Connections between classical and quantum information theory

George Androulakis

University of South Carolina

7th MATH@NTUA Summer School
in honor of Spiros Argyros

Outline

How to quantify classical information?

Definition

A **classical state** *is a probability distribution P on a finite set $\mathcal{X}$ of symbols.*

How to quantify classical information?

Definition

A **classical state** *is a probability distribution P on a finite set $\mathcal{X}$ of symbols.*

Question: How much information is contained in a classical state?

An i.i.d. classical source created from a classical state

From the classical state P create an **independent and identically distributed (i.i.d.) classical source** (or **stochastic process**) $(X_n)_{n\in\mathbb{N}}$ such that

- the random variables (r.v.) X_n's are independent, and
- the probability distribution of each X_n is equal to P.

A binary block encoding-decoding of the i.i.d. classical source generated by the classical state

A **binary block encoding** of the classical source $(X_k)_{k\in\mathbb{N}}$ each having range $\mathcal{X}$, is a family of maps

$$e : \mathcal{X}^k \to \{0,1\}^n.$$

A binary block encoding-decoding of the i.i.d. classical source generated by the classical state

A **binary block encoding** of the classical source $(X_k)_{k \in \mathbb{N}}$ each having range $\mathcal{X}$, is a family of maps

$$e : \mathcal{X}^k \to \{0,1\}^n.$$

A **binary block decoding** of the classical source $(X_k)_{k \in \mathbb{N}}$ each having range $\mathcal{X}$ is a family of maps

$$d : \{0,1\}^n \to \mathcal{X}^k.$$

Probability of error of an encoding-decoding

The **probability of error** of the encoding-decoding (e, d) is defined by

$$\mathrm{Err}(e, d) = P^k\{(x_1, \dots, x_k) \in \mathcal{X}^k : d \circ e(x_1, \dots, x_k) \neq (x_1, \dots, x_k)\}.$$

Probability of error of an encoding-decoding

The **probability of error** of the encoding-decoding (e, d) is defined by

$$\mathrm{Err}(e, d) = P^k\{(x_1, \ldots, x_k) \in \mathcal{X}^k : d \circ e(x_1, \ldots, x_k) \neq (x_1, \ldots, x_k)\}.$$

Goal: Given $\varepsilon \in (0, 1)$ find an encoding-decoding (e, d) such that

- The fraction $\frac{n}{k}$ is as small as possible, and
- $\mathrm{Err}(e, d) \leq \varepsilon$.

Asymptotic minimum number of bits per symbol

$$n(k,\varepsilon) := \min \Big\{ n \,|\, \exists e : \mathcal{X}^k \to \{0,1\}^n \text{ and } d : \{0,1\}^n \to \mathcal{X}^k \text{ such that } \mathrm{Err}(e,d) \le \varepsilon \Big\}.$$

Asymptotic minimum number of bits per symbol

$$n(k,\varepsilon) := \min\Big\{ n \,|\, \exists e : \mathcal{X}^k \to \{0,1\}^n \text{ and } d : \{0,1\}^n \to \mathcal{X}^k \\ \text{such that } \mathrm{Err}(e,d) \le \varepsilon \Big\}.$$

$\frac{n(k,\varepsilon)}{k}$ =minimum number of bits per symbol needed in order to block encode k many i.i.d. symbols emitted from the classical source, if the encoding-decoding error stays upper bounded by ε.

Asymptotic minimum number of bits per symbol

$$n(k,\varepsilon) := \min\big\{n \,|\, \exists e : \mathcal{X}^k \to \{0,1\}^n \text{ and } d : \{0,1\}^n \to \mathcal{X}^k \text{ such that } \mathrm{Err}(e,d) \le \varepsilon\big\}.$$

$\frac{n(k,\varepsilon)}{k}$ =minimum number of bits per symbol needed in order to block encode k many i.i.d. symbols emitted from the classical source, if the encoding-decoding error stays upper bounded by ε.

$\lim_{k\to\infty} \frac{n(k,\varepsilon)}{k}$ =asymptotic minimum number of bits per symbol needed to block encode i.i.d. symbols emitted from the classical source, if the encoding-decoding error stays upper bounded by ε.

