Optimum Source Resolvability Rate with Respect to f-Divergences Using the Smooth Rényi Entropy

Ryo Nomura (Waseda University, Japan)
Hideki Yagi (The Universtiy of Electro-Communications, Japan)

ISIT2020, June 21-26, 2020

Waseda University
Center for Data Science

Outline

- Preliminaries
  1. Source resolvability
  2. f-divergence
  3. Smooth Rényi entropy
- Main results (Optimum D-achievable resolvability rate)
- Specifications
- Conclusion

Preliminaries

1. Preliminaries

We focus on the source resolvability problem.

We focus on the source resolvability problem.

Given an arbitrary target source, we approximate it by using a discrete random variable which is uniformly distributed.

$$\textbf{Mapping } \phi : \mathcal{U}_M := \{1, 2, \ldots, M\} \to \mathcal{X}$$

Uniform Random Number

U_M:

i	$P(i)$
1	$1/M$
2	$1/M$
3	$1/M$
$\ldots$	$\ldots$
M	$1/M$

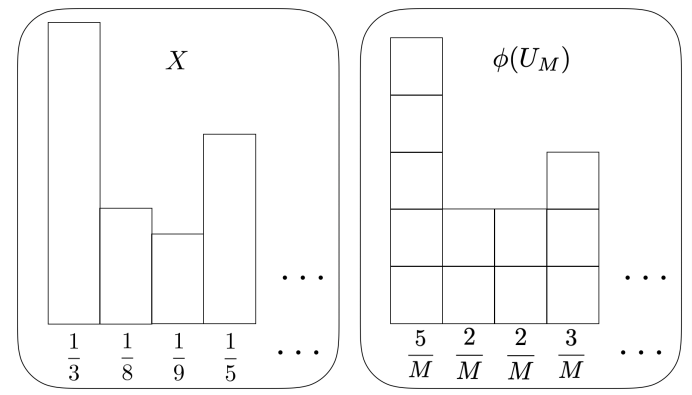
$\xrightarrow{\phi}$

Given Target Source

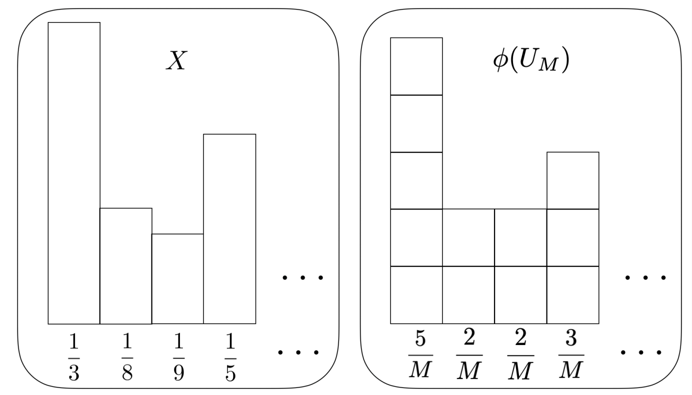
X:

x	$P(x)$
1	$1/3$
2	$1/8$
3	$1/9$
$\ldots$	$\ldots$

1. Preliminaries

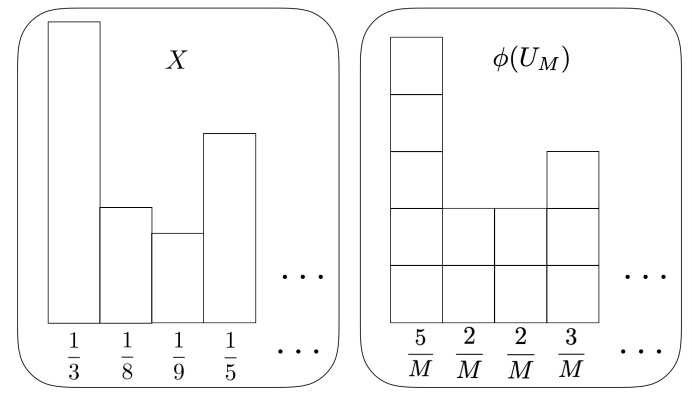


1. Preliminaries



- Distance $d(\phi(U_M), X)$: requested to be small
- Size M (rate: $\log M$): requested to be small

1. Preliminaries



- Distance $d(\phi(U_M), X)$: requested to be small ($\leq D$)
- Size M (rate: $\log M$): requested to be as small as possible

