Quantum Hypothesis Testing

Non-Commutative Chernoff and Hoeffding bounds

Institute for Mathematical Sciences

Koenraad M.R. Audenaert

Quantum Information at Imperial College London

February 16, 2007

Impressum

Based on joint work with:

- J. Calsamiglia, R. Munoz-Tapia, E. Bagan, Ll. Masanes, A. Acin (Barcelona)
- F. Verstraete (Vienna)
- M. Nussbaum (Cornell) and A. Szkola (Leipzig).

Read more about it in:

- KA et al. et F. Verstraete, “The Quantum Chernoff Bound”, **quant-ph/0610027**, to appear in PRL.
- M. Nussbaum and A. Szkola, **quant-ph/0607216**.
- ... and in a forthcoming full-length paper.

Distinguishing coins

- I have two types of coins in my pocket:
  - Coin H_0 is unbiased: heads/tails distributed according to $p = (1/2, 1/2)$
  - Coin H_1 is biased: heads/tails distributed according to $q = (q_H, q_T)$
- I take one coin, and want to know of which type it is.
- Question 1: How can I distinguish the coins, minimising the error?
- Depends on how you define the error.
- Question 2: How many throws are needed before I can tell this with near-certainty?
- This will tell me how good my "decision rule" is...
- ...but also how much H_0 and H_1 are alike.

Distinguishing coins

- What I can do is: throw the coin n times and see how much heads come up.
- What I know about the coins is:
  - With coin H_0, heads come up k times in n throws with probability

