Applications of Random Coding and Algebraic Coding Theories to Universal Lossless Source Coding Performance Bounds
Universal Lossless Coding Performance Bounds

Overview

Research Problem

Models Discussed

- Finite number of parameters parametric sources
- i.i.d. sources with large alphabets
- Patterns induced by i.i.d. sources
- Piecewise stationary sources with slowly varying statistics
- Piecewise stationary sources
- Switching sources

Research Approach

- Use Redundancy-Capacity Theorems to obtain bounds
- Lower bound the relevant capacity for given source model
- Scheme for a specific model
- Performance Lower Bounds (on Redundancy) - best possible performance of any
- Average Case Universal Lossless Compression
A sequence $x_n$ of length $n$, governed by $P$, is uniquely decipherable code for $\mu$ unknown in a known class $\mathcal{V}$. Average redundancy of code $L(\cdot)$ for $n$-sequences drawn by source $\theta$ is average redundancy.

\[
\text{Average Universality Measure of a Class } \mathcal{V} \quad \text{of code } L(\cdot) \text{ for } n\text{-sequences drawn by source } \theta.
\]

\[
\left(\mu X \right)^{\theta H} - \left(\mu X \right)^{\mathbb{E}_\mathcal{V} \left( \Theta \right)^{\underline{\mu}_H}} = \left( \Theta^\mathcal{V} \right)^{\underline{\mu}_H}
\]

Average redundancy for most sources [Rissanen, 1984] (strongest sense).

Maximin $R^-_\mathcal{V}(\cdot)$ and Minimax $R^+_\mathcal{V}(\cdot)$ averages redundancies - best code for some worst average (over $x_n$) case. [Davisson, 1973].

Unknown parameters cost redundancy.

Average redundancy for $\theta$ unknown parameters cost redundancy.

Average redundancy for $\theta$ unknown in a known class $\mathcal{V}$.

A sequence $x_n$ of length $n$, governed by $P^\theta$, unknown in a known class $\mathcal{V}$.

Problem Layout
Distiguishability

for every code \( T(\cdot) \) and almost every \( \theta \in V^r_k \).

\[
\frac{u}{\log N} \left( \varepsilon - 1 \right) \geq (\theta, T) \, u_H
\]


Strong Random Coding Version

\[
\frac{u}{\log N} \left( \varepsilon - 1 \right) \geq (\gamma V) \, u_H = (\gamma V) \, u_H
\]

[Implied from Davisson, 1973, Gallager, 1976]

Weak Version
<table>
<thead>
<tr>
<th>Compound Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of Redundancy-Capacity Theorem</td>
</tr>
</tbody>
</table>

\[ V = V_f \cup V_r \]

\textbf{Weak Version for } \( V_f \)

1. Demonstrate how to find \( \theta \).
2. Lower bound \( \mathcal{W} \).
3. Prove that \( \forall \theta \in \mathcal{W} \), all \( \mathcal{D} \in \theta \) are distinguishable by \( x^n \).

\textbf{Strong Version for } \( V_f \)

1. Demonstrate how to define \( \mathcal{V}_f \) such that \( \forall \mathcal{V}_f \) is in exactly one \( \mathcal{D} \).
2. Show that \( \forall \mathcal{V}_f \) is most of the class.
3. Show how to partition \( \mathcal{V}_f \) such that every source in \( \mathcal{V}_f \) is the uniform prior on \( \mathcal{V}_f \).

\textbf{Inter-class Redundancy Distinguishability} from \( V_r \).

- \( \forall \theta \in \mathcal{W}_r \), redundancy for \( \mathcal{V}_r \) consists of inter-class redundancy in \( V_r \) and
- \( \forall \theta \in \mathcal{W}_r \), redundancy for \( \mathcal{V}_r \) consists of intra-class redundancy in \( V_r \).

**Lower bound** \( \mathcal{W} \).

- Prove that for every valid \( \forall \theta \), all \( \mathcal{D} \in \theta \) are distinguishable by \( x^n \).
\[
\frac{u}{\log 10^u} (3 - 1) \leq (\theta, I) \log_2 P
\]

**By Theorem**, for every code and almost every \( \theta \) \( V \in \theta \).

In every all points distinguishable by \( x_n \) •

\( \phi \) is contained in a unique \( \theta \) and has equal probability to other \( \theta \).

Any \( V \notin \theta \) is contained in \( \phi \) and assumed negligible.

The volume of \( V \) outside \( V \) is negligible.

By theorem, for every code and almost every \( \theta \) \( V \in \theta \).

In every all points distinguishable by \( x_n \) •

\( \phi \) is contained in a unique \( \theta \) and has equal probability to other \( \theta \).

