The Redundancy Gains of Almost Lossless Universal Source Coding over Envelope Families

Jorge F. Silva[†] and Pablo Piantanida[††]

[†]*Universidad de Chile*
[††]*CentraleSupélec-CNRS-Université Paris Sud*

ISIT 2017 Aachen July 2017

Background: Universal Source Coding

Let $X_1, .., X_n, ..$ be a i.i.d. source with values in $\mathcal{X}$ (countable alphabet) equipped with a probability measure $\mu \in \mathcal{P}(\mathcal{X})$.

Lossless Source Coding

For $X_1^n = (X_1, .., X_n) \sim \mu^n$, the problem is to find

$$\underbrace{f_n}_{prefix\ \ free\ \ mapping} \quad \mathcal{X}^n \longrightarrow \underbrace{\cup_{k \geq 1} \{0,1\}^k}_{\{0,1\}^*}$$

that mimimizes the rate

$$r(f_n, \mu^n) = \frac{1}{n} \mathbb{E}_{X^n \sim \mu^n} (\underbrace{\mathcal{L}(f_n(X^n))}_{length\ \ of\ \ f_n(X^n)}).$$

Background: Universal Source Coding

Shannon Entropy

$$\lim_{n\to\infty} \min_{f_n \in \mathcal{F}_n} r(f_n, \mu^n) = \mathcal{H}(\mu)$$

Background: Universal Source Coding

Universality

μ is unknown, but $\mu \in \Lambda \subset \mathcal{P}(\mathcal{X})$.

A Universal Source Code

$\{f_n : n \geq 1\}$ is *strongly minimax universal* for Λ if

$$\lim_{n \to \infty} \underbrace{\sup_{\mu \in \Lambda} \left[r(f_n, \mu^n) - \mathcal{H}(\mu) \right]}_{worse\ \ case\ \ redundancy} = 0.$$

Background: Universal Source Coding

Definition

Λ is strongly minimax universal if

$$\lim_{n\to\infty} \underbrace{\min_{f_n \in \mathcal{F}_n} \sup_{\mu\in\Lambda} \left[r(f_n, \mu^n) - \mathcal{H}(\mu) \right]}_{\textit{minimax redundancy per sample}} = 0.$$

Background: Universal Source Coding

Information Radius of Λ^n

$$\underbrace{\min_{f_n \in \mathcal{F}_n} \sup_{\mu \in \Lambda} \left[r(f_n, \mu^n) - \mathcal{H}(\mu) \right]}_{mini-max \;\; redundancy} \sim \frac{1}{n} \cdot \underbrace{\min_{v^n \in \mathcal{P}(\mathcal{X}^n)} \sup_{\mu^n \in \Lambda^n} \mathcal{D}(\mu^n | v^n)}_{R^+(\Lambda^n) \equiv}$$

where $\mathcal{P}(\mathcal{X}^n)$ is the set of probability measures in $\mathcal{X}^n$ and $\Lambda^n = \{\mu^n : \mu \in \Lambda\}$.

Background: Universal Source Coding

Information Radius of Λ^n

$$\underbrace{\min_{f_n \in \mathcal{F}_n} \sup_{\mu \in \Lambda} \left[r(f_n, \mu^n) - \mathcal{H}(\mu) \right]}_{mini-max\ \ redundancy} \sim \frac{1}{n} \cdot \underbrace{\min_{v^n \in \mathcal{P}(\mathcal{X}^n)} \sup_{\mu^n \in \Lambda^n} \mathcal{D}(\mu^n | v^n)}_{R^+(\Lambda^n) \equiv}$$

where $\mathcal{P}(\mathcal{X}^n)$ is the set of probability measures in $\mathcal{X}^n$ and $\Lambda^n = \{\mu^n : \mu \in \Lambda\}$.

The class of memoryless sources indexed by Λ is strongly minimax universal, if and only if,

$$R^+(\Lambda^n) \text{ is } o(n).$$

Background: Finite Alphabet

Finite Alphabet Result

If $|\mathcal{X}| = k$ it is known that[a]:

$$\frac{1}{2}(k-1)\log n - K_1 \leq R^+(\mathcal{P}(\mathcal{X}^n)) \leq \frac{1}{2}(k-1)\log n + K_2,$$

for some $K_1, K_2 > 0$. Then,

$$R^+(\mathcal{P}(\mathcal{X}^n)) \text{ is } O(\log n).$$

[a] Csiszar and Shields, *Information theory and Statistics: A Tutorial*, 2004.

