Coding Theorems using Rényi Information Measures

Himanshu Tyagi

Indian Institute of Science

Lossless Source Coding

The lossless source coding problem

Let $L_\epsilon(X)$ be the minimum l such that there exists $\mathcal{T} \subseteq \mathcal{X}$

1. $|\mathcal{T}| \leq 2^l$
2. $\Pr\left(X \in \mathcal{T}\right) \geq 1 - \epsilon$

The lossless source coding problem

Let $L_\epsilon(X)$ be the minimum l such that there exists $\mathcal{T} \subseteq \mathcal{X}$

1. $|\mathcal{T}| \leq 2^l$
2. $\Pr(X \in \mathcal{T}) \geq 1 - \epsilon$

Characterize $L_\epsilon(X)$

An upper bound

[Han-Verdú '93] Suppose that there exists a $\lambda > 0$ such that

$$\Pr\left(X \in \{x : -\log \mathrm{P}\left(x\right) \leq \lambda\}\right) \geq 1 - \epsilon.$$

Then,

$$L_\epsilon(X) \leq \lambda.$$

An upper bound

[Han-Verdú '93] Suppose that there exists a $\lambda > 0$ such that

$$\Pr\left(X \in \{x : -\log \mathrm{P}\left(x\right) \leq \lambda\}\right) \geq 1 - \epsilon.$$

Then,

$$L_\epsilon(X) \leq \lambda.$$

A large prob. upper bound for $h(X) = -\log \mathrm{P}\left(X\right)$

is an upper bound for $L_\epsilon(X)$

Shannon's source coding theorem: Upper bound

Shannon's source coding theorem: Upper bound

Using Markov's inequality:

$$\Pr\left(h(X) \leq \mathbb{E}[h(X)]/\epsilon\right) \geq 1 - \epsilon.$$

Shannon's source coding theorem: Upper bound

Using Markov's inequality:

$$\Pr\left(h(X) \leq \mathbb{E}[h(X)]/\epsilon\right) \geq 1 - \epsilon.$$

The quantity $H(X) = \mathbb{E}[h(X)]$ is the *Shannon entropy of* X.

Shannon's source coding theorem: Upper bound

Using Markov's inequality:

$$\Pr\left(h(X) \leq \mathbb{E}[h(X)]/\epsilon\right) \geq 1 - \epsilon.$$

The quantity $H(X) = \mathbb{E}[h(X)]$ is the *Shannon entropy of X*.

Thus,

$$L_\epsilon(X) \leq H(X)/\epsilon.$$

Shannon's source coding theorem: Upper bound

Using Markov's inequality:

$$\Pr\left(h(X) \leq \mathbb{E}[h(X)]/\epsilon\right) \geq 1 - \epsilon.$$

The quantity $H(X) = \mathbb{E}[h(X)]$ is the *Shannon entropy of* X.

Thus,

$$L_\epsilon(X) \leq H(X)/\epsilon.$$

[Shannon 1948] For $X^n = (X_1, ..., X_n)$ consisting of n i.i.d. samples,

$$h(X^n) = \sum_{i=1}^{n} h(X_i).$$

Thus, by the law of large numbers

$$L_\epsilon(X^n) \leq nH(X_1) + O(\sqrt{n}).$$

A Rényi entropy based upper bound

Rényi entropy of order $0 \leq \alpha \neq 1$:

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

A Rényi entropy based upper bound

Rényi entropy of order $0 \leq \alpha \neq 1$:

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

Consider the set

$$\mathcal{T}_\alpha = \left\{ x \in \mathcal{X} : h(x) \leq H_\alpha(X) + \frac{1}{1-\alpha} \log \frac{1}{\epsilon} \right\}, \quad 0 \leq \alpha < 1.$$

A Rényi entropy based upper bound

Rényi entropy of order $0 \leq \alpha \neq 1$:

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

Consider the set

$$\mathcal{T}_\alpha = \left\{ x \in \mathcal{X} : h(x) \leq H_\alpha(X) + \frac{1}{1-\alpha} \log \frac{1}{\epsilon} \right\}, \quad 0 \leq \alpha < 1.$$

Then,

$$\begin{aligned}
1 &= \sum_{x \in \mathcal{T}_\alpha} \mathrm{P}(x) + \sum_{x \in \mathcal{T}^c} \mathrm{P}(x) \\
&\leq \Pr\left(X \in \mathcal{T}_\alpha\right) + \sum_{x \in \mathcal{T}_\alpha^c} \mathrm{P}(x)^\alpha \cdot \mathrm{P}(x)^{1-\alpha} \\
&< \Pr\left(X \in \mathcal{T}_\alpha\right) + \epsilon 2^{-(1-\alpha)H_\alpha(X)} \sum_{x \in \mathcal{T}_\alpha^c} \mathrm{P}(x)^\alpha \\
&< \Pr\left(X \in \mathcal{T}_\alpha\right) + \epsilon.
\end{aligned}$$

A Rényi entropy based upper bound

Rényi entropy of order $0 \leq \alpha \neq 1$:

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

Consider the set

$$\mathcal{T}_\alpha = \left\{ x \in \mathcal{X} : h(x) \leq H_\alpha(X) + \frac{1}{1-\alpha} \log \frac{1}{\epsilon} \right\}, \quad 0 \leq \alpha < 1.$$

Then,

$$\Pr\left(X \in \mathcal{T}_\alpha\right) \geq 1 - \epsilon.$$

A Rényi entropy based upper bound

Rényi entropy of order $0 \leq \alpha \neq 1$:

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

Consider the set

$$\mathcal{T}_\alpha = \left\{ x \in \mathcal{X} : h(x) \leq H_\alpha(X) + \frac{1}{1-\alpha} \log \frac{1}{\epsilon} \right\}, \quad 0 \leq \alpha < 1.$$

Then,

$$\Pr\left(X \in \mathcal{T}_\alpha\right) \geq 1 - \epsilon.$$

We have shown: For every $0 \leq \alpha < 1$

$$L_\epsilon(X) \leq H_\alpha(X) + \frac{1}{1-\alpha} \log \frac{1}{\epsilon}$$

A lower bound

[Han-Verdú '93]

Suppose that there exists a $\lambda > 0$ such that

$$\Pr\left(X \in \{x : -\log \mathrm{P}\left(x\right) \geq \lambda\}\right) \geq 1 - \delta.$$

Then,

$$L_\epsilon(X) \geq \lambda - \log \frac{1}{1 - \epsilon - \delta}.$$

A lower bound

[Han-Verdú '93]

Suppose that there exists a $\lambda > 0$ such that

$$\Pr\left(X \in \{x : -\log \mathrm{P}\left(x\right) \geq \lambda\}\right) \geq 1 - \delta.$$

Then,

$$L_\epsilon(X) \geq \lambda - \log \frac{1}{1 - \epsilon - \delta}.$$

A large prob. lower bound for $h(X) = -\log \mathrm{P}\left(X\right)$

is a lower bound for $L_\epsilon(X)$

Shannon's source coding theorem: Lower bound

Using Chebyshev's inequality:

$$\Pr\left(h(X) \geq \mathbb{E}[h(X)] - \sqrt{\frac{V}{\epsilon}}\right) \geq 1 - \epsilon,$$

where $V = \mathbb{V}(h(X))$. Thus,

$$L_\epsilon(X) \geq H(X) - \sqrt{\frac{V}{\epsilon}} - \log \frac{1}{1 - \epsilon - \delta}.$$

Shannon's source coding theorem: Lower bound

Using Chebyshev's inequality:

$$\Pr\left(h(X) \geq \mathbb{E}[h(X)] - \sqrt{\frac{V}{\epsilon}}\right) \geq 1 - \epsilon,$$

where $V = \mathbb{V}(h(X))$. Thus,

$$L_\epsilon(X) \geq H(X) - \sqrt{\frac{V}{\epsilon}} - \log\frac{1}{1-\epsilon-\delta}.$$

