Universal Ensembles for Sample-Wise Lossy Compression

Neri Merhav

The Viterbi Faculty of Electrical & Computer Engineering
Technion—Israel Institute of Technology
Haifa, Israel

ITA 2023, San Diego, CA, February 12–17, 2023
A Very Quick Overview

**Universal lossless coding:**

- Davisson ('73): maximin+minimax universality; mixtures.
- Rissanen ('84): a converse for ‘most’ parameter values.
- Merhav & Feder ('95): parametric class → general class.
- Many: extensions, improvements, relations to prediction, etc.
Universal lossy $d$-semifaithful coding:

- Zhang, Yang & Wei ('97): non-universal redundancy $\geq \frac{\log n}{2n}$; achievable $\leq \frac{\log n}{n}$; universality - larger constant.
- Yu & Speed ('93): weak universality.
- Ornstein & Shields ('90): stat. erg. sources, Hamming distortion.
- Kontoyiannis ('00): a.s. results – CLT, LIL, no-cost universality.
- Kontoyiannis & Zhang ('02): $-\log \Pr\{D\text{-ball}\}$.
In This Work ...

We adopt the semi-deterministic paradigm of Weinberger, Merhav & Feder ('94) for lossy compression:

Redundancy rates relative to the ‘memoryless' empirical RDF

- Random coding using a mixture (Kontoyiannis & Zhang - '02).
- Asympt. accurate evaluation of $\Pr\{D\text{-ball}\}$.
- Universality w.r.t. the distortion measure.
- Converse.

Sequences “with memory”

- Optimal length $= - \log P_{LZ}\{D\text{-ball}\}$
- The main contribution is in the converse.
- Discussion
Notation & Definitions

- **Source sequence**: \( \mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n, |\mathcal{X}| = J. \)
- **Reproduction sequence**: \( \hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n) \in \hat{\mathcal{X}}^n, |\hat{\mathcal{X}}| = K. \)
- **Distortion measure**: \( d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+; d(\mathbf{x}, \hat{\mathbf{x}}) = \sum_i d(x_i, \hat{x}_i). \)
- **Encoder**: \( \phi_n : \mathcal{X}^n \to \mathcal{G}_n \subset \{0, 1\}^n. \)
- **Decoder**: \( \psi_n : \mathcal{G}_n \to \mathcal{C}_n \subseteq \hat{\mathcal{X}}^n. \)
- **\( D \)-semifaithful code**: \( \forall \mathbf{x} \in \mathcal{X}^n, d(\mathbf{x}, \psi_n(\phi_n(\mathbf{x}))) \leq nD. \)
- **Code ensemble**: independent random selection under

\[
W(\hat{\mathbf{x}}) = (K - 1)! \cdot \int_{\mathcal{Q}} d\mathcal{Q} \prod_{i=1}^{n} \mathcal{Q}(\hat{x}_i).
\]

- **\( D \)-sphere**: \( S(\mathbf{x}, D) = \{ \hat{\mathbf{x}} : d(\mathbf{x}, \hat{\mathbf{x}}) \leq nD \}. \)
- **\( \mathcal{T}_n(\mathcal{P}) \)**: \{all \( \mathbf{x} \in \mathcal{X}^n \) with empirical distribution \( \mathcal{P} \}\).


A Key Lemma - Assessing $W[S(x, D)]$

Let $x \in T_n(P)$ and define

$$F(s, Q) \triangleq - \sum_x P(x) \ln \left[ \sum_{\hat{x}} Q(\hat{x}) e^{-s d(x, \hat{x})} \right] - sD.$$ 

Then, it is well known that

$$R_d(D, P) = \sup_{s \geq 0} \min_Q F(s, Q) = \min_Q \sup_{s \geq 0} F(s, Q).$$

Let $(s^*, Q^*)$ be the saddle-point that achieves $R_d(D, P)$ and define

$$V(P, d) = \left| \det \left\{ \operatorname{Hess} F(s^* + j\omega, Q) \right|_{(0, Q^*)} \right|, \quad j = \sqrt{-1}.$$
A Key Lemma - Assessing $W[S(\mathbf{x}, D)]$ (Cont’d)

Suppose that \( \{d(j, k), 1 \leq j \leq J, 1 \leq k \leq K\} \) are commensurable and let \( \Delta \) be their largest common divisor, and define

\[
T_n(P, d) = (K - 1)! \cdot (2\pi)^{K/2-1} \cdot \frac{\Delta \exp\{-s^*[(nD) \mod \Delta]\}}{(1 - e^{-s^*\Delta}) \sqrt{V(P, d)}} ,
\]

