The Rate of Information Destruction and f-Divergence Pinsker Inequalities

arXiv:2411.17241

Ian George[1], Alice Zheng[2], and Akshay Bansal[2]

1 National University of Singapore

2 Virginia Tech

Pinsker inequality

- Probability distributions $\mathbf{p}, \mathbf{q} \in \mathcal{P}(\mathcal{X})$

$$\frac{1}{2}\|\mathbf{p} - \mathbf{q}\|_1^2 \leq D(\mathbf{p}||\mathbf{q})$$

$$D(\mathbf{p}||\mathbf{q}) := \sum_{i \in \mathcal{X}} p_i \ln(p_i/q_i)$$

- Quantum states $\rho, \sigma \in D(\mathcal{X})$

$$\frac{1}{2}\|\rho - \sigma\|_1^2 \leq D(\rho||\sigma)$$

$$D(\rho||\sigma) := \mathrm{Tr}[\rho(\ln \rho - \ln \sigma)].$$

Pinsker inequality

- Probability distributions $\mathbf{p}, \mathbf{q} \in \mathcal{P}(\mathcal{X})$

$$\frac{1}{2}\|\mathbf{p} - \mathbf{q}\|_1^2 \leq D(\mathbf{p}||\mathbf{q})$$

$$D(\mathbf{p}||\mathbf{q}) := \sum_{i \in \mathcal{X}} p_i \ln(p_i/q_i)$$

- Quantum states $\rho, \sigma \in D(\mathcal{X})$

$$\frac{1}{2}\|\rho - \sigma\|_1^2 \leq D(\rho||\sigma)$$

$$D(\rho||\sigma) := \mathrm{Tr}[\rho(\ln \rho - \ln \sigma)].$$

Pinsker inequality

- Probability distributions $\mathbf{p}, \mathbf{q} \in \mathcal{P}(\mathcal{X})$

$$\frac{1}{2}\|\mathbf{p} - \mathbf{q}\|_1^2 \leq D(\mathbf{p}||\mathbf{q})$$

$$D(\mathbf{p}||\mathbf{q}) := \sum_{i \in \mathcal{X}} p_i \ln(p_i/q_i)$$

- Quantum states $\rho, \sigma \in \mathrm{Pos}(\mathcal{X})$

$$\frac{1}{2}\|\rho - \sigma\|_1^2 \leq \max\{\mathrm{Tr}(\rho), \mathrm{Tr}(\sigma)\}[D(\rho||\sigma) - \mathrm{Tr}(\rho - \sigma)]$$

$$D(\rho||\sigma) := \mathrm{Tr}[\rho(\ln \rho - \ln \sigma)].$$

Pinsker inequality

- Probability distributions $\mathbf{p}, \mathbf{q} \in \mathcal{P}(\mathcal{X})$

$$?\|\mathbf{p}-\mathbf{q}\|_1^2 \overset{?}{\leq} D_?(\mathbf{p}\|\mathbf{q}) \overset{?}{\leq} ?\|\mathbf{p}-\mathbf{q}\|_1^2$$
$$D_?(\mathbf{p}\|\mathbf{q}) := ?$$

- Quantum states $\rho, \sigma \in D(\mathcal{X})$

$$?\|\rho-\sigma\|_1^2 \overset{?}{\leq} D_?(\rho\|\sigma) \overset{?}{\leq} ?\|\rho-\sigma\|_1^2$$
$$D_?(\rho\|\sigma) := ?$$

Outlook

- Summary of results
  - Framework for Pinsker inequalities for f-divergences
  - Relating f-divergences to the χ^2 divergence
  - Application: finite-dimensional time-homogeneous Markov chains
    - Contraction coefficients
    - Mixing times
  - Generalization to quantum f-divergences and Petz f-divergences
- Techniques
  - Multivariate Taylor's theorem with integral remainder
  - Integral representations

Prior work: Pinsker inequalities for f-divergences

This problem has been considered before[1,2,3,4], using similar methods. Existing results typically end up with some sort of limitation:

- Coefficients being difficult to compute[1,3]
- Imposing more structure[2]
- Applying to only specific settings[3] (e.g., only probability distributions)

It is known that f-divergences can be related to each other[4].