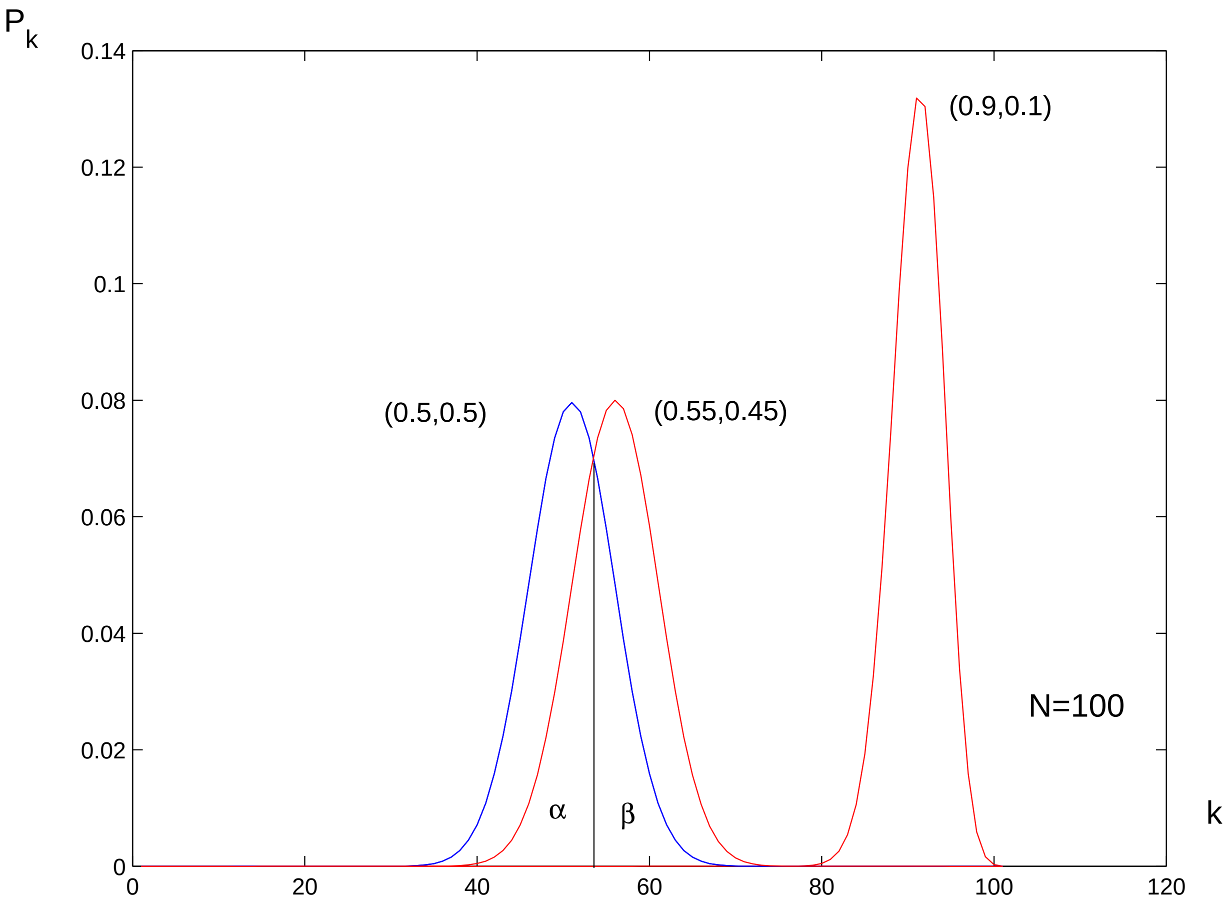
$$P_k = \binom{n}{k}\ 1/2^n.$$

  - With coin H_1 this probability is

$$Q_k = \binom{n}{k}\ q_H^k\ q_T^{n-k}.$$

- Say, in an actual experiment, heads come up k times out of n.
- Maximum Likelihood (ML) Decision rule: if $P_k > Q_k$, decide H_0, else H_1.

