

Non-Asymptotic Bounds for Optimal Fixed-to-Variable Lossless Compression Without Prefix Constraints

Igal Sason (Technion)

Joint work with Sergio Verdú (Princeton)

ISIT 2018

Vail, Colorado, USA

June 17-22, 2018

Variable-Length Lossless Binary Codes

- A variable-length lossless binary code for a discrete set \mathcal{X} is an injective mapping:

$$f: \mathcal{X} \rightarrow \{0,1\}^* = \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \dots\};$$

- $f(x)$ is the codeword which is assigned to $x \in \mathcal{X}$. Its length is denoted by $\ell(f(x))$ where $\ell: \{0,1\}^* \rightarrow \{0,1,2,\dots\}$ with $\ell(\emptyset) = 0$.

Variable-Length Lossless Binary Codes

- A variable-length lossless binary code for a discrete set \mathcal{X} is an injective mapping:

$$f: \mathcal{X} \rightarrow \{0,1\}^* = \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \dots\};$$

- $f(x)$ is the codeword which is assigned to $x \in \mathcal{X}$. Its length is denoted by $\ell(f(x))$ where $\ell: \{0,1\}^* \rightarrow \{0,1,2,\dots\}$ with $\ell(\emptyset) = 0$.
- A variable-length lossless source code is **compact** whenever it contains a codeword only if all shorter codewords also belong to the code.
- Given a probability mass function P_X on \mathcal{X} , a variable-length lossless source code is **P_X -efficient** if for all $(a,b) \in \mathcal{X}^2$,

$$\ell(f(a)) < \ell(f(b)) \implies P_X(a) \geq P_X(b).$$

- Given a probability mass function P_X on \mathcal{X} , a variable-length lossless source code is **P_X -optimal** if it is both compact and P_X -efficient.

Properties of Optimal Codes (Kontoyiannis & Verdú, 2014)

Let

- P_X be a probability mass function of X on a finite set \mathcal{X} ;
- $f_X^*: \mathcal{X} \rightarrow \{0, 1\}^*$ be a P_X -optimal code.

The following results hold:

- The distribution of $\ell(f_X^*(X))$ only depends on P_X ; it is invariant to the actual choice of f_X^* .

-

$$\sum_{x \in \mathcal{X}} 2^{-\ell(f_X^*(x))} \leq \log_2(1 + |\mathcal{X}|);$$

Kraft's inequality is replaced by a weaker condition.

Cumulant Generating Functions (Campbell, 1965)

The **cumulant generating function** of the codeword lengths of P_X -optimal binary codes $f_X^*: \mathcal{X} \rightarrow \{0, 1\}^*$ is given by

$$\Lambda^*(\rho) \triangleq \log \mathbb{E}[2^{\rho \ell(f_X^*(X))}], \quad \rho \in \mathbb{R}.$$

Connection to Average Length

$$\lim_{\rho \rightarrow 0} \frac{\Lambda^*(\rho)}{\rho} = \mathbb{E}[\ell(f_X^*(X))].$$

The Rényi Entropy

Let P_X be a probability distribution on a discrete set \mathcal{X} . The **Rényi entropy of order $\alpha \in (0, 1) \cup (1, \infty)$ of X** is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X^\alpha(x)$$

By its continuous extension, $H_1(X) = H(X)$.

Theorem 1 (Courtade and Verdú, ISIT 2014)

If $\rho \in (-\infty, -1]$, then

$$H_\infty(X) - \log \log_2(1 + |\mathcal{X}|) \leq -\Lambda^*(\rho) \leq H_\infty(X),$$

and, if $\rho \in (-1, 0) \cup (0, \infty)$, then

$$H_{\frac{1}{1+\rho}}(X) - \log \log_2(1 + |\mathcal{X}|) \leq \frac{\Lambda^*(\rho)}{\rho} \leq H_{\frac{1}{1+\rho}}(X).$$