---

[1] Gustavo L Gilardoni. "On the minimum f-divergence for given total variation". In: *Comptes rendus. Mathématique* 343.11-12 (2006), pp. 763–766.

[2] Gustavo L Gilardoni. "On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences". In: *IEEE Transactions on Information Theory* 56.11 (2010), pp. 5377–5386.

[3] Mark D Reid and Robert C Williamson. "Generalised Pinsker inequalities". In: *arXiv:0906.1244* (2009).

[4] Peter Harremoës and Igor Vajda. "On pairs of f-divergences and their joint range". In: *IEEE Transactions on Information Theory* 57.6 (2011), pp. 3230–3235.

Prior work: mixing times

- In the classical setting, our main theorem generalizes a result in[5] to a large class of f-divergences
- Mixing times measured by f-divergences considered in[6]
- Convergence of measure under different metrics considered in[7]
- In the quantum setting, mixing times have been considered in[8,9]

[5] Anuran Makur and Lizhong Zheng. "Comparison of contraction coefficients for f-divergences". In: *Problems of Information Transmission* 56 (2020), pp. 103–156.

[6] Maxim Raginsky. "Strong data processing inequalities and Φ-Sobolev inequalities for discrete channels". In: *IEEE Transactions on Information Theory* 62.6 (2016), pp. 3355–3389.

[7] Alison L Gibbs and Francis Edward Su. "On choosing and bounding probability metrics". In: *International statistical review* 70.3 (2002), pp. 419–435.

[8] Kristan Temme et al. "The χ2-divergence and mixing times of quantum Markov processes". In: *Journal of Mathematical Physics* 51.12 (2010).

[9] Alexander Müller-Hermes and Daniel Stilck Franca. "Sandwiched Rényi convergence for quantum evolutions". In: *Quantum* 2 (2018), p. 55.

f-divergence

Definition (f-divergence)

Let $f : (0, +\infty) \to \mathbb{R}$ be convex with $f(1) = 0$, and let $\mathbf{p}, \mathbf{q} \in \mathbb{R}_{\geq 0}^{|\mathcal{X}|}$. The f-divergence of $\mathbf{p}$ with respect to $\mathbf{q}$ is given by

$$D_f(\mathbf{p}||\mathbf{q}) := \sum_{x \in \mathcal{X}} q(x) f(p(x)/q(x)) \ ,$$

where we use the conventions $0f(0/0) = 0$, $0f(a/0) = a \lim_{x \downarrow 0} x f(1/x)$ for $a > 0$.

Examples:

- $f(t) = t \ln(t)$ gives KL divergence, $f(t) = -\ln(t)$ gives reverse KL divergence
- $f(t) = t^2 - 1$ gives χ^2 divergence

f-divergence

Definition (f-divergence)

Let $f : (0, +\infty) \to \mathbb{R}$ be convex with $f(1) = 0$, and let $\mathbf{p}, \mathbf{q} \in \mathbb{R}_{\geq 0}^{|\mathcal{X}|}$. The f-divergence of $\mathbf{p}$ with respect to $\mathbf{q}$ is given by

$$D_f(\mathbf{p}||\mathbf{q}) := \sum_{x \in \mathcal{X}} q(x) f(p(x)/q(x)) \ ,$$

where we use the conventions $0f(0/0) = 0$, $0f(a/0) = a \lim_{x \downarrow 0} x f(1/x)$ for $a > 0$.

Examples:

- $f(t) = t \ln(t)$ gives KL divergence, $f(t) = -\ln(t)$ gives reverse KL divergence
- $f(t) = t^2 - 1$ gives χ^2 divergence

Taylor's theorem

Let $f : [a, b] \to \mathbb{R}$ such that $f \in C^2([a, b])$.

$$f(b) = f(a) + f'(a)(b - a) + R_1(b) \ .$$

Can write the remainder as an integral.

