The Theory of Quantum Statistical Comparison and Some Applications in Quantum Information Sciences

Francesco Buscemi (Nagoya U)

The 39th Quantum Information Technologies Symposium (QIT).
RCAST, The University of Tokyo, 27 November 2018
The Original Formulation
Definition

- The **statistical model** is given by: the parameter set $\Theta$, the sample set $\mathcal{X}$, and the PDs $w(x|\theta)$.
- The **statistical decision problem**: is given by the parameter set $\Theta$, the action set $\mathcal{U}$, and the payoff function $\ell : \Theta \times \mathcal{U} \rightarrow \mathbb{R}$. 
How Much Is an Experiment Worth?

- the experiment is given, i.e., it is the “resource”
- the decision instead can be optimized

Definition (Expected Payoff)

The expected payoff of statistical model \( w = \langle \Theta, \mathcal{X}, w(x|\theta) \rangle \) w.r.t. a decision problem \( \ell = \langle \Theta, \mathcal{U}, \ell(\theta,u) \rangle \) is given by

\[
\mathbb{E}_\ell[w] \overset{\text{def}}{=} \max \sum \ell(\theta,u) d(u|x) w(x|\theta) |\Theta|^{-1}.
\]
Comparing Statistical Models 1/2

First model: $w = \langle \Theta, \mathcal{X}, w(x|\theta) \rangle$

$\Theta \xrightarrow{\text{experiment}} \mathcal{X} \xrightarrow{\text{decision}} \mathcal{U}$

$\downarrow \quad \downarrow \quad \downarrow$

$\theta \quad x \quad d(u|x) \quad u$

Second model: $w' = \langle \Theta, \mathcal{Y}, w'(y|\theta) \rangle$

$\Theta \xrightarrow{\text{experiment}} \mathcal{Y} \xrightarrow{\text{decision}} \mathcal{U}$

$\downarrow \quad \downarrow \quad \downarrow$

$\theta \quad y \quad d'(u|y) \quad u$

Given a statistical decision problem $\ell = \langle \Theta, \mathcal{U}, \ell(\theta, u) \rangle$, if $\mathbb{E}_\ell[w] \geq \mathbb{E}_\ell[w']$, then one says that model $w$ is *more informative than* model $w'$ with respect to problem $\ell$. 
**Definition (Information Preorder)**

If model $w = \langle \Theta, X, w(x|\theta) \rangle$ is more informative than model $w' = \langle \Theta, Y, w'(y|\theta) \rangle$ for all decision problems $\ell = \langle \Theta, U, \ell(\theta, u) \rangle$, then we say that $w$ is (always) more informative than $w'$, and write

$$w \succeq w'.$$

**Problem.** The information preorder is operational, but not really “concrete”. Can we visualize this better?
A Fundamental Result

**Blackwell-Sherman-Stein (1948-1953)**

Given two models with the same parameter space, \( w = \langle \Theta, \mathcal{X}, w(x|\theta) \rangle \) and \( w' = \langle \Theta, \mathcal{Y}, w'(y|\theta) \rangle \), the condition \( w \succeq w' \) holds iff \( w \) is sufficient for \( w' \), that is, iff there exists a conditional PD \( \varphi(y|x) \) such that \( w'(y|\theta) = \sum_x \varphi(y|x) w(x|\theta) \).

\[
\begin{align*}
\Theta & \rightarrow \mathcal{Y} & \Theta & \rightarrow \mathcal{X} & \text{noise} & \rightarrow \mathcal{Y} \\
\downarrow & = & \downarrow & & \downarrow & \downarrow \\
\theta & \rightarrow y & \theta & \rightarrow x & \varphi(y|x) & \rightarrow y \\
& w'(y|\theta) & & w(x|\theta) & & \\
\end{align*}
\]

David H. Blackwell (1919-2010)
The Precursor: Majorization
Lorenz Curves and Majorization

