Quantum Multi-Hypothesis Testing
A new take on an old problem

Koenraad M.R. Audenaert

June 26, 2012

Joint unpublished work with Milan Mosonyi (Bristol).

Preprint with lots of [TODO] markers available on request.

Hypothesis testing

- The problem:
  - Given a system in an unknown quantum state;
  - I tell you it comes from a given set of r density matrices $\{\rho_i\}_{i=1}^r$;
  - Figure out which of the r density matrices is the true one, using a well-chosen measurement.
- What is the optimal measurement, what is the optimal error probability?

1. Classical communication over a quantum channel

The most natural setting for the hypothesis testing problem

Allows me to stay in the closet

Classical communication

- Classical Communication = passing a message from one point (space/time) to another using a physical information carrier.
- Message: sequence of symbols from an alphabet $\{1, 2, \ldots, r\}$.
- Basic communication system:

$$\boxed{\text{Source}} \longrightarrow \boxed{\text{Xmitter}} \longrightarrow \boxed{\text{Channel}} \longrightarrow \boxed{\text{Receiver}} \longrightarrow \boxed{\text{Recipient}}$$

- Source: produces symbols
- Transmitter: symbol values $\rightarrow$ values of physical property
- Channel: passes information carrier from sender to receiver; adds noise
- Receiver: values of physical property $\rightarrow$ symbol values

Quantum communication

- Quantum Communication = passing a message from one point (space/time) to another using a quantum state.
- Message: sequence of symbols from an alphabet $\{1, 2, \ldots, r\}$.
- Basic communication system:

$$\boxed{\text{Source}} \longrightarrow \boxed{\text{Xmitter}} \longrightarrow \boxed{\text{Quantum Channel}} \longrightarrow \boxed{\text{Receiver}} \longrightarrow \boxed{\text{Recipient}}$$

- Source: produces symbols
- Transmitter: symbol values $\rightarrow$ quantum states: *signal states*
- Quantum Channel: CPTP map
- Receiver: quantum measurement

The problem of communication: Noise

- Noise = Nature modifying the physical properties of the carrier, causing communication errors.
- Noise makes the symbols less distinguishable, destroying information.
- Good communication design: reduce amount of noise (technology issues), and *amount of errors caused by noise*: optimal choice of signal states, optimal measurement
- Optimality expressed in terms of a mathematical model.
- In this talk: signal states given, measurement to be optimised.

Mathematical model

$$\boxed{\text{Source}} \to I \to \boxed{\text{Xmitter}} \to \sigma \to \boxed{\text{Channel}} \to \rho \to \boxed{\text{Receiver}} \to J \to \boxed{\text{Recipient}}$$

- **Source:**
  - Emits symbols from the alphabet $\{1, 2 \ldots, r\}$.
  - Each symbol modelled as i.i.d. random variable I with probability distribution $p_I(i)$, $i \in \{1, 2 \ldots, r\}$.
- **Transmitter:**
  - Translates symbols i to signal states σ_i.

Mathematical model

$$\boxed{\text{Source}} \to I \to \boxed{\text{Xmitter}} \to \sigma \to \boxed{\text{Channel}} \to \rho \to \boxed{\text{Receiver}} \to J \to \boxed{\text{Recipient}}$$

- **Noisy quantum channel:**
  - State coming out of the channel (ρ) will be different from what got in (σ).
  - Channel modelled as CPTP map Φ.
- **Transmitter + noisy channel:**
  - Combined model: $i \mapsto \Phi(\sigma_i) =: \rho_i$.

Mathematical model

$$\boxed{\text{Source}} \rightarrow I \rightarrow \boxed{\text{Xmitter}} \rightarrow \sigma \rightarrow \boxed{\text{Channel}} \rightarrow \rho \rightarrow \boxed{\text{Receiver}} \rightarrow J \rightarrow \boxed{\text{Recipient}}$$

- **Receiver:**
  - Has to "guess" i from state ρ_i.
  - Should be designed so that the probability of guessing incorrectly is minimal.
  - Need to incorporate receiver imperfections as well, e.g. receiver noise.
  - Model: POVM $\{E_j\}_{j=1}^{r}$
  - Produces outcome j with probability (Born rule) $p_j(\rho) = \operatorname{Tr} E_j \rho$.

