Simulation under Rényi Divergences

Lei Yu

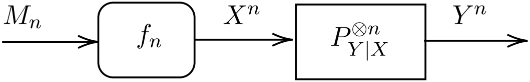
Nankai University

Presented at the University of Hong Kong

Outline

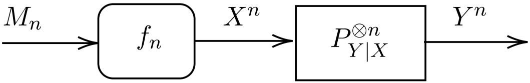
Outline

Simulate a Source through a Channel



- How much information is needed to simulate a source through a given channel?
- Called the channel resolvability problem.
- It was studied by Han-Verdú in 1993, but a similar simulation problem was first studied by Wyner in 1975.

Channel Resolvability



- M_n is uniformly distributed over $[\mathrm{e}^{nR}] := \{1, \ldots, \mathrm{e}^{nR}\}$.
- An (n, R) code $f_n : [\mathrm{e}^{nR}] \to \mathcal{X}^n$.
- The output distribution:

$$Q_{Y^n}(y^n) := \mathrm{e}^{-nR} \sum_{m \in [\mathrm{e}^{nR}]} P_{Y|X}^{\otimes n}(y^n | f_n(m)). \tag{1}$$

- We want Q_{Y^n} approximates a target $P_Y^{\otimes n}$.

Relative Entropy and Rényi Divergence

- The Rényi divergence of order $q \geq 0$ is

$$D_q(Q\|P) := \frac{1}{q-1} \log \sum_{x\in\mathcal{X}} Q(x)^q P(x)^{1-q}.$$

- The relative entropy (Kullback-Leibler divergence) is a special case:

$$\lim_{q\to 1} D_q(Q\|P) = D(Q\|P) := \sum_{x\in\mathcal{X}} Q(x) \log \frac{Q(x)}{P(x)}.$$

- The conditional version:

$$D_q(Q_{Y|X}\|P_{Y|X}|Q_X) := D_q(Q_X Q_{Y|X}\|Q_X P_{Y|X}).$$

- Rényi divergence is nondecreasing in its order.

Rényi Resolvability

- We minimize $D_q(Q_{Y^n} \| P_Y^{\otimes n})$ over all (n, R) codes f_n.
- The q-Rényi resolvability rate is defined as

$$\mathrm{R}_q := \inf\{R : D_q(Q_{Y^n} \| P_Y^{\otimes n}) \to 0\}.$$

What is R_q?

Existing Results

- Define $\mathcal{P}(P_{Y|X}, P_Y) := \left\{P_X : \sum_x P_{Y|X}(\cdot|x)P_X(x) = P_Y\right\}$.

Theorem ([Han-Verdú'93, Hayashi'06,'11])

For $q = 1$ (relative entropy),

$$\mathrm{R}_1 = \min_{P_X \in \mathcal{P}(P_{Y|X}, P_Y)} I(X;Y).$$

- Converse: standard IT techniques.
- Achievability: soft-covering lemma [Wyner'75].

Soft-covering Lemma

- Let $\mathcal{C} = \{X^n(m)\}_{m \in [e^{nR}]}$ with $X^n(m) \sim Q_X^{\otimes n}, m \in [e^{nR}]$ drawn independently.
- Randomly and uniformly choose one codeword from $\mathcal{C}$.
- The output distribution:

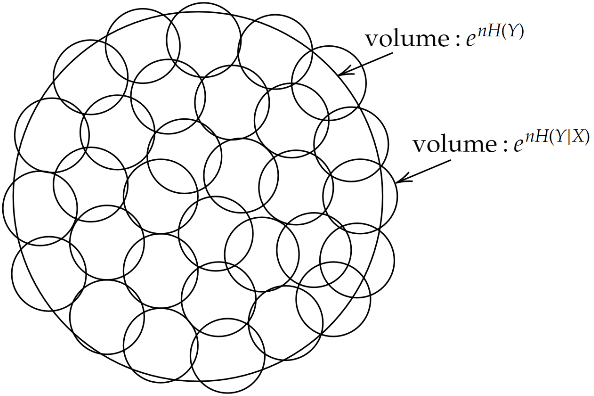
$$Q_{Y^n|\mathcal{C}}(y^n|\mathcal{C}) := e^{-nR} \sum_{m \in [e^{nR}]} P_{Y|X}^{\otimes n}(y^n|X^n(m)). \tag{2}$$

Lemma ([Wyner'75])

