han, *Information-Spectrum Methods in Information Theory*. Springer, 2002, vol. 50.

[7] W. Kumagai and M. Hayashi, "A new family of probability distributions and asymptotics of classical and locc conversions," *arXiv preprint arXiv:1306.4166*, 2013.

[8] R. Chou and M. Bloch, "Uniform distributed source coding for the multiple access wiretap channel," in *Proc. IEEE Conf. on Communications and Network Security (CNS)*, 2014.

[9] R. Chou, B. Vellambi, M. Bloch, and J. Kliewer, "Coding schemes for achieving strong secrecy at negligible cost," *arXiv preprint arXiv:1508.07920*, 2015.

[10] G. Basharin, "On a statistical estimate for the entropy of a sequence of independent random variables," *Theory of Probability & Its Applications*, vol. 4, no. 3, pp. 333–336, 1959.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge Univ Pr, 1981.

[12] Y. Oohama and T. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1908–1919, 1994.

[13] Y. Oohama, "Intrinsic randomness problem in the framework of slepian-wolf separate coding system," *IEICE Trans. On Fundamentals Of Electronics Communications And Computer Sciences E Series A*, vol. 90, no. 7, p. 1406, 2007.

[14] M. Yassaee, M. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6760–6786, 2014.

[15] I. Sason and S. Verdú, "f-divergence inequalities," *arXiv preprint arXiv:1508.00335*, 2015.

[16] R. Chou, M. Bloch, and J. Kliewer, "Polar coding for empirical and strong coordination via distribution approximation," in *Proc. of IEEE Int. Symp. Inf. Theory*, 2015.

[17] R. Chou and M. Bloch, "Separation of reliability and secrecy in rate-limited secret-key generation," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4941–4957, 2014.

[18] R. Rockafellar, *Convex Analysis*. Princeton University Press, Princeton, NJ, 2011.