[Shannon 1948] For $X^n = (X_1, ..., X_n)$ consisting of n i.i.d. samples,

$$h(X^n) = \sum_{i=1}^{n} h(X_i).$$

Thus, by the law of large numbers

$$L_\epsilon(X^n) \geq nH(X_1) - \sqrt{\frac{nV}{\epsilon}} - \log\frac{1}{1-\epsilon-\delta}.$$

Shannon's source coding theorem: Lower bound

Using Chebyshev's inequality:

$$\Pr\left(h(X) \geq \mathbb{E}[h(X)] - \sqrt{\frac{V}{\epsilon}}\right) \geq 1 - \epsilon,$$

where $V = \mathbb{V}(h(X))$. Thus,

$$L_\epsilon(X) \geq H(X) - \sqrt{\frac{V}{\epsilon}} - \log \frac{1}{1 - \epsilon - \delta}.$$

[Shannon 1948] For $X^n = (X_1, ..., X_n)$ consisting of n i.i.d. samples,

$$\lim_{n \to \infty} \frac{L_\epsilon(X^n)}{n} = H(X).$$

A Rényi entropy based lower bound

Consider the set

$$\mathcal{T}_\beta = \left\{ x \in \mathcal{X} : h(x) \geq H_\beta(X) - \frac{1}{\beta - 1} \log \frac{1}{\delta} \right\}, \quad 1 < \beta.$$

A Rényi entropy based lower bound

Consider the set

$$\mathcal{T}_\beta = \left\{ x \in \mathcal{X} : h(x) \geq H_\beta(X) - \frac{1}{\beta - 1} \log \frac{1}{\delta} \right\}, \quad 1 < \beta.$$

Then,

$$\begin{aligned}
1 &= \sum_{x \in \mathcal{T}_\beta} \mathrm{P}(x) + \sum_{x \in \mathcal{T}_\beta^c} \mathrm{P}(x) \\
&\leq \Pr\left(X \in \mathcal{T}_\beta\right) + \sum_{x \in \mathcal{T}_\beta^c} \mathrm{P}(x)^\beta \cdot \mathrm{P}(x)^{1-\beta} \\
&< \Pr\left(X \in \mathcal{T}_\beta\right) + \delta 2^{-(1-\beta) H_\beta(X)} \sum_{x \in \mathcal{T}_\beta^c} \mathrm{P}(x)^\beta \\
&< \Pr\left(X \in \mathcal{T}_\beta\right) + \delta.
\end{aligned}$$

A new source coding theorem

Theorem 1. For every $\epsilon \in (0, 1)$ and $0 \leq \alpha < 1$

$$L_\epsilon(X) \leq H_\alpha(X) + \frac{1}{1-\alpha} \log \frac{1}{\epsilon}.$$

Conversely, for every $\delta < 1 - \epsilon$ and $1 < \beta$

$$L_\epsilon(X) \geq H_\beta(X) - \frac{1}{\beta - 1} \log \frac{1}{1 - \epsilon - \delta}.$$

The strong converse property

For a sequence $X^{(n)}$, define

$$R_\epsilon^* = \limsup_n \frac{L_\epsilon(X^{(n)})}{n},$$

and

$$R^* = \lim_{\epsilon \to 0} R_\epsilon^*.$$

The sequence is said to satisfy the *strong converse property* if

$$R_\epsilon^* = R^*, \text{ for all } \epsilon \in (0,1).$$

The strong converse property

For a sequence $X^{(n)}$, define

$$R_\epsilon^* = \limsup_n \frac{L_\epsilon(X^{(n)})}{n},$$

and

$$R^* = \lim_{\epsilon \to 0} R_\epsilon^*.$$

The sequence is said to satisfy the *strong converse property* if

$$R_\epsilon^* = R^*, \text{ for all } \epsilon \in (0, 1).$$

We saw earlier that for $X^{(n)} = X^n$ i.i.d. P_X,

$$R_\epsilon^* = R^* = H(X).$$

Thus, i.i.d. source satisfies the strong converse property.

