The Role of Statistical Comparison Theory in the Study of Open Quantum Systems

Francesco Buscemi (Nagoya U)

Advances in open systems and fundamental tests of quantum mechanics
684th WE-Heraeus-Seminar, Bad Honnef (Germany), 4 December 2018
Classical Markov chains: some nomenclature

Time convention: \( t_0 \leq t_1 \leq \cdots \leq t_N \).

- classical Markov chain:
  \[
  P(x_{t_i} | x_{t_{i-1}}, x_{t_{i-2}}, \ldots, x_{t_0}) = P(x_{t_i} | x_{t_{i-1}}), \quad \forall i \in [1, N]
  \]

- physical divisibility (Markov equation):
  \[
  P(x_{t_N}, x_{t_{N-1}}, \ldots, x_{t_0}) = P(x_{t_N} | x_{t_{N-1}}) \cdots P(x_{t_1} | x_{t_0}) P(x_{t_0})
  \]

- stochastic divisibility (Chapman-Kolmogorov equation):
  \[
  P(x_{t_k} | x_{t_i}) = \sum_{x_{t_j}} P(x_{t_k} | x_{t_j}) P(x_{t_j} | x_{t_i}), \quad \forall k \geq j \geq i
  \]

- physical divisibility \( \iff \) stochastic divisibility
The Problem with Quantum Systems

Quantum stochastic processes are like sealed black boxes: an observation at some time $t_1$ generally disturbs the process thus “spoiling” any subsequent observation made at later times $t_2 \geq t_1$.

![Figure 1](image)

**Figure 1:** Here $t_0$ is an initial time, at which the quantum system can be prepared (fully controlled). There is no direct quantum analogue of the $N$-time joint distribution $P(x_{t_N}, \ldots, x_{t_0})$. 
Quantum Dynamical Mappings

How to describe *quantum* stochastic processes then?

- time convention: \( t_0 \leq t_1 \leq \cdots \leq t_N \)
- open quantum systems formalism:
  \[
  \rho_S(t_i) := \text{Tr}_E \left\{ U_{t_0 \rightarrow t_i} \left[ \rho_S(0) \otimes \rho_E(0) \right] U_{t_0 \rightarrow t_i}^\dagger \right\}
  \]
- if the system is fully controlled at \( t_0 \), we obtain a sequence of CPTP linear maps by discarding the bath:
  \[
  \Phi_i(\bullet_S) := \text{Tr}_E \left\{ U_{t_0 \rightarrow t_i} \left[ \bullet_S \otimes \rho_E(0) \right] U_{t_0 \rightarrow t_i}^\dagger \right\}
  \]

**Definition**

A *quantum dynamical mapping (QDM)* is a sequence of CPTP linear maps \( (\Phi_i)_{0 \leq i \leq N} \) satisfying \( \Phi_0 = \text{id}_S \) (consistency condition).
Two Approaches to Quantum Markovianity

- **Global (extrinsic) picture**: Markovianity is a property of the whole system+bath compound (like, e.g., singular coupling regime, approximate factorizability, etc)

- **Reduced (intrinsic) picture**: Markovianity is a property of the resulting quantum dynamical mapping alone (like, e.g., information decrease, divisibility, etc)
A “Zoo” of Quantum Markovianities

A “Zoo” of Quantum Markovianities

Decreasing System Distinguishability (DSD)

- introduced in [Breuer, Laine, Piilo; PRL 2009], it provides the bridge between physical and information-theoretic Markovianity

- for any pair of possible initial states of the system, say, $\rho^1_S(0)$ and $\rho^2_S(0)$, consider the same pair evolved at later times $t_i > t_0$:

$$\rho^{1,2}_S(t_i) := \Phi_i[\rho^{1,2}_S(0)]$$

- DSD condition:

$$\|\rho^1_S(t_i) - \rho^2_S(t_i)\|_1 \geq \|\rho^1_S(t_j) - \rho^2_S(t_j)\|_1, \quad \forall i \leq j$$

- interpretation: revival of distinguishability $\implies$ back-flow of information $\implies$ memory effects $\implies$ non-Markovianity
Divisibility (DIV)

- extends the idea of dynamical semigroups: $t \mapsto \Phi_t$ such that $\Phi_s \circ \Phi_t = \Phi_{t+s}$
- a QDM $(\Phi_i)_i$ is **CPTP divisible** if there exist CPTP linear maps $(\mathcal{E}_{i \rightarrow j})_{i \leq j}$, which we call **propagators**, such that $\Phi_j = \mathcal{E}_{i \rightarrow j} \circ \Phi_i$, for all $0 \leq i \leq j \leq N$