Taylor's theorem

Let $f : [a, b] \to \mathbb{R}$ such that $f \in C^2([a, b])$.

$$f(b) = f(a) + f'(a)(b - a) + \int_a^b f''(t)(b - t)dt \ .$$

Can write the remainder as an integral.

A simple derivation

By DPI, we can consider 2-dimensional vectors

$$\mathbf{p} = \begin{bmatrix} p \\ c - p \end{bmatrix} , \qquad \mathbf{q} = \begin{bmatrix} q \\ c - q \end{bmatrix} , \qquad p, q \in [0, c] ,$$

We have $\mathrm{TV}(\mathbf{p}, \mathbf{q})^2 = (p - q)^2$ and

$$D_f(\mathbf{p}||\mathbf{q}) = qf\left(\frac{p}{q}\right) + (c - q)f\left(\frac{c-p}{c-q}\right) .$$

Fix q, differentiate $g : p \mapsto D(\mathbf{p}||\mathbf{q})$ with respect to p

$$g''(p) = \frac{1}{q}f''\left(\frac{p}{q}\right) + \frac{1}{c-q}f''\left(\frac{c-p}{c-q}\right) .$$

A simple derivation

By DPI, we can consider 2-dimensional vectors

$$\mathbf{p} = \begin{bmatrix} p \\ c - p \end{bmatrix} , \qquad \mathbf{q} = \begin{bmatrix} q \\ c - q \end{bmatrix} , \qquad p, q \in [0, c] ,$$

We have $\mathrm{TV}(\mathbf{p}, \mathbf{q})^2 = (p - q)^2$ and

$$D_f(\mathbf{p}||\mathbf{q}) = qf\left(\frac{p}{q}\right) + (c - q)f\left(\frac{c - p}{c - q}\right) .$$

Fix q, differentiate $g : p \mapsto D(\mathbf{p}||\mathbf{q})$ with respect to p

$$g''(p) = \frac{1}{q}f''\left(\frac{p}{q}\right) + \frac{1}{c - q}f''\left(\frac{c - p}{c - q}\right) .$$

A simple derivation

By DPI, we can consider 2-dimensional vectors

$$\mathbf{p} = \begin{bmatrix} p \\ c - p \end{bmatrix} , \qquad \mathbf{q} = \begin{bmatrix} q \\ c - q \end{bmatrix} , \qquad p, q \in [0, c] ,$$

We have $\mathrm{TV}(\mathbf{p}, \mathbf{q})^2 = (p - q)^2$ and

$$D_f(\mathbf{p}||\mathbf{q}) = qf\left(\frac{p}{q}\right) + (c - q)f\left(\frac{c - p}{c - q}\right) .$$

Fix q, differentiate $g : p \mapsto D(\mathbf{p}||\mathbf{q})$ with respect to p

$$g''(p) = \frac{1}{q}f''\left(\frac{p}{q}\right) + \frac{1}{c - q}f''\left(\frac{c - p}{c - q}\right) .$$

Condition for bounding

$$D_f(\mathbf{p}||\mathbf{q}) = \int_q^p g''(t)(p-t)dt$$

Need to lower-bound $g''(t)$. Suppose

$$L_f \leq \frac{1}{y} f''\left(\frac{x}{y}\right) + \frac{1}{1-y} f''\left(\frac{1-x}{1-y}\right) .$$

- Sanity check: using $f(t) = t\ln(t)$ recovers Pinsker's inequality for KL divergence
- Only needs twice continuously differentiable f
- Does not give a tight bound for reverse KL divergence

Condition for bounding

$$D_f(\mathbf{p}||\mathbf{q}) = \int_q^p g''(t)(p-t)dt \geq \frac{L_f}{c}\int_q^p (p-t)dt$$
$$= \frac{L_f}{2c}(p-q)^2 = \frac{L_f}{2c}\mathrm{TV}(\mathbf{p},\mathbf{q})^2 \;.$$

Need to lower-bound $g''(t)$. Suppose

$$L_f \leq \frac{1}{y}f''\left(\frac{x}{y}\right) + \frac{1}{1-y}f''\left(\frac{1-x}{1-y}\right) \;.$$

