Quantum Statistical Comparison and Majorization

(and their applications to generalized resource theories)

Francesco Buscemi (Nagoya University)

MAQIT 2019, Seoul National University
21 May 2019

Guiding idea:
generalized resource theories as order theories for stochastic (probabilistic) structures

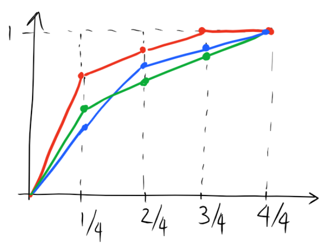
The Precursor: Majorization

Lorenz Curves and Majorization

- two probability distributions, $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, \ldots, q_n)$
- truncated sums $P(k) = \sum_{i=1}^{k} p_i^{\downarrow}$ and $Q(k) = \sum_{i=1}^{k} q_i^{\downarrow}$, for all $k = 1, \ldots, n$
- $\boldsymbol{p}$ majorizes $\boldsymbol{q}$, i.e., $\boldsymbol{p} \succeq \boldsymbol{q}$, whenever $P(k) \geq Q(k)$, for all k
- minimal element: uniform distribution $\boldsymbol{e} = n^{-1}(1, 1, \cdots, 1)$

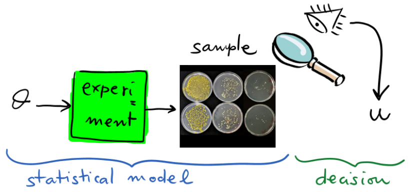
Hardy, Littlewood, and Pólya (1929): $\boldsymbol{p} \succeq \boldsymbol{q} \iff \boldsymbol{q} = M\boldsymbol{p}$, for some bistochastic matrix M.

Lorenz curve for probability distribution $\boldsymbol{p} = (p_1, \cdots, p_n)$:



$(x_k, y_k) = (k/n, P(k)), \quad 1 \leq k \leq n$

Blackwell’s Extensions



$$\begin{array}{ccccc} \Theta & \xrightarrow{\text{experiment}} & \mathcal{X} & \xrightarrow{\text{decision}} & \mathcal{U} \\ \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\ \theta & \xrightarrow[w(x|\theta)]{} & x & \xrightarrow[d(u|x)]{} & u \end{array}$$

payoff is $\ell(\theta, u) \in \mathbb{R}$

Definition

A **statistical model** (or *experiment*) is a triple $\mathbf{w} = \langle \Theta, \mathcal{X}, w \rangle$, a **statistical decision problem** (or *game*) is a triple $\mathbf{g} = \langle \Theta, \mathcal{U}, \ell \rangle$.

Playing Games with Experiments

- the experiment (model) *is given*, i.e., it is the "resource"
- the decision instead *can be optimized*

$$\begin{array}{ccccc} \Theta & \xrightarrow{\text{experiment}} & \mathcal{X} & \xrightarrow{\text{decision}} & \mathcal{U} \\ \rightsquigarrow & & \rightsquigarrow & & \rightsquigarrow \\ \theta & \xrightarrow[w(x|\theta)]{} & x & \xrightarrow[d(u|x)]{} & u \end{array}$$

Definition

The **expected payoff** of a statistical model $\mathbf{w} = \langle \Theta, \mathcal{X}, w \rangle$ w.r.t. a decision problem $\mathbf{g} = \langle \Theta, \mathcal{U}, \ell \rangle$ is given by

$$\mathbb{E}_{\mathbf{g}}[\mathbf{w}] \stackrel{\text{def}}{=} \max_{d(u|x)} \sum_{u,x,\theta} \ell(\theta, u) d(u|x) w(x|\theta) |\Theta|^{-1} .$$

Comparing Statistical Models 1/2

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

$$\begin{array}{ccccc} \Theta & \xrightarrow{\text{experiment}} & \mathcal{X} & \xrightarrow{\text{decision}} & \mathcal{U} \\ \rightsquigarrow & & \rightsquigarrow & & \rightsquigarrow \\ \theta & \xrightarrow[w(x|\theta)]{} & x & \xrightarrow[d(u|x)]{} & u \end{array}$$

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

$$\begin{array}{ccccc} \Theta & \xrightarrow{\text{experiment}} & \mathcal{Y} & \xrightarrow{\text{decision}} & \mathcal{U} \\ \rightsquigarrow & & \rightsquigarrow & & \rightsquigarrow \\ \theta & \xrightarrow[w'(y|\theta)]{} & x & \xrightarrow[d'(u|y)]{} & u \end{array}$$

For a fixed decision problem $\mathbf{g} = \langle \Theta, \mathcal{U}, \ell \rangle$, the expected payoffs $\mathbb{E}_{\mathbf{g}}[\mathbf{w}]$ and $\mathbb{E}_{\mathbf{g}}[\mathbf{w}']$ can always be ordered.