- DIV constitutes a **quantum analogue of the Chapman-Kolmogorov equation** (i.e., stochastic divisibility)
- very well captures the property of being **memoryless**, which is a crucial (**the crucial?**) property of Markovian processes
can we make these equivalent?
• DIV is equivalent to the property of **degradability**: channel $\Phi$ is said to be degradable into channel $\Phi'$ whenever there exists a third channel $\mathcal{E}$ such that $\Phi' = \mathcal{E} \circ \Phi$

• hence, “DIV $\implies$ DSD” is a consequence of the **data-processing inequality for the trace norm**: for any pair of states $(\rho^1_S, \rho^2_S)$,

$$\|\Phi'(\rho^1_S) - \Phi'(\rho^2_S)\|_1 = \|(\mathcal{E} \circ \Phi)(\rho^1_S) - (\mathcal{E} \circ \Phi)(\rho^2_S)\|_1 \leq \|\Phi(\rho^1_S) - \Phi(\rho^2_S)\|_1$$

• in fact, the data-processing inequality is satisfied by most (all?) distinguishability measures

• hence, it is interesting to seek for possible **alternative (stronger) definitions of DSD**, maintaining the same “intuitive meaning”, but leading to the sought after equivalence: $\text{DSD} \iff \text{DIV}$
Theorem (Chruściński, Kossakowski, and Rivas, 2011; Chruściński and Maniscalco, 2014; Wißman, Breuer, Vacchini, 2015)

Let $\Phi : A \rightarrow B$ and $\Phi' : A \rightarrow B'$ be two quantum channels, with $\Phi$ invertible (as a linear map). Then, $\Phi$ is degradable into $\Phi'$ with a $k$-positive TP map $E : B \rightarrow B'$, if and only if, for all $p \in [0, 1]$ and all pairs of $k$-extended states $\rho_{k,A}^1, \rho_{k,A}^2 \in L(\mathbb{C}^k \otimes \mathcal{H}_A)$,

\[
\|p(\text{id}_k \otimes \Phi')(\rho_{k,A}^1) - (1 - p)(\text{id}_k \otimes \Phi')(\rho_{k,A}^2)\|_1 \\
\leq \|p(\text{id}_k \otimes \Phi)(\rho_{k,A}^1) - (1 - p)(\text{id}_k \otimes \Phi)(\rho_{k,A}^1)\|_1
\]
Statistical Distinguishability in Mathematical Statistics
A statistical model is given by: a parameter set $\Theta$, a sample set $\mathcal{X}$, and a family of PDs $\{w_\theta(x)\} \equiv w(x|\theta)$.

A statistical decision problem is given by: a parameter set $\Theta$, an “action” set $\mathcal{U}$, and a payoff function $\ell : \Theta \times \mathcal{U} \rightarrow \mathbb{R}$. 

Formal Definitions
How Much Is a Statistical Model Worth?

Each decision problem implicitly defines a **statistical distinguishability measure** for the PDs \( \{w_\theta(x)\} \).

- the model \( w = \{w_\theta(x)\} \) represents info in \( X \) about \( \theta \)
- the decision \( d(u|x) \) optimally extracts from \( X \) information about \( \theta \), and uses this to decide the best action

**Definition (Expected Payoff)**

The expected payoff of statistical model \( w = \{w_\theta(x)\} \) w.r.t. decision problem \( \ell = \{\ell(\theta, u)\} \) is given by

\[
\mathbb{E}_\ell[w] \triangleq \max_{d(u|x)} \sum_{u,x,\theta} \ell(\theta, u)d(u|x)w(x|\theta)|\Theta|^{-1}.
\]

**Intuition:** the higher the payoff, the more information about \( \theta \) the PDs \( w_\theta(x) \) carry, the more “distinguishable” they are.
Comparing Statistical Models 1/2

First model: \( w = \{ w_\theta(x) \} \)

<table>
<thead>
<tr>
<th>( \Theta )</th>
<th>experiment</th>
<th></th>
<th>( \chi )</th>
<th>decision</th>
<th>( U )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
</tr>
<tr>
<td>( \theta )</td>
<td></td>
<td>( x )</td>
<td></td>
<td>( d(u</td>
<td>x) )</td>
</tr>
<tr>
<td>( w(x</td>
<td>\theta) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Second model: \( w' = \{ w'_\theta(y) \} \)

