Distance measures between states

Def: The trace distance between two states ρ and σ in $S(A)$ is defined by

$$\Delta(\rho,\sigma) = \frac{1}{2}\underbrace{\|\rho-\sigma\|_1}_{\sum_i |\lambda_i|,\ \lambda_i \text{ eigenvalues of } \rho-\sigma} = \frac{1}{2}\mathrm{Tr}|\rho-\sigma| .$$

Rk: - $\Delta(\rho,\rho) = 0$, $\Delta(\rho,\sigma) \leq \frac{1}{2}(\|\rho\|_1 + \|\sigma\|_1) = 1$.

- Invariant under unitary

$$\Delta(U\rho U^*, U\sigma U^*) = \Delta(\rho,\sigma)$$

- For $\rho = \sum_a{}' P(a)|a\rangle\langle a|$

$$\sigma = \sum_a{}' Q(a)|a\rangle\langle a|$$

$$\Delta(\rho,\sigma) = \underbrace{\frac{1}{2}\sum_a{}' |P(a) - Q(a)|}_{\text{called total variation distance between } P \text{ and } Q}$$

- Data processing $\leftarrow$ important property for any distance measure.

$\mathcal{E}$ quantum channel

$$\Delta(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq \Delta(\rho,\sigma)$$

Operational interpretation: distinguishing states.

Hypotheses System A is either in:

$H_0 : \rho_0$ $\qquad\qquad$ $H_1 : \rho_1$

Question: Minimum probability of error?

Strategy given by a POVM: E_0, E_1

$E_0 \Rightarrow H_0$, $E_1 \Rightarrow H_1$

Prior: H_0 with probability $1/2$.
H_1 with probability $1/2$.

$$\text{Error probability} = \frac{1}{2}\underbrace{\mathrm{Tr}(E_1\rho_0)}_{\substack{\text{state is } \rho_0 \\ \text{but we wrongly say } H_1 \\ \text{"Type I" error}}} + \frac{1}{2}\underbrace{\mathrm{Tr}(E_0\rho_1)}_{\substack{\text{state is } \rho_1 \\ \text{but wrongly say } H_0 \\ \text{"Type II" error}}}$$

Often H_0 and H_1 play asymetric role.

Proposition: The minimum error probability over all possible strategies is

$$\frac{1}{2} - \frac{1}{2}\Delta(\rho_0, \rho_1).$$

Rk: Hypothesis testing can also be considered in different regimes e.g. fix Type I error $\leq \varepsilon$ and minimize Type II error.

(also called divergence).

Def: The hypothesis testing relative entropy with parameter $\varepsilon \in [0,1)$ is defined by

$$D_H^{\varepsilon}(\rho \| \sigma) = \max_{\substack{0 \le E \le I \\ \mathrm{Tr}(E\rho) \ge 1-\varepsilon}} -\log \mathrm{Tr}(E\sigma)$$

E corresponds to E_0, $E_1 = I - E_0 = I - E$.

Rk:

- $D_H^{\varepsilon}(\rho\|\sigma) \in [0,+\infty]$.
- $2^{-D_H^{\varepsilon}(\rho\|\sigma)}$ is the minimum Type II error if Type I error $\le \varepsilon$.
- For $\varepsilon = 1$, $D_H^{1}(\rho\|\sigma) = +\infty$ (not interesting).
- For $\varepsilon = 0$, $D_H^{0}(\rho\|\sigma) = -\log \mathrm{Tr}(\Pi_\rho \sigma)$

  where $\Pi_\rho :=$ projection onto the support of ρ

  $:= \sum_{i:\lambda_i \neq 0} |e_i\rangle\langle e_i|$ where $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$

  eigendecomposition.
- For $\rho = \sigma$, $D_H^{\varepsilon}(\rho\|\rho) = -\log(1-\varepsilon)$.

  $\simeq 0$ if ε small.
- For ρ and σ having orthogonal supports, i.e., $\rho\sigma = 0$

  $D_H^{\varepsilon}(\rho\|\sigma) = +\infty$

Further remarks about $D_H^\varepsilon(\rho\|\sigma)$

- In general, no closed form expression but

$$\min_{0\leq E\leq I} \mathrm{Tr}(E\sigma) \quad \text{subject to } \mathrm{Tr}(E\rho)\geq 1-\varepsilon .$$

is a convex optimization program, more specifically it is a semi-definite program.
can be computed efficiently for small dimension.