Background: Countable Alphabet

Infeasibility Result

No weak universal source coding scheme is available for the family of memoryless processes[a].

[a] J.C. Kieffer, *A unified approach to weak universal source coding*, IEEE Trans. on IT., 1978.

Gyorfi et al.[a] prove that for any $f_n : \mathcal{X}^n \to \{0.1\}^*$ there exists $\mu \in \mathcal{P}_{\mathcal{H}}(\mathcal{X})$ such that $r(f_n, \mu^n) - \mathcal{H}(\mu) = \infty$

$$\Rightarrow R^+(\mathcal{P}_{\mathcal{H}}(\mathcal{X})^n) = \infty,$$

where $\mathcal{P}_{\mathcal{H}}(\mathcal{X}) \equiv \{\mu : \mathcal{H}(\mu) < \infty\} \subset \mathcal{P}(\mathcal{X})$.

[a] Gyorfi, Pali, and van der Meulen, *There is no universal source code for an infinite source alphabet*, IEEE Trans. on IT.,1994.

Background: Countable Alphabet

Infeasibility Result

No weak universal source coding scheme is available for the family of memoryless processes[a].

[a] J.C. Kieffer, *A unified approach to weak universal source coding*, IEEE Trans. on IT., 1978.

Gyorfi et al.[a] prove that for any $f_n : \mathcal{X}^n \to \{0.1\}^*$ there exists $\mu \in \mathcal{P}_{\mathcal{H}}(\mathcal{X})$ such that $r(f_n, \mu^n) - \mathcal{H}(\mu) = \infty$

$$\Rightarrow R^+(\mathcal{P}_{\mathcal{H}}(\mathcal{X})^n) = \infty,$$

where $\mathcal{P}_{\mathcal{H}}(\mathcal{X}) \equiv \{\mu : \mathcal{H}(\mu) < \infty\} \subset \mathcal{P}(\mathcal{X})$.

[a] Gyorfi, Pali, and van der Meulen, *There is no universal source code for an infinite source alphabet*, IEEE Trans. on IT.,1994.

Background: The Envelope Family

$$\Lambda_f \equiv \left\{ \mu \in \mathcal{P}(\mathcal{X}) : \mu(x) \leq \underbrace{f(x)}_{\textit{envelop function}} \quad \forall x \right\}$$

Theorem (Boucheron et al. 2009)

Let f be non-negative mapping from $\mathcal{X}$ to $[0,1]$[a]:

1. if $f \in \ell_1(\mathcal{X})$ then $\mathcal{R}^+(\Lambda_f^n)$ is $o(n)$.
2. otherwise, $\mathcal{R}^+(\Lambda_f^n) = \infty$ for all $n \geq 1$.

[a] Boucheron, Garivier and Gassiat, *Coding on countably infinite alphabets*, IEEE Trans. on Inf. Th.., 2009

Background: Weak Source Coding

Question

Can we relax the lossless criterion to make universal coding "feasible" for $\mathcal{P}_{\mathcal{H}}(\mathcal{X})$?[a]

[a] T. S. Han, *Weak variable-lenght source coding*, IEEE Tran. on IT.., 2000.

Idea: From lossless to "almost" lossless:

- we relax the lossless block-wise condition.
- we consider a zero asymptotic distortion (per sample), using the *Hamming distance* as the distortion.

Background: Weak Source Coding

Question

Can we relax the lossless criterion to make universal coding "feasible" for $\mathcal{P}_{\mathcal{H}}(\mathcal{X})$?[a]

[a] T. S. Han, *Weak variable-lenght source coding*, IEEE Tran. on IT.., 2000.

Idea: From lossless to "almost" lossless:

- we relax the lossless block-wise condition.
- we consider a zero asymptotic distortion (per sample), using the *Hamming distance* as the distortion.