- two probability distributions, \( p \) and \( q \), of the same dimension \( n \)
- truncated sums \( P(k) = \sum_{i=1}^{k} p_i \) and \( Q(k) = \sum_{i=1}^{k} q_i \), for all \( k = 1, \ldots, n \)
- \( p \) majorizes \( q \), i.e., \( p \succeq q \), whenever \( P(k) \geq Q(k) \), for all \( k \)
- minimal element: uniform distribution \( e = n^{-1}(1, 1, \cdots, 1) \)

**Hardy, Littlewood, and Pólya (1934)**

\( p \succeq q \iff q = Mp \), for some bistochastic matrix \( M \).

Lorenz curve for probability distribution \( p = (p_1, \cdots, p_n) \):

\[
(x_k, y_k) = \left(\frac{k}{n}, P(k)\right), \quad 1 \leq k \leq n
\]
Generalization: Relative Majorization

- two pairs of probability distributions, \((p_1, p_2)\) and \((q_1, q_2)\), of dimension \(m\) and \(n\), respectively

- relabel entries such that ratios \(p_i^1/p_i^2\) and \(q_i^1/q_i^2\) are nonincreasing

- construct the truncated sums \(P_{1,2}(k) = \sum_{i=1}^{k} p_{1,2}^i\) and \(Q_{1,2}(k)\)

- \((p_1, p_2) \succeq (q_1, q_2)\) iff the relative Lorenz curve of the former is never below that of the latter

**Blackwell (Theorem for Dichotomies), 1953**

\[(p_1, p_2) \succeq (q_1, q_2) \iff q_i = M p_i, \text{ for some stochastic matrix } M.\]
Formulation in Terms of Channels
Statistics vs Information Theory

Statistical theory
Nature does not bother with coding

\[ \Theta \xrightarrow{\text{experiment}} X \xrightarrow{\text{decision}} U \]

Communication theory
A sender, instead, does code

\[ M \xrightarrow{\text{encoding}} \Theta \xrightarrow{\text{channel}} X \xrightarrow{\text{decoding}} U \]
Statistics vs Information Theory

**Statistical theory**
Nature does not bother with coding

\[ \Theta \xrightarrow{\text{experiment}} X \xrightarrow{\text{decision}} U \]
\[ \theta \xrightarrow{w(x|\theta)} x \xrightarrow{d(u|x)} u \]

**Communication theory**
a sender, instead, *does code*

\[ M \xrightarrow{\text{encoding}} X \xrightarrow{\text{channel}} Y \xrightarrow{\text{decoding}} M \]
\[ m \xrightarrow{e(x|m)} x \xrightarrow{w(y|x)} y \xrightarrow{d(m'|y)} m' \]
Sufficiency vs Degradability

Sufficiency relation
for statistical experiments

\[ w'(y|\theta) = \sum_x \varphi(y|x)w(x|\theta) \]

Degradability relation
for noisy channels

\[ w'(z|x) = \sum_y \varphi(z|y)w(y|x) \]

Only the labeling convention changes, but the two conditions are absolutely equivalent.
When dealing with communication channels, it is natural to restrict to particular decision problems that we name “decoding problems”.

\[ E = \{ e(x|m) \}, \quad N = \{ w(y|x) \}, \quad D = \{ d(m'|y) \} \]

**Code Fidelity**

Given a noisy channel \( N : X \rightarrow Y \), its code fidelity, for any set \( M \) and any coding channel \( \mathcal{E} : M \rightarrow X \), is defined as

\[
\mathbb{E}_{\mathcal{E}}[N] \overset{\text{def}}{=} \max_{D:Y \rightarrow M} \frac{1}{|M|} \sum_{m,x,y,m'} e(x|m)w(y|x)d(m'|y)\delta_{m,m'}
\]
Theorem (Coding Problems Are Complete)

Given two noisy channels $\mathcal{N} : \mathcal{X} \rightarrow \mathcal{Y}$ and $\mathcal{N}' : \mathcal{X} \rightarrow \mathcal{Y}'$, $\mathcal{N}$ is degradable into $\mathcal{N}'$ if and only if

$$\mathbb{E}_{\mathcal{E}}[\mathcal{N}] \geq \mathbb{E}_{\mathcal{E}}[\mathcal{N'}],$$

for all codes $\mathcal{E} : \mathcal{M} \rightarrow \mathcal{X}$, with $\mathcal{M} \cong \mathcal{Y'}$. 