<table>
<thead>
<tr>
<th>( \Theta )</th>
<th>experiment</th>
<th></th>
<th>( \chi )</th>
<th>decision</th>
<th>( U )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
<td>( \downarrow )</td>
</tr>
<tr>
<td>( \theta )</td>
<td></td>
<td>( y )</td>
<td></td>
<td>( d'(u</td>
<td>y) )</td>
</tr>
<tr>
<td>( w'(y</td>
<td>\theta) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given a statistical decision problem \( \ell = \{ \ell(\theta, u) \} \), if \( \mathbb{E}_\ell[w] \geq \mathbb{E}_\ell[w'] \), then one says that model \( w \) is “more informative” (or “broader”, or “more distinguishable”) than model \( w' \), with respect to problem \( \ell \).
Definition (Information Preorder)

If model \( w = \{w_\theta(x)\} \) is more informative than model \( w' = \{w'_\theta(y)\} \) for all decision problems \( \ell = \{\ell(\theta, u)\} \), then we say that \( w \) is (always) more informative than \( w' \), and write

\[
w \succeq w'.
\]

Intuition: \( w \succeq w' \) means that the PDs \( \{w_\theta(x)\} \) are always more distinguishable than \( \{w'_\theta(y)\} \).

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

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

Given two statistical models \( w = \{ w_\theta(x) \} \) and \( w' = \{ w'_\theta(y) \} \), the following are equivalent:

1. \( w \) is more informative than \( w' \), i.e.,\( w \succeq w' \);
2. \( w \) is sufficient for \( w' \), i.e., there exists a conditional PD \( \varphi(y|x) \) such that
   \[
   w'(y|\theta) = \sum_x \varphi(y|x) w(x|\theta).
   \]

**Diagram:**

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

David H. Blackwell
(1919-2010)
statistical sufficiency

≈

better distinguishability w.r.t. all operational distinguishability measures
Paramount Example: Majorization and Lorenz Curves

- two probability distributions, $p$ and $q$, of the same dimension $n$
- truncated sums $P(k) = \sum_{i=1}^{k} p_i \downarrow$ and $Q(k) = \sum_{i=1}^{k} q_i \downarrow$, for all $k = 1, \ldots, n$
- $p$ 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 $M$.

Intuition: $p \succeq q$ means that $p$ is always more distinguishable than $q$ from the uniform $e$.  

$(x_k, y_k) = (k/n, P(k)), \quad 1 \leq k \leq n$
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 \(p_1^i/p_2^i\) and \(q_1^j/q_2^j\) are nonincreasing

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

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

Blackwell Thm for Dichotomies, 1953

\[(p_1, p_2) \succeq (q_1, q_2) \iff q_i = Mp_i, \text{ for some stochastic matrix } M.\]

Intuition: \((p_1, p_2) \succeq (q_1, q_2)\) means that \((p_1, p_2)\) are always more distinguishable than \((q_1, q_2)\).
observation: discrete noisy channels

\[ \Phi : \begin{cases} \mathcal{X} \to \mathcal{Y} \\ x \mapsto p_x(y) \end{cases} \]

are equivalent to statistical models
<table>
<thead>
<tr>
<th>guessing problems</th>
<th>quantum decoding problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>simple guess</td>
<td>singlet extraction</td>
</tr>
<tr>
<td><img src="image1" alt="Simple guess diagram" /></td>
<td><img src="image2" alt="Singlet extraction diagram" /></td>
</tr>
<tr>
<td>extended guess</td>
<td>encoding-decoding</td>
</tr>
<tr>
<td><img src="image3" alt="Extended guess diagram" /></td>
<td><img src="image4" alt="Encoding-decoding diagram" /></td>
</tr>
</tbody>
</table>
• given two quantum channels (CPTP linear maps) $\Phi : A \rightarrow B$ and $\Phi' : A \rightarrow B'$, we say that $\Phi$ is less noisy than $\Phi'$, i.e., $\Phi \supseteq \Phi'$, whenever, for any input ensemble $\{p_x, \rho_A^x\}$,

$$P_{\text{guess}}(\{p_x, \Phi(\rho_A^x)\}) \geq P_{\text{guess}}(\{p_x, \Phi'(\rho_A^x)\})$$