1. Preliminaries

Table 1: Previous results

	Approximation measure (or distance)		
	Variational distance	KL-divergence	f-divergence
Information Spectrum	Han & Verdú, '93 Steinberg & Verdú, '96 Yagi & Han, '17	Steinberg & Verdú, '96 Nomura, '18	Nomura, ISIT'19
Rényi Entropy	Uyematsu, '10		

1. Preliminaries

Table 2: Previous results

	Approximation measure (or distance)		
	Variational distance	KL-divergence	f-divergence
Information Spectrum	Han & Verdú, ’93 Steinberg & Verdú, ’96 Yagi & Han, ’17	Steinberg & Verdú, ’96 Nomura, ’18	Nomura, ISIT’19
Rényi Entropy	Uyematsu, ’10		This study

1. Preliminaries

Notation

- $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$: general source with values in countable sets $\mathcal{X}^n$.
- P_{X^n}: probability distribution of X^n
- U_M: random variable uniformly distributed on $\mathcal{U}_M := \{1, 2, \cdots, M\}$,

$$P_{U_M}(i) = \frac{1}{M}, \qquad 1 \leq i \leq M$$

1. Preliminaries

Notation

- $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$: general source with values in countable sets $\mathcal{X}^n$.
- P_{X^n}: probability distribution of X^n
- U_M: random variable uniformly distributed on $\mathcal{U}_M := \{1, 2, \cdots, M\}$,

$$P_{U_M}(i) = \frac{1}{M}, \qquad 1 \le i \le M$$

Assumption:

$$\underline{H}(\mathbf{X}) := \sup\left\{ R \,\middle|\, \lim_{n\to\infty} \Pr\left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(X^n)} \ge R \right\} = 1 \right\} < +\infty$$

1. Preliminaries

We define a class of f-divergences between P_Z and $P_{\overline{Z}}$.

1. Preliminaries

We define a class of f-divergences between P_Z and $P_{\overline{Z}}$.

Let $f(t)$ be a convex function defined for $t > 0$ and $f(1) = 0$.

We define a class of f-divergences between P_Z and $P_{\overline{Z}}$.

Let $f(t)$ be a convex function defined for $t > 0$ and $f(1) = 0$.

Definition ([Csiszár and Shields, '04])

Let P_Z and $P_{\overline{Z}}$ denote probability distributions over a finite set $\mathcal{Z}$. The f-divergence between P_Z and $P_{\overline{Z}}$ is defined by

$$D_f(Z||\overline{Z}) := \sum_{z\in\mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right),$$

where we set $0f\left(\frac{0}{0}\right) = 0$, $f(0) = \lim_{t\to 0} f(t)$, $0f\left(\frac{a}{0}\right) = a \lim_{u\to\infty} \frac{f(u)}{u}$. $\square$

We define a class of f-divergences between P_Z and $P_{\overline{Z}}$.

Let $f(t)$ be a convex function defined for $t > 0$ and $f(1) = 0$.

Definition ([Csiszár and Shields, '04])

Let P_Z and $P_{\overline{Z}}$ denote probability distributions over a finite set $\mathcal{Z}$. The f-divergence between P_Z and $P_{\overline{Z}}$ is defined by

$$D_f(Z||\overline{Z}) := \sum_{z \in \mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right),$$

where we set $0f\left(\frac{0}{0}\right) = 0$, $f(0) = \lim_{t \to 0} f(t)$, $0f\left(\frac{a}{0}\right) = a \lim_{u \to \infty} \frac{f(u)}{u}$. □

We next give some examples of f-divergences.

1. Preliminaries

Definition

$$D_f(Z||\overline{Z}) := \sum_{z\in\mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right)$$

Examples [Csiszár and Shields, '04][Sason and Verdú, '16]

- $f(t) = t\log t$: (KL divergence)

$$D_f(Z||\overline{Z}) = \sum_{z\in\mathcal{Z}} P_Z(z)\log\frac{P_Z(z)}{P_{\overline{Z}}(z)} =: D(Z||\overline{Z}).$$

- $f(t) = -\log t$: (Reverse KL divergence)

$$D_f(Z||\overline{Z}) = \sum_{z\in\mathcal{Z}} P_{\overline{Z}}(z)\log\frac{P_{\overline{Z}}(z)}{P_Z(z)} = D(\overline{Z}||Z).$$

Definition

$$D_f(Z\|\overline{Z}) := \sum_{z\in\mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right).$$

Examples [Csiszár and Shields, ’04][Sason and Verdú, ’16]