Comparison of Noisy Channels
Extensions to the Quantum Case
Extending Decoding Problems

**Decoding Problems**

**Quantum Decoding Problems**

**Quantum “Realignment” Problems**
Quantum Decoding Problems

\[
|\Phi_{M'M}'\rangle = \frac{1}{\sqrt{d_M}} \sum_{i=1}^{d_M} |i\rangle_M |i\rangle_M
\]

Quantum Code Fidelity

Given a quantum channel (i.e., CPTP linear map) \( \mathcal{N} : A \rightarrow B \), its quantum code fidelity, for any Hilbert space \( \mathcal{H}_M \cong \mathcal{H}_{M'} \) and any quantum coding channel \( \mathcal{E} : M \rightarrow A \), is defined as

\[
\mathbb{E}_{\langle \mathcal{E} \rangle}[\mathcal{N}] \overset{\text{def}}{=} \max_{\mathcal{D} : B \rightarrow M} \langle \Phi^+_{M'M} | (\text{id}_{M'} \otimes \mathcal{D} \circ \mathcal{N} \circ \mathcal{E})(\Phi^+_{M'M}) | \Phi^+_{M'M} \rangle = d_M^{-1} 2^{-H_{\text{min}}(M'|B)}
\]
Quantum Realignment Problems

Given a quantum channel $\mathcal{N} : A \to B$, for any Hilbert space $\mathcal{H}_C \cong \mathcal{H}_{C'}$ and any “misaligning” channel $\mathcal{F} : A' \to C'$, its quantum realignment fidelity is defined as

$$F^q_{\langle \mathcal{F} \rangle} [\mathcal{N}] \overset{\text{def}}{=} \max_{D : B \to C} \langle \Phi^+_{C'C} | (\mathcal{F}_{A'} \otimes D \circ \mathcal{N}) (\Phi^+_{A'A}) | \Phi^+_{C'C} \rangle = d_C^{-1/2} 2^{-H_{\text{min}}(C'|B)}$$
Theorem (Quantum Coding and Realignment Problems Are Complete)

Given two quantum channels $\mathcal{N} : A \to B$ and $\mathcal{N}' : A \to B'$, the following are equivalent:

1. $\mathcal{N}$ is degradable into $\mathcal{N}'$;
2. for any quantum coding channel $\mathcal{E} : M \to A$, with $\mathcal{H}_M \cong \mathcal{H}_{B'}$, one has $E^q_{\mathcal{E}}[\mathcal{N}] \geq E^q_{\mathcal{E}}[\mathcal{N}']$, or, equivalently, $H_{\min}(M'|B) \leq H_{\min}(M'|B')$;
3. for any quantum misaligning channel $\mathcal{F} : A' \to C'$, with $\mathcal{H}_{C'} \cong \mathcal{H}_{B'}$, one has $F^q_{\mathcal{F}}[\mathcal{N}] \geq F^q_{\mathcal{F}}[\mathcal{N}']$, or, equivalently, $H_{\min}(C'|B) \leq H_{\min}(C'|B')$. 