Information contained in a classical state

Definition

Information contained in a classical state := Asymptotic minimum number of bits per symbol needed for the block encoding of the corresponding i.i.d. classical source, if the probability of error is arbitrarily small

$$= \lim_{\varepsilon \to 0} \lim_{k \to \infty} \frac{n(k, \varepsilon)}{k}.$$

Classical Noiseless Coding Theorem

Definition (The entropy of a classical state P)

$$H(P) = -\sum_i p_i \log_2 p_i.$$

Classical Noiseless Coding Theorem

Definition (The entropy of a classical state P)

$$H(P) = -\sum_i p_i \log_2 p_i.$$

Theorem (Classical Noiseless Coding Theorem, C. Shannon 1948)

The information contained in a classical state P is equal to $H(P)$,

Classical Noiseless Coding Theorem

Definition (The entropy of a classical state P)

$$H(P) = -\sum_i p_i \log_2 p_i.$$

Theorem (Classical Noiseless Coding Theorem, C. Shannon 1948)

The information contained in a classical state P is equal to $H(P)$, i.e.

$$\lim_{\varepsilon \to 0} \lim_{k \to \infty} \frac{n(k,\varepsilon)}{k} = H(P),$$

Classical Noiseless Coding Theorem

Definition (The entropy of a classical state P)

$$H(P) = -\sum_i p_i \log_2 p_i .$$

Theorem (Classical Noiseless Coding Theorem, C. Shannon 1948)

The information contained in a classical state P is equal to $H(P)$, i.e.

$$\lim_{\varepsilon \to 0} \lim_{k \to \infty} \frac{n(k,\varepsilon)}{k} = H(P),$$

i.e. **Achievability:** *For every $\varepsilon > 0$ and $k \in \mathbb{N}$ there exists a block encoding-decoding of k many emissions of the classical source generated by P into $kH(P)$ many bits with probability or error at most ε.*
Converse: *If fewer than $kH(P)$ bits are used to encode k many emissions of the classical source generated by P as $k \to \infty$, then the probability of error will stay bounded from below by a positive number.*

Main ingredient of the proof

Define the **typical sets**:

$$T_{k,\delta} = \left\{ (x_{i_1}, \ldots, x_{i_k}) \in \mathcal{X}^k : 2^{-k(H(P)+\delta)} \leq p_{i_1} \cdots p_{i_k} \leq 2^{-k(H(P)-\delta)} \right\}.$$

Main ingredient of the proof

Define the **typical sets**:

$$T_{k,\delta} = \left\{ (x_{i_1}, \dots, x_{i_k}) \in \mathcal{X}^k : 2^{-k(H(P)+\delta)} \leq p_{i_1} \cdots p_{i_k} \leq 2^{-k(H(P)-\delta)} \right\}.$$

Then,

- $P^k(T_{k,\delta}) \to 1$ as $k \to \infty$.

Main ingredient of the proof

Define the **typical sets**:

$$T_{k,\delta} = \left\{ (x_{i_1}, \ldots, x_{i_k}) \in \mathcal{X}^k : 2^{-k(H(P)+\delta)} \leq p_{i_1} \cdots p_{i_k} \leq 2^{-k(H(P)-\delta)} \right\}.$$

Then,

- $P^k(T_{k,\delta}) \to 1$ as $k \to \infty$.
- $\#(T_{k,\delta}) \leq 2^{k(H(P)+\delta)}$.

How do you quantify quantum information?

Definition (Dirac Notation)

Ket *denotes a (column) vector* $|y\rangle = \begin{pmatrix} y_1 \\ \vdots \\ y_D \end{pmatrix} \in \mathbb{C}^D$. **Bra** *denotes the complex conjugate and transpose of the ket, i.e.* $\langle y| = \left(\overline{y_1} \cdots \overline{y_D}\right)$.

Definition

A quantum state ρ *contains a probability distribution on a finite set of rank-*1 *projections, i.e.* $\{p_i, |x_i\rangle\langle x_i|\}_{i=1}^{\ell}$, *where* $|x_i\rangle$*'s are normalized (not necessarily linearly independent) vectors in the Hilbert space* $\mathbb{C}^D$. *Thus,*

$$\rho = \sum_{i=1}^{\ell} p_i \, |x_i\rangle\langle x_i| \, .$$

Question: How much information is contained in a quantum state?

An i.i.d. quantum source created from a quantum state

From the quantum state $\rho = \sum_{i=1}^{\ell} p_i \, |x_i\rangle\langle x_i|$ create an i.i.d. **quantum source** $(X_k)_{k\in\mathbb{N}}$ such that

- the X_k's are independent, and
- X_k takes the value $|x_i\rangle\langle x_i|$ with probability p_i for all $i = 1, \ldots, \ell$ and $k \in \mathbb{N}$.