- Sanity check: using $f(t) = t\ln(t)$ recovers Pinsker's inequality for KL divergence
- Only needs twice continuously differentiable f
- Does not give a tight bound for reverse KL divergence

Condition for bounding

$$D_f(\mathbf{p}||\mathbf{q}) = \int_q^p g''(t)(p-t)dt \geq \frac{L_f}{c}\int_q^p (p-t)dt$$
$$= \frac{L_f}{2c}(p-q)^2 = \frac{L_f}{2c}\mathrm{TV}(\mathbf{p},\mathbf{q})^2 \ .$$

Need to lower-bound $g''(t)$. Suppose

$$L_f \leq \frac{1}{y} f''\left(\frac{x}{y}\right) + \frac{1}{1-y} f''\left(\frac{1-x}{1-y}\right) \ .$$

- Sanity check: using $f(t) = t\ln(t)$ recovers Pinsker's inequality for KL divergence
- Only needs twice continuously differentiable f
- Does not give a tight bound for reverse KL divergence

Differentiating with respect to q

Gives a similar condition

$$L_f \leq \frac{x^2}{y^3} f''\left(\frac{x}{y}\right) + \frac{(1-x)^2}{(1-y)^3} f''\left(\frac{1-x}{1-y}\right) .$$

- Gives a tight bound for KL divergence with $f(t) = -\ln(t)$
- Does not give a tight bound for KL divergence
- Neither of the conditions give a tight bound for Jeffrey's divergence with $f(t) = (t-1)\ln(t)$

Differentiating over both variables

Use multivariate Taylor's theorem.
Let $F : \mathbb{R}^n \to \mathbb{R}$ be C^2 on an open convex set S. If $\mathbf{a} \in S$ and $\mathbf{a} + \mathbf{h} \in S$ then

$$F(\mathbf{a}+\mathbf{h}) = F(\mathbf{a}) + \langle \nabla F(\mathbf{a}), \mathbf{h} \rangle + \int_0^1 (1-t)\mathbf{h}^T H_F|_{\mathbf{a}+t\mathbf{h}} \mathbf{h} \, dt \ ,$$

where $x^{\odot n}$ applies power to the n in an entry-wise fashion.

Condition for lower bound

Theorem

Let $f:(0,+\infty)\to\mathbb{R}$ with $f(1)=0$ be convex and twice continuously differentiable. Suppose there exists $\lambda\in[0,1]$ such that L_f satisfies for all $x,y\in(0,1)$

$$L_f \le \left[(1-\lambda)+\lambda\frac{x}{y}\right]^2 \frac{1}{y} f''\left(\frac{x}{y}\right) + \left[(1-\lambda)+\lambda\frac{1-x}{1-y}\right]^2 \frac{1}{1-y} f''\left(\frac{1-x}{1-y}\right) .$$

Then for all $\mathbf{p},\mathbf{q}\ge 0$ such that $\|\mathbf{p}\|_1=\|\mathbf{q}\|_1=c>0$,

$$\frac{L_f}{2c}\mathrm{TV}(\mathbf{p},\mathbf{q})^2 \le D_f(\mathbf{p}\|\mathbf{q}) .$$

When $\lambda = 1/2$,

$$L_f \le \left[1+\frac{x}{y}\right]^2 \frac{1}{4y} f''\left(\frac{x}{y}\right) + \left[1+\frac{1-x}{1-y}\right]^2 \frac{1}{4(1-y)} f''\left(\frac{1-x}{1-y}\right) .$$

Condition for lower bound

Theorem

Let $f : (0, +\infty) \to \mathbb{R}$ with $f(1) = 0$ be convex and twice continuously differentiable. Suppose there exists $\lambda \in [0,1]$ such that L_f satisfies for all $x, y \in (0,1)$

$$L_f \leq \left[(1-\lambda) + \lambda \frac{x}{y}\right]^2 \frac{1}{y} f''\left(\frac{x}{y}\right) + \left[(1-\lambda) + \lambda \frac{1-x}{1-y}\right]^2 \frac{1}{1-y} f''\left(\frac{1-x}{1-y}\right) .$$