Background: Weak Source Coding

- Let us consider a encoder-decoder pair (f_n, g_n), with $f_n : \mathcal{X}^n \to \{0,1\}^\star$ and $g_n : \{0,1\}^\star \to \mathcal{X}^n$ and

$$\{x^n : g_n(f_n(x^n)) \neq x^n\} \neq \emptyset.$$

- For an information source $X_1^n \sim \mu^n$, the distortion is given by:

$$d(f_n, g_n, \mu) \equiv \mathbb{E}_{X^n \sim \mu^n} \left\{ \rho^{(n)}(X^n, g_n(f_n(X^n))) \right\}$$

where $\rho^{(n)}(x^n, y^n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}}$.

- The rate is:

$$r(f_n, \mu^n) \equiv \frac{1}{n} \mathbb{E}_{X^n \sim \mu^n} \left\{ \mathcal{L}(f_n(X^n)) \right\}.$$

Background: Weak Source Coding

- Let us consider a encoder-decoder pair (f_n, g_n), with $f_n : \mathcal{X}^n \to \{0,1\}^\star$ and $g_n : \{0,1\}^\star \to \mathcal{X}^n$ and

$$\{x^n : g_n(f_n(x^n)) \neq x^n\} \neq \emptyset.$$

- For an information source $X_1^n \sim \mu^n$, the distortion is given by:

$$d(f_n, g_n, \mu) \equiv \mathbb{E}_{X^n \sim \mu^n} \left\{ \rho^{(n)}(X^n, g_n(f_n(X^n))) \right\}$$

where $\rho^{(n)}(x^n, y^n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}}$.

- The rate is:

$$r(f_n, \mu^n) \equiv \frac{1}{n} \mathbb{E}_{X^n \sim \mu^n} \{\mathcal{L}(f_n(X^n))\}.$$

Background: Weak Source Coding

- Let us consider a encoder-decoder pair (f_n, g_n), with $f_n : \mathcal{X}^n \to \{0,1\}^\star$ and $g_n : \{0,1\}^\star \to \mathcal{X}^n$ and

$$\{x^n : g_n(f_n(x^n)) \neq x^n\} \neq \emptyset.$$

- For an information source $X_1^n \sim \mu^n$, the distortion is given by:

$$d(f_n, g_n, \mu) \equiv \mathbb{E}_{X^n \sim \mu^n} \left\{ \rho^{(n)}(X^n, g_n(f_n(X^n))) \right\}$$

where $\rho^{(n)}(x^n, y^n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}}$.

- The rate is:

$$r(f_n, \mu^n) \equiv \frac{1}{n} \mathbb{E}_{X^n \sim \mu^n} \left\{ \mathcal{L}(f_n(X^n)) \right\}.$$

Background: Weak Source Coding

Definition 1

$\Lambda \subset \mathcal{P}(\mathcal{X})$ admits an almost lossless universal source coding scheme, if there is $\{(f_n, g_n) : n \geq 1\}$ such that[a]

$$\underbrace{\sup_{\mu \in \Lambda} \lim_{n \to \infty} d(f_n, g_n, \mu) = 0}_{point-wise\ \ zero\ \ distortion\ \ condition} \quad \text{and} \tag{1}$$

$$\lim_{n \to \infty} \underbrace{\sup_{\mu \in \Lambda} \left[r(f_n, \mu^n) - H(\mu) \right]}_{worse\ \ case\ \ redundancy} = 0. \tag{2}$$

---

[a] $H(\mu)$ is the minimum achievable rate for the almost lossless source coding, Silva and Piantanida, Th.1, ISIT2016.

Background: Weak Source Coding

Theorem (Silva and Piantanida, 2016)

$\mathcal{P}_{\mathcal{H}}(\mathcal{X})$ admits an almost lossless universal source coding scheme[a].

[a] Silva and Piantanida, Th.2, ISIT2016.

Contribution of this Work

- Revisit the problem of weak USC adopting a stringent condition on the distortion.
- Explore the rates of convergence for the worst-case distortion and redundancy.

We will focus on Λ_f.

Contribution of this Work

- Revisit the problem of weak USC adopting a stringent condition on the distortion.
- Explore the rates of convergence for the worst-case distortion and redundancy.