Comparing Statistical Models 2/2

Definition (Information Preorder)

If the model $\mathbf{w} = \langle \Theta, \mathcal{X}, w \rangle$ is better than model $\mathbf{w}' = \langle \Theta, \mathcal{Y}, w' \rangle$ **for all** decision problems $\mathbf{g} = \langle \Theta, \mathcal{U}, \ell \rangle$, then we say that $\mathbf{w}$ is *more informative* than $\mathbf{w}'$, and write

$$\mathbf{w} \succeq \mathbf{w}' \ .$$

Problem. Can we visualize the information morphism $\succeq$ more concretely?

Information Morphism = Statistical Sufficiency

Blackwell-Sherman-Stein Theorem (1948-1953)

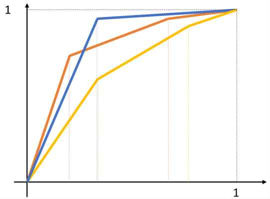
Given two experiments with the same parameter space, $\mathbf{w} = \langle \Theta, \mathcal{X}, w \rangle$ and $\mathbf{w}' = \langle \Theta, \mathcal{Y}, w' \rangle$, the condition $\mathbf{w} \succeq \mathbf{w}'$ holds *iff* there exists a conditional probability $\varphi(y|x)$ such that $w'(y|\theta)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

$$
\begin{array}{ccccccccc}
\Theta & \longrightarrow & \mathcal{Y} & & \Theta & \longrightarrow & \mathcal{X} & \xrightarrow{\text{noise}} & \mathcal{Y} \\
\wr\!\downarrow & & \wr\!\downarrow & = & \wr\!\downarrow & & \wr\!\downarrow & & \wr\!\downarrow \\
\theta & \underset{w'(y|\theta)}{\longrightarrow} & y & & \theta & \underset{w(x|\theta)}{\longrightarrow} & x & \underset{\varphi(y|x)}{\longrightarrow} & y
\end{array}
$$

David H. Blackwell (1919-2010)

Special Case: Dichotomies

- two *pairs* of probability distributions, i.e., two *dichotomies*, $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ and $(\boldsymbol{q}_1, \boldsymbol{q}_2)$, of dimension m and n, respectively
- relabel entries such that ratios p_1^i / p_2^i and q_1^j / q_2^j are nonincreasing
- construct the truncated sums $P_\alpha(k) = \sum_{i=1}^k p_\alpha^i$ and $Q_\beta(k) = \sum_{j=1}^k q_\beta^j$
- $(\boldsymbol{p}_1, \boldsymbol{p}_2) \succeq (\boldsymbol{q}_1, \boldsymbol{q}_2)$ iff the relative Lorenz curve of the former is never below that of the latter



Relative Lorenz curves:

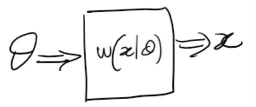
$(x_k, y_k) = (P_2(k), P_1(k))$

Blackwell's Theorem for Dichotomies (1953):
$(\boldsymbol{p}_1, \boldsymbol{p}_2) \succeq (\boldsymbol{q}_1, \boldsymbol{q}_2) \iff \boldsymbol{q}_\alpha = M \boldsymbol{p}_\alpha$, for some stochastic matrix M.

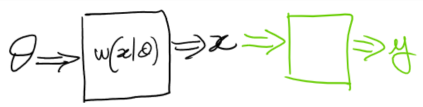
The Viewpoint of Communication Theory

Statistics vs Information Theory

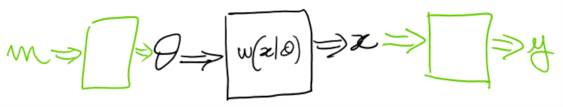
- Statistical models are mathematically equivalent to noisy channels:



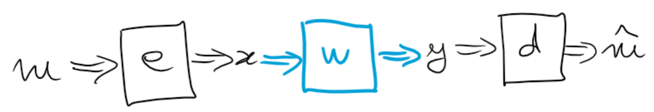
- However: while in statistics the input is inaccessible (Nature does not bother with coding!)



- in communication theory a sender *does code*!