Then for all $\mathbf{p}, \mathbf{q} \geq 0$ such that $\|\mathbf{p}\|_1 = \|\mathbf{q}\|_1 = c > 0$,

$$\frac{L_f}{2c} \mathrm{TV}(\mathbf{p}, \mathbf{q})^2 \leq D_f(\mathbf{p} \| \mathbf{q}) .$$

When $\lambda = 1/2$,

$$L_f \leq \left[1 + \frac{x}{y}\right]^2 \frac{1}{4y} f''\left(\frac{x}{y}\right) + \left[1 + \frac{1-x}{1-y}\right]^2 \frac{1}{4(1-y)} f''\left(\frac{1-x}{1-y}\right) .$$

Results from theorem

Divergence	$f(t)$	L_f
KL-divergence	$t \ln t$	4
Reverse KL-divergence	$-\ln t$	4
Pearson χ^2-divergence	$t^2 - 1$	8
Neyman χ^2-divergence	$\frac{1}{t} - 1$	8
Symmetric χ^2-divergence	$\frac{(t-1)^2(t+1)}{t}$	16
Arithmetic-geometric mean	$\left(\frac{t+1}{2}\right) \ln \left(\frac{t+1}{2\sqrt{t}}\right)$	1
Jeffrey's divergence	$(t-1) \ln t$	8
Rényi's information gain	$\frac{t^\alpha - 1}{\alpha(\alpha-1)}$	$\begin{cases} 4 & \alpha \in [-1, 2] \\ 1 & \text{otherwise} \end{cases}$
Squared Hellinger distance	$\frac{1}{2}(\sqrt{t} - 1)^2$	1
Lin's measure	$\theta t \ln t - (\theta t + 1 - \theta) \ln(\theta t + 1 - \theta)$	$4\theta(1-\theta)$
Jensen-Shannon divergence	$\frac{1}{2}\left(t \ln t - (t+1) \ln \left(\frac{t+1}{2}\right)\right)$	1
Triangular discrimination	$\frac{(t-1)^2}{t+1}$	4

Lower bound with χ^2 divergence

Corollary

Let f and L_f satisfy the conditions of the above theorem and let $\widetilde{q}_{\min} := \min_{i:q_i>0} q_i$. Then, for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}(\mathcal{X})$,

$$\frac{L_f \widetilde{q}_{\min}}{4} \chi^2(\mathbf{p} \| \mathbf{q}) \leq D_f(\mathbf{p} \| \mathbf{q}) \ .$$

Note that this is data-dependent, as it involves $\widetilde{q}_{\min}$. This is possible to avoid through an alternative proof method.

Pinsker inequality for quantum f-divergences

This result works for a general class of quantum f-divergences that a) satisfy DPI and b) reduce to classical f-divergences.

Corollary

Let $\mathbb{D}_f(\rho\|\sigma)$ be any quantum f-divergence satisfying the data processing inequality for some f that is continuously twice differentiable and ρ, σ be quantum states. Then

$$\frac{L_f}{8}\|\rho - \sigma\|_1^2 \leq \mathbb{D}_f(\rho\|\sigma) \ ,$$

where L_f satisfies the condition of the above theorem.

Sanity check: recovers quantum Pinsker's inequality.

Contraction coefficients

Rate at which information is lost due to data processing

Definition (Input-independent contraction coefficient)

Let D_f be an f-divergence, $\mathcal{W}_{X\to Y}$ be a channel, and $\mathbf{q}_X \in \mathcal{P}(\mathcal{X})$ be a probability distribution. The input-dependent contraction coefficient of $\mathcal{W}$ with respect to $\mathbf{q}$ is

$$\eta_f(\mathcal{W},\mathbf{q}) = \sup_{\substack{\mathbf{p}\in\mathcal{P}(\mathcal{X}) \text{ s.t.} \\ 0<D_f(\mathbf{p}\|\mathbf{q})<+\infty}} \frac{D_f(\mathcal{W}(\mathbf{p})\|\mathcal{W}(\mathbf{q}))}{D_f(\mathbf{p}\|\mathbf{q})}\,.$$

It is known that $\eta_f(\mathcal{W},\mathbf{q}) \geq \eta_{\chi^2}(\mathcal{W},\mathbf{q})$ and $\eta_{\chi^2}(\mathcal{W},\mathbf{q}) = \rho_m(\mathbf{q},\mathcal{W}(\mathbf{q}))^2$. Moreover, the maximal correlation coefficient $\rho_m(X,Y)$ is efficient to compute.