If \( \{d(j, k), 1 \leq j \leq J, 1 \leq k \leq K\} \) are incommensurable, take \( \Delta \to 0 \):

\[
T_n(P, d) = \frac{(K - 1)! \cdot (2\pi)^{K/2-1}}{s^* \sqrt{V(P, d)}} .
\]

Lemma:

\[
W[S(\mathbf{x}, D)] = \frac{T_n(P, d)}{n^{K/2}} \cdot \exp\{-nR_d(D, P)\} \cdot [1 - \epsilon_{P,d}(n)].
\]

The exact pre-exponent is essential for an exact characterization of the code-length redundancy in the sequel.
Main Analysis Tool - the Saddlepoint Method

Representing the unit step function $U(t)$ as the inverse Laplace transform of $1/z$:

$$U(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{zt}}{z} \, dz, \quad c > 0,$$

we have:

$$W[S(x, D)] = (K - 1)! \sum_{\{\hat{x} : d(x, \hat{x}) \leq nD\}} \int_{Q} Q(\hat{x}) \, dQ$$

$$= (K - 1)! \sum_{\hat{x} \in \hat{x}^n} U \left( nD - \sum_{i=1}^{n} d(x_i, \hat{x}_i) \right) \int_{Q} Q(\hat{x}) \, dQ$$

$$= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \int_{Q} \frac{e^{-nF(z, Q)}}{z} \cdot dQ \, dz,$$

and we select $c = s^*$ to pass thru all saddle-points.
A Universal Coding Scheme

- Generate $\mathcal{C}_n$ with $A^n$ independent random codewords ($A > K$), $\hat{X}_i \sim W$, $i = 1, 2, \ldots, A^n$.
- Reveal the codebook to both parties.
- Given $x$ and $d$, find $I_d(x) = \min\{i : \hat{X}_i \in \mathcal{S}(x, D)\}$.
- Encode $I_d(x)$ using a Shannon code w.r.t. the distribution $u[i] \propto 1/i$, $i = 1, 2, \ldots, A^n$.
- The decoder decodes $I_d(x)$ and outputs the $I_d(x)$-th reproduction vector from $\mathcal{C}_n$.

Note that the codebook is the same for every (bounded) $d$ – distortion-universality.
Coding Theorem

∀ε > 0, ∃ a sequence of codebooks, \{C_n\}_{n \geq 1}, and \{\psi_n\}, such that
∀d ∈ \bigcup_{k \geq 1} \left\{0, \frac{d_{\text{max}}}{k}, \ldots, \frac{d_{\text{max}}}{k}\right\}^{JK}, ∃ \{\phi_n\}, such that ∀P ∈ \bigcup_{k \geq 1} P_k,
n ∈ \mathcal{N} \triangleq \{\hat{n} : d ∈ D_{\hat{n}}, P ∈ P_{\hat{n}}\} and \textbf{x} ∈ \mathcal{T}_n(P):

(a)

\[ L_d(\textbf{x}) \leq nR_d(D, P) + \left(\frac{K}{2} + 2 + \epsilon\right) \cdot \ln n + \beta_{P,d}(n) + \log(\log A + 1) + O(J^n e^{-n^{1+\epsilon}}). \]

(b) The code is d-semifaithful: \(d(\textbf{x}, \psi_n(\phi_n(\textbf{x}))) \leq nD.\)

\(C_n\) and \(\psi_n\) do not depend on \(P\) and \(d\), but \(\phi_n\) does.

Mahmood & Wagner (’22): 3 schemes with log \(n\)-coefficients: \(2JK + J + 3\), \(J(K + 1)\) and \(J^2K^2 + J - 2.\)
Let $P$ and $d$ be given. $\forall \epsilon > 0$ and sufficiently large $n$, $\forall$ codebook that covers $\mathcal{T}_n(P)$ and every one-to-one variable-length code applied to that codebook, the following lower bound applies to a fraction of at least $(1 - 2n^{-\epsilon})$ of the codewords that cover $\mathcal{T}_n(P)$:

$$L_d(\hat{x}) \geq nR_d(D, P) + \left(\frac{1}{2} - \epsilon\right) \log n + c - c' \log(\log n),$$

where $c$ and $c'$ are constants that depend on $P$. 
Converse Theorem (Cont’d)

The proof is based on a sphere-covering argument:

$$\log |\mathcal{T}_n(P)| \geq nH(P) - \frac{J - 1}{2} \log n + c(P)$$

and

$$\ln \left| \mathcal{T}_n(P) \cap \{x : d(x, \hat{x}) \leq nD\} \right|$$

$$\leq \max_{\{P_{\hat{X}|X} : \mathbf{E}\{d(X, \hat{X})\} \leq D\}} H(X | \hat{X}) - \frac{J}{2} \log n + c' \log(\log n), \quad P_X = P$$

and so,

$$|\mathcal{C}_n| \geq \exp_2 \left\{ nR_d(D, P) + \frac{\log n}{2} + \ldots \right\}.$$ 