Combined emissions from the quantum source

Definition (Tensor product of vectors)

$$|y\rangle \otimes |z\rangle = |yz\rangle = (y_1, \cdots, y_D)^\intercal \otimes (z_1, \cdots, z_D)^\intercal$$
$$= (y_1 z_1, \cdots, y_1 z_D, y_2 z_1, \cdots, y_2 z_D, \cdots, y_D z_D)^\intercal \in \mathbb{C}^{D^2}.$$

Combined emissions from the quantum source

Definition (Tensor product of vectors)

$$|y\rangle \otimes |z\rangle = |yz\rangle = (y_1, \cdots, y_D)^\intercal \otimes (z_1, \cdots, z_D)^\intercal$$
$$= (y_1 z_1, \cdots, y_1 z_D, y_2 z_1, \cdots, y_2 z_D, \cdots, y_D z_D)^\intercal \in \mathbb{C}^{D^2}.$$

Definition (Tensor product of matrices)

$$(a_{i,j})_{i,j} \otimes (b_{k,l})_{k,l} = (a_{i,j} B)_{i,j} = \begin{pmatrix} a_{1,1} b_{1,1} & a_{1,1} b_{1,2} & \cdots \\ a_{1,1} b_{2,1} & a_{1,1} b_{2,2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Combined emissions from the quantum source

Definition (Tensor product of vectors)

$$|y\rangle \otimes |z\rangle = |yz\rangle = (y_1, \cdots, y_D)^\intercal \otimes (z_1, \cdots, z_D)^\intercal$$
$$= (y_1 z_1, \cdots, y_1 z_D, y_2 z_1, \cdots, y_2 z_D, \cdots, y_D z_D)^\intercal \in \mathbb{C}^{D^2}.$$

Definition (Tensor product of matrices)

$$(a_{i,j})_{i,j} \otimes (b_{k,l})_{k,l} = (a_{i,j} B)_{i,j} = \begin{pmatrix} a_{1,1} b_{1,1} & a_{1,1} b_{1,2} & \cdots \\ a_{1,1} b_{2,1} & a_{1,1} b_{2,2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Definition (Tensor product of rank-1 projections)

Total emission from the quantum source after k-many emissions:
$|x_{i_1}\rangle\langle x_{i_1}| \otimes \cdots \otimes |x_{i_k}\rangle\langle x_{i_k}| = |x_{i_1} \cdots x_{i_k}\rangle\langle x_{i_1} \cdots x_{i_k}|$, *(it is a $D^k \times D^k$ matrix).*

A qubit block encoding-decoding of the i.i.d. quantum source generated by the quantum state ρ

A **qubit block encoding** of the quantum source $(X_k)_{k\in\mathbb{N}}$ is a family of maps

$$e : \{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \to \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n}$$

which extend linearly from Span $\left(\{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k}\right)$ to $\mathbb{C}^{2^n}$.

Notation: $|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2$ and $|1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{C}^2$ are called **qubits**.

A qubit block encoding-decoding of the i.i.d. quantum source generated by the quantum state ρ

A **qubit block encoding** of the quantum source $(X_k)_{k\in\mathbb{N}}$ is a family of maps

$$e : \{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \to \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n}$$

which extend linearly from $\mathrm{Span}\left(\{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k}\right)$ to $\mathbb{C}^{2^n}$.

Notation: $|0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2$ and $|1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{C}^2$ are called **qubits**.

A **qubit block decoding** of the quantum source $(X_k)_{k\in\mathbb{N}}$ is a family of maps

$$d : \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n} \to \{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k}$$

which extend linearly from $\mathrm{Span}(\{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n})$ to $\mathrm{Span}\left(\{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k}\right)$.

Probability of error of an encoding-decoding

The **probability of error** of the encoding-decoding (e, d) is defined by

$$\mathrm{Err}(e, d) =$$
$$1 - \sum_{i_1,\ldots,i_k} p_{i_1} \cdots p_{i_k} F\left(|x_{i_1} \cdots x_{i_k}\rangle\langle x_{i_1} \cdots x_{i_k}|\,, d \circ e(|x_{i_1} \cdots x_{i_k}\rangle\langle x_{i_1} \cdots x_{i_k}|)\right)$$

where the **fidelity** between two rank-1 projections is defined as

$$F\left(|u\rangle\langle u|\,, |v\rangle\langle v|\right) = |\,\langle u|v\rangle\,|^2.$$