Success Probability

- The quality of a receiver is determined by the probability that it chooses the correct answer; i.e. $j = i$.
- Probability that receiver outputs j given that source produced i:

$$p_{J|I}(j|i) \;=\; \operatorname{Tr} \rho_i E_j.$$

- Success probability given symbol i is

$$p_{J|I}(i|i) \;=\; \operatorname{Tr} \rho_i E_i.$$

- *Average success probability* p_s:

$$p_s = \sum_{i=1}^{r} p_I(i)\; p_{J|I}(i|i).$$

All together now...

- Quantum state: density operator (matrix) ρ_i, satisfying

$$\boxed{\forall i : \rho_i \geq 0, \text{ with } \operatorname{Tr} \rho_i = 1.}$$

- Quantum measurement: r measurement operators E_j, satisfying

$$\boxed{\forall j : E_j \geq 0, \text{ with } \sum_{j=1}^{r} E_j = \mathbb{I}.}$$

- Quantum success probability (to be optimised):

$$\boxed{p_s = \sum_{i=1}^{r} p_I(i) \ \operatorname{Tr} \rho_i E_i.}$$

- Looks like an inner product between the ρ_i and the E_i. However, optimisation is difficult because ρ_i and E_i have different normalisations.

2. Optimal Quantum Receiver

What is the optimal POVM? What is the optimal success probability?

- For a binary alphabet ($r = 2$) this has an analytical answer.
- For larger alphabets, no analytical answer is known, but it can be solved using numerical methods.

Optimal Quantum Receiver

- To keep notations simpler, I will absorb the source symbol probabilities $p_I(i)$ into the received states ρ_i and write $A_i := p_I(i)\rho_i$ (signal operators).
- Furthermore, I will write $A_0 = \sum_i A_i$.
- Thus:

$$A_i \geq 0, \ \mathrm{Tr}\, A_0 = 1.$$

- Formula for p_s, for a given POVM $\{E_k\}$:

$$p_s(\{E_k\}) = \sum_{k=1}^{r} \mathrm{Tr}\, A_k E_k.$$

SDP Formulation

- Optimal success rate, for optimal POVM:

$$p_s^* = \max_{\{E_k\}} \left\{ \sum_{k=1}^{r} \mathrm{Tr}\, A_k E_k; \quad E_k \geq 0, \sum_{k=1}^{r} E_k = \mathbb{I} \right\}.$$

- This type of maximisation is known as a semidefinite program (SDP):
  - maximisation of a linear functional
  - over variables subject to linear constraints and semidefiniteness constraints.
- (The set of feasible values of the variables is convex.)
- Efficient numerical methods exist for solving SDPs.

Dual Formulation

- For every SDP there exists a *dual* SDP, which gives the same answer but via a *minimisation* over a new variable (essentially a Lagrange multiplier).
- Primal SDP:

$$p_s^* = \max_{\{E_k\}} \left\{ \sum_{k=1}^{r} \operatorname{Tr} A_k E_k; \quad E_k \geq 0, \sum_{k=1}^{r} E_k = \mathbb{I} \right\}.$$

- Dual SDP:

$$p_s^* = \min_{Y} \left\{ \operatorname{Tr} Y; \quad \forall k : Y \geq A_k \right\}.$$

- See, e.g. Eldar, Megretski and Verghese, IEEE IT-49 (2003), 1007–1012.

Proof

Introduce r operators $X_k \geq 0$, one more operator Y, and define

$$\mathcal{L} = \sum_k \mathrm{Tr}(A_k E_k) + \sum_k \mathrm{Tr}(X_k E_k) + \mathrm{Tr}\, Y(\mathbb{I} - \sum_k E_k) = \mathrm{Tr}\, Y + \sum_k \mathrm{Tr}\, E_k(A_k + X_k - Y).$$

Then, $\mathcal{L} \geq \sum_k \mathrm{Tr}(A_k E_k) = p_s(\{E_k\})$, so that

$$p_s^* \leq \max_{E_k} \left\{ \mathcal{L}; E_k \geq 0, \sum_{k=1}^{r} E_k = \mathbb{I} \right\} \leq \max_{E_k} \{\mathcal{L}\} = \begin{cases} \mathrm{Tr}\, Y, & Y = A_k + X_k \\ +\infty, & \text{otherwise.} \end{cases}$$