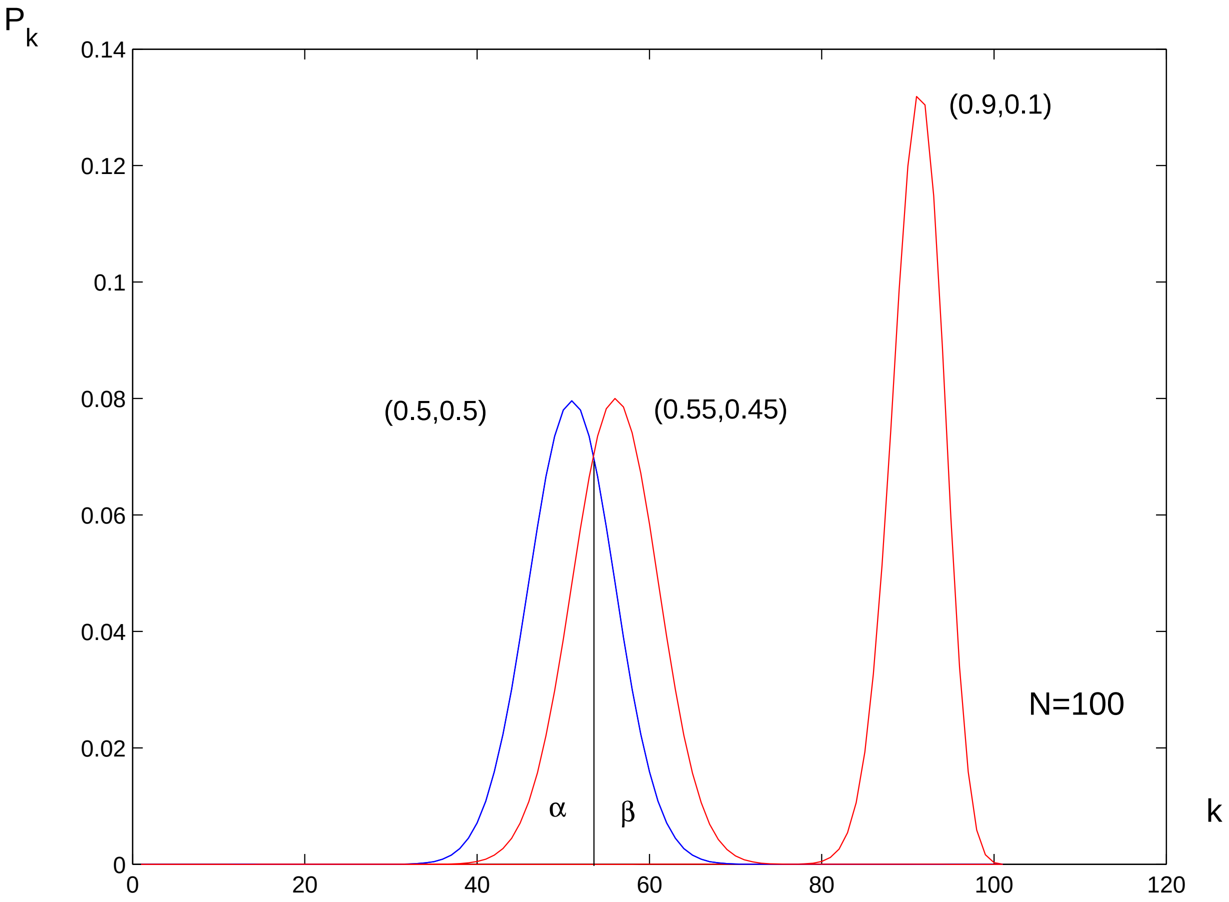
ML Decision rule



Error probabilities

- Type-I error: decide on H_0 while H_1 is true; probability α
- Type-II error: decide on H_1 while H_0 is true; probability β
- In symmetric hypothesis testing, type-I and type-II errors treated equally, via:
- "Total" or Bayesian error probability: $P_e = (\alpha + \beta)/2$ (assuming equal priors).
- Quantifies the "cost" of making a mistake.
- This P_e is what we want to minimise.

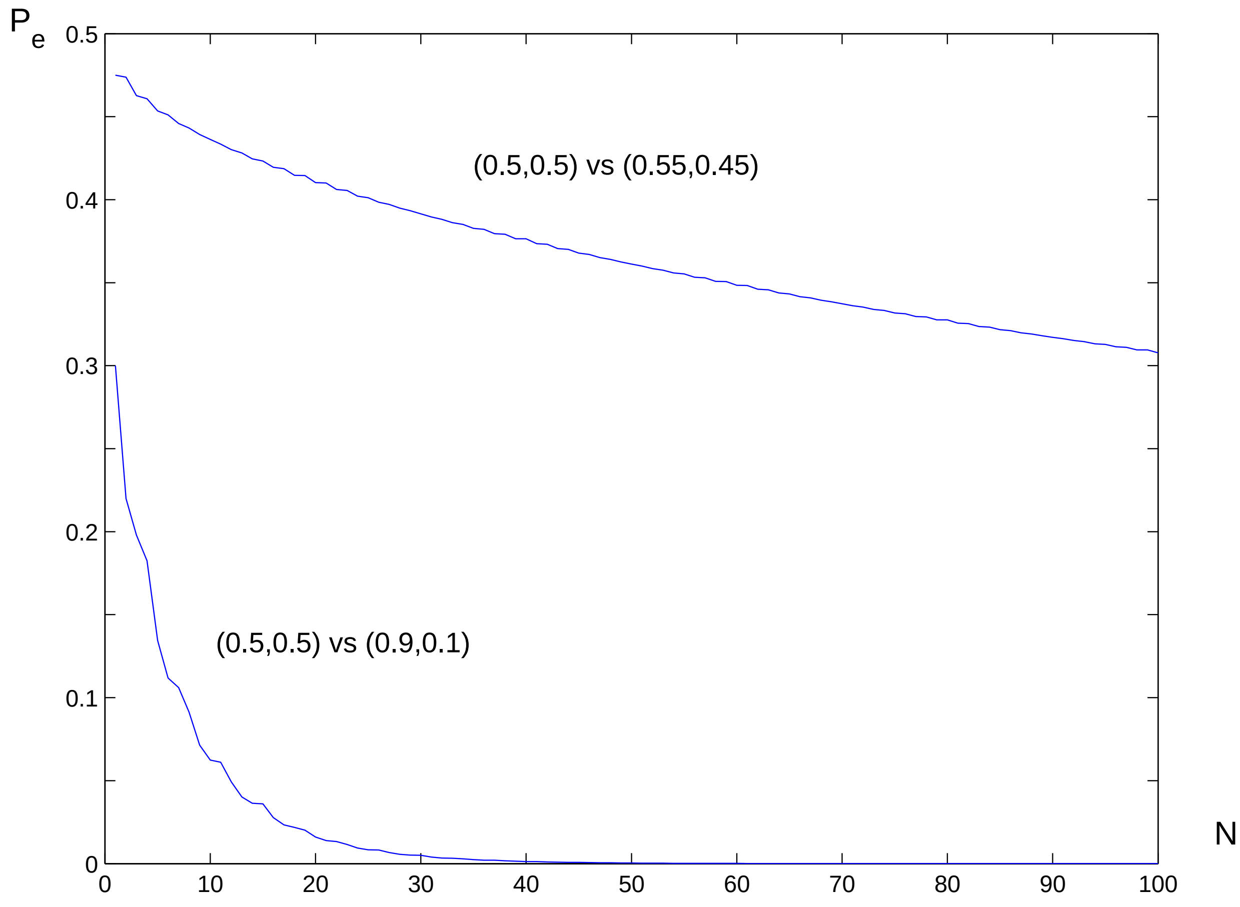
Error probabilities



Answer 1

- Answer to Question 1: total error minimised by ML decision rule.
- What about Question 2? How big must n be to get "negligible" error?
- Depends on definition of "negligible".
- Let's look at how total error behaves in terms of n.