We will focus on Λ_f.

Outline

1. Uniform Convergence of the Distortion

2. Redundancy Gain Analysis for Summable Envelope Families

A stronger notion of week universality is explored....

Definition 2

$\Lambda \subset \mathcal{P}(\mathcal{X})$ admits a strong almost lossless source coding scheme, if there is $\{(f_n, g_n) : n \geq 1\}$ such that

$$\lim_{n \to \infty} \underbrace{\sup_{\mu \in \Lambda} d(f_n, g_n, \mu)}_{worse-case\ \ distortion} = 0 \text{ and } \quad (3)$$

$$\lim_{n \to \infty} \sup_{\mu \in \Lambda} \left[r(f_n, \mu^n) - \mathcal{H}(\mu) \right] = 0. \quad (4)$$

Result for Envelope Families

Theorem 1

There exists a strong almost lossless coding scheme for Λ_f if, and only if, $f \in \ell_1(\mathcal{X})$.

Remarks...

1. $f \in \ell_1(\mathcal{X})$ is the same necessary and sufficient condition for the lossless USC (Boucheron et al., 2009).
2. ... then if $f \in \ell_1(\mathcal{X})$, the achievability part can be obtained with a lossless scheme, i.e., $\sup_{\mu \in \Lambda} d(f_n, g_n, \mu) = 0$ for all $n \geq 1$.

Question

Can it be gains in redundancy by using a non-zero distortion when $f \in \ell_1(\mathcal{X})$?

Remarks...

1. $f \in \ell_1(\mathcal{X})$ is the same necessary and sufficient condition for the lossless USC (Boucheron et al., 2009).
2. ... then if $f \in \ell_1(\mathcal{X})$, the achievability part can be obtained with a lossless scheme, i.e., $\sup_{\mu \in \Lambda} d(f_n, g_n, \mu) = 0$ for all $n \geq 1$.

Question

Can it be gains in redundancy by using a non-zero distortion when $f \in \ell_1(\mathcal{X})$?

Redundancy Gain Analysis for Summable Envelope Families

We say that an almost lossless scheme $\{(f_n, g_n) : n \geq 1\}$ for Λ, in the sense that

$$\lim_{n \to \infty} \sup_{\mu \in \Lambda} d(f_n, g_n, \mu) = 0,$$

offers a gain in minimax redundancy if:

$$\lim_{n \to \infty} \frac{\sup_{\mu \in \Lambda} \left[r(f_n, \mu^n) - \mathcal{H}(\mu) \right]}{\frac{1}{n} R^+(\Lambda^n)} = 0.$$

Constraining the Source Alphabet

Let us define the finite set $\Gamma_k \equiv \{1, \ldots, k\}$.

A two-stage lossy code of length n and size k_n is given by:

1. First stage: a lossy mapping (ϕ_n, ψ_n) of size k_n, where
   - $\phi_n : \mathcal{X} \to \Gamma_{k_n}$ and
   - $\psi_n : \Gamma_{k_n} \to \mathcal{X}$.
2. Second stage: a fixed to variable length prefix-free pair of lossless coder-decoder $(\mathcal{C}_n, \mathcal{D}_n)$, where:
   - $\mathcal{C}_n : \Gamma^n_{k_n} \to \{0,1\}^\star$ and
   - $\mathcal{D}_n : \{0,1\}^\star \to \Gamma^n_{k_n}$.

Given a source $\mathbf{X} = \{X_i\}_{i=1}^{\infty}$ and a (n, k_n) lossy code $(\phi_n, \psi_n, \mathcal{C}_n, \mathcal{D}_n)$ operates as:

$$Y^n = \underbrace{(\phi_n(X_1), \cdots, \phi_n(X_n))}_{\text{Lossy description } \in \Gamma^n_{k_n} \text{ (letter by letter)}} \longrightarrow \underbrace{\mathcal{C}_n(Y^n)}_{\in \{0,1\}^\star}$$

Constraining the Source Alphabet

Let us define the finite set $\Gamma_k \equiv \{1, \ldots, k\}$.