Probability of error of an encoding-decoding

The **probability of error** of the encoding-decoding (e, d) is defined by

$$\mathrm{Err}(e,d) =$$
$$1 - \sum_{i_1,\ldots,i_k} p_{i_1}\cdots p_{i_k} F\left(|x_{i_1}\cdots x_{i_k}\rangle\langle x_{i_1}\cdots x_{i_k}|\,, d\circ e(|x_{i_1}\cdots x_{i_k}\rangle\langle x_{i_1}\cdots x_{i_k}|)\right)$$

where the **fidelity** between two rank-1 projections is defined as

$$F\left(|u\rangle\langle u|\,, |v\rangle\langle v|\right) = |\,\langle u|v\rangle\,|^2.$$

Goal: Given $\varepsilon \in (0,1)$ find an encoding-decoding (e, d) such that

- The fraction $\frac{n}{k}$ is as small as possible, and
- $\mathrm{Err}(e,d) \le \varepsilon$.

Asymptotic minimum number of qubits/symbol

$$n(k,\varepsilon) := \min\Big\{ n \,\big|\, \exists e : \{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \to \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n} \text{ and } d : \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n} \to \{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \text{ such that } \mathrm{Err}(e,d) \le \varepsilon \Big\}.$$

Asymptotic minimum number of qubits/symbol

$$n(k,\varepsilon) := \min\Big\{n \,|\, \exists e : \{|x_1\rangle\langle x_1|, \ldots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \to \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n} \\ \text{and } d : \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n} \to \{|x_1\rangle\langle x_1|, \ldots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \\ \text{such that } \mathrm{Err}(e,d) \leq \varepsilon\Big\}.$$

$\frac{n(k,\varepsilon)}{k}$ =minimum number of qubits per symbol needed for encoding k many i.i.d. symbols emitted from the quantum source, if the encoding-decoding error stays upper bounded by ε.

Asymptotic minimum number of qubits/symbol

$$n(k,\varepsilon) := \min\Big\{ n \,|\, \exists e : \{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \to \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n} \text{ and } d : \{|0\rangle\langle 0|, |1\rangle\langle 1|\}^{\otimes n} \to \{|x_1\rangle\langle x_1|, \dots, |x_\ell\rangle\langle x_\ell|\}^{\otimes k} \text{ such that } \mathrm{Err}(e,d) \le \varepsilon \Big\}.$$

$\frac{n(k,\varepsilon)}{k}$ = minimum number of qubits per symbol needed for encoding k many i.i.d. symbols emitted from the quantum source, if the encoding-decoding error stays upper bounded by ε.

$\lim_{k\to\infty} \frac{n(k,\varepsilon)}{k}$ = asymptotic minimum number of qubits per symbol needed for encoding i.i.d. symbols emitted from the quantum source, if the encoding-decoding error stays upper bounded by ε.

Information contained in a quantum state

Definition

Information contained in a quantum state $:=$ *Asymptotic minimum number of qubits per symbol needed for a block encoding of the corresponding i.i.d. quantum source, while the probability of error is arbitrarily small*

$$= \lim_{\varepsilon \to 0} \lim_{k \to \infty} \frac{n(k, \varepsilon)}{k}.$$

Quantum Noiseless Coding Theorem

Definition

The **quantum entropy** $S(\rho)$ *of a quantum state* ρ *is given by*

$$S(\rho) = -Tr(\rho \log_2 \rho).$$

Quantum Noiseless Coding Theorem

Definition

The **quantum entropy** $S(\rho)$ *of a quantum state* ρ *is given by*

$$S(\rho) = -Tr(\rho \log_2 \rho).$$

Theorem (Quantum Noiseless Coding Theorem, B. Schumacher 1995)

The information contained in a quantum state ρ *is equal to* $S(\rho)$*,*

Quantum Noiseless Coding Theorem

Definition

The **quantum entropy** $S(\rho)$ *of a quantum state* ρ *is given by*

$$S(\rho) = -Tr(\rho \log_2 \rho).$$

Theorem (Quantum Noiseless Coding Theorem, B. Schumacher 1995)

The information contained in a quantum state ρ *is equal to* $S(\rho)$*, i.e.*

$$\lim_{\varepsilon \to 0} \lim_{k \to \infty} \frac{n(k, \varepsilon)}{k} = S(\rho).$$