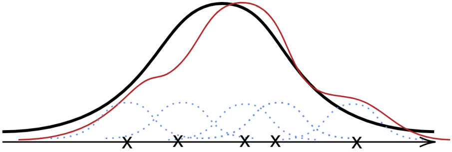
If $R > I(X;Y)$, then

$$D(Q_{Y^n|\mathcal{C}} \| P_Y^{\otimes n} | Q_{\mathcal{C}}) \to 0.$$

Covering



Soft-covering



Outline

Rényi Resolvability Rates

Theorem

For $q \in [0, \infty]$, we have

$$\mathrm{R}_q = \Gamma_q(P_{Y|X}, P_Y), \tag{3}$$

where

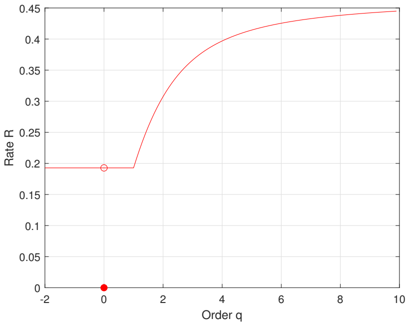
$$\Gamma_q(P_{Y|X}, P_Y) := \begin{cases} \min_{P_X \in \mathcal{P}(P_{Y|X}, P_Y)} \mathbb{E}_{P_X}[D_q(P_{Y|X} \| P_Y)], & q \in (1, \infty] \\ \min_{P_X \in \mathcal{P}(P_{Y|X}, P_Y)} I(X;Y), & q \in (0, 1] \\ 0, & q = 0. \end{cases}.$$

Binary Example

- For $P_{Y|X} = \mathrm{BSC}(\epsilon)$ and $P_Y = \mathrm{Bern}\,(1/2)$, it holds that

$$\Gamma_q(P_{Y|X}, P_Y) = \begin{cases} 1 - H_q(\epsilon), & q \in (1, \infty] \\ 1 - H(\epsilon), & q = (0, 1] \\ 0, & q = 0 \end{cases},$$

where H_q is the q-Rényi entropy.



Typical Code is Optimal

- Let $\mathcal{C} = \{X^n(m)\}_{m \in [e^{nR}]}$ with $X^n(m) \sim P_X^{\otimes n}(\cdot|\mathcal{T}_\epsilon^{(n)}(P_X)), m \in [e^{nR}]$ drawn independently, where $\mathcal{T}_\epsilon^{(n)}(P_X)$ is the ϵ-typical set.
- Randomly and uniformly choose one codeword from $\mathcal{C}$.
- The output distribution:

$$Q_{Y^n|\mathcal{C}}(y^n|\mathcal{C}) := e^{-nR} \sum_{m \in [e^{nR}]} P_{Y|X}^{\otimes n}(y^n|X^n(m)). \tag{4}$$

Lemma (Rényi-Covering Lemma)

If

$$R > \begin{cases} \mathbb{E}_{P_X}[D_q(P_{Y|X} \| P_Y)], & q \in (1, \infty] \\ I(X;Y), & q \in (0, 1] \\ 0, & q = 0. \end{cases},$$

then

$$D_q(Q_{Y^n|\mathcal{C}} \| P_Y^{\otimes n} | Q_{\mathcal{C}}) \to 0.$$

Intuition

$$D_q(Q_{Y^n} \| P_Y^{\otimes n}) = \frac{1}{q-1} \log\Bigg(\sum_{y^n \in \mathcal{T}_\epsilon^{(n)}(Q_Y)} Q(y^n)^q P(y^n)^{1-q} + \sum_{y^n \notin \mathcal{T}_\epsilon^{(n)}(Q_Y)} Q(y^n)^q P(y^n)^{1-q} \Bigg).$$

- To reduce $D_q(Q_{Y^n} \| P_Y^{\otimes n})$, the tail part should be as small as possible——truncation to typical sets!

What if $R > \Gamma_q(P_{Y|X}, P_Y)$?

Theorem (Exponential Behavior)

Given $q \in (0, \infty)$, if $R > \Gamma_q(P_{Y|X}, P_Y)$, then there exists a sequence of typical codes such that $D_q(Q_{Y^n} \| P_Y^{\otimes n})$ decays at least exponentially fast.

- Characterize exact exponent?—Difficult!

Theorem (Exponential Behavior of i.i.d. Codes)

For the i.i.d. code, if the rate $R > D_q(P_{Y|X} \| P_Y | P_X)$, then we have

$$\lim_{n\to\infty} -\frac{1}{n} \log D_q(Q_{Y^n|\mathcal{C}_n} \| P_Y^{\otimes n} | P_{\mathcal{C}_n}) = \begin{cases} \min\{\gamma(2), \gamma(q)\}, & q \in (2, \infty) \\ \max_{t\in[q,2]} \gamma(t), & q \in (1, 2] \\ \max_{t\in[1,2]} \gamma(t), & q \in (0, 1] \end{cases}, \tag{5}$$

where $\gamma(q) := (q-1)(R - D_q(P_{Y|X} \| P_Y | P_X))$.