The strong converse property

For a sequence $X^{(n)}$, define

$$R^*_\epsilon = \limsup_n \frac{L_\epsilon(X^{(n)})}{n},$$

and

$$R^* = \lim_{\epsilon \to 0} R^*_\epsilon.$$

The sequence is said to satisfy the *strong converse property* if

$$R^*_\epsilon = R^*, \text{ for all } \epsilon \in (0,1).$$

We saw earlier that for $X^{(n)} = X^n$ i.i.d. P_X,

$$R^*_\epsilon = R^* = H(X).$$

Thus, i.i.d. source satisfies the strong converse property.

What other sources satisfy the strong converse property?

A sufficient condition for strong converse

Theorem 2. Consider a source sequence $X^{(n)}$ such that

$$\limsup_n \frac{1}{n}\mathbb{V}\Big(h\left(X^{(n)}\right)\Big) < \infty.$$

Then, $X^{(n)}$ satisfies the strong converse property.

Application: For X^n i.i.d. P_X,

$$\mathbb{V}(h\left(X^n\right)) = n\mathbb{V}(h(X_1)).$$

Thus, we recover the strong converse property for the i.i.d. case.

A sufficient condition for strong converse

Proof: Using our coding theorem, for every $0 \leq \alpha < 1$ and $1 < \beta$

$$\limsup_n \frac{1}{n} H_\alpha\left(X^{(n)}\right) \leq R^*_\epsilon \leq \limsup_n \frac{1}{n} H_\beta\left(X^{(n)}\right)$$

A sufficient condition for strong converse

Proof: Using our coding theorem, for every $0 \le \alpha < 1$ and $1 < \beta$

$$\limsup_n \frac{1}{n} H_\alpha\left(X^{(n)}\right) \le R_\epsilon^* \le \limsup_n \frac{1}{n} H_\beta\left(X^{(n)}\right)$$

Note that

$$\lim_{\alpha \uparrow 1} H_\alpha(X) = \lim_{\beta \downarrow 1} H_\beta(X) = H(X).$$

Our proof will be complete if we can exchange the limits in order and n.

A sufficient condition for strong converse

Proof: Using our coding theorem, for every $0 \leq \alpha < 1$ and $1 < \beta$

$$\limsup_n \frac{1}{n} H_\alpha\left(X^{(n)}\right) \leq R^*_\epsilon \leq \limsup_n \frac{1}{n} H_\beta\left(X^{(n)}\right)$$

Note that

$$\lim_{\alpha \uparrow 1} H_\alpha(X) = \lim_{\beta \downarrow 1} H_\beta(X) = H(X).$$

Our proof will be complete if we can exchange the limits in order and n.

1. $H_\alpha(X)$ is nonincreasing in α.
2. $\lim_{\alpha \to 1} \frac{d}{d\alpha} H_\alpha(X) = \mathbb{V}(h(X))$.

A sufficient condition for strong converse

Proof: Using our coding theorem, for every $0 \le \alpha < 1$ and $1 < \beta$

$$\limsup_n \frac{1}{n} H_\alpha\left(X^{(n)}\right) \le R^*_\epsilon \le \limsup_n \frac{1}{n} H_\beta\left(X^{(n)}\right)$$

Note that

$$\lim_{\alpha \uparrow 1} H_\alpha(X) = \lim_{\beta \downarrow 1} H_\beta(X) = H(X).$$

Our proof will be complete if we can exchange the limits in order and n.

1. $H_\alpha(X)$ is nonincreasing in α.
2. $\lim_{\alpha \to 1} \frac{d}{d\alpha} H_\alpha(X) = \mathbb{V}(h(X))$.

Thus, under the condition of the theorem,

the convergence of $H_\alpha(X^{(n)})$ to $H(X^{(n)})$ is uniform in n.