- $f(t) = 1 - \sqrt{t}$: (Hellinger distance)

$$D_f(Z\|\overline{Z}) = 1 - \sum_{z\in\mathcal{Z}} \sqrt{P_Z(z)P_{\overline{Z}}(z)}.$$

- $f(t) = (t-1)^+ = (1-t)^+ := \max\{1-t, 0\}$: (Variational distance)

$$D_f(Z\|\overline{Z}) = \frac{1}{2}\sum_{z\in\mathcal{Z}} |(P_Z(z) - P_{\overline{Z}}(z))|.$$

Definition

$$D_f(Z||\overline{Z}) := \sum_{z \in \mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right).$$

Examples [Csiszár and Shields, '04][Sason and Verdú, '16]

- $f(t) = (t-\gamma)^+$: (E_γ-divergence) For any given $\gamma \geq 1$,

$$D_f(Z||\overline{Z}) = \sum_{z \in \mathcal{Z}: P_Z(z) > \gamma P_{\overline{Z}}(z)} (P_Z(z) - \gamma P_{\overline{Z}}(z)) =: E_\gamma(Z||\overline{Z}).$$

- It is not difficult to check that $f(t) = (\gamma - t)^+ + 1 - \gamma$ leads E_γ-divergence.
- $\gamma = 1$: variational distance

Definition

$$D_f(Z\|\overline{Z}) := \sum_{z\in\mathcal{Z}} P_{\overline{Z}}(z) f\left(\frac{P_Z(z)}{P_{\overline{Z}}(z)}\right).$$

Examples [Csiszár and Shields, '04][Sason and Verdú, '16]

- $f(t) = (t-\gamma)^+$: (E_γ-divergence) For any given $\gamma \geq 1$,

$$D_f(Z\|\overline{Z}) = \sum_{z\in\mathcal{Z}:P_Z(z)>\gamma P_{\overline{Z}}(z)} (P_Z(z) - \gamma P_{\overline{Z}}(z)) =: E_\gamma(Z\|\overline{Z}).$$

- It is not difficult to check that $f(t) = (\gamma - t)^+ + 1 - \gamma$ leads E_γ-divergence.
- $\gamma = 1$: variational distance

1. Preliminaries

In this study, we assume the following conditions on the function f.

C1) The function $f(t)$ is a strictly decreasing function of t for $t \in (0, 1]$ and a decreasing function for $t \geq 1$.

1. Preliminaries

In this study, we assume the following conditions on the function f.

C1) The function $f(t)$ is a strictly decreasing function of t for $t \in (0, 1]$ and a decreasing function for $t \geq 1$.

C2) For any pair of positive numbers (a, b), it holds that

$$\lim_{n\to\infty} \frac{f\left(e^{-nb}\right)}{e^{na}} = 0.$$

1. Preliminaries

In this study, we assume the following conditions on the function f.

C1) The function $f(t)$ is a strictly decreasing function of t for $t \in (0, 1]$ and a decreasing function for $t \geq 1$.

C2) For any pair of positive numbers (a, b), it holds that

$$\lim_{n \to \infty} \frac{f\left(e^{-nb}\right)}{e^{na}} = 0.$$

C3) For any number $a \in [0, 1]$, it holds that

$$0 f\left(\frac{a}{0}\right) = 0.$$

1. Preliminaries

In this study, we assume the following conditions on the function f.

- **C1)** The function $f(t)$ is a strictly decreasing function of t for $t \in (0, 1]$ and a decreasing function for $t \geq 1$.
- **C2)** For any pair of positive numbers (a, b), it holds that