Contraction coefficients

Rate at which information is lost due to data processing

Definition (Input-independent contraction coefficient)

Let D_f be an f-divergence, $\mathcal{W}_{X\to Y}$ be a channel, and $\mathbf{q}_X \in \mathcal{P}(\mathcal{X})$ be a probability distribution. The input-dependent contraction coefficient of $\mathcal{W}$ with respect to $\mathbf{q}$ is

$$\eta_f(\mathcal{W},\mathbf{q}) = \sup_{\substack{\mathbf{p}\in\mathcal{P}(\mathcal{X}) \text{ s.t.} \\ 0<D_f(\mathbf{p}\|\mathbf{q})<+\infty}} \frac{D_f(\mathcal{W}(\mathbf{p})\|\mathcal{W}(\mathbf{q}))}{D_f(\mathbf{p}\|\mathbf{q})}.$$

It is known that $\eta_f(\mathcal{W},\mathbf{q}) \geq \eta_{\chi^2}(\mathcal{W},\mathbf{q})$ and $\eta_{\chi^2}(\mathcal{W},\mathbf{q}) = \rho_m(\mathbf{q},\mathcal{W}(\mathbf{q}))^2$. Moreover, the maximal correlation coefficient $\rho_m(X,Y)$ is efficient to compute.

Characterizations of Markov chains

Let $\mathcal{W}$ be a Markov chain and $W \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ its matrix such that $\mathcal{W}(\mathbf{p}) = W\mathbf{p}$.

1. *Irreducible:* can get from anywhere to everywhere

$$\forall x, x' \in \mathcal{X}, \exists t \in \mathbb{N} \text{ s.t. } W^t \text{ has no non-zero entries}$$

2. *Aperiodic:* do not go around in loops

$$\forall x \in \mathcal{X}, 1 = d(x) := gcd\{t \geq 1 : W^t(x,x) > 0\}$$

3. *Scrambling:* no two columns of W are orthogonal
4. *Indecomposable:* has a unique stationary distribution π and is connected

$$\text{Joint distribution } p_{X',X} \text{ generated by } \pi \text{ is indecomposable}$$
$$\neg\exists A \subset \mathcal{X}', B \subset \mathcal{X}, 0 < \Pr[x' \in A], \Pr[x \in B] < +\infty,\ x' \in A \iff x \in B$$

Contraction coefficient result

Theorem

Let L_f be as defined previously and strictly positive, and let W be a Markov chain with a stationary distribution π Suppose that one of the following holds:

1. *W is irreducible and aperiodic,*
2. *W is scrambling and either (a) $\boldsymbol{\pi}$ full rank or (b) $f'(+\infty) = +\infty$,*
3. *W is indecomposable and $\boldsymbol{\pi}$ is full rank.*

Then, any distribution $\mathbf{p} \in \mathcal{P}(\mathcal{X})$ converges to $\boldsymbol{\pi}$ at a rate of at most

$$\lim_{n\to\infty} \eta_f(W^n, \boldsymbol{\pi})^{1/n} \leq \eta_{\chi^2}(\mathcal{W}, \boldsymbol{\pi}) \ .$$

Moreover, if $\mathcal{W}$ is reversible, the above bound is known to be tight.