- Classical case $\rho = \sum_{x\in\mathcal{X}}' P(x)\, |x\rangle\langle x|$

$$\sigma = \sum_{x\in\mathcal{X}} Q(x)\, |x\rangle\langle x| .$$

A natural test:

For sample x, compute $\underbrace{\dfrac{P(x)}{Q(x)}}_{\text{called likelihood ratio.}}$ $\to$ If ≥ 1 output "P"; If ≤ 1 output "Q"

Prop: D_H^ε satisfies the data processing inequality

i.e. for any quantum channel $\mathcal{E}$, we have

$$D_H^\varepsilon(\mathcal{E}(\rho)\,\|\,\mathcal{E}(\sigma)) \leq D_H^\varepsilon(\rho\|\sigma)$$

Proof: Very intuitive: If I have a strategy to distinguish $\mathcal{E}(\rho)$ from $\mathcal{E}(\sigma)$, can distinguish ρ and σ by first applying $\mathcal{E}$ then the strategy.

Let E be such that

$$D_H^\varepsilon(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) = -\log \mathrm{Tr}(E\mathcal{E}(\sigma)) \text{ and } \mathrm{Tr}(E\mathcal{E}(\rho)) \geq 1-\varepsilon$$

Note that $L(\mathcal{H})$ is itself a Hilbert space with inner product $\langle S,T\rangle = \mathrm{Tr}(S^*T)$.

So $\mathcal{E} \in L(L(\mathcal{H}))$ has an adjoint denoted $\mathcal{E}^*$, it satisfies:

$$\mathrm{Tr}(E\,\mathcal{E}(\sigma)) \underset{\uparrow}{=} \mathrm{Tr}(E^*\mathcal{E}(\sigma)) = \mathrm{Tr}(\mathcal{E}^*(E)\sigma)$$

E is Hermitian

and $\mathrm{Tr}(E\,\mathcal{E}(\rho)) = \mathrm{Tr}(\mathcal{E}^*(E)\rho)$

Fact: $\mathcal{E}$ completely positive $\iff$ $\mathcal{E}^*$ completely positive.

$\mathcal{E}$ trace preserving $\iff$ $\mathcal{E}^*$ is unital i.e. $\mathcal{E}^*(I) = I$.

As a result, $\mathcal{E}^*(E)$ satisfies

$$0 = \mathcal{E}^*(0) \leq \mathcal{E}^*(E) \leq \mathcal{E}^*(I) = I.$$

and it is a feasible solution for the program for $D_H^\varepsilon(\rho\|\sigma)$

So $D_H^\varepsilon(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq D_H^\varepsilon(\rho\|\sigma)$ ∎

Special states of interest: $\rho^{\otimes n}$, $\sigma^{\otimes n}$ with $n \to \infty$.

Th (Quantum Stein Lemma)

Let $\varepsilon \in (0,1)$ and $\rho, \sigma \in S(A)$.

Then

$$\lim_{n \to \infty} \frac{1}{n} D_H^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho \| \sigma).$$

The quantum relative entropy

(ρ state but σ not necessarily normalized)

Def: For $\rho \in S(A)$, $\sigma \in \mathrm{Pos}(A)$ where A is a finite dimensional Hilbert space, the quantum relative entropy is defined by:

$\mathrm{supp}(\rho) = \mathrm{Span}\{|e_i\rangle : \lambda_i \neq 0\}$ if $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$

$$D(\rho \| \sigma) = \begin{cases} \mathrm{Tr}(\rho(\log \rho - \log \sigma)) & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma) \\ +\infty & \text{else} \end{cases}$$

Rk: * $\log \rho = \sum_i{}' (\log \lambda_i) |e_i\rangle\langle e_i|$ for $\rho = \sum_i{}' \lambda_i |e_i\rangle\langle e_i|$, $\lambda_i \neq 0$.

* Classical case, i.e. ρ and σ commute

$$\rho = \sum_x{}' P(x) |x\rangle\langle x|, \quad \sigma = \sum_x Q(x) |x\rangle\langle x|.$$

$$D(\rho \| \sigma) = \underbrace{\sum_x{}' P(x) \log \frac{P(x)}{Q(x)}}_{\text{called relative entropy or Kullback-Leibler divergence}}$$

Quantum relative entropy can be seen as a noncommutative generalization of KL divergence (there are others as well)

Th (Properties of the quantum relative entropy)

- We have $D(\rho \| \sigma) \geq 0$ for $\rho, \sigma \in S(A)$ with equality iff $\rho = \sigma$.
- Data processing for D: for a quantum channel $\mathcal{E}$
$$D(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \leq D(\rho \| \sigma)$$

Will prove this later.