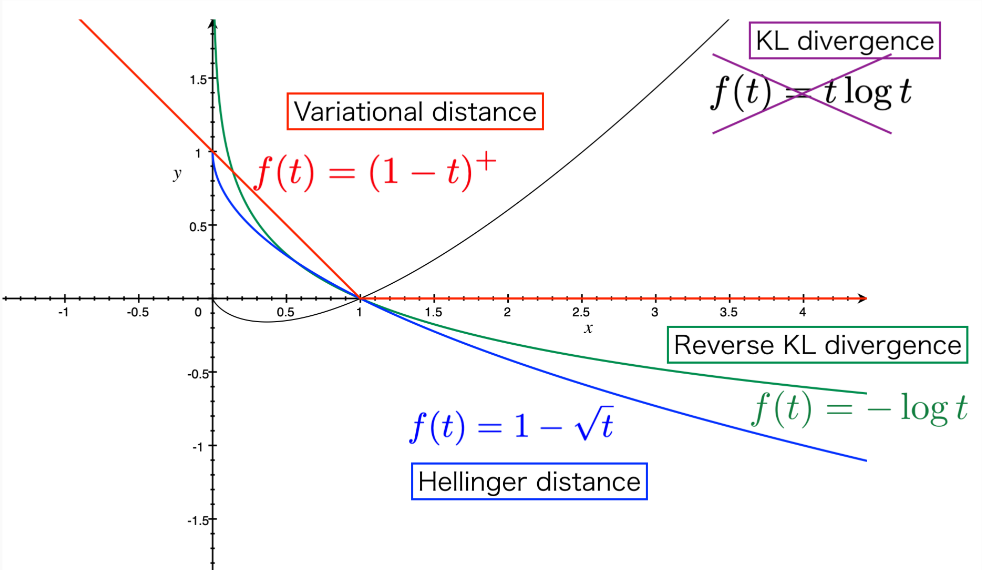
$$\lim_{n \to \infty} \frac{f\left(e^{-nb}\right)}{e^{na}} = 0.$$

- **C3)** For any number $a \in [0, 1]$, it holds that

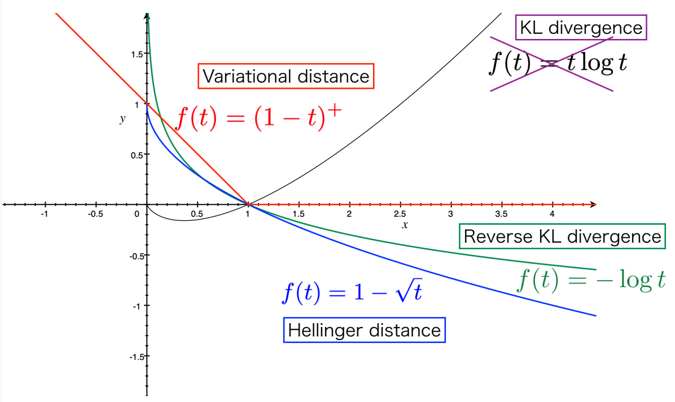
$$0 f\left(\frac{a}{0}\right) = 0.$$

Note
We only consider the f-divergence with the function f satisfying C1)–C3).

1. Preliminaries

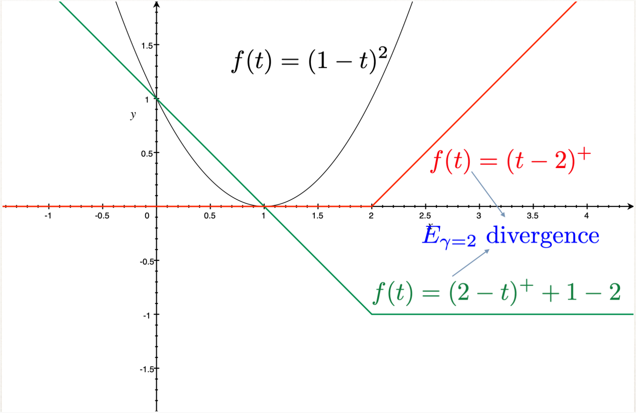


2. Main result



Note

$f(t) = -\log t$, $f(t) = 1 - \sqrt{t}$, and $f(t) = (1-t)^+$ satisfy three conditions, while $f(t) = t \log t$ does not satisfy C1).



Note

$f(t) = (\gamma - t)^+ + 1 - \gamma$ satisfy three conditions,

while $f(t) = (t - \gamma)^+$ does not satisfy C1).

Key property

$$\sum_{z \in \mathcal{Z}'} b(z) f\left(\frac{a(z)}{b(z)}\right) \geq \left(\sum_{z \in \mathcal{Z}'} b(z)\right) f\left(\frac{\sum_{z \in \mathcal{Z}'} a(z)}{\sum_{z \in \mathcal{Z}'} b(z)}\right).$$

Together with the fact that $f(1) = 0$, we have $D_f(Z||\overline{Z}) \geq 0$.

- This is an analogue of the log-sum inequality in the KL divergence.