So, we can exchange the limits indeed.

A sufficient condition for strong converse

Theorem 2. Consider a source sequence $X^{(n)}$ such that

$$\limsup_n \frac{1}{n} \mathbb{V}\Big(h\left(X^{(n)}\right)\Big) < \infty.$$

Then, $X^{(n)}$ satisfies the strong converse property.

In fact,

$$R_\epsilon^* = \lim_{n \to \infty} \frac{H\left(X^{(n)}\right)}{n}, \quad \forall \epsilon \in (0,1).$$

Other Coding Theorems

What do the following coding problems have in common?

- Channel coding
- Optimal exponent for missed detection probability
- Lossy source coding
- Multiterminal lossless source coding

What do the following coding problems have in common?

- Channel coding
- Optimal exponent for missed detection probability
- Lossy source coding
- Multiterminal lossless source coding

The optimal lengths are determined by

large prob. bounds on $\log \frac{\mathrm{P}(X)}{\mathrm{Q}(X)}$ for appropriate P and Q and $X \sim \mathrm{P}$.

Rényi divergence and typical sets

Rényi divergence of order $0 \leq \alpha \neq 1$:

$$D_\alpha(\mathrm{P}, \mathrm{Q}) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \mathrm{P}(x)^\alpha \mathrm{Q}(x)^{1-\alpha}.$$

Rényi divergence and typical sets

Rényi divergence of order $0 \le \alpha \ne 1$:

$$D_\alpha(\mathrm{P}, \mathrm{Q}) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \mathrm{P}(x)^\alpha \mathrm{Q}(x)^{1-\alpha}.$$

For every $0 \le \alpha < 1$, the set

$$\mathcal{T}_\alpha = \left\{ x \in \mathcal{X} : \log \frac{\mathrm{P}(x)}{\mathrm{Q}(x)} \ge D_\alpha(\mathrm{P}, \mathrm{Q}) - \frac{1}{1-\alpha} \log \frac{1}{\epsilon} \right\}$$

satisfies $\Pr(X \in \mathcal{T}_\alpha) \ge 1 - \epsilon$.

Rényi divergence and typical sets

Rényi divergence of order $0 \leq \alpha \neq 1$:

$$D_\alpha(\mathrm{P}, \mathrm{Q}) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \mathrm{P}(x)^\alpha \mathrm{Q}(x)^{1-\alpha}.$$

For every $0 \leq \alpha < 1$, the set

$$\mathcal{T}_\alpha = \left\{ x \in \mathcal{X} : \log \frac{\mathrm{P}(x)}{\mathrm{Q}(x)} \geq D_\alpha(\mathrm{P}, \mathrm{Q}) - \frac{1}{1-\alpha} \log \frac{1}{\epsilon} \right\}$$

satisfies $\Pr\left(X \in \mathcal{T}_\alpha\right) \geq 1 - \epsilon$.

For every $1 < \beta$, the set

$$\mathcal{T}_\beta = \left\{ x \in \mathcal{X} : \log \frac{\mathrm{P}(x)}{\mathrm{Q}(x)} \leq D_\beta(\mathrm{P}, \mathrm{Q}) + \frac{1}{\beta - 1} \log \frac{1}{\epsilon} \right\}$$

satisfies $\Pr\left(X \in \mathcal{T}_\beta\right) \geq 1 - \epsilon$.

What could we do with our Rényi typical sets?

Could obtain asymptotically tight single-shot bounds for

- length of channel codes
- exponent of missed detection error in binary hypothesis testing
- length of multiterminal lossless source codes
- length of lossy source codes

For all but the last problem,

we could recover the strong converse property for the i.i.d. case.

What would we like to do with our Rényi typical sets?

- A sufficient condition for strong converse to hold for channel coding;
- A suff. condition for strong converse to hold for lossy source coding;
- Studying the implications for source and channels with memory;
- Application to multiterminal coding theorems
  - similar to the program started by Oohama.