A specific kind of quantum f-divergence

Definition (Petz f-divergence[10])

For $P, Q \in \mathrm{Pd}(A)$ the Petz f-divergence is defined as

$$\overline{D}_f(P\|Q) := \mathrm{Tr}[Q^{-1/2} f(L_P R_{Q^{-1}}) Q^{-1/2}] \ ,$$

where $L_W(X) = WX$, $R_W(X) = XW$ for all $X, W \in \mathrm{L}(A)$. Extend to $P, Q \in \mathrm{Pos}(A)$ by

$$\overline{D}_f(P\|Q) := \lim_{\varepsilon \downarrow 0} S_f(P + \varepsilon I \| Q + \varepsilon I) \ .$$

[10] Fumio Hiai and Milán Mosonyi. "Different quantum f-divergences and the reversibility of quantum operations". In: *Reviews in Mathematical Physics* 29.07 (2017), p. 1750023.

Quantum extension of contraction coefficient result

Theorem

Let L_f be as defined previously (with f operator convex) and strictly positive, and let $\mathcal{E}$ be a quantum channel with unique full-rank stationary state π such that $\lim_{n\to\infty} \|\mathcal{E}^{\circ n}(\rho) - \pi\|_1 = 0$ for all quantum states ρ. Any quantum state ρ converges to the stationary state π at a rate of

$$\lim_{n\to\infty} \eta_f(\mathcal{E}^{\circ n}, \pi)^{1/n} \leq \eta_{\overline{\chi}^2}(\mathcal{E}, \pi) \ ,$$

Moreover, we know the above bound can be tight.

Mixing times

Number of times a channel needs to be applied to make its distribution δ-indistinguishable from π under some dissimilarity measure Δ.

$$t_{\text{mix}}^{\Delta}(\mathcal{W}, \delta) := \min\{n \in \mathbb{N} : \sup_{\mathbf{p} \in \mathcal{P}(\mathcal{X})} \Delta(\mathcal{W}^n(\mathbf{p}), \boldsymbol{\pi}) \leq \delta\} \ .$$

Lemma

Let $f : (0, \infty) \to \mathbb{R}$ be convex, differentiable at unity with $f(1) = 0$, $f(0) < \infty$, and $g(t) := \frac{f(t) - f(0)}{t}$ concave on $(0, \infty)$. Let W be a Markov chain with unique full rank stationary distribution $\boldsymbol{\pi}$. Then, whenever $\eta_{\chi^2}(W, \boldsymbol{\pi}) < 1$,

$$t_{mix}^{D_f}(\mathcal{W}, \delta) \leq \frac{\log(2/[\delta \pi_{\min}]) + \log(f'(1) + f(0))}{\log(1/\eta_{\chi^2}(W, \boldsymbol{\pi}))} \ .$$

Mixing times

Number of times a channel needs to be applied to make its distribution δ-indistinguishable from π under some dissimilarity measure Δ.

$$t^{\Delta}_{\mathrm{mix}}(\mathcal{W},\delta) := \min\{n \in \mathbb{N} : \sup_{\mathbf{p}\in\mathcal{P}(\mathcal{X})} \Delta(\mathcal{W}^n(\mathbf{p}),\boldsymbol{\pi}) \le \delta\} \ .$$

Lemma

Let $f:(0,\infty)\to\mathbb{R}$ be convex, differentiable at unity with $f(1)=0$, $f(0)<\infty$, and $g(t) := \frac{f(t)-f(0)}{t}$ concave on $(0,\infty)$. Let W be a Markov chain with unique full rank stationary distribution $\boldsymbol{\pi}$. Then, whenever $\eta_{\chi^2}(W,\boldsymbol{\pi})<1$,

$$t^{D_f}_{mix}(\mathcal{W},\delta) \le \frac{\log(2/[\delta\pi_{\mathrm{min}}]) + \log(f'(1)+f(0))}{\log(1/\eta_{\chi^2}(W,\boldsymbol{\pi}))} \ .$$

Conclusion

Summary

- Pinsker inequalities for f-divergences
- Apply to Markov chains to obtain contraction coefficients, mixing times
- Generalize to quantum f-divergences

Other results

- Integral representations of f-divergences and Bregman divergences
- More inequalities relating f-divergences to the χ^2 divergence
- More inequalities for contraction coefficients and mixing times
- Quantum extension of the mixing times result

2411.17241

2411.17241

Thank you!

Contact: alicezheng@vt.edu