Definition: Smooth Rényi entropy of order α ([Renner and Wolf, '04])

$$H_\alpha(\delta|X^n) := \frac{1}{1-\alpha} \inf_{P_{\overline{X}^n} \in B^\delta(P_{X^n})} \log\left(\sum_{\mathbf{x} \in \mathcal{X}^n} P_{\overline{X}^n}(\mathbf{x})^\alpha \right),$$

where

$$B^\delta(P_{X^n}) := \left\{ P_{\overline{X}^n} \in \mathcal{P}^n \,\middle|\, \frac{1}{2} \sum_{\mathbf{x} \in \mathcal{X}^n} |P_{X^n}(\mathbf{x}) - P_{\overline{X}^n}(\mathbf{x})| \leq \delta \right\}.$$

1. Preliminaries

Definition: Smooth Rényi entropy of order α ([Renner and Wolf, '04])

$$H_\alpha(\delta|X^n) := \frac{1}{1-\alpha} \inf_{P_{\overline{X}^n} \in B^\delta(P_{X^n})} \log \left(\sum_{\mathbf{x} \in \mathcal{X}^n} P_{\overline{X}^n}(\mathbf{x})^\alpha \right),$$

where

$$B^\delta(P_{X^n}) := \left\{ P_{\overline{X}^n} \in \mathcal{P}^n \,\middle|\, \frac{1}{2} \sum_{\mathbf{x} \in \mathcal{X}^n} |P_{X^n}(\mathbf{x}) - P_{\overline{X}^n}(\mathbf{x})| \le \delta \right\}.$$

Theorem ([Uyematsu, '10])

$$H_0(\delta|X^n) = \min_{\substack{A_n \subset \mathcal{X}^n \\ \Pr\{X^n \in A_n\} \ge 1-\delta}} \log |A_n|.$$

Main results

2. Main Result

Theorem (Direct Theorem)

Under conditions C1)–C3), for any $\gamma > 0$ and any M_n satisfying

$$\frac{1}{n}\log M_n \geq \frac{1}{n}H_0(1 - f^{-1}(D)|X^n) + \gamma,$$

there exists a mapping ϕ_n which satisfies

$$D_f(X^n||\phi_n(U_{M_n})) \leq D + \gamma$$

for sufficiently large n.

$H_0(\delta|X^n)$: smooth Rényi entropy of order 0

2. Main result

Theorem (Converse Theorem)

Under conditions C1) and C2), for any mapping ϕ_n satisfying

$$D_f(X^n||\phi_n(U_{M_n})) \leq D,$$

it holds that

$$\frac{1}{n}\log M_n \geq \frac{1}{n}H_0(1 - f^{-1}(D)|X^n).$$

2. Main result

The optimum D-achievable rate with respect to a class of f-divergences:

2. Main result

The optimum D-achievable rate with respect to a class of f-divergences:

Definition

R is said to be D-achievable with the given f-divergence if there exists a mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that

$$\limsup_{n\to\infty} D_f(X^n||\phi_n(U_{M_n})) \leq D, \quad \limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R.$$

2. Main result

The optimum D-achievable rate with respect to a class of f-divergences:

Definition

R is said to be D-achievable with the given f-divergence if there exists a mapping $\phi_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ such that

$$\limsup_{n\to\infty} D_f(X^n||\phi_n(U_{M_n})) \leq D, \quad \limsup_{n\to\infty} \frac{1}{n}\log M_n \leq R.$$

We want here to make R as small as possible.

Definition (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) := \inf\left\{R \,|\, R \text{ is } D\text{-achievable with the given } f\text{-divergence}\right\}.$$

2. Main result

Then, we have the main theorem:

Theorem

Under conditions C1)–C3), it holds that

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H_0(1 - f^{-1}(D + \nu)|X^n)$$
$$= \lim_{\nu \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H_0(1 - f^{-1}(D) + \nu|X^n).$$

Proof: Omitted. □

3. Specifications

3. Specifications

We apply our general formula to three cases

$f(t) = (1-t)^+$ → Variational Distance

$$D_f(Z||\overline{Z}) = \frac{1}{2}\sum_{z\in\mathcal{Z}} |(P_{\overline{Z}}(z) - P_Z(z))|$$

3. Specifications

We apply our general formula to three cases

$f(t) = (1-t)^+$ → Variational Distance

$$D_f(Z||\overline{Z}) = \frac{1}{2}\sum_{z\in\mathcal{Z}} |(P_{\overline{Z}}(z) - P_Z(z))|$$

$f(t) = -\log t$ → Reverse KL Divergence

$$D_f(X^n||\phi_n(U_{M_n})) = \sum_{\mathbf{x}\in\mathcal{X}^n} P_{\phi_n(U_{M_n})}(\mathbf{x}) \log \frac{P_{\phi_n(U_{M_n})}(\mathbf{x})}{P_{X^n}(\mathbf{x})}$$