From Decision Problems to Decoding Problems



Definition (Decoding Problems)

Given a channel $\mathbf{w} = \langle \mathcal{X}, \mathcal{Y}, w(y|x) \rangle$, a decoding problem is defined by an encoding $\mathbf{e} = \langle \mathcal{M}, \mathcal{X}, e(x|m) \rangle$ and the payoff function is the optimum guessing probability:

$$\mathbb{E}_{\mathbf{e}}[\mathbf{w}] \stackrel{\text{def}}{=} \max_{d(m|y)} \sum_{m,x,y} d(m|y) w(y|x) e(x|m) |\mathcal{M}|^{-1} = 2^{-H_{\min}(M|Y)}$$

Comparison of Classical Noisy Channels

Theorem

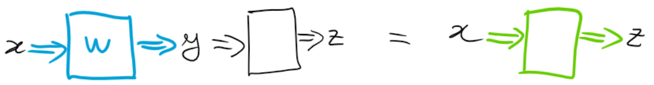
Given two discrete noisy channels $\mathbf{w} : X \to Y$ and $\mathbf{w}' : X \to Z$, consider the following pre-orders:

1. degradability: there exists $\varphi(z|y)$: $w'(x|z) = \sum_y \varphi(z|y) w(y|x)$
2. noisiness: for all encodings $\mathbf{e} = \langle \mathcal{M}, \mathcal{X}, e(x|m) \rangle$, $H(M|Y) \leq H(M|Z)$
3. ambiguity: for all encodings $\mathbf{e} = \langle \mathcal{M}, \mathcal{X}, e(x|m) \rangle$, $H_{\min}(M|Y) \leq H_{\min}(M|Z)$

we have: (1) $\Longrightarrow$ (2) (data-processing inequality), (2) $\not\Longrightarrow$ (1) (Körner and Marton, 1977), but (1) $\Longleftrightarrow$ (3) (FB, 2016).

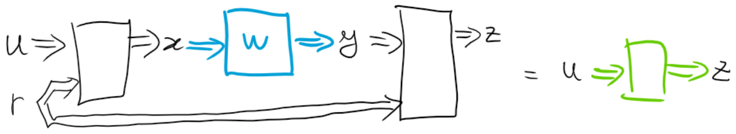
Some Classical Channel Morphisms

Output degrading:



Input degrading:

Full coding (Shannon's "channel inclusion", 1958):



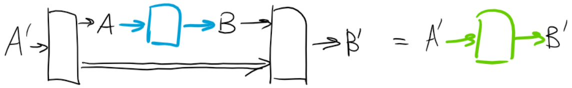
Extensions to the Quantum Case

Some Quantum Channel Morphisms

Output degrading:

Input degrading:

Quantum coding with forward CC:



Output Degradability

Comparison of Quantum Statistical Models 1/2

Quantum statistical models as cq-channels:

Formulation below from: A.S. Holevo, *Statistical Decision Theory for Quantum Systems*, 1973.

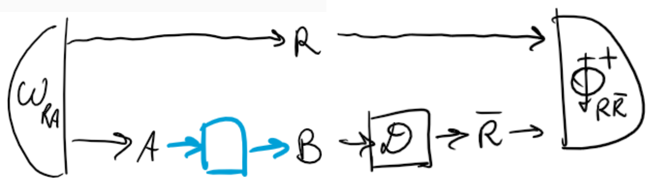
classical case	**quantum case**
• decision problems $\mathbf{g} = \langle \Theta, \mathcal{U}, \ell \rangle$	• decision problems $\mathbf{g} = \langle \Theta, \mathcal{U}, \ell \rangle$
• models $\mathbf{w} = \langle \Theta, \mathcal{X}, \{w(x\|\theta)\} \rangle$	• quantum models $\mathcal{E} = \left\langle \Theta, \mathcal{H}_S, \{\rho_S^\theta\} \right\rangle$
• decisions $d(u\|x)$	• POVMs $\{P_S^u : u \in \mathcal{U}\}$
• $p_c(u, \theta) = \sum_x d(u\|x) w(x\|\theta) \|\Theta\|^{-1}$	• $p_q(u, \theta) = \mathrm{Tr}\left[\rho_S^\theta \; P_S^u\right] \|\Theta\|^{-1}$
• $\mathbb{E}_{\mathbf{g}}[\mathbf{w}] = \max_{d(u\|x)} \sum \ell(\theta, u) p_c(u, \theta)$	• $\mathbb{E}_{\mathbf{g}}[\mathcal{E}] = \max_{\{P_S^u\}} \sum \ell(\theta, u) p_q(u, \theta)$