Theorem 2 (Courtade and Verdú, ISIT 2014)

For all $H(X) < R < \log |\mathcal{X}|$

$$\log \frac{1}{\mathbb{P}[\ell(f_X^*(X)) \geq R]} \geq \sup_{\rho > 0} \left\{ \rho R - \rho H_{\frac{1}{1+\rho}}(X) \right\}$$

$$= D(X_\alpha \| X)$$

where $\alpha \in (0, 1)$ is a function of R chosen so that

$$R = H(X_\alpha),$$

and X_α has the scaled probability mass function

$$P_{X_\alpha}(x) = \frac{P_X^\alpha(x)}{\sum_{a \in \mathcal{X}} P_X^\alpha(a)}, \quad x \in \mathcal{X}.$$

Theorem: Key Result

Given a discrete random variable X taking values on a set \mathcal{X} , a function $g: \mathcal{X} \rightarrow (0, \infty)$, and a scalar $\rho \neq 0$, then

$$\begin{aligned} & \sup_{\beta \in (-\rho, +\infty) \setminus \{0\}} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log \sum_{x \in \mathcal{X}} g^{-\beta}(x) \right] \\ & \leq \frac{1}{\rho} \log \mathbb{E}[g^\rho(X)] \\ & \leq \inf_{\beta \in (-\infty, -\rho) \setminus \{0\}} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log \sum_{x \in \mathcal{X}} g^{-\beta}(x) \right]. \end{aligned}$$

Letting $\beta = 1$ yields the lower bound by Courtade and Verdú (ISIT '14).

Application of Key Result

Let X be a random variable taking values on a finite set \mathcal{X} , and let $\rho \neq 0$. Then, for an optimal binary code,

$$\frac{1}{\rho} \log \mathbb{E}[2^{\rho \ell(f_X^*(X))}] \geq \sup_{\beta \in (-\rho, \infty) \setminus \{0\}} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log t(\beta, |\mathcal{X}|) \right],$$

where $t(\cdot, \cdot)$ is defined as

$$t(\beta, |\mathcal{X}|) \triangleq \sum_{x \in \mathcal{X}} 2^{-\beta \ell(f_X^*(x))} = \begin{cases} (2^\Delta - 1)s_\beta^m + \frac{1 - s_\beta^m}{1 - s_\beta}, & \beta \neq 1 \\ m + 2^\Delta - 1, & \beta = 1, \end{cases}$$

and

$$s_\beta = 2^{1-\beta}, \quad m = \lfloor \log_2(1 + |\mathcal{X}|) \rfloor, \quad \Delta = \log_2(1 + |\mathcal{X}|) - m \in [0, 1).$$

Guessing and Ranking functions

- X is a discrete random variable, taking values on $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$.
- One wishes to guess the value of X by repeatedly asking questions of the form “Is X equal to x ?” until the value of X is correctly guessed.
- A **guessing function** is a 1-to-1 function $g: \mathcal{X} \rightarrow \mathcal{X}$ where the number of guesses is equal to $g(x)$ if $X = x \in \mathcal{X}$.
- For $\rho > 0$, the ρ -th moment of the number of guesses is minimized by selecting the guessing function to be a **ranking function** g_X , for which $g_X(x) = k$ if $P_X(x)$ is the k -th largest mass.

Connection Between Guessing and Cumulant Generating Functions

Let

- X be a random variable taking values on a finite set \mathcal{X} ;
- g_X be a ranking function of X ;
- $f_X^*: \mathcal{X} \rightarrow \{0, 1\}^*$ be an optimal binary P_X -code with a cumulant generating function Λ^* .

Then, for every $\rho > 0$,

$$\log(\mathbb{E}[g_X^\rho(X)]) - \rho \log 2 < \Lambda^*(\rho) \leq \log \mathbb{E}[g_X^\rho(X)].$$

Improved Upper Bound on Optimal Guessing Moment

Let

- X be a discrete random variable taking values on a set \mathcal{X} ;
- g_X be the ranking function according to P_X .