• $\Phi \supseteq_k \Phi' \iff \text{id}_k \otimes \Phi \supseteq \text{id}_k \otimes \Phi'$ ($\text{id}_k$: identity channel on $\mathbb{C}^k$)

• $\Phi \supseteq_\infty \Phi' \iff \text{id}_{B'} \otimes \Phi \supseteq \text{id}_{B'} \otimes \Phi'$

• the identity channel can be replaced by any fixed, though arbitrary, invertible channel (possibly entanglement-breaking)

**Theorem**

• $\Phi \supseteq_k \Phi' \iff \exists$ $k$-statistical morphism $\mathcal{M}$: $\Phi' = \mathcal{M} \circ \Phi$

• $\Phi \supseteq_\infty \Phi' \iff \exists$ quantum channel $\mathcal{E}$: $\Phi' = \mathcal{E} \circ \Phi$
• given a quantum channel \( \Phi : A \rightarrow B \), for any input bipartite state \( \omega_{RA} \), the transmitted singlet fraction is defined as

\[
\mathcal{F}(\omega_{RA}|\Phi_A) := \sup_{D: \text{CPTP}} \langle \Phi_{RR'}^+ | (\text{id}_R \otimes D_B \circ \Phi)(\omega_{RA}) | \Phi_{RR'}^+ \rangle,
\]

where \( |\Phi_{RR'}^+\rangle \) denotes the maximally entangled state (\( R' \cong R \))

• given two quantum channels \( \Phi : A \rightarrow B \) and \( \Phi' : A \rightarrow B' \), we write \( \Phi \geq \Phi' \), whenever \( \mathcal{F}(\omega_{RA}|\Phi_A) \geq \mathcal{F}(\omega_{RA}|\Phi'_A) \) for all \( \omega_{RA} \) (\( R \cong B' \) is enough)

**Theorem**

• \( \Phi \geq \Phi' \iff \Phi \geq \Phi' \text{ only for separable } \omega_{RA} \iff \exists \text{ quantum channel } \mathcal{E} : \Phi' = \mathcal{E} \circ \Phi \)

• \( \Phi \geq \Phi' \text{ only for classical-quantum} \)

\[
\omega_{RA} = \sum_x p_x |x\rangle\langle x|_R \otimes \rho_A^x \iff \exists \text{ statistical morphism } \mathcal{M} : \Phi' = \mathcal{M} \circ \Phi
\]
Application to Quantum Dynamical Mappings

Figure 2: The varying thickness of the green lines depict the singlet fraction at any time.

- a QDM \((\Phi_i)_i\) is **CP-divisible** iff \(\Phi_i \succeq \Phi_j\) for all \(j \geq i\) and all initial **separable** states
- a QDM \((\Phi_i)_i\) **P-divisible** iff \(\Phi_i \succeq \Phi_j\) for all \(j \geq i\) and all initial **classical-quantum** states
- in terms of entropies: \(H_{\min}(R|S_i) \leq H_{\min}(R|S_j), \text{ for all } j \geq i\)
Some Final Remarks
Meaning of DIV

Why the propagators $(\mathcal{E}_{i \to j})_{i \leq j}$ are assumed to be CPTP?

CP-divisibility is equivalent to saying that the open evolution is “collisional,” in the sense that it can be realized by summoning a “fresh environment” at each time step.
Do the propagators \((\mathcal{E}_{i \rightarrow j})_{i \leq j}\) really need to be linear CPTP?

- linearity is necessary (QDMs are linear)
- trace-preservation (a linear constraint) also

**CP perhaps not**: propagators could be just P or even less (e.g., statistical morphisms), and yet be related to important physical/computational/thermodynamical properties (like, e.g., the “locality” or “causality” of the evolution)
Possible Ideas to Think About

- classical correlations can witness P-indivisibility but not CP-indivisibility
- for that, separable states are required: discord/coherence, anyone?
- it is known that CP-DIV can be decided by SDP: way to design efficient tests?
- robustness to small deviations ($\epsilon$-DIV $\iff$ $\epsilon$-DSD)
- to impose extra properties to DIV, e.g., thermality or group-covariance
- to understand P-DIV in a generalized circuit formalism (no extension possible, however no problem, because not in the black-box picture)
- to understand the information-theoretic and computational capabilities of such generalized circuit models, e.g., data-processing inequalities, computational/thermodynamical aspects, etc