A two-stage lossy code of length n and size k_n is given by:

1. First stage: a lossy mapping (ϕ_n, ψ_n) of size k_n, where
   - $\phi_n : \mathcal{X} \to \Gamma_{k_n}$ and
   - $\psi_n : \Gamma_{k_n} \to \mathcal{X}$.
2. Second stage: a fixed to variable length prefix-free pair of lossless coder-decoder $(\mathcal{C}_n, \mathcal{D}_n)$, where:
   - $\mathcal{C}_n : \Gamma_{k_n}^n \to \{0,1\}^\star$ and
   - $\mathcal{D}_n : \{0,1\}^\star \to \Gamma_{k_n}^n$.

Given a source $\mathbf{X} = \{X_i\}_{i=1}^\infty$ and a (n, k_n) lossy code $(\phi_n, \psi_n, \mathcal{C}_n, \mathcal{D}_n)$ operates as:

$$Y^n = \underbrace{(\phi_n(X_1), \cdots, \phi_n(X_n))}_{Lossy\ description\ \in \Gamma_{k_n}^n\ (letter\ by\ letter)} \longrightarrow \underbrace{\mathcal{C}_n(Y^n)}_{\in\{0,1\}^\star}$$

Constraining the Source Alphabet

Let us define the finite set $\Gamma_k \equiv \{1, \ldots, k\}$.

A two-stage lossy code of length n and size k_n is given by:

1. First stage: a lossy mapping (ϕ_n, ψ_n) of size k_n, where
   - $\phi_n : \mathcal{X} \to \Gamma_{k_n}$ and
   - $\psi_n : \Gamma_{k_n} \to \mathcal{X}$.
2. Second stage: a fixed to variable length prefix-free pair of lossless coder-decoder $(\mathcal{C}_n, \mathcal{D}_n)$, where:
   - $\mathcal{C}_n : \Gamma_{k_n}^n \to \{0,1\}^\star$ and
   - $\mathcal{D}_n : \{0,1\}^\star \to \Gamma_{k_n}^n$.

Given a source $\mathbf{X} = \{X_i\}_{i=1}^\infty$ and a (n, k_n) lossy code $(\phi_n, \psi_n, \mathcal{C}_n, \mathcal{D}_n)$ operates as:

$$Y^n = \underbrace{(\phi_n(X_1), \cdots, \phi_n(X_n))}_{\text{Lossy description } \in \Gamma_{k_n}^n \text{ (letter by letter)}} \longrightarrow \underbrace{\mathcal{C}_n(Y^n)}_{\in\{0,1\}^\star}$$

The Induced Partition of the First Stage

Associated to the first stage (ϕ_n, ψ_n)

we have an induced partition of $\mathcal{X}$

$$\underbrace{\pi_n}_{\textit{key design object!}} \equiv \left\{ \mathcal{A}_{n,i} \equiv \phi_n^{-1}(\{i\}) : i \in \Gamma_{k_n} \right\}$$

and a set of prototypes

$$\mathcal{Y}_n \equiv \{\psi_n(i) : i \in \Gamma_{k_n}\}.$$

Approximation Quality of $(\pi_n, \mathcal{Y}_n)$

Distortion of $(\phi_n, \psi_n, \mathcal{C}_n, \mathcal{D}_n)$ for $X_1^n \sim \mu^n$

Assuming that $\psi_n(i) \in \mathcal{A}_{n,i}$ then:

$$d(\underbrace{\phi_n, \psi_n}_{\textit{first stage}}, \mu) = \mathbb{E}_{X^n \sim \mu^n}\left\{\rho^{(n)}(X^n, \Psi_n(\Phi_n(X^n)))\right\}$$

$$= \mathbb{P}\big(X \neq \psi_n(\phi_n(X))\big) = 1 - \mu(\underbrace{\mathcal{Y}_n}_{\textit{prototypes}}).$$

Complexity of π_n

Redundancy of $(\phi_n, \psi_n, \mathcal{C}_n, \mathcal{D}_n)$

The redundancy over $\Lambda \subset \mathcal{P}(\mathcal{X})$ is:

$$R(\phi_n, \mathcal{C}_n, \Lambda^n) \equiv \sup_{\mu \in \Lambda} \left(r(\phi_n, \mathcal{C}_n, \mu^n) - H(\mu) \right).$$