Then, for all $\rho > 0$,

$$\mathbb{E}[g_X^\rho(X)] \leq \frac{1}{1+\rho} \left[\exp\left(\rho H_{\frac{1}{1+\rho}}(X)\right) - 1 \right] + \exp\left((\rho - 1)^+ H_{\frac{1}{\rho}}(X)\right)$$

where $(x)^+ \triangleq \max\{x, 0\}$ for $x \in \mathbb{R}$.

Arikan's Result

This improves the bound by Arikan:

$$\mathbb{E}[g_X^\rho(X)] \leq \exp\left(\rho H_{\frac{1}{1+\rho}}(X)\right).$$

Improved Bounds on the Cumulant for Optimal Binary Codes

Let X be a random variable taking values on a finite set \mathcal{X} . Then, for every optimal binary code, the cumulant generating function satisfies

$$\begin{aligned} & \sup_{\beta \in (-\rho, \infty) \setminus \{0\}} \frac{1}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \log t(\beta, |\mathcal{X}|) \right] \\ & \leq \frac{\Lambda^*(\rho)}{\rho} \end{aligned} \quad (1)$$

$$\begin{aligned} & \leq H_{\frac{1}{1+\rho}}(X) + \frac{1}{\rho} \log \left(\frac{1}{1+\rho} \left[1 - \exp \left(-\rho H_{\frac{1}{1+\rho}}(X) \right) \right] \right. \\ & \quad \left. + \exp \left((\rho - 1)^+ H_{\frac{1}{\rho}}(X) - \rho H_{\frac{1}{1+\rho}}(X) \right) \right), \end{aligned} \quad (2)$$

for all $\rho > 0$. Moreover, (1) also holds for $\rho < 0$.

Optimal Fixed-to-Variable-Length Lossless Binary Codes

- Consider the object to be compressed $X^n = (X_1, \dots, X_n) \in \mathcal{A}^n$ is a string of length n (n is known to both encoder and decoder).
- The letters $\{X_i\}$ are drawn i.i.d. from a finite alphabet \mathcal{X} according to the probability mass function $P_{X^n}(x^n) = \prod_{i=1}^n P_X(x_i)$ for $x^n \in \mathcal{A}^n$.

Optimal Fixed-to-Variable-Length Lossless Binary Codes

- Consider the object to be compressed $X^n = (X_1, \dots, X_n) \in \mathcal{A}^n$ is a string of length n (n is known to both encoder and decoder).
- The letters $\{X_i\}$ are drawn i.i.d. from a finite alphabet \mathcal{X} according to the probability mass function $P_{X^n}(x^n) = \prod_{i=1}^n P_X(x_i)$ for $x^n \in \mathcal{A}^n$.
- Consider the following non-asymptotic measures for optimal fixed-to-variable lossless compression of X^n :
 - ▶ The **cumulant generating function of the codeword lengths** is given by

$$\Lambda_n(\rho) := \frac{1}{n} \log \mathbb{E} \left[2^{\rho \ell(f_{X^n}^*(X^n))} \right], \quad \rho \in \mathbb{R}.$$

- ▶ The **non-asymptotic version of the source reliability function** is given by

$$E_n(R) = \frac{1}{n} \log \left(\frac{1}{\mathbb{P} \left[\frac{1}{n} \ell(f_{X^n}^*(X^n)) \geq R \right]} \right).$$

