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.
- Weinberger, Merhav & Feder ('94): 'semi-deterministic' analogue.
- Merhav & Feder ('95): parametric class $\rightarrow$ general class.
- Many: extensions, improvements, relations to prediction, etc.

A Very Quick Overview (Cont'd)

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): $-\log \text{Pr}\{D\text{-ball}\}$.
- Mahmood & Wagner ('22): minimax distortion-universality.

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$\{D$-ball$\}$.
- Universality w.r.t. the distortion measure.
- Converse.

Sequences "with memory"

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

Notation & Definitions

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

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

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

A Key Lemma - Assessing $W[\mathcal{S}(\boldsymbol{x}, D)]$

Let $\boldsymbol{x} \in \mathcal{T}_n(P)$ and define

$$F(s, Q) \stackrel{\triangle}{=} -\sum_{x} P(x) \ln \left[\sum_{\hat{x}} Q(\hat{x}) e^{-sd(x,\hat{x})} \right] - sD.$$

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

$$V(P, d) = \left| \det \left\{ \text{Hess} F(s^* + j\omega, Q) \Big|_{(0, Q^*)} \right\} \right|, \qquad j = \sqrt{-1}.$$

A Key Lemma - Assessing $W[\mathcal{S}(\boldsymbol{x}, D)]$ (Cont'd)

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

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

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:

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

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 $1/z$:

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

we have:

$$\begin{aligned}
W[\mathcal{S}(\boldsymbol{x}, D)] &= (K-1)! \sum_{\{\hat{\boldsymbol{x}}:\ d(\boldsymbol{x},\hat{\boldsymbol{x}}) \leq nD\}} \int_{\mathcal{Q}} Q(\hat{\boldsymbol{x}}) \mathrm{d}Q \\
&= (K-1)! \sum_{\hat{\boldsymbol{x}} \in \hat{\mathcal{X}}^n} U\left(nD - \sum_{i=1}^{n} d(x_i, \hat{x}_i)\right) \int_{\mathcal{Q}} Q(\hat{\boldsymbol{x}}) \mathrm{d}Q \\
&= \frac{(K-1)!}{2\pi j} \int_{c-j\infty}^{c+j\infty} \int_{\mathcal{Q}} \frac{e^{-nF(z,Q)}}{z} \cdot \mathrm{d}Q \mathrm{d}z,
\end{aligned}$$

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

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

Coding Theorem

$\forall \epsilon > 0$, $\exists$ a sequence of codebooks, $\{\mathcal{C}_n\}_{n\geq 1}$, and $\{\psi_n\}$, such that $\forall d \in \bigcup_{k\geq 1}\{0, d_{\max}/k, \ldots, d_{\max}\}^{JK}$, $\exists$ $\{\phi_n\}$, such that $\forall P \in \bigcup_{k\geq 1} \mathcal{P}_k$, $n \in \mathcal{N} \stackrel{\triangle}{=} \{\hat{n} :\ d \in \mathcal{D}_{\hat{n}},\ P \in \mathcal{P}_{\hat{n}}\}$ and $\boldsymbol{x} \in \mathcal{T}_n(P)$:

(a)

$$\begin{aligned} L_d(\boldsymbol{x}) \quad &\leq \quad nR_d(D,P) + \left(\frac{K}{2} + 2 + \epsilon\right) \cdot \ln n + \\ & \qquad \beta_{P,d}(n) + \log(\log A + 1) + O(J^n e^{-n^{1+\epsilon}}). \end{aligned}$$

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

$\mathcal{C}_n$ and ψ_n do not depend on P and d, but ϕ_n does.

Mahmood & Wagner ('22): 3 schemes with $\log n$-coefficients: $2JK + J + 3$, $J(K+1)$ and $J^2K^2 + J - 2$.