It is more tractable to use

$$\bar{R}(\phi_n, \mathcal{C}_n, \Lambda^n) \equiv \sup_{\mu \in \Lambda} \left[r(\phi_n, \mathcal{C}_n, \mu^n) - \mathcal{H}_{\sigma(\pi_n)}(\mu) \right] \geq R(\phi_n, \mathcal{C}_n, \Lambda^n).$$

Complexity of π_n

Redundancy of $(\phi_n, \psi_n, \mathcal{C}_n, \mathcal{D}_n)$

The redundancy over $\Lambda \subset \mathcal{P}(\mathcal{X})$ is:

$$R(\phi_n, \mathcal{C}_n, \Lambda^n) \equiv \sup_{\mu \in \Lambda} \left(r(\phi_n, \mathcal{C}_n, \mu^n) - H(\mu) \right).$$

It is more tractable to use

$$\bar{R}(\phi_n, \mathcal{C}_n, \Lambda^n) \equiv \sup_{\mu \in \Lambda} \left[r(\phi_n, \mathcal{C}_n, \mu^n) - \mathcal{H}_{\sigma(\pi_n)}(\mu) \right] \geq R(\phi_n, \mathcal{C}_n, \Lambda^n).$$

Complexity of π_n

Fixing the first stage (i.e., π_n), the minimax code is

$$\mathcal{C}_n^\star \equiv \arg \min_{\mathcal{C}^n:\Gamma^n_{k_n}\to\{0,1\}^\star} \bar{R}(\phi_n, \mathcal{C}_n, \Lambda^n), \text{ for all } n \geq 1.$$

in this Lossy Setting

$$\min_{\mathcal{C}^n:\Gamma^n_{k_n}\to\{0,1\}^\star} \sup_{\mu\in\Lambda} \left[r(\phi_n, \mathcal{C}_n, \mu^n) - \mathcal{H}_{\sigma(\pi_n)}(\mu) \right]$$
$$\sim \frac{1}{n} \times \underbrace{\min_{v^n\in\mathcal{P}(\mathcal{X}^n)} \sup_{\mu^n\in\Lambda^n} \mathcal{D}_{\sigma(\pi_n\times..\times\pi_n)}(\mu^n|v^n)}_{R^+(\Lambda^n,\sigma(\pi_n))\equiv}$$

where $\mathcal{D}_{\sigma(\pi)}(\mu|v) \equiv \sum_{\mathcal{A}\in\pi} \mu(\mathcal{A}) \log_2 \frac{\mu(\mathcal{A})}{v(\mathcal{A})}$.

... at the end

distortion

$$\sup_{\mu \in \Lambda} d(\phi_n, \psi_n, \mu) = \sup_{\mu \in \Lambda} (1 - \mu(\mathcal{Y}_n))$$

overhead

$$\sup_{\mu \in \Lambda} \left[r(\phi_n, \mathcal{C}_n^\star, \mu^n) - \mathcal{H}_{\sigma(\pi_n)}(\mu) \right] \sim \frac{1}{n} R^+(\Lambda^n, \sigma(\pi_n))$$

The Tail-based partition

$$\tilde{\pi}_{k_n} = \left\{ \{1\}, \{2\}, ..., \{k_n - 1\}, \Gamma^c_{k_n - 1} \right\}$$

$$\tilde{\mathcal{Y}}_n = \{1, .., k_n\}$$

The Tail-based partition

distortion

$$\sup_{\mu\in\Lambda_f} d(\tilde{\phi}_n, \tilde{\psi}_n, \mu^n) \leq \underbrace{\sup_{\mu\in\Lambda_f} \mu(\Gamma^c_{k_n}) = \sum_{x>k_n} f(x)}_{eventually\ \ in\ \ n\ \ if\ \ k_n\to\infty}$$

overhead

$$minimax\ \ redundacy \sim \frac{1}{n} R^{+}(\Lambda^n_f, \sigma(\tilde{\pi}_{k_n}))$$

Determine regimes on $(k_n)_{n\geq 1}$ that guarantee a gain in minimax redundancy in the sense that:

$$\lim_{n\to\infty} \frac{R^+(\Lambda_f^n, \sigma(\tilde{\pi}_{k_n}))}{R^+(\Lambda_f^n)} = 0$$

subject to $(1/k_n)$ being $o(1)$ $\Leftrightarrow$ $\lim_n \sup_{\mu\in\Lambda_f} d(\tilde{\phi}_n, \tilde{\psi}_n, \mu^n) = 0$.

