Nussbaum-Szkoła distributions and their use

George Androulakis

University of South Carolina

University of Missouri-Columbia, Thurs. Oct. 5th, 2023

Based on joint works with Tiju Cherian John
arXiv:2308.02929, (RMP), arXiv:2203.01964, (IDAQP), and arXiv.org:2303.03380.

Outline

Statistical Hypothesis Testing

A statistician challenges the general belief that a certain hypothesis H_0 is correct. The statistician suggests that H_1 is in fact correct.

- H_0 **Null Hypothesis**.
- H_1 **Alternative Hypothesis**.

The statistician collects data, makes measurements on these data, and then either

- accepts the alternative hypothesis, or
- rejects the alternative hypothesis.

This is not a mathematical proof! There are two types of errors:

- **Type I Error** H_0 is correct, but H_1 is accepted.
- **Type II Error** H_1 is correct, but it is rejected.

$$\alpha = \mathbb{P}(\text{Type I Error}) = \mathbb{P}(H_1|H_0), \quad \beta = \mathbb{P}(\text{Type II Error}) = \mathbb{P}(H_0|H_1).$$

Example 1

In 19th century it was believed that the average human body temperature for ages 11-65 is 98.6 F. Later, doctors challenged that:

- H_0: The average human body temperature for ages 11-65 is 98.6 F.
- H_1: The average human body temperature for ages 11-65 is lower than 98.6 F.

After collecting enough data, the Alternative Hypothesis was accepted. Today it is believed that it is 97.6 F.
Evaluation of severity of errors:

- Type I Error: The average human body temperature is equal to 98.6 F, but people accept that the average human body temperature is lower than that. Some sick people with hypothermia will not be diagnosed.
- Type II Error: The average human body temperature is below 98.6 F, but people reject that. Some sick people with fever will not be diagnosed.

Example 2

A person is suspected for committing a crime.

- H_0: The suspect is innocent.
- H_1: The suspect is guilty.

Data will be collected and measured. The Alternative Hypothesis will be rejected or accepted.

Evaluation of severity of errors:

- Type I Error: The suspect is innocent, but people believe that the suspect is guilty. An innocent person goes to jail.
- Type II Error: The suspect is guilty, but people reject that. A criminal is released in the society.

Example 3

An individual has symptoms of COVID-19 and decides to be tested.

- H_0: The individual has COVID-19.
- H_1: The individual is healthy.

The individual will be tested. The Alternative Hypothesis will be accepted or rejected.
Evaluation of severity of errors:

- Type I Error: The individual has COVID-19, but test is negative. More people will be infected with COVID-19 by coming in contact with the individual.
- Type II Error: The individual is healthy, but the test is positive. The individual will be quarantined unnecessarily.

In this example the severity of Type I Error is very large, so we want a guarantee that α is small.

Basics of quantum state discrimination

- A **quantum state** on $\mathcal{B}(\mathcal{H})$ is a positive semi-definite operator on $\mathcal{H}$ of trace equal to 1, (a.k.a. **density operator**). It is the quantum analogue of a classical probability distribution.
- A **measurement (POVM)** is a finite collection of positive semi-definite operators on a Hilbert space whose sum is equal to the identity operator. It is the quantum analogue of a finite partition of the classical probability sample space.
- If ρ is a quantum state and we measure it using the POVM $\{A_0, \ldots, A_{n-1}\}$, then we obtain $i (\in \{0, \ldots n-1\})$ with probability $\mathrm{Tr}\,(\rho A_i)$.
- You are presented with one of two states ρ or σ which are known to you in advance, but you do not know which of the two states you are presented with. The task is to design a measurement $\{A_0, A_1\}$ that will be able to reveal the presented state. If $\mathrm{Tr}\,(\rho A_0)$, $\mathrm{Tr}\,(\sigma A_1)$ are large, (hence $\mathrm{Tr}(\rho A_1)$ and $\mathrm{Tr}\,(\sigma A_0)$ are small), then if the measurement yields 0 then you guess ρ, and if the measurement yields 1 then you guess σ.