D can be used to define the von Neumann entropy:

Def: For a state $\rho_{AB} \in S(A \otimes B)$ we define

- $H(A)_\rho := -D(\rho_A \| I_A)$

entropy

sign.

not normalized.

Recall $\rho_A = \mathrm{Tr}_B \rho_{AB}$

- $H(A|B)_\rho := -D(\rho_{AB} \| I_A \otimes \rho_B)$.

conditional entropy

von Neumann entropies.

- $I(A:B)_\rho := D(\rho_{AB} \| \rho_A \otimes \rho_B)$.

mutual information

Proof of Stein lemma:

Recall statement:

$$\lim_{n\to\infty} \frac{1}{n} D_H^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho\|\sigma)$$

* Achievability: $\geq$ (have to give a strategy)

Let us start with the case where $\boxed{\rho \text{ and } \sigma \text{ commute}}$.

$$\rho = \sum_x P(x)\, |x\rangle\langle x| \qquad \sigma = \sum_x Q(x)\, |x\rangle\langle x| .$$

$$\rho^{\otimes n} = \sum_{x_1 \dots x_n} P(x_1)\cdots P(x_n) |x_1\rangle\langle x_1| \otimes \dots \otimes |x_n\rangle\langle x_n|$$

$$\sigma^{\otimes n} = \sum_{x_1 \dots x_n} Q(x_1)\cdots Q(x_n)\, |x_1\rangle\langle x_1| \otimes \dots \otimes |x_n\rangle\langle x_n| .$$

Will define a test for this hypothesis testing problem.

Given $X_1, \dots, X_n$

Compute $R = \dfrac{P(X_1)P(X_2)\cdots P(X_n)}{Q(X_1)Q(X_2)\cdots Q(X_n)}$

If $\frac{1}{n}\log R \geq D(P\|Q) - \delta$

Return "Samples from P".

Else Return "Samples from Q".

$\delta > 0$ is a parameter, will let $\delta \to 0$ at the end.

In quantum notation corresponds to:

$$E = \sum_{x_1 \dots x_n :\, \frac{P(x_1)\cdots P(x_n)}{Q(x_1)\cdots Q(x_n)} \geq 2^{n(D(P\|Q)-\delta)}} |x_1 \dots x_n\rangle\langle x_1 \dots x_n|$$

it clearly depends on n

Analysis of this test.

* If samples are from P. (Hypothesis 0)

$$\mathbb{P}_{X_1 \dots X_n \sim P} \left\{ \frac{1}{n} \log R \geq D(P\|Q) - \delta \right\} \quad \left(= \mathrm{Tr}(E\rho^{\otimes n}) \right)$$

$$= \mathbb{P}\left\{ \frac{1}{n} \sum_{i=1}^{n} \log \frac{P(X_i)}{Q(X_i)} \geq D(P\|Q) - \delta \right\}$$

But $\mathbb{E}_{X_i \sim P} \left\{ \log \frac{P(X_i)}{Q(X_i)} \right\} = \sum_{x} P(x) \log \frac{P(x)}{Q(x)} = D(P\|Q)$.

So by the law of large numbers

$$\xrightarrow[n \to \infty]{} 1$$

The constraint $\mathrm{Tr}(E\rho^{\otimes n}) \geq 1 - \varepsilon$ satisfied for large enough n.

* If samples are from Q. (Hypothesis 1)

$$\underbrace{\mathbb{P}_{X_1 \dots X_n \sim Q} \left\{ \frac{1}{n} \log R \geq D(P\|Q) - \delta \right\}}_{= \mathrm{Tr}(E\sigma^{\otimes n})} = \sum_{\substack{x_1, \dots, x_n \\ \frac{1}{n}\sum_i \log \frac{P(x_i)}{Q(x_i)} \geq (D(P\|Q) - \delta)}} Q(x_1) \dots Q(x_n)$$

$$= \sum_{x_1 \dots x_n}' Q(x_1) \dots Q(x_n).$$

$$Q(x_1) \dots Q(x_n) \leq 2^{-n(D(P\|Q) - \delta)} P(x_1) \dots P(x_n)$$

$$\leq 2^{-n(D(P\|Q) - \delta)} \sum_{x_1 \dots x_n}' P(x_1) \dots P(x_n)$$

$$\leq 2^{-n(D(P\|Q) - \delta)}$$

So

$$-\log \mathrm{Tr}(E\sigma^{\otimes n}) \geq n\big(D(P\|Q)-\delta\big)$$

and $\frac{1}{n} D_H^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \geq D(P\|Q) - \delta$ for large enough n.