Quantum Noiseless Coding Theorem

Definition

The **quantum entropy** $S(\rho)$ *of a quantum state* ρ *is given by*

$$S(\rho) = -Tr(\rho \log_2 \rho).$$

Theorem (Quantum Noiseless Coding Theorem, B. Schumacher 1995)

The information contained in a quantum state ρ *is equal to* $S(\rho)$, *i.e.*

$$\lim_{\varepsilon \to 0} \lim_{k \to \infty} \frac{n(k, \varepsilon)}{k} = S(\rho).$$

i.e. **Achievability:** *For every* $\varepsilon > 0$ *and* $k \in \mathbb{N}$ *there exists a block encoding of* k *many emissions of the quantum source generated by* ρ *into* $kS(\rho)$ *many qubits with probability or error at most* ε.

Converse: *If fewer than* $kS(\rho)$ *qubits are used to encode* k *many symbols emitted by the quantum source generated by* ρ, *as* $k \to \infty$, *then the probability of error will stay bounded below by a positive number.*

Main ingredient of the proof

Define the **typical sets**:

$$T_{k,\delta} = \left\{ |x_{i_1}, \ldots, x_{i_k}\rangle\langle x_{i_1}, \ldots, x_{i_k}| : 2^{-k(S(\rho)+\delta)} \leq p_{i_1} \cdots p_{i_k} \leq 2^{-k(S(\rho)-\delta)} \right\}.$$

and

$$\Pi_{k,\delta} = \text{the orthogonal projection to the span of } |x_{i_1} \cdots x_{i_k}\rangle \text{'s for all } |x_{i_1} \cdots x_{i_k}\rangle\langle x_{i_1} \cdots x_{i_k}| \in T_{k,\delta}.$$

Main ingredient of the proof

Define the **typical sets**:

$$T_{k,\delta} = \left\{ |x_{i_1}, \ldots, x_{i_k}\rangle\langle x_{i_1}, \ldots, x_{i_k}| : 2^{-k(S(\rho)+\delta)} \leq p_{i_1} \cdots p_{i_k} \leq 2^{-k(S(\rho)-\delta)} \right\}.$$

and

$$\Pi_{k,\delta} = \text{the orthogonal projection to the span of } |x_{i_1} \cdots x_{i_k}\rangle\text{'s for all } |x_{i_1} \cdots x_{i_k}\rangle\langle x_{i_1} \cdots x_{i_k}| \in T_{k,\delta}.$$

Then,

- $\mathrm{Tr}\left(\Pi_{k,\delta}\rho^{\otimes k}\right) \to 1$ as $k \to \infty$.

Main ingredient of the proof

Define the **typical sets**:

$$T_{k,\delta} = \left\{ |x_{i_1}, \ldots, x_{i_k}\rangle\langle x_{i_1}, \ldots, x_{i_k}| : 2^{-k(S(\rho)+\delta)} \leq p_{i_1} \cdots p_{i_k} \leq 2^{-k(S(\rho)-\delta)} \right\}.$$

and

$$\Pi_{k,\delta} = \text{the orthogonal projection to the span of } |x_{i_1} \cdots x_{i_k}\rangle \text{'s for all } |x_{i_1} \cdots x_{i_k}\rangle\langle x_{i_1} \cdots x_{i_k}| \in T_{k,\delta}.$$

Then,

- $\mathrm{Tr}\left(\Pi_{k,\delta}\rho^{\otimes k}\right) \to 1$ as $k \to \infty$.
- $\dim\left(\Pi_{k,\delta}\right) \leq 2^{k(S(P)+\delta)}$.

How to distinguish two classical states?

Consider two known classical states P, Q. You are presented with an n many i.i.d. draws of a random variable X such that either $X \sim P$ or $X \sim Q$, and you need to decide the probability distribution of X.

Assume that the random variable X takes values in a set $\mathcal{X}$. You choose a subset A_n of $\mathcal{X}^n$ which aligns with P^n. If the n draws that you are presented with belong to A_n, then you decide that $X \sim P$. Otherwise, you decide that $X \sim Q$.

Classical Asymmetric hypothesis testing

As in the previous page, consider two classical states P, Q and a random variable X such that $X \sim P$ or $X \sim Q$. We are presented with n many i.i.d. draws of X and we would like to compute the smallest probability of error while trying to figure out the distribution of X.

Classical Asymmetric hypothesis testing

As in the previous page, consider two classical states P, Q and a random variable X such that $X \sim P$ or $X \sim Q$. We are presented with n many i.i.d. draws of X and we would like to compute the smallest probability of error while trying to figure out the distribution of X.