Asymmetric Quantum Hypothesis Testing

You are presented with one of two states ρ or σ which are known to you in advance, but you do not know which of the two states you are presented with.

- H_0: You are presented with ρ.
- H_1: You are presented with σ.

You design measurements $A = \{A_0, A_1\}$ that will help you recognize the presented state. **Assume that the severity of Type I Error is very large**. Hence, for some $\epsilon \in (0, 1)$ and you only consider measurements $A = \{A_0, A_1\}$ such that $\alpha(A) = \mathbb{P}(H_1|H_0) = \mathrm{Tr}\,(\rho A_1) \leq \epsilon$. Quantity of interest:

$$\beta^*(\epsilon) = \min_{\{A:\alpha(A)\leq\epsilon\}} \beta(A) = \min_{\{A=(A_0,A_1):\mathrm{Tr}\,(\rho A_1)\leq\epsilon\}} \mathrm{Tr}\,(\sigma A_0).$$

Hiai-Petz (1991), Ogawa-Nagaoka (2000)

You are presented with n many independent copies of one of two states ρ or σ (**asymptotic i.i.d. case**) which are known to you in advance, but you do not know which of the two states you are presented with.

- H_0: You are presented with $\rho^{\otimes n}$.
- H_1: You are presented with $\sigma^{\otimes n}$.

Then,

$$\lim_{n\to\infty} -\frac{1}{n}\log \min_{\{A^{(n)}=(A_0^{(n)},A_1^{(n)}):\mathrm{Tr}\,(\rho^{\otimes n}A_1^{(n)})\leq\epsilon\}} \mathrm{Tr}\,(\sigma^{\otimes n}A_0^{(n)})$$

$= \text{(optimal error exponent)} = \textbf{the Umegaki relative entropy } D(\rho||\sigma)$

$$= \begin{cases} \mathrm{Tr}(\rho(\log\rho - \log\sigma)) & \text{if supp}\,(\rho)\subseteq\text{supp}\,(\sigma) \\ \infty & \text{otherwise} \end{cases}$$

(Quantum Stein's Lemma).

Symmetric Quantum Hypothesis Testing

Given two known states ρ and σ, and a known $p \in (0,1)$, a quantum source emits the state ρ with probability p and the state σ with probability $1-p$. You are trying to design a measurement $A = \{A_0, A_1\}$ which minimizes the **probability of error**

$$\begin{aligned} \mathbb{P}_{\text{err}}(p,\rho,\sigma) &:= \min_{A} p\alpha(A) + (1-p)\beta(A) \\ &= \min_{A=(A_0,A_1)} p\mathrm{Tr}\,(\rho A_1) + (1-p)\mathrm{Tr}\,(\sigma A_0). \end{aligned}$$

Nussbaum-Szkoła (2009), Audenaert et al. (2007)

(**Asymptotic i.i.d. case.**) Given two known states ρ and σ, and a known $p \in (0,1)$, a quantum source emits the state $\rho^{\otimes n}$ with probability p and the state $\sigma^{\otimes n}$ with probability $1-p$. Then,

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}_{\mathrm{err}}(p, \rho^{\otimes n}, \sigma^{\otimes n}) = \text{optimal error exponent}$$

$$= C(\rho||\sigma) := \sup_{0\le t\le 1} -\log \mathrm{Tr}\,(\rho^t \sigma^{1-t})$$

(**quantum Chernoff divergence**).

The proof of Nussbaum-Szkoła (2009) uses a corresponding classical result of Chernoff (1952).