Improved Bounds for Optimal FV Lossless Binary Codes

Consider a memoryless and stationary source of finite alphabet \mathcal{A} , and let $f_{X^n}^* : \mathcal{A}^n \rightarrow \{0, 1\}^*$ be an optimal compression code. Then, for all $\rho > 0$,

$$\begin{aligned} & \sup_{\beta \in (-\rho, \infty) \setminus \{0\}} \frac{\rho}{\beta} \left[H_{\frac{\beta}{\beta+\rho}}(X) - \frac{1}{n} \log t(\beta, |\mathcal{A}|^n) \right] \\ & \leq \Lambda_n(\rho) \\ & \leq \rho H_{\frac{1}{1+\rho}}(X) + \frac{1}{n} \log \left(\frac{1}{1+\rho} \left[1 - \exp \left(-n \rho H_{\frac{1}{1+\rho}}(X) \right) \right] \right. \\ & \quad \left. + \exp \left(n \left[(\rho - 1)^+ H_{\frac{1}{\rho}}(X) - \rho H_{\frac{1}{1+\rho}}(X) \right] \right) \right). \end{aligned}$$

Asymptotic Result

$$\lim_{n \rightarrow \infty} \Lambda_n(\rho) = \rho H_{\frac{1}{1+\rho}}(X).$$

Improved Bounds for Optimal FV Lossless Binary Codes (Cont.)

Furthermore, for $R < \log |\mathcal{A}|$,

$$E_n(R) \geq \sup_{\rho > 0} \left\{ \rho R - \rho H_{\frac{1}{1+\rho}}(X) - \frac{1}{n} \log \left(\frac{1}{1+\rho} \left[1 - \exp \left(-n \rho H_{\frac{1}{1+\rho}}(X) \right) \right] + \exp \left(n \left[(\rho - 1)^+ H_{\frac{1}{\rho}}(X) - \rho H_{\frac{1}{1+\rho}}(X) \right] \right) \right) \right\}.$$

Bounds (Courtade and Verdú, ISIT 2014)

$$\rho H_{\frac{1}{1+\rho}}(X) - \frac{\rho}{n} \log \log_2(1 + |\mathcal{A}|^n) \leq \Lambda_n(\rho) \leq \rho H_{\frac{1}{1+\rho}}(X), \quad \rho > 0,$$

$$E_n(R) \geq \sup_{\rho > 0} \left\{ \rho R - \rho H_{\frac{1}{1+\rho}}(X) \right\}, \quad R < \log |\mathcal{A}|.$$

Bounds (Courtade and Verdú, ISIT 2014)

$$\rho H_{\frac{1}{1+\rho}}(X) - \frac{\rho}{n} \log \log_2(1 + |\mathcal{A}|^n) \leq \Lambda_n(\rho) \leq \rho H_{\frac{1}{1+\rho}}(X), \quad \rho > 0,$$

$$E_n(R) \geq \sup_{\rho > 0} \left\{ \rho R - \rho H_{\frac{1}{1+\rho}}(X) \right\}, \quad R < \log |\mathcal{A}|.$$

Comparison of Bounds

- The improvement is noticeable for small to moderate values of n .
- They are all asymptotically tight as we let $n \rightarrow \infty$.

Bounds (Courtade and Verdú, ISIT 2014)

$$\rho H_{\frac{1}{1+\rho}}(X) - \frac{\rho}{n} \log \log_2(1 + |\mathcal{A}|^n) \leq \Lambda_n(\rho) \leq \rho H_{\frac{1}{1+\rho}}(X), \quad \rho > 0,$$

$$E_n(R) \geq \sup_{\rho > 0} \left\{ \rho R - \rho H_{\frac{1}{1+\rho}}(X) \right\}, \quad R < \log |\mathcal{A}|.$$