What if $R < \Gamma_q(P_{Y|X}, P_Y)$?

Theorem (Linear Behavior)

1. *For* $q \in [1, \infty]$,

$$\lim_{n\to\infty} \frac{1}{n} \inf_{f:[e^{nR}]\to\mathcal{X}^n} D_q(Q_{Y^n} \| P_Y^{\otimes n})$$
$$= \min_{Q_X} \max\{\mathbb{E}_{Q_X}[D_q(P_{Y|X} \| P_Y)] - R,$$
$$\max_{Q_{Y|X}} -q' D(Q_{Y|X} \| P_{Y|X} | Q_X) + D(Q_Y \| P_Y)\}. \quad (6)$$

2. *For* $q \in (0, 1)$,

$$\lim_{n\to\infty} \frac{1}{n} \inf_{f:[e^{nR}]\to\mathcal{X}^n} D_q(Q_{Y^n} \| P_Y^{\otimes n})$$
$$= \min_{Q_{XY}} \max\{-q' D(Q_{Y|X} \| P_{Y|X} | Q_X) + D(Q_{Y|X} \| P_Y | Q_X) - R,$$
$$- q' D(Q_{Y|X} \| P_{Y|X} | Q_X) + D(Q_Y \| P_Y)\}. \quad (7)$$

Here, $q' := \frac{q}{q-1}$ is the Holder conjugate of q.

Key Lemma

Constant composition codes: Let $\mathcal{C} = \{X^n(m)\}_{m\in[\mathrm{e}^{nR}]}$ with $X^n(m) \sim \mathrm{Unif}(\mathcal{T}(T_X)), m \in [\mathrm{e}^{nR}]$ drawn independently, where $\mathcal{T}(T_X)$ is the type class w.r.t. T_X.

Lemma (Strong Packing-Covering Lemma)

Given T_X, $R > 0$, and any $\epsilon > 0$, with high probability it holds that

$$|\mathcal{T}_{T_{X|Y}}(y^n) \cap \mathcal{C}| \in \frac{e^{nR}|\mathcal{T}_{T_{XY}}|}{|\mathcal{T}_{T_X}||\mathcal{T}_{T_Y}|}(1 \pm e^{-n\epsilon/3}) = e^{n(R-I_T(X;Y)+o(1))}$$

for all $T_{Y|X}$ s.t. $I_T(X;Y) \le R - \epsilon$, and

$$0 \le |\mathcal{T}_{T_{X|Y}}(y^n) \cap \mathcal{C}| \le e^{n\epsilon}$$

for all $T_{Y|X}$ s.t. $I_T(X;Y) > R - \epsilon$.

Outline

Hypercontractivity

- Let $(X,Y) \sim P_{XY} = \mathrm{DSBS}(\epsilon)$, i.e., $X \sim \mathrm{Bern}(1/2)$ and $Y|X \sim \mathrm{BSC}(\epsilon)$.
- Let (X^n, Y^n) be n i.i.d. copies of (X,Y).
- Denote $P_{X|Y}^{\otimes n}(f)(y^n) = \mathbb{E}[f(X^n)|Y^n = y^n]$.

Theorem ([Bonami '70][Beckner '75])

For $q \geq 1$ *and* $p \geq 1 + (1-2\epsilon)^2(q-1)$,

$$\|P_{X|Y}^{\otimes n}(f)\|_q \leq \|f\|_p, \forall f \geq 0, \tag{8}$$

and for $q \leq 1$ *and* $p \leq 1 + (1-2\epsilon)^2(q-1)$,

$$\|P_{X|Y}^{\otimes n}(f)\|_q \geq \|f\|_p, \forall f \geq 0. \tag{9}$$

Anti-contractivity

Theorem ([Y. 2024])

For $q \geq p \geq 1$,

$$\|P_{X|Y}^{\otimes n}(f)\|_q \geq e^{-nH_q(\epsilon)/p'}\|f\|_p, \forall f \geq 0, \tag{10}$$

and for $0 \leq p \leq 1$ and $q \leq p$,

$$\|P_{X|Y}^{\otimes n}(f)\|_q \leq e^{-nH_p(\epsilon)/p'}\|f\|_p, \forall f \geq 0, \tag{11}$$

where $p' := \frac{p}{p-1}$ is the Holder conjugate of p. The exponents above cannot be further improved.