Any \( V \notin \theta \) is contained in \( \phi \) and assumed negligible.

The volume of \( V \) outside \( V \) is negligible.

\[
M = 10
\]

\[
\begin{align*}
M &= 10 \\
\phi &= 3 \\
\phi &= 12 \\
\phi &= 10 \\
\phi &= 13 \\
\phi &= 1
\end{align*}
\]
Finite \( k \)-dimensional Parametric Sources

Universal Lossless Coding Performance Bounds

[Prasunen, 1984] For every code \((\cdot, T) \supset \theta\) and almost every \(\theta \in \Theta\), almost every \(u \in [0, 1]n\)

\[
\frac{u}{u \log_2 (\varepsilon - 1)} \leq (\theta, T)^u_H
\]

\((\varepsilon - 1)^{-u} \geq n - 0.5(1 - e)\)

Initial shifts that define set \(\Phi \equiv \theta\)

Distinguishable if \(\Phi \equiv \theta\)

Determined by initial shift \(n\) in a grid (one sufficient for maximum)
Choose a random grid \( \mathcal{D} \) (as in random coding).

Generate \( x_n \) by a given \( \theta \).

Let \( \hat{\theta}_i \) be the Maximum Likelihood estimator of \( \theta \) from \( x_n \).

Let \( \mathcal{A}_i \) be the event that \( \hat{\theta}_i \neq \theta \).

Let \( \mathcal{D} \) be the grid point whose components are nearest \( \hat{\theta}_i \) to \( \theta \).

Prove that \( \Pr \{ \mathcal{D} \} = \frac{1}{\mathcal{D}} \).

\( \mathcal{D} \) is the event that \( \hat{\theta} \neq \theta \).

Distinguishability

\( \forall i \in \mathcal{I}, \forall \theta \in \Theta \) for \( \frac{\| \hat{\theta} - \theta \|}{\sqrt{2n}} \leq \left( \frac{1}{\mathcal{D}} \right)^{\frac{1}{2n}} \).

\( \forall i \in \mathcal{I}, \forall \theta \in \Theta \) for \( \mathcal{D} \).

\( \forall i \in \mathcal{I}, \forall \theta \in \Theta \).

\( \forall i \in \mathcal{I}, \forall \theta \in \Theta \).
Good for minimax/maximin redundancies.

This structure violates the requirements of the strong version, and thus is only

**Drawback**

- Number of grid points preceding proportional to $\frac{n^u}{\theta^x}$
- Spacing near $\frac{n^u}{\theta^x}$ proportional to $\frac{n^u}{\theta^x}$
- Build non-uniform grids.

**Solution**

- Too small spacing in grid $u_{(n-\varepsilon)-\varepsilon}$ results in loose bound.
- Too large spacing in grid $u_{(n-\varepsilon)+\varepsilon}$ results in lack of distinguishability in grids.

$$I(x) \leq \sum_{y=1}^{\Theta} \theta^y$$

Volume of $V^y$ is $I(x)$. 

Problems with Large $k$

[Shamir, 2003]

I.I.D. Sources - Large Alphabet $k$ - Minimax
Minimax/Maximin Redundancy - I.I.D. Large $\gamma$

Bounding number of points in grid results in

\[
\frac{\gamma}{u} \log \frac{u/2}{(1 - \gamma)} (\delta - 1) \leq (\gamma V) uH = (\gamma V) uH
\]

Bounding number of points in grid results in grid below, distinguishing by above definition (proven as in finite parametric case),

\[ \delta \in \Theta \]

is grid below.
Most Sources - I.I.D. Large $k$

Note: Second order term is lower than that of minimax/maximum bound.

[Shamir, 2003] for every code $T(\cdot)$ and almost every $\theta \in \mathcal{V}_g$

\[
\frac{u}{u} \log \frac{u}{(1 - \gamma)} (\varepsilon - 1) \leq (\theta, T)^{u/2}
\]

Result

\[
\frac{(1 - \gamma) \varepsilon (1 - \gamma) \log (1 - \gamma)}{I} \leq W
\]

Method

Pack as many as possible spheres with radius $r$ and volume $V^{k-1}(r)$ in the $k-1$ dimensional space $\mathcal{V}$ of volume $I$.

Place centers of the spheres (whole grid shifted for random selection).

To reduce number of points.

Factor in packing density $2 - (1 - \gamma) - \gamma$.

Factors in distance of distinguishability from $\theta$.

All sources outside a $\gamma - 1$ dimensional sphere with radius $r = (\varepsilon - 0.5)(1 - \gamma)$ are not useful here.

Key Realizations

Non-uniform grid above is not used here.
Motivation

Patterns Induced by I.I.D. Sources

Examples: The strings: $x_n = \text{lossless}, \text{sellsoll, 12331433, 76887288}$ all have the same pattern $f(x_n) = 12331433$.