Works for any $\delta > 0$ ~~so~~ we have

$$\lim_{n\to\infty} \frac{1}{n} D_H^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \geq D(P\|Q).$$

Considered general case ρ, σ not commuting.

Pinching general technique to reduce general case to commuting.

given $\sigma = \sum_{\lambda \in \mathrm{Spec}(\sigma)} \lambda \Pi_\lambda$, let $\mathcal{P}_\sigma(S) = \sum_\lambda \Pi_\lambda S \Pi_\lambda$.

<u>Two important properties:</u>
- $\mathcal{P}_\sigma(S)$ commutes with σ
- $\rho \geq 0$, $\mathcal{P}_\sigma(\rho) \geq \frac{1}{|\mathrm{Spec}(\sigma)|} \rho$ number of distinct eigenvalues

$$D_H^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \geq D_H^{\varepsilon}\big(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \| \mathcal{P}_{\sigma^{\otimes n}}(\sigma^{\otimes n})\big)$$
$$= D_H^{\varepsilon}\big(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \| \sigma^{\otimes n}\big)$$

commute!

Common eigenbasis $\rightarrow$ corresponds to distributions $P :=$ eigenvalues of $\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})$

$Q :=$ eigenvalues of $\sigma^{\otimes n}$.

$$D_H^{\varepsilon}\big(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \| \sigma^{\otimes n}\big) = D_H^{\varepsilon}(P\|Q)$$

We have $\lim_{m\to\infty} \frac{1}{m} D_H^{\varepsilon}(P^{\otimes m} \| Q^{\otimes m}) \geq D(P\|Q)$

As a result

$$\frac{1}{nm} D_H^{\varepsilon}\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})^{\otimes m} \,\middle\|\, (\sigma^{\otimes n})^{\otimes m}\right) = \frac{1}{n}\cdot\frac{1}{m} D_H^{\varepsilon}(P^{\otimes m} \| Q^{\otimes m})$$

$$\xrightarrow[m\to\infty]{} \frac{1}{n} D(P\|Q)$$

$$= \frac{1}{n} D\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \,\middle\|\, \sigma^{\otimes n}\right)$$

Now use $\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \geq \frac{1}{|\mathrm{spec}(\sigma^{\otimes n})|}\, \rho^{\otimes n}$

Elements in $\mathrm{spec}(\sigma^{\otimes n})$ are of the form $\prod_{i=1}^{n} \lambda_i$ with $\lambda_i \in \mathrm{spec}(\sigma)$

But $|\mathrm{spec}(\sigma)| \leq d$ $(= \dim A)$

So $|\mathrm{spec}(\sigma^{\otimes n})| \leq (n+1)^{d-1}$ (each eigenvalue appears a number of times between 0 and n)

$$D\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \,\middle\|\, \sigma^{\otimes n}\right) = \mathrm{Tr}\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \log \mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) - \mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \log \sigma^{\otimes n}\right)$$

$$= \mathrm{Tr}\left(\rho^{\otimes n} \mathcal{P}_{\sigma^{\otimes n}}\left(\log \mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\right) - \rho^{\otimes n} \log \sigma^{\otimes n}\right)$$

$$= \mathrm{Tr}\left(\rho^{\otimes n} \log \mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) - \rho^{\otimes n} \log \sigma^{\otimes n}\right)$$

log is operator monotone $\rightarrow$ $\geq \mathrm{Tr}\left(\rho^{\otimes n} \log \rho^{\otimes n} - \rho^{\otimes n} \log \sigma^{\otimes n}\right) - \log|\mathrm{spec}(\sigma^{\otimes n})|$

Thus:

$$\frac{1}{n} D\left(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}) \,\middle\|\, \sigma^{\otimes n}\right) \geq \frac{1}{n} D(\rho^{\otimes n} \| \sigma^{\otimes n}) - \frac{1}{n} \log|\mathrm{spec}(\sigma^{\otimes n})|$$

$$\geq D(\rho\|\sigma) - \underbrace{\frac{1}{n}\log (n+1)^{d-1}}_{\xrightarrow[n\to\infty]{} 0}.$$

Note that $D(\rho^{\otimes n} \| \sigma^{\otimes n}) = n D(\rho \| \sigma)$.