Definition of Nussbaum-Szkoła distributions

In the course of their proof, Nussbaum and Szkoła (2009) observed the following:

For every two states ρ and σ on a finite dimensional Hilbert space there exist two probability distributions P and Q such that

$$\mathrm{Tr}\left(\rho^t \sigma^{1-t}\right) = \sum_k P_k^t Q_k^{1-t}, \text{ for every } t \in (0,1).$$

Definition of P and Q (**Nussbaum-Szkoła distributions**): Let

$$\rho = \sum_{i=1}^{n} r_i \left|u_i\right\rangle\!\left\langle u_i\right| \text{ and } \sigma = \sum_{j=1}^{n} s_j \left|v_j\right\rangle\!\left\langle v_j\right|$$

be the spectral decompositions of ρ and σ. Then

$$P(i,j) = r_i |\left\langle u_i | v_j \right\rangle|^2 \text{ and } Q(i,j) = s_j |\left\langle u_i | v_j \right\rangle|^2 \text{ for } i,j \in \{1,\ldots n\}.$$

Uses of the Nussbaum-Szkoła distributions in literature

- Nussbaum and Szkoła (2010) studied the symmetric quantum hypothesis testing (quantum Chernoff bounds) for **finitely many states** on a finite dimensional Hilbert space. They conjectured that the optimal error exponent is equal to the minimum of all pairwise error exponents. This conjecture was finally proved by Li (2016).
- Audenaert-Mosonyi-Verstraete (2012) study both the symmetric hypothesis testing, as well as the asymmetric hypothesis testing for finite sampling size (**without taking limits to infinity**).
- Mosonyi (2009) uses Nussbaum-Szkoła distributions to discriminate **certain Gaussian states**.
- Hiai-Mosonyi-Petz-Bény (2011) use Nussbaum-Szkoła distributions for general $\boldsymbol{f}$**-divergences** between states on finite dimensional C^*-algebras.

More uses of the Nussbaum-Szkoła distributions

- Tomamichel-Hayashi (2013) studied **second order asymptotics** of certain entropic quantities of i.i.d. copies of classical and quantum states. Datta-Mosonyi-Hsieh-Brandao (2013) and (independently) Li (2014) also studied second order asymptotics in quantum hypothesis testing. Datta-Pautrat-Rouzé compute second order asymptotics in non-i.i.d. quantum hypothesis testing.
- The **classical capacity of a quantum channel** is defined as the maximum number of bits that can be reliably transmitted via a quantum channel. It is described via the Holevo-Schumacher-Westmoreland Theorem. The question of whether one can transmit reliably classical information via a quantum channel at a rate asymptotically close to its classical capacity has been studied using Nussbaum-Szkoła distributions by several authors such as: Hayashi (2007), Dalai (2013), Tomamichel-Tan (2015), Chung-Guha-Zheng (2016), Chubb-Tan-Tomamichel (2017), Cheng-Hsieh (2018), Cheng-Hsieh-Tomamichel (2019).

Classical f-divergences

Definition (Cziszár (1963))

Let P, Q be probability distributions in a measure space $(X, \mathcal{F})$. Let μ be a σ-finite measure with $P \ll \mu$ and $Q \ll \mu$. Let $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$. Let $f : (0, \infty) \to \mathbb{R}$ be a convex or concave function. Define the f-divergence by

$$D_f(P||Q) = \int_{\{pq>0\}} f(\frac{p}{q})dQ + f(0)Q(p=0) + f'(\infty)P(q=0),$$

where $f'(\infty) := \lim_{t\to\infty} \frac{f(t)}{t}$, and "natural" conventions about 0 and ∞.

Special cases of f-divergences

Assume that P and Q are discrete.

- $f(t) = t \log t$ gives the **Kullback-Leibler relative entropy**

$$D_f(P||Q) = D(P||Q) = \begin{cases} \sum_i P(i) \log \frac{P(i)}{Q(i)} & \text{if } P \ll Q \\ \infty & \text{otherwise} \end{cases}$$

- $f_\alpha(t) = t^\alpha$ for $\alpha \in (0,1) \cup (1,\infty)$ gives the **Rényi α-relative entropy** $D_\alpha(P||Q) = \frac{1}{\alpha-1} \log D_{f_\alpha}(P||Q)$ with

$$D_\alpha(P||Q) = \begin{cases} \frac{1}{\alpha-1} \log \sum_i P(i)^\alpha Q(i)^{1-\alpha} & \text{if } P \ll Q \\ \infty & \text{otherwise} \end{cases}$$