Comparison of Bounds

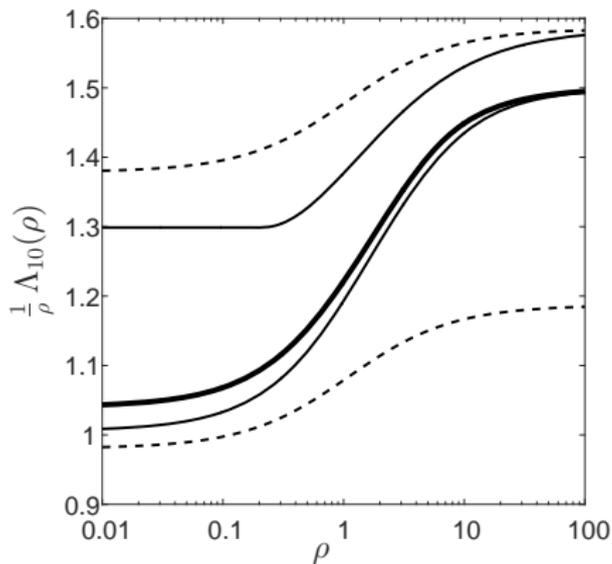
- The improvement is noticeable for small to moderate values of n .
- They are all asymptotically tight as we let $n \rightarrow \infty$.

Example: Ternary Memoryless Source

$$P_X(a) = \frac{4}{7}, \quad P_X(b) = \frac{2}{7}, \quad P_X(c) = \frac{1}{7}.$$

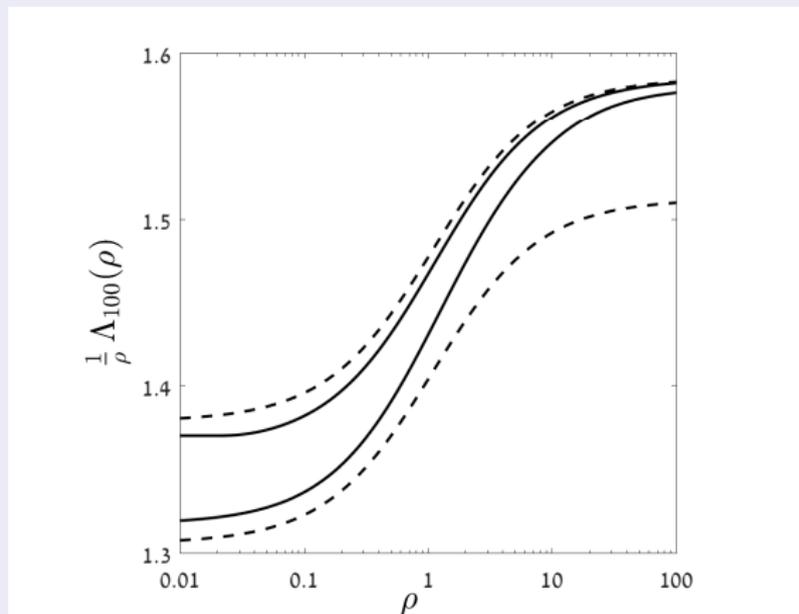
Numerical Results ($n = 10$)

Figure: Bounds on the normalized cumulant generating function, $\frac{\Lambda_n(\rho)}{\rho}$ (in bits), of the codeword lengths of optimal lossless compression of strings of length n emitted by the discrete memoryless source in the example. The dashed lines are the old bounds, and the thin solid lines refer to the improved bounds. The thick solid line refers to the exact normalized cumulant (thick solid curve).



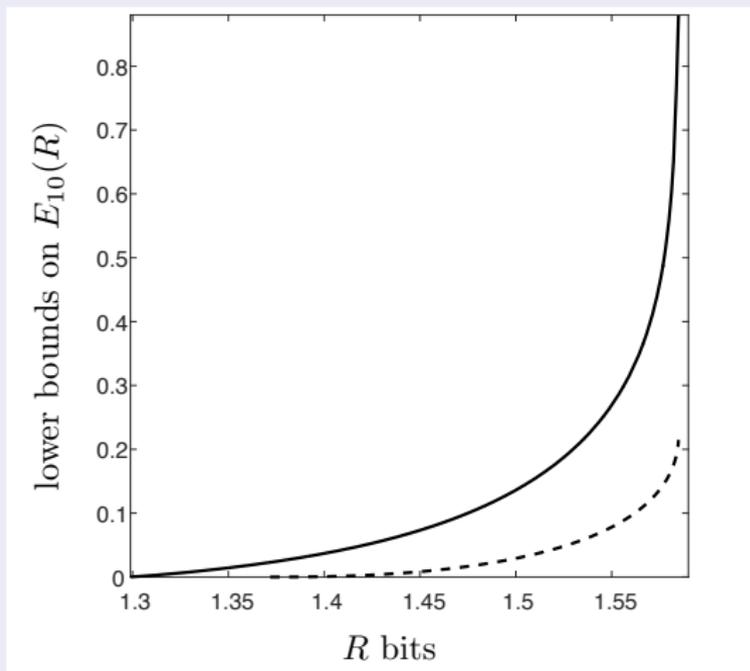
Numerical Results ($n = 100$)