Most codewords cannot have code-length much less than $\log |\mathcal{C}_n|$.
Beyond the Memoryless Structure

Consider the universal distribution

\[ U(\hat{x}) = \frac{2^{-LZ(\hat{x})}}{\sum \hat{x}' 2^{-LZ(\hat{x}')}} \]

and let

\[ U[S(x, D)] = \sum_{\hat{x} \in S(x, D)} U(\hat{x}). \]

**Converse theorem:** Let \( \ell \) divide \( n \) and let \( \mathcal{T}_n(\hat{P}^\ell) \) be any \( \ell \)-th order type of source sequences. Let \( d \) be a distortion function that depends on \((x, \hat{x})\) only via \( \hat{P}^1_x \hat{x} \hat{x} \). Then, \( \forall d \)-semifaithful variable-length block code, and \( \forall \epsilon > 0 \), the following lower bound applies to a fraction of at least \((1 - 2n^{-\epsilon})\) of the codewords, \( \{\phi_n(x), \ x \in \mathcal{T}_n(\hat{P}^\ell)\} \):

\[ L(\phi_n(x)) \geq -\log(U[S(x, D)]) - n\Delta_n(\ell) - \epsilon \log n, \]

where \( \lim_{n \to \infty} \Delta_n(\ell) = 1/\ell \).
Main Ideas of the Proof

Relating sphere-covering and $U[S(x, D)]$ in a few steps.
First, observe that

$$N(D) \triangleq \sum_{\mathbf{x}, \hat{\mathbf{x}}} \mathbb{I}\{\mathbf{x} \in \mathcal{T}_n(P^\ell), \hat{\mathbf{x}} \in \mathcal{T}_n(Q^\ell), d(\mathbf{x}, \hat{\mathbf{x}}) \leq nD\}$$

$$\begin{align*}
N(D) &= \left|\mathcal{T}_n(P^\ell)\right| \cdot \left|\mathcal{T}_n(Q^\ell) \cap S(x, D)\right| \\
&= \left|\mathcal{T}_n(Q^\ell)\right| \cdot \left|\mathcal{T}_n(P^\ell) \cap \hat{S}(\hat{x}, D)\right|, \quad \hat{S}(\hat{x}, D) \triangleq \{\mathbf{x}: d(\mathbf{x}, \hat{\mathbf{x}}) \leq nD\}
\end{align*}$$

and so,

$$\frac{|T_n(P^\ell)|}{|T_n(P^\ell) \cap \hat{S}(\hat{x}, D)|} = \frac{|T_n(Q^\ell)|}{|T_n(Q^\ell) \cap S(x, D)|}$$

LHS = sphere-covering ratio;
RHS = $1/U_Q[S(x, D)] \geq 1/U[S(x, D)] \to$ use $U$ for random coding!
Direct Theorem

Let \( d : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to \mathbb{R}^+ \) be an arbitrary distortion function. Then, \( \forall \epsilon > 0, \exists \) sequence of \( d \)-semifaithful, variable-length block codes of block length \( n \), such that \( \forall x \in \mathcal{X}^n \), the code length for \( x \) is upper bounded by

\[
L(x) \leq -\log(U[S(x, D)]) + (2 + \epsilon) \log n + c + \delta_n,
\]

where \( c > 0 \) is a constant and \( \delta_n = O(nJ^n e^{-n^{1+\epsilon}}) \).

The proof is very similar to that of the previous direct theorem.
Related to the Kontoyiannis-Zhang converse:
\[ \forall x, \mathcal{C}_n \exists Q : L(x) \geq -\log Q[S(x, D)]. \]

\[ -\log(U[S(x, D)]) \sim \min_L \{ L - \log |\{ \hat{x} : LZ(\hat{x}) = L \} \cap S(x, D)| \}, \text{ analogous to } \min_{P_{\hat{X}}} [H(\hat{X}) - \max\{H(\hat{X}|X) : Ed(X, \hat{X}) \leq D\}]. \]

Easy to see that the proposed scheme is better than \( \min_{\hat{x} \in S(x, D)} LZ(\hat{x}) \).

Complexity of both schemes depend on \( D \).

Universality w.r.t. a wide (continuous, parametric) class of distortion measures can also be proved. Here, the class distortion measures is quite arbitrary.