- $f(t) = \frac{1}{2}(\sqrt{t} - 1)^2$ gives the **squared Hellinger distance**

$$D_f(P||Q) = \mathcal{H}^2(P||Q) = \frac{1}{2} \sum_i (\sqrt{P(i)} - \sqrt{Q(i)})^2$$

More special cases of f-divergences

- $f_\alpha(t) = \frac{t^\alpha - 1}{\alpha - 1}$ for $\alpha \in (0,1) \cup (1,\infty)$ gives the **Hellinger α-divergence** $D_{f_\alpha}(P\|Q) = \mathcal{H}_\alpha(P\|Q)$, with

$$\mathcal{H}_\alpha(P||Q) = \begin{cases} \frac{1}{\alpha-1}\left(\left(\sum_i P(i)^\alpha Q(i)^{1-\alpha}\right) - 1\right), & \text{if } \alpha < 1 \text{ or } P \ll Q; \\ \infty, & \text{otherwise.} \end{cases}$$

- $f(t) = |t-1|$ gives the **total variation distance**

$$D_f(P\|Q) = V(P\|Q) = \sum_i |P(i) - Q(i)|.$$

- $f(t) = (t-1)^2$ gives the χ^2**-divergence**,

$$\chi^2(P||Q) = \begin{cases} \sum_{\{i|Q(i)>0\}} \frac{(P(i)-Q(i))^2}{Q(i)}, & \text{if } P \ll Q; \\ \infty, & \text{otherwise.} \end{cases}$$

The relative modular operator

Notation

- $\mathcal{B}(\mathcal{H})$: *bounded operators on* $\mathcal{H}$.
- $\mathcal{B}_2(\mathcal{H})$: *Hilbert-Schmidt operators on* $\mathcal{H}$.
- Π_ρ: *the projection on the* $\operatorname{supp}(\rho)$, *(if* ρ *is a state).*

Definition (Araki (1977))

$$D(S) = \{X\sqrt{\sigma} \ : \ X \in \mathcal{B}(\mathcal{H})\} + \{Y(I - \Pi_\sigma) \ : \ Y \in \mathcal{B}_2(\mathcal{H})\} \subseteq \mathcal{B}_2(\mathcal{H}).$$

Define the antilinear operator $S : D(S) \to \mathcal{B}_2(\mathcal{H})$ *by*

$$S\left(X\sqrt{\sigma} + Y(I - \Pi_\sigma)\right) = \Pi_\sigma X^\dagger \sqrt{\rho}. \tag{1}$$

Then, the **relative modular operator** $\Delta_{\rho,\sigma}$ *is defined by*

$$\Delta_{\rho,\sigma} = S^\dagger \overline{S}.$$

The relative modular operator in a simplified case

Remark

Assume that $\mathcal{H}$ is a finite dimensional Hilbert space, ρ, σ are density operators on $\mathcal{H}$, and σ is invertible. Then

$$\Delta_{\rho,\sigma} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$$

is given by

$$\Delta_{\rho,\sigma}(X) = \rho X \sigma^{-1}.$$

Quantum f-divergences

Definition

Let ρ, σ be states on $\mathcal{H}$. Let $f : (0, \infty) \to \mathbb{R}$ be a convex or concave function. Then the **quantum f-divergence** $D_f(\rho||\sigma)$ *is defined by*

$$D_f(\rho||\sigma) = \int_{0^+}^{\infty} f(\lambda) \left\langle \sqrt{\sigma} \middle| \xi^{\Delta_{\rho,\sigma}}(\mathrm{d}\lambda) \middle| \sqrt{\sigma} \right\rangle_2 + f(0) \operatorname{tr} \left(\sigma \Pi_\rho^\perp \right) + f'(\infty) \operatorname{tr} \left(\rho \Pi_\sigma^\perp \right)$$

where $\xi^{\Delta_{\rho,\sigma}}$ is the spectral measure of the relative modular operator $\Delta_{\rho,\sigma}$ and $\langle\cdot|\cdot\rangle_2$ denotes the inner product in $\mathcal{B}_2(\mathcal{H})$.