There are two types of errors:

- **Type I error:** $X \sim P$, but we erroneously decide that $X \sim Q$.
- **Type II error:** $X \sim Q$, but we erroneously decide that $X \sim P$.

Classical Asymmetric hypothesis testing

As in the previous page, consider two classical states P, Q and a random variable X such that $X \sim P$ or $X \sim Q$. We are presented with n many i.i.d. draws of X and we would like to compute the smallest probability of error while trying to figure out the distribution of X.

There are two types of errors:

- **Type I error:** $X \sim P$, but we erroneously decide that $X \sim Q$.
- **Type II error:** $X \sim Q$, but we erroneously decide that $X \sim P$.

Let $\varepsilon > 0$. Consider all decision strategies that satisfy $\mathbb{P}(\text{Type I error}) \leq \varepsilon$.
Goal: Among all these decision strategies compute $\inf \mathbb{P}(\text{Type II error})$.

Classical Asymmetric hypothesis testing

As in the previous page, consider two classical states P, Q and a random variable X such that $X \sim P$ or $X \sim Q$. We are presented with n many i.i.d. draws of X and we would like to compute the smallest probability of error while trying to figure out the distribution of X.

There are two types of errors:

- **Type I error:** $X \sim P$, but we erroneously decide that $X \sim Q$.
- **Type II error:** $X \sim Q$, but we erroneously decide that $X \sim P$.

Let $\varepsilon > 0$. Consider all decision strategies that satisfy $\mathbb{P}(\text{Type I error}) \leq \varepsilon$.
Goal: Among all these decision strategies compute $\inf \mathbb{P}(\text{Type II error})$.

Let $\text{ran}\,(X) = \mathcal{X}$. A **decision strategy** is a subset A_n of $\mathcal{X}^n$ such that when the sequence of n draws of X belongs to A_n, then we decide that $X \sim P$; otherwise we decide that $X \sim Q$.

Classical Asymmetric hypothesis testing

As in the previous page, consider two classical states P, Q and a random variable X such that $X \sim P$ or $X \sim Q$. We are presented with n many i.i.d. draws of X and we would like to compute the smallest probability of error while trying to figure out the distribution of X.

There are two types of errors:

- **Type I error:** $X \sim P$, but we erroneously decide that $X \sim Q$.
- **Type II error:** $X \sim Q$, but we erroneously decide that $X \sim P$.

Let $\varepsilon > 0$. Consider all decision strategies that satisfy $\mathbb{P}(\text{Type I error}) \leq \varepsilon$.
Goal: Among all these decision strategies compute $\inf \mathbb{P}(\text{Type II error})$.

Let $\operatorname{ran}(X) = \mathcal{X}$. A **decision strategy** is a subset A_n of $\mathcal{X}^n$ such that when the sequence of n draws of X belongs to A_n, then we decide that $X \sim P$; otherwise we decide that $X \sim Q$.

$\mathbb{P}(\text{Type I error}) = P^n(\mathcal{X}^n \setminus A_n)$, $\mathbb{P}(\text{Type II error}) = Q^n(A_n)$.

The Question

Question

Compute the smallest "average" probability of Type II error for the asymmetric classical hypothesis testing, i.e.

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log_2 \inf_{\substack{A_n\subseteq \mathcal{X}^n,\\ P^n(\mathcal{X}^n\setminus A_n)\le\varepsilon}} Q^n(A_n).$$

Stein's Lemma

Definition (Kullback-Leibler Divergence (1951))

$$D(P||Q) = \begin{cases} \sum_i P(i) \log_2 \frac{P(i)}{Q(i)} & \textit{if } P \ll Q \\ \infty & \textit{otherwise} \end{cases}$$

Stein's Lemma

Definition (Kullback-Leibler Divergence (1951))

$$D(P||Q) = \begin{cases} \sum_i P(i) \log_2 \frac{P(i)}{Q(i)} & \textit{if } P \ll Q \\ \infty & \textit{otherwise} \end{cases}$$

Theorem ("Stein's Lemma", R. Blahut ($\leq$) (1974), T.S. Han, K. Kobayashi ($\geq$) (1989))

Let P, Q be two probability distributions on a set $\mathcal{X}$, and you are presented with a sequence of i.i.d. draws of a r.v. X such that $X \sim P$ or $X \sim Q$ and the range of X is equal to $\mathcal{X}$, then