The positivity condition on the X_k can be replaced by requiring that $Y \geq A_k$, for all k. Hence

$$p_s^* \leq \begin{cases} \mathrm{Tr}\, Y, & Y \geq A_k \\ +\infty, & \text{otherwise,} \end{cases} \qquad \forall Y.$$

Minimising over Y yields the best upper bound on p_s^*:

$$p_s^* \leq \min_{Y \geq A_k} \{\mathrm{Tr}\, Y; Y \geq A_k\}.$$

One can show that equality can be achieved (under certain mild technical conditions). □

LUB

- The optimal Y in

$$p_s^* = \min_Y \{\operatorname{Tr} Y; \quad \forall k : Y \geq A_k\}\,.$$

  can be interpreted as some sort of *least upper bound* (LUB) on the A_k.

- Henceforth I will write

$$p_s^* = \operatorname{Tr} \operatorname{LUB}(A_1, \ldots, A_r),$$

  with

$$\operatorname{LUB}(A_1, \ldots, A_r) = \operatorname{argmin}_Y \{\operatorname{Tr} Y; \quad \forall k : Y \geq A_k\}\,.$$

LUB: properties

- For diagonal A_i, LUB is just the entrywise maximum.
- $\mathrm{LUB}(A, B, C) \neq \mathrm{LUB}(A, \mathrm{LUB}(B, C))$;
- However, $\operatorname{Tr} \mathrm{LUB}(A, B, C) \leq \operatorname{Tr} \mathrm{LUB}(A, \mathrm{LUB}(B, C))$.
- $\operatorname{Tr}(\mathrm{LUB}(A, B) + \mathrm{LUB}(C, D)) \geq \operatorname{Tr} \mathrm{LUB}(A + C, A + D, B + C, B + D)$.
- $\mathrm{LUB}(A_1, \ldots, A_r) + B = \mathrm{LUB}(A_1 + B, \ldots, A_r + B)$.
- $\mathrm{LUB}(A, 0) = A_+$.

GLB

- I will also need the *greatest lower bound* (GLB):

$$\mathrm{GLB}(A_1, \ldots, A_r) = \mathrm{argmax}_Y \left\{ \mathrm{Tr}\, Y; \quad \forall k : Y \leq A_k \right\}.$$

- Property: $\mathrm{GLB}(A_1, \ldots, A_r) = -\,\mathrm{LUB}(-A_1, \ldots, -A_r)$.
- $A, B \geq 0$ does *not* imply $\mathrm{GLB}(A, B) \geq 0$, only $\mathrm{Tr}\,\mathrm{GLB}(A, B) \geq 0$.

Solution of binary case

- The case $r = 2$ can be solved analytically (Holevo, Helstrom 1973).

$$\begin{aligned} p_s(\{E, \mathbb{I} - E\}) &= \mathrm{Tr}(A_1 E + A_2(\mathbb{I} - E)) \\ &= \mathrm{Tr}\, A_2 + \mathrm{Tr}((A_1 - A_2)E). \end{aligned}$$

- Choose as E the *projector* on the support of $(A_1 - A_2)_+$.
- The p_s of this POVM, which is a lower bound on p_s^*, is:

$$p_s^* \geq p_s(\{E, \mathbb{I} - E\}) = \mathrm{Tr}\, A_2 + \mathrm{Tr}(A_1 - A_2)_+.$$

Solution of binary case

- Now consider the dual problem, $p_s^* = \min_Y \{\operatorname{Tr} Y : Y \geq A_k\}$.
- Let us choose

$$Y = (A_1 + A_2 + |A_1 - A_2|)/2 = A_2 + (A_1 - A_2)_+ = A_1 + (A_2 - A_1)_+.$$

- This is clearly an upper bound ($Y \geq A_k$), hence

$$p_s^* \leq \operatorname{Tr} Y = \operatorname{Tr} A_2 + \operatorname{Tr}(A_1 - A_2)_+.$$

- Now note: this is the same expression as in the lower bound.