Theorem 2

For Λ_f with $f \in \ell_1(\mathcal{X})$ there exists a critical sequence of non decreasing integers $u_f^*(n)$ such that:

- if $k_n \geq u_f^*(n)$, eventually with n, then

$$\lim_{n \to \infty} \frac{R^+(\Lambda_f^n, \sigma(\tilde{\pi}_{k_n}))}{R^+(\Lambda_f^n)} = 1.$$

- conversely, if $k_n / u_f^*(n) \longrightarrow 0$, then

$$\lim_{n \to \infty} \frac{R^+(\Lambda_f^n, \sigma(\tilde{\pi}_{k_n}))}{R^+(\Lambda_f^n)} = 0.$$

Remarks about $u_f^*(n)$:[a]

1. $u_f^*(n)$ is fully determined by the envelope function f.
2. $u_f^*(n) \to \infty$ and $u_f^*(n)$ is $o(n/logn)$
3. furthermore $R^+(\Lambda_f^n) \approx (u_f^*(n) - 1) \log n$.

[a] Bontemps, Boucheron, and Gassiat, About adaptive coding on countable alphabets, IEEE Trans. on Inf. Th., 2014.

Remarks about Λ_f

1. the complexity of $\{\Lambda_f^n : n \geq 1\}$ is captured asymtotically by a finite (but dynamic) alphabet projection $\{\Lambda_f^n/\sigma(\tilde{\pi}_{k_n}), n \geq 1\}$ where

$$\Lambda_f^n/\sigma(\tilde{\pi}_{k_n}) = \{\mu^n/\sigma(\tilde{\pi}_{k_n} \times \cdots \times \tilde{\pi}_{k_n}) : \mu \in \Lambda_f\},$$

   with $\mu/\sigma(\pi) = \{\mu(A) : A \in \sigma(\pi)\}$.

2. $\{\tilde{\pi}_{k_n}\}$ is optimal in the sense of achieving the information radius of $\{\Lambda_f^n : n \geq 1\}$ with minimum size.

Power Law Envelope

$f_\alpha(x) = \min\{1, 1/x^\alpha\}$ for $\alpha > 2$ then $u_f^*(n)$ is $O(n^{1/(\alpha-1)})$.

Exponential Envelope

$f_\alpha(x) = \min\{1, Ce^{-\alpha x}\}$ for $\alpha > 0$ then $u_f^*(n)$ is $\frac{1}{\alpha}\log n + O(1)$.

Elements of the Proof

The gain regime: the argument follows from Bontemps et al. [Th. 2] and Boucheron et al. [Th. 4][1,2].

Elements:

- $(1+o(1))\frac{(u_f^*(n)-1)}{4}\log n \leq R^+(\Lambda_f^n) \leq 2+\log e + \frac{u_f^*(n)-1}{2}\log n$
- $R^+(\Lambda_f^n, \sigma(\tilde{\pi}_{k_n})) \leq \frac{k_n-1}{2}\log n + K$

[1] Bontemps, Boucheron, and Gassiat, About adaptive coding on countable alphabets, IEEE Trans. on Inf. Th., 2014

[2] Boucheron, Garivier, and Gassiat, Codign on countable infininite alphabets, IEEE Trans on Inf. Th, 2009.