Indices assigned to original sequence letters in order of first occurrence.

Approach

Use the inevitable cost to improve compression.

Coding cost of unknown alphabet is inevitable.

Sometimes alphabet is unknown and possibly large.

Classical compression considers known small alphabets.

Classical compression considers known small alphabets.

Motivation

Patterns Induced by I.I.D. Sources

Universal Lossless Coding Performance Bounds
Note: For \( k = 3 \) this is true for any combination of 2 out of 3 letters.

\[ \{0, 0.2, 0.7\} = \theta, \quad \{0, 1, 0.2\} = \theta \]

Example: Typical sequences - similar patterns

Any \( \theta \) which is a permutation of \( \theta \) appears to be the same source:

\( k = 2 \)

\( k = 3 \)
Averagence most-sources lower bound

\[
\begin{align*}
\frac{\varepsilon}{1 - \varepsilon} \cdot \frac{1}{u} &< \gamma \quad \text{for} \\
\frac{\varepsilon}{1 - \varepsilon} \cdot \frac{1}{u} &\geq \gamma \quad \text{for} \\
\left(\frac{u}{1 - u}\right) O - \frac{\varepsilon}{(1 - \varepsilon)\gamma} O &\geq \frac{\varepsilon}{1 - \varepsilon} \cdot \log \left\lfloor \frac{u}{1 - u}\right\rfloor \frac{u - \varepsilon}{1 - \varepsilon} \cdot \log \left\lfloor \frac{u}{1 - u}\right\rfloor \frac{u - \varepsilon}{1 - \varepsilon}
\end{align*}
\]

\[\gamma \in \left\{(\theta, \Phi, T)^u H \right\} \subseteq \left\{(h, V)^u H \right\}
\]

Averagence min-max lower bound

\[
\begin{align*}
\frac{\varepsilon}{1 - \varepsilon} \cdot \frac{1}{u} &< \gamma \quad \text{for} \\
\frac{\varepsilon}{1 - \varepsilon} \cdot \frac{1}{u} &\geq \gamma \quad \text{for} \\
\left(\frac{u}{1 - u}\right) O - \frac{\varepsilon}{(1 - \varepsilon)\gamma} O &\geq \frac{\varepsilon}{1 - \varepsilon} \cdot \log \left\lfloor \frac{u}{1 - u}\right\rfloor \frac{u - \varepsilon}{1 - \varepsilon} \cdot \log \left\lfloor \frac{u}{1 - u}\right\rfloor \frac{u - \varepsilon}{1 - \varepsilon}
\end{align*}
\]

\[\gamma \in \left\{(h, V)^u H \right\} \subseteq \left\{(\gamma, V, \Phi, T)^u H \right\}
\]

[Shamir, 2003]

Bounds

There is no reduction in both maximum and most source cases (relevant for large $u$)

Pattern Redundancy

Universal Lossless Coding Performance Bounds
for every $\mathcal{L} \in \Theta$, for almost every $\phi \in \Phi$.

\[
\frac{u}{(b/u)\log_2 \left( 1 - b + b \gamma \frac{2}{1} \right) (\bar{\varepsilon} - 1)} \leq (\phi, \mathcal{T})^u \mathcal{H}
\]

Redundancy bound [Shamir, 2000]

\begin{align*}
\text{(TP) transition path} & - \{t_1, t_2, \ldots, t_l\} \supseteq \mathfrak{t} \bullet \\
\text{sequential parameters} & - \{b \theta, \ldots, \bar{\varepsilon} \theta, t_1 \theta, t_2 \theta \} \supseteq \mathfrak{\theta} \bullet \\
\text{All PSS's in } \mathcal{V} \text{ with } b \text{ segments} & - \mathcal{V} \supseteq b \mathcal{V} \bullet \\
\text{y-dimensional parameters for } n \text{-sequences} & - \mathcal{V} \supseteq b \mathcal{V} \bullet \\
\text{with order class of PSS's (contains all possible combinations of the} & \text{ abrupt changes in statistics} \\
\text{PSS's) emits data divided into independent stationary segments separated by} & \mathcal{V} \supseteq b \mathcal{V} \supseteq (\mathfrak{t}, \mathfrak{\theta}) \supseteq \mathfrak{\phi} \bullet
\end{align*}
By definition of the grids, the bound for finite $b$ [Merhav, 1993] results:

1. Given $\theta$, estimate transitions from respective grids.
   - Use phrases entirely inside segments to estimate $\theta$.