Solution of binary case

- Thus, we have equality; the chosen E is optimal, and so is Y:

$$Y = \mathrm{LUB}(A_1, A_2) = A_2 + (A_1 - A_2)_+.$$

- Hence, we have

$$p_s^* = \mathrm{Tr}\,\mathrm{LUB}(A_1, A_2) = \mathrm{Tr}(A_2 + (A_1 - A_2)_+).$$

- The optimal error probability $p_e^* = 1 - p_s^*$ can be expressed in terms of GLB.

GLB

- Define the *complementary* states

$$\overline{A}_k := \sum_{j:j\neq k} A_j = A_0 - A_k.$$

- Then

$$\begin{aligned} p_e^* = 1 - p_s^* &= \operatorname{Tr} A_0 - \operatorname{Tr} \operatorname{LUB}(A_1, \ldots, A_r) \\ &= \operatorname{Tr} A_0 + \operatorname{Tr} \operatorname{GLB}(-A_1, \ldots, -A_r) \\ &= \operatorname{Tr} \operatorname{GLB}(A_0 - A_1, \ldots, A_0 - A_r) \\ &= \operatorname{Tr} \operatorname{GLB}(\overline{A}_1, \ldots, \overline{A}_r) \\ &= \min_{\{E_k\}} \sum_k \operatorname{Tr} \overline{A}_k E_k \end{aligned}$$

- For $r = 2$, $\overline{A}_1 = A_2$ and $\overline{A}_2 = A_1$, hence $P_e^* = \operatorname{Tr} \operatorname{GLB}(A_2, A_1)$ and

$$\operatorname{GLB}(A_1, A_2) = A_2 - (A_1 - A_2)_+.$$

General case

- No analytical solutions for $\mathrm{LUB}(A_1, \ldots, A_r)$, nor for p_s^* are known for $r > 2$.
- Can be solved in specific cases. Numerical methods needed for the general case.
- What if we want to answer questions of a general nature? What if the dimension d becomes too high for numerics?
- In particular, what can we say about the *asymptotic behaviour* of the error probability?

3. Asymptotic error rate

Asymptotic error rate

- In the case of classical communication, we can reduce the error by sending the same symbol n times, make n measurements and then do a majority vote on the n outcomes.
- We can do something similar in the quantum case.
- Here the transmitter-channel produces, for input symbol i, n identical quantum systems each described by the same density operator ρ_i, i.e. it produces n *copies* of ρ_i.
- In the quantum case, the optimal strategy is *not* to perform measurements on each copy of ρ_i, but to store these n copies and make a single measurement on the composite system as a whole.
- This composite system is described by the n-fold Kronecker product

$$\rho_i^{\otimes n} := \rho_i \otimes \ldots \otimes \rho_i.$$

Asymptotic error rate

- The formula for the optimal success probability p_s^* remains the same, but the operators A_k are replaced by $A_{k,n} = p_k \rho_k^{\otimes n}$.

$$p_{s,n}^* = \max_{\{E_k\}} \sum_k \operatorname{Tr} A_{k,n} E_k = \operatorname{Tr} \operatorname{LUB}(\{A_{k,n}\}).$$

- The error probability $p_{e,n}^* = 1 - p_{s,n}^*$ decreases exponentially with n:

$$p_{e,n}^* \approx \exp(-Rn).$$

- One defines the *asymptotic error rate* as follows:

$$R = \lim_{n\to\infty} -\frac{1}{n} \log p_{e,n}^*.$$

Quantum Chernoff Divergence

- There is a simple formula in the case of $r = 2$, (generalising a similar formula valid in the classical setting, due to Chernoff (1952)):

$$R = -\log \min_{0 \le s \le 1} \operatorname{Tr} \rho_1^s \, \rho_2^{1-s}.$$

- This was recently generalised to vN algebras by Y Ogata.
- This quantity is now known as the Chernoff divergence of ρ_1 and ρ_2, $C(\rho_1, \rho_2)$.
- It quantifies the distinguishability of ρ_1 and ρ_2.

Quantum Chernoff Divergence

- For larger r, it is believed that the asymptotic error rate is equal to the *smallest pairwise* quantum Chernoff divergence.

$$R = C := \min_{j<k} C(\rho_j, \rho_k).$$

- In the classical case, this was proved by Salikhov (1973).
- Nussbaum and Szkola (2010) have shown this to be true in the quantum case when all density operators have rank 1 (pure states).
- The general case is still an open problem.
- Nussbaum and Szkola showed that $R \leq C$.
- To show also that $R \geq C$, good upper bounds on the optimal error probability are required.