Comparison of Quantum Channels
Application to Open Quantum Systems Dynamics
Discrete-Time Stochastic Processes

- Let $x_i$, for $i = 0, 1, \ldots$, index the state of a system at time $t = t_i$

- if the system can be initialized at time $t = t_0$, the process is fully described by the conditional distribution $p(x_N, \ldots, x_1 | x_0)$

- if the system evolving is quantum, we only have a quantum dynamical mapping \( \{ N^{(i)}_{Q_0 \rightarrow Q_i} \}_{i=1, \ldots, N} \)

- the process is divisible if there exist channels $D^{(i)}$ such that $N^{(i+1)} = D^{(i)} \circ N^{(i)}$ for all $i$
A “Zoo of Quantum Markovianities”

A “Zoo of Quantum Markovianities”

Divisibility as “Information Flow”

Theorem

Given a mapping \( \{ N_{Q_0 \rightarrow Q_i}^{(i)} \}_{i \geq 1} \), the following are equivalent to divisibility:

1. for any quantum code, its fidelity is monotonically non-increasing in time
2. for any misaligning channel, its quantum realignment fidelity is monotonically non-increasing in time
Application to Quantum Thermodynamics
3.5 years ago I presented some ideas (arXiv:1505.00535) that eventually led to (arXiv:1708.04302)
Formulation of quantum thermodynamics as a “resource theory of out-of-thermal-equilibrium-ness”

- **thermal (or Gibbs) distribution**: \( \gamma = \gamma(\hat{H}, T) = Z^{-1}e^{-\hat{H}/kT} \)
- **thermal processes** (Janzing et al, 2000; Horodecki and Oppenheim, 2013) use free thermal ancillas, total energy-preserving interactions, partial traces: 
  \[ \mathcal{E}_{th}(\rho_S) = \text{Tr}[U(\rho_S \otimes \gamma_E)U^\dagger] \]
- **thermal** \( \implies \) **Gibbs-preserving** (but \( \not\Leftarrow \))
- **thermal accessibility**: \( \rho \xrightarrow{th} \sigma \) whenever there exists thermal process \( \mathcal{E}_{th} \) such that \( \mathcal{E}_{th}(\rho) = \sigma \)
- **thermal monotone**: any function \( f(\rho) \) such that \( f(\rho) \geq f(\mathcal{E}_{th}(\rho)) \) for any thermal process \( \mathcal{E}_{th} \) (e.g., the free energy)
Free Energy as Statistical Distinguishability

- **fact**: the free energy $F(\rho) = \text{Tr}[\rho H] - kT S(\rho)$ can be expressed in terms of the quantum relative entropy as $(kT)^{-1} F(\rho) + \log Z = D(\rho\|\gamma)$
- hence, $F(\rho)$ measures the statistical distinguishability of $\rho$ from a thermal background...
- ...and the Second Law is nothing but a data-processing inequality: since $E_{th}(\gamma) = \gamma$, one has

$$\rho \xrightarrow{th} \sigma \implies D(\rho\|\gamma) \geq D(\sigma\|\gamma) \iff F(\rho) \geq F(\sigma)$$

- **problem**: to find a complete set of generalized “free energies” $F_f(\rho)$ (i.e., generalized divergences $D_f(\rho\|\gamma)$) such that

$$\rho \xrightarrow{th} \sigma \iff D_f(\rho\|\gamma) \geq D_f(\sigma\|\gamma), \forall D_f$$
**Second Laws as Quantum Relative Majorization**

- **idea**: to characterize “all” statistical distinguishability measures at once...
- ...in terms of one relative majorization relation: \((\rho, \gamma) \succeq (\sigma, \gamma)\)
- as the **majorization** preorder captures the statistical distinguishability of a given PD from a uniform background...
- ...so that **thermo-majorization** preorder captures thermal accessibility
- however: known results only concern **classical PDs**
- our result: extension of thermo-majorization to deal with **non-commuting density operators**
Conclusions
Conclusions

- statistical comparison was formulated as a generalization of the “majorization” order
- it constitutes an important foundational tool in mathematical statistics
- in this talk, I argue that it can play an important role also in other areas where statistical predictions are involved
- most importantly, information theory and quantum mechanics
- applications found in quantum statistical mechanics and quantum foundations, but more are waiting!

Thank you