2. Partition $\mathcal{A}$ into sets as follows:
   - $A \in \mathcal{A}$ and $t \in \theta$ must be from grids with identical initial shifts.
   - $A \in \mathcal{A}$, $t \in \theta$ is a point in a grid as defined for stationary sources.
   - For all $\mathcal{A} \in \mathcal{A}$, $t \in \theta$ is a point in the same phrase in a grid with spacing $l$.
   - Parse n-tuple to phrases of length $l = \lfloor n_{1-\varepsilon} \rfloor$.

3. Partition $\mathcal{V}$ into sets as follows:
   - $V \in \mathcal{V}$ is most class for fixed $b$.
   - Transitions are large.
   - $V \in \mathcal{V}$ contains all for which all segments long (longer than $n_{1-\varepsilon}$) and all

Finite Number of Segments $b$

Bound Derivation - PSS
Set $\Phi$ contains all $\theta^1$ is one grid point $\theta^2$ is one grid point $\theta^3$ is one grid point $\{y_j^1 : \phi^1 \in \theta^1, \phi^2 \in \theta^2, \phi^3 \in \theta^3\}$ are only red points $\{y_j^1 : \phi^1 \in \theta^1, \phi^2 \in \theta^2, \phi^3 \in \theta^3\}$ are only blue points $\{y_j^1 : \phi^1 \in \theta^1, \phi^2 \in \theta^2, \phi^3 \in \theta^3\}$ are only green points $y_j^1, y_j^2, y_j^3$ are only blue points.

Phrase Set $\{y_j^1 : \phi^1 \in \theta^1, \phi^2 \in \theta^2, \phi^3 \in \theta^3\}$ contains all combinations with one point from each of the five grids.
General Bound Derivation - PSS's

Large $q$

1. $A_e$ defined is not most class.
2. For very large $q$, probability of error in at least one of the source parameters significantly increases the overall error probability.

Solutions to Asymptotic Problems

1. $A_e$ contains sources for which most segments are long and most transitions are large.
2. Reduce sets $\varphi$ to improve distinguishability for very large $q$.

Two different Cases

- $q \gg n/q$ - almost similar to fixed $q$ (modified according to modification 1 above).
- $q \gg n/q$ - requires additional algebraic coding techniques for distinguishability.
universal-lossless-coding-performance-bounds

Second Case: $b/n$ is now larger

Guarantees distinguishability even for $b/n \ll b$; resulting in the same asymptotic

codes designed to correct up to any errors (exist: Gilbert-Varshamov).

Each grid point is assigned an element in the proper Galois Field.

Grids’ resolutions chosen to yield Galois Fields.

Remaining parameters are parity checks.

Remaining parameters are parity checks.

remaining parameters and $(1 - \rho) b (l - 1)$ transitions chosen from grids.

$\rho$ - number of free transitions.

Let $\rho > 0$ be arbitrarily small.

Solution - Reduce $\phi$ by Linear Block Codes:

$\phi$ - Error in estimating one results in error in estimating $\phi$.

Too many parameters.

$\frac{b}{n} \ll b$
\[ \frac{u}{b/u} \log \left( I - b + b \frac{\zeta}{u} \right) (1 - b) \leq (\phi, T)^u H \]

\[ \frac{u}{b/u} \log \left( I - b + b \frac{\zeta}{u} \right) (1 - b) \leq (\phi, T)^u H \]

\[ \left[ s \log (s - b) + (b/u) \log \left( I - b \right) + \frac{s}{u} \log \left( \frac{\zeta}{s} \right) \right] \frac{u}{(1 - b)} \leq (\phi, T)^u H \]

**Switching Sources**

Significant cost above PSS's.

Hierarchical version of redundancy-capacity for compound class must be used.

\[ \frac{u}{b/u} \log \left( I - b + b \frac{\zeta}{u} \right) (1 - b) \leq (\phi, T)^u H \]

\[ \frac{u}{b/u} \log \left( \frac{\zeta}{\phi} - I \right) (1 - b) + \frac{\zeta}{b/y} \right] (1 - b) \leq (\phi, T)^u H \]

PSS's with Slowly Linearly Varying Statistics

Additional Source Classes

[Shamir, 2001]
Summary and Conclusions

1. The redundancy-capacity theorem is very useful to derive lower bounds on
   ● algebraic code distance bounds.
   ● sphere packing.
   ● random coding.

3. Different techniques from coding theory were used:
   ● switching sources.
   ● piecewise stationary sources with slowly varying statistics.
   ● piecewise stationary sources.
   ● patterns induced by i.i.d. sources.
   ● i.i.d. sources with large alphabets.
   ● finite number of parameters parametric sources.

2. Lower bounds on redundancy in both cases were obtained for
   ● redundancy for most sources in universal coding.
   ● minimum/maximum redundancy in universal coding.

Universal Lossless Coding Performance Bounds