Special cases of quantum f-divergences

- $f(t) = t \log t$ gives the **Umegaki Relative Entropy** $D(\rho\|\sigma) := D_f(\rho\|\sigma)$.
- $f_\alpha(t) = t^\alpha$ for $\alpha \in (0,1) \cup (1,\infty)$ gives the **Petz-Rényi α-relative entropy** $D_\alpha(\rho||\sigma) := \frac{1}{\alpha-1} \log D_{f_\alpha}(\rho\|\sigma)$.
- $f_\alpha(t) = \frac{t^\alpha - 1}{\alpha-1}$ for $\alpha \in (0,1) \cup (1,\infty)$ gives the **quantum Hellinger α-divergence**.
- $f(t) = |t-1|$ gives the **quantum total variation** $V(\rho||\sigma) := D_f(\rho\|\sigma)$.
- $f(t) = (t-1)^2$ gives the **quantum χ^2-divergence** $\chi^2(\rho||\sigma) := D_f(\rho\|\sigma)$.

The Nussbaum-Szkoła distributions

Definition (The Nussbaum-Szkoła distributions)

Let $\mathcal{H}$ be a Hilbert space. Let ρ and σ be states on $\mathcal{B}(\mathcal{H})$ with spectral decompositions

$$\rho = \sum_{i \in \mathcal{I}} r_i \, |u_i\rangle\langle u_i| \quad \text{and} \quad \sigma = \sum_{j \in \mathcal{I}} s_j \, |v_j\rangle\langle v_j| \, .$$

Define the **Nussbaum-Szkoła distributions** *P and Q associated with ρ and σ on $\mathcal{I} \times \mathcal{I}$ by,*

$$P(i,j) = r_i |\langle u_i | v_j \rangle|^2 \text{ and } Q(i,j) = s_j |\langle u_i | v_j \rangle|^2, \quad \forall (i,j) \in \mathcal{I} \times \mathcal{I}.$$

More uses of the Nussbaum-Szkoła distributions

Theorem (G.A., T.C.John)

Let $\mathcal{H}$ be a Hilbert space and ρ, σ be states on $\mathcal{B}(\mathcal{H})$. Let P, Q be the Nussbaum-Szkoła distributions associated with ρ and σ. Let $f : (0, \infty) \to \mathbb{R}$ be a convex or concave function. Then

$$D_f(\rho||\sigma) = D_f(P||Q).$$

Applications: f-divergence inequalities (G.A. T.C. John)

- $$\mathcal{H}^2(P\|Q) \leq V(P\|Q)^2 \leq \mathcal{H}(P\|Q)\sqrt{2-\mathcal{H}^2(P\|Q)},$$
  hence
  $$\mathcal{H}^2(\rho\|\sigma) \leq V(\rho\|\sigma)^2 \leq \mathcal{H}(\rho\|\sigma)\sqrt{2-\mathcal{H}^2(\rho\|\sigma)}.$$
- $$D(P\|Q) \leq \log\left(1+\chi^2(P\|Q)\right),$$
  hence
  $$D(\rho\|\sigma) \leq \log\left(1+\chi^2(\rho\|\sigma)\right).$$
- For $0 < \alpha < 1 < \beta < \infty$,
  $$\mathcal{H}_\alpha(P\|Q) \leq D_\alpha(P\|Q) \leq D(P\|Q) \leq D_\beta(P\|Q) \leq \mathcal{H}_\beta(P\|Q),$$
  hence
  $$\mathcal{H}_\alpha(\rho\|\sigma) \leq D_\alpha(\rho\|\sigma) \leq D(\rho\|\sigma) \leq D_\beta(\rho\|\sigma) \leq \mathcal{H}_\beta(\rho\|\sigma).$$
- etc.