4,3,2,...

- There are some weaker results that indicate that the pairwise Chernoff divergence can not be too different from R.
- First, Nussbaum and Szkola proved $R \geq C/(r(r-1)/2)$.
- We were then able to improve the lower bound to $R \geq C/4$.
- After that, Nussbaum and Szkola improved it further to $R \geq C/3$.
- Earlier this year we showed $R \geq C/2$.
- Proof involves: Tyson's bound, Lieb's theorem, Hölder's inequality

Tyson's bound

- Bound on the optimal error probability: $P_e^* \leq 2(1 - \mathrm{Tr}(\sum_i A_i^2)^{1/2})$.
- *Proof.* Choose a special POVM in the primal SDP: $E_k = S^{-1/2} A_k^2 S^{-1/2}$, with $S := \sum_j A_j^2$. By Hölder's inequality:

$$\begin{aligned}
\mathrm{Tr}\, A_k^2 S^{-1/2} &= ||A_k S^{-1/2} A_k||_1 \\
&= ||(A_k S^{-1/2} A_k^{1/2})\, A_k^{1/2}||_1 \\
&\leq ||A_k S^{-1/2} A_k^{1/2}||_2 \, ||A_k^{1/2}||_2 \\
&= (\mathrm{Tr}\, E_k A_k)^{1/2} (\mathrm{Tr}\, A_k)^{1/2} \\
&\leq (\mathrm{Tr}\, A_k + \mathrm{Tr}\, E_k A_k)/2.
\end{aligned}$$

Sum over k yields $\mathrm{Tr}\, S^{1/2} \leq (1 + P_s(E))/2$, hence $2(1 - \mathrm{Tr}\, S^{1/2}) \geq P_e(E) \geq P_e^*$. □

Proof of C/2 bound

- Lieb's theorem: $\mathrm{Tr}(B^r C^{1-r})$ is jointly concave in B and C, for $0 < r \leq 1$; that is, for n PSD matrices B_j and C_j,

$$\mathrm{Tr} \sum_j B_j^r C_j^{1-r} \leq \mathrm{Tr}(\sum_j B_j)^r (\sum_j C_j)^{1-r}$$

- Put $r = 2/3$, $B_j = A_j^{1/2}, C = A_j^2$:

$$1 = \mathrm{Tr} \sum_j A_j^{1/3} A_j^{2/3} \leq \mathrm{Tr}(\sum_j A_j^{1/2})^{2/3} \ (\sum_j A_j^2)^{1/3}$$

Proof of C/2 bound

- Applying Hölder's inequality with well-chosen parameters:

$$\operatorname{Tr} XY \leq ||XY||_1 \leq ||X||_3 ||Y||_{3/2},$$

$$\begin{aligned} 1 &\leq ||(\sum_j A_j^{1/2})^{2/3}||_3 \quad ||(\sum_j A_j^2)^{1/3}||_{3/2} \\ &= \left(\operatorname{Tr}(\sum_j A_j^{1/2})^2\right)^{1/3} \left(\operatorname{Tr}(\sum_j A_j^2)^{1/2}\right)^{2/3}. \end{aligned}$$

Proof of C/2 bound

- Taking the $3/2$ power and rearranging then yields

$$\begin{aligned}\mathrm{Tr}(\sum_j A_j^2)^{1/2} \;&\geq\; \left(\mathrm{Tr}(\sum_j A_j^{1/2})^2\right)^{-1/2} \;=\; \left(\sum_{j,k} \mathrm{Tr}\, A_j^{1/2} A_k^{1/2}\right)^{-1/2} \\ &=\; \left(1 + 2\sum_{j<k} \mathrm{Tr}\, A_j^{1/2} A_k^{1/2}\right)^{-1/2} \;\geq\; 1 - \sum_{j<k} \mathrm{Tr}\, A_j^{1/2} A_k^{1/2},\end{aligned}$$

  where in the last line we exploited the inequality $(1+x)^{-1/2} \geq 1 - x/2$.