Indeed: if $\rho = \sum_x \lambda_x |x\rangle\langle x|$ for a basis $\{|x\rangle\}$

$$\rho^{\otimes n} = \sum_{x_1 \dots x_n} \lambda_{x_1} \lambda_{x_2} \dots \lambda_{x_n} |x_1 \dots x_n\rangle\langle x_1 \dots x_n|$$

$$\log(\rho^{\otimes n}) = \sum_{x_1 \dots x_n} \sum_{i=1}^{n} \log \lambda_i \; |x_1 \dots x_n\rangle\langle x_1 \dots x_n|$$

$$= \sum_{i=1}^{n} I_{A_1} \otimes \dots \otimes I_{A_{i-1}} \otimes (\log \rho_{A_i}) \otimes I_{A_{i+1}} \otimes \dots \otimes I_{A_n}$$

* **Converse**:

Will only prove

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} D_H^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \le D(\rho \| \sigma)$$

General quantum case

The statement is that it holds for any $\varepsilon \in (0,1)$

Let E be such that $\mathrm{Tr}(E \rho^{\otimes n}) \ge 1 - \varepsilon$.

We apply the data processing inequality for the quantum channel

$$\mathcal{E} : L(A^{\otimes n}) \longrightarrow L(\mathbb{C}^2)$$

$$T \longmapsto |0\rangle\langle 0| \mathrm{Tr}(ET) + |1\rangle\langle 1| \mathrm{Tr}((I - E)T).$$

$$\mathcal{E}(\rho^{\otimes n}) = |0\rangle\langle 0| \mathrm{Tr}(E \rho^{\otimes n}) + |1\rangle\langle 1| (1 - \mathrm{Tr}(E \rho^{\otimes n}))$$

$$\mathcal{E}(\sigma^{\otimes n}) = |0\rangle\langle 0| \mathrm{Tr}(E \sigma^{\otimes n}) + |1\rangle\langle 1| (1 - \mathrm{Tr}(E \sigma^{\otimes n}))$$

We have on one side:

$$D(\rho^{\otimes n} \| \sigma^{\otimes n}) = n D(\rho \| \sigma).$$

- But

$$D(\rho^{\otimes n} \| \sigma^{\otimes n}) \overset{\text{data processing}}{\geq} D(\mathcal{E}(\rho^{\otimes n}) \| \mathcal{E}(\sigma^{\otimes n}))$$

$$= \mathrm{Tr}(E\rho^{\otimes n}) \log \frac{\mathrm{Tr}(E\rho^{\otimes n})}{\mathrm{Tr}(E\sigma^{\otimes n})} + (1 - \mathrm{Tr}(E\rho^{\otimes n})) \log\left(\frac{1-\mathrm{Tr}(E\rho^{\otimes n})}{1-\mathrm{Tr}(E\sigma^{\otimes n})}\right)$$

$$\geq -1 - \mathrm{Tr}(E\rho^{\otimes n}) \log \mathrm{Tr}(E\sigma^{\otimes n})$$

↑ elementary inequalities

So $-\log \mathrm{Tr}(E\sigma^{\otimes n}) \leq \dfrac{n D(\rho\|\sigma) + 1}{\mathrm{Tr}(E\rho^{\otimes n})} \leq \dfrac{n D(\rho\|\sigma) + 1}{1-\varepsilon}$

$$\frac{1}{n} D_H^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \leq \frac{D(\rho\|\sigma)}{1-\varepsilon} + \frac{1}{(1-\varepsilon)\cdot n}$$

letting $n \to \infty$ then $\varepsilon \to 0$, we get the desired result. $\boxtimes$

Rk: • D_H^{ε} is called a "one-shot entropy" measure as it has an operational interpretation for any states. Many others: $H_{min}^{\varepsilon} \leftarrow$ Cryptography.
⋮ "worst case" entropy.

• The usual relative entropy D and corresponding von Neumann entropy H only has an operational interpretation in an iid (independent identically distributed) or average setting.