The Fock space in 1-mode

Definition (The Fock space of 1-mode and its basis)

The **Fock space of 1-mode** *is* ℓ_2*. Its basis* $(e_n)_{n\in\mathbb{N}\cup\{0\}}$ *is called* **the particle basis***.*

Definition (The exponential vectors)

Define the **exponential map** $e : \mathbb{C} \to \ell_2$ *by*

$$e(h) = \sum_{n=0}^{\infty} \frac{h^n}{\sqrt{n!}} e_n.$$

The collection $(e(h))_{h\in\mathbb{C}}$ *is normalized, linearly independent, and total in* ℓ_2*.*

Definition of the Weyl map

Definition (The Weyl map in 1-mode)

Define the **Weyl map** $W : \mathbb{C} \to \mathcal{B}(\ell_2)$ *by*

$$W(\mu)e(h) = e^{-\frac{1}{2}|\mu|^2 - \langle \mu | h \rangle} e(\mu + h).$$

Then, $W(\mu)$ *is a unitary operator and* $W(\mu_1)W(\mu_2) = e^{-i \operatorname{Im}\langle \mu_1 | \mu_2 \rangle} W(\mu_1 + \mu_2)$ *(projective representation).*

The definition of Gaussian states

Definition (The quantum characteristic function of a state in 1-mode)

$$\widehat{\rho}(h) = Tr(W(h)\rho) \quad \textit{for} \quad h \in \mathbb{C}.$$

Definition (Gaussian states in 1-mode)

The state ρ *is* **Gaussian** *with* **mean** $\mu \in \mathbb{C}$ *and* **covariance** S *(symmetric operator acting on the real vector space* $\mathbb{C}$*) if*

$$\widehat{\rho}(h) = e^{-2i\,\mathrm{Im}\langle h|\mu\rangle - \mathrm{Re}\langle h|Sh\rangle}, \textit{ for all } h \in \mathbb{C}.$$

Examples of Gaussian states

Definition

Thermal state *with inverse temperature* $t \in (0, \infty]$ *in 1-mode:*

$$\gamma(t) = (1 - e^{-t}) \sum_{k=0}^{\infty} e^{tk} |k\rangle\langle k| = (1 - e^{-t}) \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & e^{-t} & 0 & 0 & \cdots \\ 0 & 0 & e^{-2t} & 0 & \cdots \\ 0 & 0 & 0 & e^{-3t} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Definition (Displaced faithful thermal states)

$$\rho = W(\mu)\gamma(t)W(\mu)^{\dagger}$$

for some $\mu \in \mathbb{C}$ *and* $t \in (0, \infty)$.

Applications to Gaussian states

Theorem (G.A., T.C.John)

Petz-Rényi divergence of two 1-mode displaced faithful thermal states

$$\rho = W(\mu_1)\gamma(r)W(\mu_1)^\dagger \text{ and } \sigma = W(\mu_2)\gamma(s)W(\mu_2)^\dagger.$$

Then

$$D_\alpha(\rho||\sigma) < \infty \Leftrightarrow \alpha \in \begin{cases} (0,1) \cup (1, \frac{s}{s-r}) \text{ if } r < s \\ (0,1) \cup (1, \infty) \text{otherwise} \end{cases}.$$

A similar result holds in *n*-modes as well. This helped us to prove a special case of a conjecture of Seshadreesan-Lami-Wilde (2018).

Open Questions

- Given two Gaussian states ρ and σ determine the exact range of α such that $D_\alpha(\rho||\sigma) < \infty$. It is known that this set always contains $(0,1]$.
- Construct states in von Neumann algebras whose "Nussbaum-Szkoła distributions" are continuous. I.e. for any two states ρ and σ in a family of states of a von Neumann algebra construct continuous probability distributions P and Q such that

$$D_f(\rho||\sigma) = D_f(P||Q)$$

for every convex or concave function $f : (0,\infty) \to \mathbb{R}$.

Thank you!