- In combination with Tyson's bound this yields

$$P_e^* \leq 2(1 - \mathrm{Tr}(\sum_i A_i^2)^{1/2}) \leq 2\sum_{j<k} \mathrm{Tr}\, A_j^{1/2} A_k^{1/2}$$

Proof of C/2 bound

- It can be shown that

$$\operatorname{Tr} \rho_j^{1/2} \rho_k^{1/2} \leq \sqrt{\min_s \operatorname{Tr} \rho_j^s \rho_k^{1-s}} = \exp(-C(\rho_j, \rho_k)/2).$$

- Thus

$$P_e^* \leq 2 \sum_{j<k} \sqrt{p_j p_k} \exp(-C(\rho_j, \rho_k)/2).$$

- The asymptotic error rate is therefore bounded below by the minimal Chernoff distance *times one half*. □

4. Other bounds on the optimal error probability

Pairwise discrimination

- We need bounds on the optimal error probability for discriminating between r operators A_k in terms of the *pairwise* optimal error probabilities.
- Define

$$P^*_{s,2} = \frac{1}{r-1} \sum_{k,l:\, k<l} \operatorname{Tr} \operatorname{LUB}(A_k, A_l)$$

$$P^*_{e,2} = \frac{1}{r-1} \sum_{k,l:\, k<l} \operatorname{Tr} \operatorname{GLB}(A_k, A_l).$$

Pairwise discrimination

- The importance of the quantity $P^*_{e,2}$ comes from the fact that its asymptotic rate equals the smallest pairwise Chernoff divergence.
- Indeed, the rate of the term $\operatorname{Tr} \operatorname{GLB}(A_k, A_l)$ is given by $C(\rho_k, \rho_l)$, hence the rate of the sum $\sum_{k,l}$ is the smallest of these rates.
- It is very easily shown that

$$P^*_s \leq P^*_{s,2}, \qquad \text{so that } P^*_e \geq P^*_{e,2}.$$

- This provides an alternative proof that $R \leq C := \min_{k<l} C(\rho_k, \rho_l)$.

Nottingham, here is a problem

- Numerical results indicate that there is a constant c (which may depend on r) such that

$$P_e^* \leq c\, P_{e,2}^*.$$

- Open question: prove this!
- If P_e^* differs from $P_{e,2}^*$ only by a constant, independent of the dimension, hence independent of n, then it must have the same rate.
- **Problem 1.** Show that there is a constant c, independent of the dimension of the A_i, but possibly depending on r, such that

$$1 - \operatorname{Tr} \operatorname{LUB}(A_1, \ldots, A_r) = \operatorname{Tr} \operatorname{GLB}(\overline{A}_1, \ldots, \overline{A}_r) \leq c \sum_{k,l:\, k<l} \operatorname{Tr} \operatorname{GLB}(A_k, A_l) \qquad ?$$

Problem 2

- I was able to show:

$$1 - \operatorname{Tr} \operatorname{LUB}(A_1, \ldots, A_r) \leq \sum_k \operatorname{Tr} \operatorname{GLB}(A_k, \overline{A}_k).$$

- Proof relies on the following proposition:
  Let $\{P_i\}$ be a set of r projectors and let $P_0 = \sum_i P_i$, then the set $(2P_i - P_0)_+$ forms an incomplete POVM.
- **Problem 2.** Show that there is a constant c' such that

$$\sum_k \operatorname{Tr} \operatorname{GLB}(A_k, \overline{A}_k) \leq c' \sum_{k,l:\, k<l} \operatorname{Tr} \operatorname{GLB}(A_k, A_l) \qquad ?$$

- This would imply a positive answer to Problem 1.

Thank You!

Homework questions:

Prove

$$1 - \operatorname{Tr} \operatorname{LUB}(A_1, \ldots, A_r) = \operatorname{Tr} \operatorname{GLB}(\overline{A}_1, \ldots, \overline{A}_r) \leq c \sum_{k,l:\, k<l} \operatorname{Tr} \operatorname{GLB}(A_k, A_l)$$

$$\sum_k \operatorname{Tr} \operatorname{GLB}(A_k, \overline{A}_k) \leq c' \sum_{k,l:\, k<l} \operatorname{Tr} \operatorname{GLB}(A_k, A_l).$$