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log_2 \inf_{\substack{A_n \subseteq \mathcal{X}^n, \\ P^n(\mathcal{X}^n \setminus A_n) \leq \varepsilon}} Q^n(A_n) = -D(P||Q).$$

Main ingredient of the proof

Define the **typical sets**:

$$T_{n,\delta} = \left\{ (x_{i_1}, \ldots, x_{i_n}) \in \mathcal{X}^n : 2^{n(D(P||Q)-\delta)} \leq \frac{p_{i_1} \cdots p_{i_n}}{q_{i_1} \cdots q_{i_n}} \leq 2^{n(D(P||Q)+\delta)} \right\}.$$

Main ingredient of the proof

Define the **typical sets**:

$$T_{n,\delta} = \left\{ (x_{i_1}, \ldots, x_{i_n}) \in \mathcal{X}^n : 2^{n(D(P||Q)-\delta)} \leq \frac{p_{i_1} \cdots p_{i_n}}{q_{i_1} \cdots q_{i_n}} \leq 2^{n(D(P||Q)+\delta)} \right\}.$$

Then,

- $P^n(T_{n,\delta}) \to 1$ as $n \to \infty$.

Main ingredient of the proof

Define the **typical sets**:

$$T_{n,\delta} = \left\{ (x_{i_1}, \dots, x_{i_n}) \in \mathcal{X}^n : 2^{n(D(P||Q)-\delta)} \leq \frac{p_{i_1} \cdots p_{i_n}}{q_{i_1} \cdots q_{i_n}} \leq 2^{n(D(P||Q)+\delta)} \right\}.$$

Then,

- $P^n(T_{n,\delta}) \to 1$ as $n \to \infty$.
- $Q^n(T_{n,\delta}) \leq 2^{-n(D(P||Q)-\delta)}$.

How to distinguish two quantum states?

Postulate (Postulate of Quantum Mechanics)

Given a quantum state τ*, and* $0 \leq A \leq 1$*, then* $Tr(\tau A)$ *is equal to the probability that when we measure the state* τ *we find that it aligns with* A*.*

How to distinguish two quantum states?

Postulate (Postulate of Quantum Mechanics)

Given a quantum state τ, and $0 \le A \le 1$, then $Tr(\tau A)$ is equal to the probability that when we measure the state τ we find that it aligns with A.

Consider two known $(D \times D)$ quantum states ρ, σ. You are presented with an unknown $D^n \times D^n$ matrix ? which is either equal to $\rho^{\otimes n}$ or $\sigma^{\otimes n}$, and you need to decide whether $? = \rho^{\otimes n}$ or $? = \sigma^{\otimes n}$. Even though you do not know the matrix ? you can evaluate $\mathrm{Tr}(?A_n)$ (probabilities!) for any $D^n \times D^n$ matrix A_n that satisfies $0 \le D_n \le 1$.

How to distinguish two quantum states?

Postulate (Postulate of Quantum Mechanics)

Given a quantum state τ, and $0 \leq A \leq 1$, then $Tr(\tau A)$ is equal to the probability that when we measure the state τ we find that it aligns with A.

Consider two known $(D \times D)$ quantum states ρ, σ. You are presented with an unknown $D^n \times D^n$ matrix ? which is either equal to $\rho^{\otimes n}$ or $\sigma^{\otimes n}$, and you need to decide whether $? = \rho^{\otimes n}$ or $? = \sigma^{\otimes n}$. Even though you do not know the matrix ? you can evaluate $\mathrm{Tr}(?A_n)$ (probabilities!) for any $D^n \times D^n$ matrix A_n that satisfies $0 \leq D_n \leq 1$.

You choose a $D^n \times D^n$ matrix A_n with $0 \leq A_n \leq 1$ and aligns with $\rho^{\otimes n}$ and evaluate $\mathrm{Tr}(?A_n)$ in order to check whether ? aligns with $\rho^{\otimes n}$. If it does, you decide that $? = \rho^{\otimes n}$. Otherwise, you decide that $? = \sigma^{\otimes n}$.

Quantum Asymmetric hypothesis testing

As in the previous page, consider an unknown $D^n \times D^n$ matrix ? which is either equal to $\rho^{\otimes n}$ or $\sigma^{\otimes n}$ and you are trying to decide which of the two cases is correct by choosing appropriate matrix A_n with $0 \leq A_n \leq 1$ (decision strategy) and evaluating $\mathrm{Tr}(?A_n)$.