3. Specifications

We apply our general formula to three cases

$f(t) = (1-t)^+$ → Variational Distance

$$D_f(Z||\overline{Z}) = \frac{1}{2}\sum_{z\in\mathcal{Z}} |(P_{\overline{Z}}(z) - P_Z(z))|$$

$f(t) = -\log t$ → Reverse KL Divergence

$$D_f(X^n||\phi_n(U_{M_n}) = \sum_{\mathbf{x}\in\mathcal{X}^n} P_{\phi_n(U_{M_n})}(\mathbf{x}) \log \frac{P_{\phi_n(U_{M_n})}(\mathbf{x})}{P_{X^n}(\mathbf{x})}$$

$f(t) = (\gamma - t)^+ + 1 - \gamma$ → E_γ-divergence

$$D_f(Z||\overline{Z}) = \sum_{z\in\mathcal{Z}:P_Z(z)>\gamma P_{\overline{Z}}(z)} (P_Z(z) - \gamma P_{\overline{Z}}(z))$$

3. Specifications

Theorem (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H_0(1 - f^{-1}(D + \nu)|X^n).$$

Variational distance: $f(t) = (1-t)^+ \iff f^{-1}(D) = 1 - D \quad (0 \leq D \leq 1)$

Theorem (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n\to\infty} \frac{1}{n} H_0(1 - f^{-1}(D+\nu)|X^n).$$

Variational distance: $f(t) = (1-t)^+ \iff f^{-1}(D) = 1 - D \quad (0 \le D \le 1)$

Corollary

For $f(t) = (1-t)^+$ it holds that

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \lim_{n\to\infty} \frac{1}{n} H_0(D+\nu|X^n).$$

3. Specifications

Theorem (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H_0(1 - f^{-1}(D+\nu)|X^n).$$

Variational distance: $f(t) = (1-t)^+ \iff f^{-1}(D) = 1 - D \quad (0 \le D \le 1)$

Corollary

For $f(t) = (1-t)^+$ it holds that

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \lim_{n \to \infty} \frac{1}{n} H_0(D + \nu|X^n).$$

This result coincides with the result given by [Uyematsu, '10].

Theorem (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H_0(1 - f^{-1}(D+\nu)|X^n).$$

Reverse KL divergence: $f(t) = -\log t \iff f^{-1}(D) = e^{-D}$

Theorem (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H_0(1 - f^{-1}(D + \nu)|X^n).$$

Reverse KL divergence: $f(t) = -\log t \iff f^{-1}(D) = e^{-D}$

Corollary

For $f(t) = -\log t$, it holds that

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \lim_{n \to \infty} \frac{1}{n} H_0(1 - e^{-(D+\nu)}|X^n)$$

Theorem (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n\to\infty} \frac{1}{n} H_0(1 - f^{-1}(D+\nu)|X^n)$$

E_γ-divergence: $f(t) = (\gamma - t)^+ + 1 - \gamma \iff f^{-1}(D) = 1 - D \qquad (\gamma \geq 1)$

3. Specifications

Theorem (Optimum D-achievable rate)

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} H_0(1 - f^{-1}(D+\nu)|X^n)$$

E_γ-divergence: $f(t) = (\gamma - t)^+ + 1 - \gamma \iff f^{-1}(D) = 1 - D \qquad (\gamma \geq 1)$

Noticing that $\gamma \geq 1$, substituting $f(t) = (\gamma - t)^+ + 1 - \gamma$ yields

$$\left(\gamma - \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} \leq R\right\}\right)^+ + 1 - \gamma = \Pr\left\{\frac{1}{n}\log\frac{1}{P_{X^n}(X^n)} > R\right\}$$

3. Specifications

Thus, we have:

Corollary

For $f(t) = (\gamma - t)^+ + 1 - \gamma$, it holds that

$$S_f(D|\mathbf{X}) = \lim_{\nu \downarrow 0} \lim_{n \to \infty} \frac{1}{n} H_0(D + \nu | X^n)$$

Remark

$S_f(D|\mathbf{X})$ with the E_γ-divergence does not depend on γ, which shows that it coincides with $S_f(D|\mathbf{X})$ with the variational distance.

4. Conclusion

We have considered the source resolvability problem with respect to f-divergences.

Contributions

- We have characterized the general formula of the optimum D-achievable rate by using the smooth Rényi entropy including the function f.
- It is easy to derive the optimum D-achievable rate with respect to the specified function f.

Thank you!

科研費
KAKENHI