Converse Theorem

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 $\mathcal{T}_n(P)$:

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

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

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

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

and so,

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

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

Beyond the Memoryless Structure

Consider the universal distribution

$$U(\hat{\boldsymbol{x}}) = \frac{2^{-LZ(\hat{\boldsymbol{x}})}}{\sum_{\hat{\boldsymbol{x}}'} 2^{-LZ(\hat{\boldsymbol{x}}')}}$$

and let

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

Converse theorem: Let ℓ divide n and let $\mathcal{T}_n(\hat{P}^\ell)$ be any ℓ-th order type of source sequences. Let d be a distortion function that depends on $(\boldsymbol{x}, \hat{\boldsymbol{x}})$ only via $\hat{P}^1_{\boldsymbol{x}\hat{\boldsymbol{x}}}$. 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, $\{\phi_n(\boldsymbol{x}),\ \boldsymbol{x} \in \mathcal{T}_n(\hat{P}^\ell)\}$:

$$L(\phi_n(\boldsymbol{x})) \geq -\log(U[\mathcal{S}(\boldsymbol{x}, D)]) - n\Delta_n(\ell) - \epsilon \log n,$$

where $\lim_{n\to\infty} \Delta_n(\ell) = 1/\ell$.

Main Ideas of the Proof

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

$$
\begin{aligned}
N(D) &\stackrel{\triangle}{=} \sum_{\boldsymbol{x},\hat{\boldsymbol{x}}} \mathcal{I}\{\boldsymbol{x} \in \mathcal{T}_n(P^\ell),\ \hat{\boldsymbol{x}} \in \mathcal{T}_n(Q^\ell),\ d(\boldsymbol{x},\hat{\boldsymbol{x}}) \le nD\} \\
&= |\mathcal{T}_n(P^\ell)| \cdot \left| T_n(Q^\ell) \bigcap \mathcal{S}(\boldsymbol{x}, D) \right| \\
&= |\mathcal{T}_n(Q^\ell)| \cdot \left| T_n(P^\ell) \bigcap \hat{\mathcal{S}}(\hat{\boldsymbol{x}}, D) \right|, \qquad \hat{\mathcal{S}}(\hat{\boldsymbol{x}}, D) \stackrel{\triangle}{=} \{\boldsymbol{x}:\ d(\boldsymbol{x},\hat{\boldsymbol{x}}) \le nD\}
\end{aligned}
$$

and so,

$$
\frac{|T_n(P^\ell)|}{\left| T_n(P^\ell) \bigcap \hat{\mathcal{S}}(\hat{\boldsymbol{x}}, D) \right|} = \frac{|T_n(Q^\ell)|}{\left| T_n(Q^\ell) \bigcap \mathcal{S}(\boldsymbol{x}, D) \right|}
$$

LHS $=$ sphere-covering ratio;
RHS $= 1/U_Q[\mathcal{S}(\boldsymbol{x}, D)] \stackrel{\cdot}{\ge} 1/U[\mathcal{S}(\boldsymbol{x}, D)]$ $\rightarrow$ 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 $\forall \boldsymbol{x} \in \mathcal{X}^n$, the code length for $\boldsymbol{x}$ is upper bounded by

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

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.

Discussion

♠ Related to the Kontoyiannis-Zhang converse:
$\forall \boldsymbol{x}, \mathcal{C}_n \exists Q : L(\boldsymbol{x}) \geq -\log Q[\mathcal{S}(\boldsymbol{x}, D)]$.

♠ $-\log(U[\mathcal{S}(\boldsymbol{x}, D)]) \sim \min_L \{L - \log |\{\hat{\boldsymbol{x}} : \ LZ(\hat{\boldsymbol{x}}) = L\} \cap \mathcal{S}(\boldsymbol{x}, D)|\}$, analogous to $\min_{P_{\hat{X}}} [H(\hat{X}) - \max\{H(\hat{X}|X) : \ \mathbf{E} d(X, \hat{X}) \leq D\}]$.

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