Quantum Asymmetric hypothesis testing

As in the previous page, consider an unknown $D^n \times D^n$ matrix ? which is either equal to $\rho^{\otimes n}$ or $\sigma^{\otimes n}$ and you are trying to decide which of the two cases is correct by choosing appropriate matrix A_n with $0 \leq A_n \leq 1$ (decision strategy) and evaluating $\mathrm{Tr}(?A_n)$.

- **Type I error:** $? = \rho^{\otimes n}$, but we erroneously decide that $? = \sigma^{\otimes n}$.
- **Type II error:** $? = \sigma^{\otimes n}$, but we erroneously decide that $? = \rho^{\otimes n}$.

Quantum Asymmetric hypothesis testing

As in the previous page, consider an unknown $D^n \times D^n$ matrix ? which is either equal to $\rho^{\otimes n}$ or $\sigma^{\otimes n}$ and you are trying to decide which of the two cases is correct by choosing appropriate matrix A_n with $0 \leq A_n \leq 1$ (decision strategy) and evaluating $\mathrm{Tr}(?A_n)$.

- **Type I error:** $? = \rho^{\otimes n}$, but we erroneously decide that $? = \sigma^{\otimes n}$.
- **Type II error:** $? = \sigma^{\otimes n}$, but we erroneously decide that $? = \rho^{\otimes n}$.

Given $\varepsilon > 0$ you consider all $D^n \times D^n$ matrices (decision strategies) A_n which satisfy

$$0 \leq A_n \leq 1 \text{ and } \mathrm{Tr}\left(\rho^{\otimes n}(1 - A_n)\right) \leq \varepsilon, \text{ i.e. } \mathbb{P}(\text{Type I error}) \leq \varepsilon.$$

Among all of these matrices A_n compute the

$$\inf \mathrm{Tr}\left(\sigma^{\otimes n} A_n\right) \text{ i.e. } \inf \mathbb{P}(\text{Type II error}).$$

The Question

Question

Compute the smallest "average" probability of Type II error for the asymmetric quantum hypothesis testing, i.e.

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log_2 \inf_{\substack{0\leq A_n\leq 1,\\ Tr(\rho^{\otimes n}(1-A_n))\leq\varepsilon}} Tr(\sigma^{\otimes n}A_n).$$

Quantum Stein's Lemma

Definition (Umegaki relative entropy (1962))

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

Theorem ("Quantum Stein's Lemma", Hiai-Petz ($\leq$) (1991), Ogawa-Nagaoka ($\geq$) (2000))

For the quantum asymmetric hypothesis testing between two states ρ and σ, the asymptotic smallest "average" Type II error is given by:

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log_2 \inf_{\substack{0\leq A_n\leq 1,\\ Tr(\rho^{\otimes n}(1-A_n))\leq\varepsilon}} Tr(\sigma^{\otimes n}A_n) = -D(\rho||\sigma).$$

Main ingredient of the proof

By "sandwiching" $\rho^{\otimes n}$ with the eigenprojections of $\sigma^{\otimes n}$, one may assume that the two states have the same eigenvectors, thus they are simultaneously diagonalizable. Hence, the (classical) Stein's Lemma can be used.

The Nussbaum-Szkoła distributions

Definition (Nussbaum-Szkoła distributions P and Q)

Let

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

be the spectral decompositions of ρ *and* σ*. Then*

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

The Nussbaum-Szkoła distributions

Definition (Nussbaum-Szkoła distributions P and Q)

Let

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

be the spectral decompositions of ρ and σ. Then

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

Theorem (Nussbaum and Szkoła (2009))

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

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

Classical f-divergences

Definition (Cziszár (1963))

Let P, Q be probability distributions on a common measure space. 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 probability distributions.

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

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

Special cases of f-divergences

Assume that P and Q are discrete probability distributions.

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

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

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

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

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

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

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

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

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

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

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

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

- $f(t) = |t-1|$ 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_σ: *the projection on the* $\mathrm{supp}(\sigma)$, *(if* σ *is a quantum state).*

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_σ: the projection on the $\text{supp}(\sigma)$, (if σ is a quantum state).*

Definition (Araki (1977))

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

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

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

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 quantum states 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** $\chi^2(\rho||\sigma) := D_f(\rho\|\sigma)$.

Generalized Nussbaum-Szkoła distributions

Definition (Generalized 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,*

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

The use of the generalized 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

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

An Open Question

Question

Are there "continuous Nussbaum-Szkoła distributions" and what are their applications?

Thank you!