Figure: Bounds on the normalized cumulant generating function, $\frac{\Lambda_n(\rho)}{\rho}$ (in bits), of the codeword lengths of optimal lossless compression of strings of length n emitted by the discrete memoryless source in the example. The dashed lines are the old bounds, and the thin solid lines refer to the improved bounds.



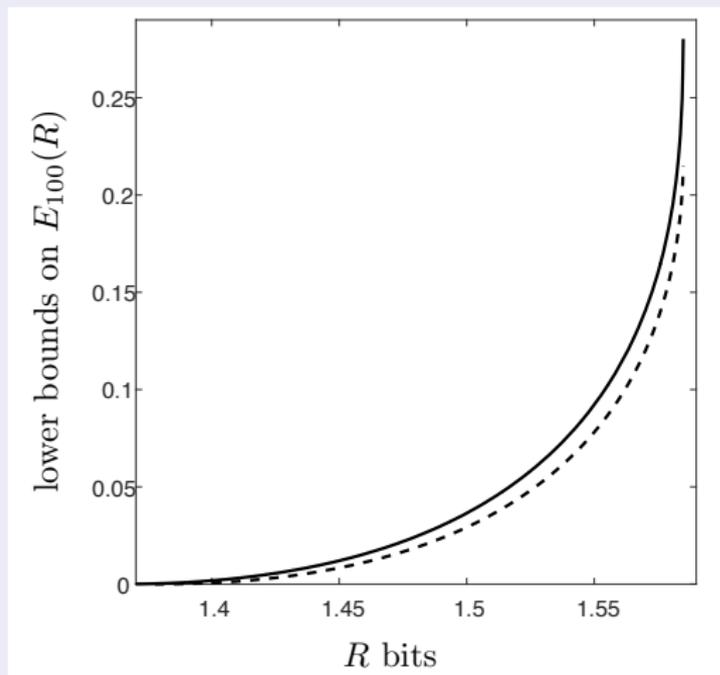
Numerical Results ($n = 10$)

Figure: Lower bounds on the non-asymptotic reliability function, $E_n(R)$ (base 2), for the discrete memoryless source in the example. The dashed curve refers to the old lower bound on $E_n(R)$, and the solid curve refers to the tighter lower bound.



Numerical Results ($n = 100$)

Figure: Lower bounds on the non-asymptotic reliability function, $E_n(R)$ (base 2), for the discrete memoryless source in the example. The dashed curve refers to the old lower bound on $E_n(R)$, and the solid curve refers to the tighter lower bound.



Summary

- Our bounds on guessing moments serve to analyze non-prefix 1-to-1 binary optimal codes, which do not satisfy Kraft's inequality.
- We derived
 - ▶ improved bounds on the distribution of optimal codeword lengths;
 - ▶ improved non-asymptotic bounds for fixed-to-variable codes without prefix constraints.
- In the full journal paper:
Relying on these techniques, lower bounds on the cumulant generating function of the codeword lengths are derived, by means of the smooth Rényi entropy, for source codes that allow decoding errors.

Journal Paper

I.S. and S. Verdú, "Improved bounds on lossless source coding and guessing moments via Rényi measures," *IEEE Trans. on Information Theory*, vol. 64, no. 6, pp. 4323–4346, June 2018.