Elements of the Proof

$k_n \geq u_f^*(n)$: We use ideas and results from Haussler and Opper[3].
Elements:

- Metric entropy: $H_\epsilon(\Lambda) \equiv \ln \mathcal{D}_\epsilon(\Lambda)$.
- Haussler et al. (Lemma 8) shows that:

$$R^+(\Lambda^n) \geq \log(e) \cdot \sup_{\epsilon>0} \min \left\{ \mathcal{H}_\epsilon(\Lambda), \frac{n\epsilon^2}{8} \right\} - 1.$$

- if $\epsilon^*_{\Lambda,n} = \inf \left\{ \epsilon > 0 : \mathcal{H}_\epsilon(\Lambda) \leq \frac{n\epsilon^2}{8} \right\}$, we have that

$$R^+(\Lambda^n) \geq \log(e) \cdot \mathcal{H}_{\epsilon^*_{\Lambda,n}}(\Lambda) - 1$$

[3] Haussler and Opper, Mutual information, metric entropy and comulative relative entropy rsik, The Annals of Statistics, 1997.

Elements of the Proof

$k_n \geq u_f^*(n)$: We use ideas and results from Haussler and Opper[4].
Elements:

- Metric entropy: $H_\epsilon(\Lambda) \equiv \ln \mathcal{D}_\epsilon(\Lambda)$.
- Haussler et al. (Lemma 8) shows that:

$$R^+(\Lambda^n) \geq \log(e) \cdot \sup_{\epsilon>0} \min \left\{ \mathcal{H}_\epsilon(\Lambda), \frac{n\epsilon^2}{8} \right\} - 1.$$

- if $\epsilon^*_{\Lambda,n} = \inf \left\{ \epsilon > 0 : \mathcal{H}_\epsilon(\Lambda) \leq \frac{n\epsilon^2}{8} \right\}$, we have that

$$\liminf_n R^+(\Lambda^n) \geq \log(e) \cdot \liminf_n \mathcal{H}_{\epsilon^*_{\Lambda,n}}(\Lambda) - 1$$

[4] Haussler and Opper, Mutual information, metric entropy and cumulative relative entropy risk, The Annals of Statistics, 1997.

Elements of the Proof

$k_n \geq u_f^*(n)$: We use ideas and results from Haussler and Opper[5].
Elements:

- Metric entropy: $H_\epsilon(\Lambda) \equiv \ln \mathcal{D}_\epsilon(\Lambda)$.
- Haussler et al. (Lemma 8) shows that:

$$R^+(\Lambda^n) \geq \log(e) \cdot \sup_{\epsilon>0} \min \left\{ \mathcal{H}_\epsilon(\Lambda), \frac{n\epsilon^2}{8} \right\} - 1.$$

- if $\epsilon_{n,k}^* = \inf \left\{ \epsilon > 0 : \mathcal{H}_\epsilon(\Lambda_{\tilde{f}_k}) \leq \frac{n\epsilon^2}{8} \right\}$, we have that

$$\liminf_n R^+(\Lambda_{\tilde{f}_{k_n}}^n) \geq \log(e) \cdot \liminf_n \mathcal{H}_{\epsilon_{n,k_n}^*}(\Lambda_{\tilde{f}_{k_n}}) - 1$$

[5] Haussler and Opper, Mutual information, metric entropy and cumulative relative entropy risk, The Annals of Statistics, 1997.

Elements of the Proof

- Proposition 1 presents sufficient conditions on (ϵ_n) and (k_n) such that:

$$\lim_{n\to\infty} \frac{\mathcal{H}_{\epsilon_n}(\Lambda_{\tilde{f}_{k_n}})}{\mathcal{H}_{\epsilon_n}(\Lambda_f)} = 1$$

  for $(k_n) \to \infty$ and $(\epsilon_n) \to 0$

- Proposition 2 presents sufficient conditions on (k_n) such that:

$$\lim_{n\to\infty} \frac{R^+(\Lambda^n_{\tilde{f}_{k_n}})}{R^+(\Lambda^n_f)} = 1.$$

Summary

- We revisit the problem of almost lossless universal source coding.
- Using a uniform convergence of the distortion to zero, we obtain the same necessary and sufficient condition of the lossless case.
- For $f \in \ell_1(\mathcal{X})$, it is feasible to obtain redundancy gain with respect to the minimax lossless solution.
- The complexity of $\{\Lambda_f^n, n \geq 1\}$ is achieved by finite alphabet projections $\{\Lambda_f^n/\sigma(\tilde{\pi}_{k_n}), n \geq 1\}$, where the critical size (dimension) of the projected family is given by $\{u_f^*(n) : n \geq 1\}$.

Thanks