Comparison of Quantum Statistical Models 2/2

What follows is from: FB, Comm. Math. Phys., 2012

- consider two quantum statistical models $\mathcal{E} = \left\langle \Theta, \mathcal{H}_S, \{\rho_S^\theta\} \right\rangle$ and $\mathcal{E}' = \left\langle \Theta, \mathcal{H}_{S'}, \{\sigma_{S'}^\theta\} \right\rangle$
- **information ordering**: $\mathcal{E} \succeq \mathcal{E}'$ iff $\mathbb{E}_{\mathbf{g}}[\mathcal{E}] \geq \mathbb{E}_{\mathbf{g}}[\mathcal{E}']$ for all $\mathbf{g}$
- $\mathcal{E} \succeq \mathcal{E}'$ iff there exists a **quantum statistical morphism** (essentially, a PTP map) $\mathcal{M} : \mathsf{L}(\mathcal{H}_S) \to \mathsf{L}(\mathcal{H}_{S'})$ such that $\mathcal{M}(\rho_S^\theta) = \sigma_{S'}^\theta$ for all θ
- **complete information ordering**: $\mathcal{E} \succeq_c \mathcal{E}'$ iff $\mathcal{E} \otimes \mathcal{F} \succeq \mathcal{E}' \otimes \mathcal{F}$ for all ancillary models $\mathcal{F}$ (in fact, one informationally complete model suffices)
- $\mathcal{E} \succeq_c \mathcal{E}'$ iff there exists a **CPTP map** $\mathcal{N} : \mathsf{L}(\mathcal{H}_S) \to \mathsf{L}(\mathcal{H}_{S'})$ such that $\mathcal{N}(\rho_S^\theta) = \sigma_{S'}^\theta$ for all θ
- if $\mathcal{E}'$ is **abelian**, then $\mathcal{E} \succeq_c \mathcal{E}'$ iff $\mathcal{E} \succeq \mathcal{E}'$

Comparison of Quantum Channels 1/2



Definition (Quantum Decoding Problems)

Given a quantum channel (CPTP map) $\mathcal{N} : A \to B$, a quantum decoding problem is defined by a bipartite state ω_{RA} and the payoff function is the optimum singlet fraction:

$$\mathbb{E}_\omega[\mathcal{N}] \stackrel{\text{def}}{=} \max_{\mathcal{D}} \langle \Phi^+_{R\bar{R}} | (\text{id}_R \otimes \mathcal{D}_{B\to\bar{R}} \circ \mathcal{N}_{A\to B})(\omega_{RA}) | \Phi^+_{R\bar{R}} \rangle$$

Comparison of Quantum Channels 2/2

Theorem (FB, 2016)

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

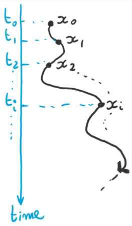
1. **output degradability**: *there exists CPTP map $\mathcal{C}$: $\mathcal{N}' = \mathcal{C} \circ \mathcal{N}$;*
2. **coherence preorder**: *for any bipartite state ω_{RA}, $\mathbb{E}_\omega[\mathcal{N}] \geq \mathbb{E}_\omega[\mathcal{N}']$, that is, $H_{\min}(R|B)_{(\mathsf{id}\otimes\mathcal{N})(\omega)} \leq H_{\min}(R|B')_{(\mathsf{id}\otimes\mathcal{N}')(\omega)}$.*

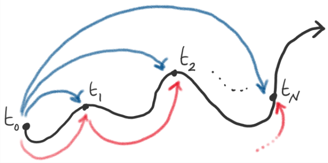
$\rightsquigarrow$ applications to the theory of open quantum systems dynamics and, by adding symmetry constraints, to quantum thermodynamics

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 $\left\{\mathcal{N}^{(i)}_{Q_0 \to Q_i}\right\}_{i=1,\ldots,N}$
- the process is divisible if there exist channels $\mathcal{D}^{(i)}$ such that $\mathcal{N}^{(i+1)} = \mathcal{D}^{(i)} \circ \mathcal{N}^{(i)}$ for all i
- problem: to provide a fully information-theoretic characterization of divisibility



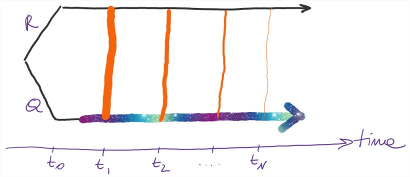


Divisibility as "Quantum Information Flow"

Theorem (2016-2018)

Given an initial open quantum system Q_0, *a quantum dynamical mapping* $\left\{\mathcal{N}^{(i)}_{Q_0\to Q_i}\right\}_{i\geq 1}$ *is divisibile if and only if, for any initial state* ω_{RQ_0},

$$H_{\min}(R|Q_1) \leq H_{\min}(R|Q_2) \leq \cdots \leq H_{\min}(R|Q_N) \ .$$



Conclusions

Conclusions

- the theory of statistical comparison studies morphisms (preorders) of one "statistical structure" X into another "statistical structure" Y
- equivalent conditions are given in terms of (finitely or infinitely many) *monotones*, e.g., $f_i(X) \geq f_i(Y)$
- such monotones shed light on the "resources" at stake in the operational framework at hand
- in a sense, statistical comparison is complementary to SDP, which instead searches for *efficiently computable* functions like $f(X, Y)$
- however, SDP does not provide much insight into the resources at stake (and not all statistical comparisons are equivalent to SDP!)

Thank you