- The special case of (10) with $p = q$ was first proven by Samorodnitsky 2022.

Key Point

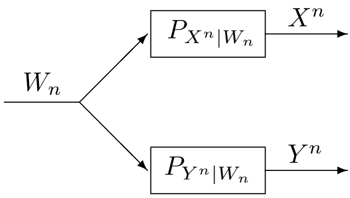
- If we write $f = \frac{\mathrm{d}Q_X}{\mathrm{d}P_X}$, then

$$\log \|f\|_p = \frac{1}{p'} D_p(Q_X \| P_X),$$
$$\log \|P_{X|Y}(f)\|_q = \frac{1}{q'} D_q(Q_Y \| P_Y),$$

  where $Q_Y := Q_X \circ P_{Y|X}$.
- Connecting Rényi resolvability and anti-contractivity.

Outline

Wyner's Common Information



- M_n is uniformly distributed over $\mathcal{M}_n = [e^{nR}]$
- An (n, R)-synthesis code consists of $(Q_{X^n|M_n}, Q_{Y^n|M_n})$.
- The induced distribution is

$$Q_{X^nY^n}(x^n, y^n) := \frac{1}{|\mathcal{M}_n|} \sum_{m \in \mathcal{M}_n} Q_{X^n|M_n}(x^n|m) Q_{Y^n|M_n}(y^n|m)$$

- Goal:

$$Q_{X^nY^n} \approx P_{XY}^{\otimes n} \quad \text{(target distribution)}$$

Wyner's Common Information

- In 1975, Wyner used normalized relative entropy $\frac{1}{n}D(Q_{X^nY^n}\|P_{XY}^{\otimes n})$ to measure the "distance" between $Q_{X^nY^n}$ and $P_{XY}^{\otimes n}$.
- He showed the minimum rate such that this "distance" vanishes is

$$C_{\mathrm{W}} := \min_{Q_W Q_{X|W} Q_{Y|W}: Q_{XY}=P_{XY}} I_Q(XY;W).$$

- The same result still holds for $D(Q_{X^nY^n}\|P_{XY}^{\otimes n})$ [Hayashi 2006].

Rényi Common Information

Define Rényi common information $C_q := \inf\left\{R : D_q(Q_{X^nY^n}\|P_{XY}^{\otimes n}) \to 0\right\}$.

Theorem (Rényi CI for DSBS [Y.–Tan'20][Y.'24])

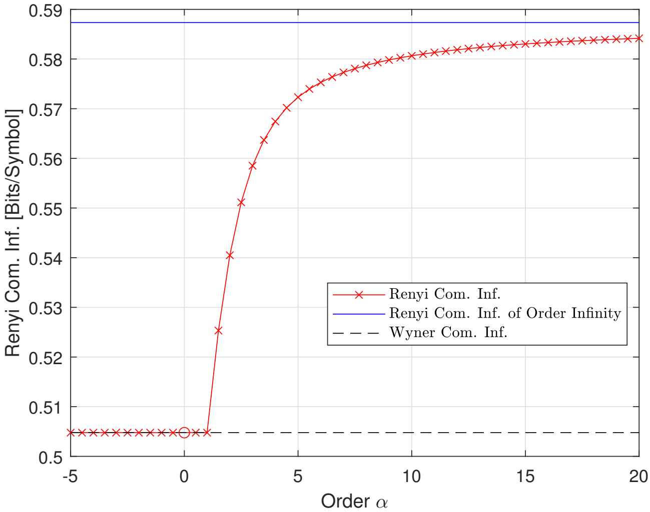
Let $P_{XY} = \mathrm{DSBS}(\epsilon)$ *with* $\epsilon \in [0, 1/2]$. *Then, for* $q \in [0, \infty]$,

$$C_q = \begin{cases} 0, & q = 0 \\ 1 + H(\epsilon) - 2H(a), & q \in (0, 1] \\ 1 - (1 + 2p^* - 2a)\log(1-\epsilon) - (2a - 2p^*)\log\epsilon & \\ \qquad -\frac{1+s}{s}2H(a) + \frac{1}{s}H(p^*, a-p^*, a-p^*, 1+p^*-2a), & q \in (1, \infty) \\ 1-(1-2a)\log(1-\epsilon) - 2a\log\epsilon - 2H(a), & q = \infty \end{cases},$$

with $s = q - 1$, $a = \frac{1-\sqrt{1-2\epsilon}}{2}$, $p^* = \frac{-1+\sqrt{\kappa^2(1-2a)^2+4\kappa a(1-a)}}{2(\kappa-1)} - (\frac{1}{2} - a)$, $\kappa = \left(\frac{1-\epsilon}{\epsilon}\right)^{2s}$, *and* $H(a_1, a_2, a_3, a_4) = -\sum_{i=1}^{4} a_i \log a_i$.

- Achievability: Typical codes, similar to Rényi resolvability problem.

Rényi Common Information



Exact Common Information?

- What if we require $Q_{X^nY^n} = P_{XY}^{\otimes n}$ and the rate of W_n is measured by $\frac{1}{n}H(W_n)$ (variable-length coding)? [Kumar–Li–El Gamal'14]
- Exact common information:

$$C_{\text{Ex}} = \inf\{\frac{1}{n}H(W_n) : X^n \leftrightarrow W_n \leftrightarrow Y^n\}.$$

- A surprising equivalence:

Theorem ([Y.–Tan'20])

For P_{XY} on a finite alphabet,

$$C_{\text{Ex}} = C_{\infty}.$$

Common Information for Gaussian Sources

Theorem ([Y.–Tan'20])

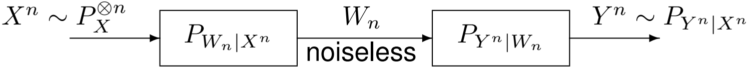
For a Gaussian source (X, Y) *with correlation coefficient* $\rho \in [0, 1)$, *we have*

$$\frac{1}{2}\log\left[\frac{1+\rho}{1-\rho}\right] \le C_\infty = C_{\mathrm{Ex}} \le \frac{1}{2}\log\left[\frac{1+\rho}{1-\rho}\right] + \frac{\rho}{1+\rho}. \quad (12)$$

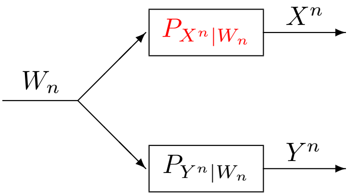
- The lower bound is just Wyner's CI.
- We conjecture the upper bound is tight—still open!

Outline

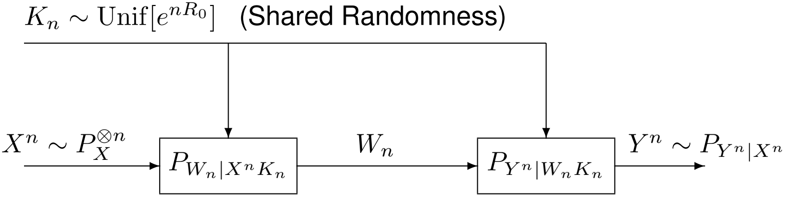
Channel Simulation



By flipping $P_{X^n|W_n}$ to $P_{W_n|X^n}$, it is equivalent to the common information problem:



General Version (with Shared Randomness)



- Goal: Ensure that

$$P_{X^nY^n} \approx P_{XY}^{\otimes n} \text{ (Approximate)} \quad \text{or} \quad P_{X^nY^n} = P_{XY}^{\otimes n} \text{ (Exact)}.$$

- Equivalently,

$$P_{Y^n|X^n} \approx P_{Y|X}^{\otimes n} \text{ (Approximate)} \quad \text{or} \quad P_{Y^n|X^n} = P_{Y|X}^{\otimes n} \text{ (Exact)}.$$

General Version (with Shared Randomness)

- Known as reverse Shannon coding problem [Bennett et. al. ’02], compression of sources of distributions [Winter ’02], or distributed channel synthesis/simulation [Cuff ’12].
- The solution for the TV-distance version was given by Cuff 2012.
- The solution for the TV-distance version in quantum setting was given by [Bennett et. al. ’02][Bennett et. al. ’14].
- The solution for exact channel simulation using fixed-length codes when $R_0 = \infty$ was given by [Cubitt et. al. ’02].
- The solution for Rényi channel simulation using fixed-length codes when $R_0 = \infty$ was given by [Li–Li–Y. ’24].
  - Interestingly, ∞-Rényi simulation rate $=$ exact simulation rate.
- Exact channel simulation using variable-length codes was studied by [Y.–Tan ’20], and the solution for the DSBS was given.
  - Interestingly, ∞-Rényi simulation rate $=$ exact simulation rate.

Concluding Remarks

Why simulation under Rényi divergences?

- ∞-Rényi simulation $\Longleftrightarrow$ exact simulation.
- Rényi divergences $\Longleftrightarrow$ norms of a function.

Thank you!