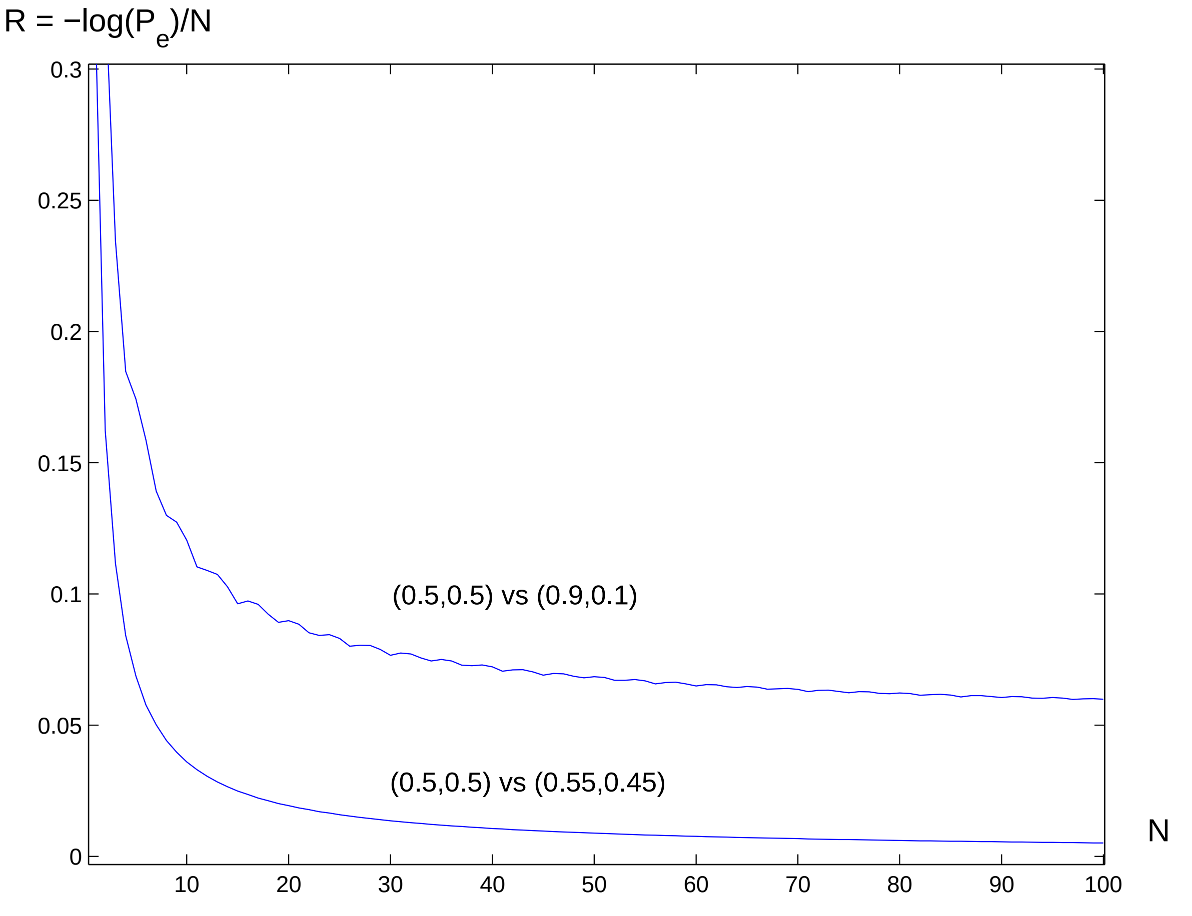
Total Error Probability v n



Answer 2

- Total Error Probability goes roughly as $\exp(-nR)$.
- Exponent R is the error rate (error exponent).
- We can take R as a qualitative answer to Question 2.
- It quantifies how well we're doing, given p and q: efficiency of the decision rule
- In turn quantifies how alike p and q are: gives a distance measure on distributions
- Well, almost...

We need the Asymptotic Error Rate



Asymptotic Error Rate

- Asymptotic error rate hard to calculate directly: large n
- H. Chernoff (1952): Simple formula for asymptotic error rate:

$$\lim_{n\to\infty} -\frac{1}{n}\log P_e = -\log Q(p,q),$$

where Q is defined as

$$Q(p,q) = \min_{0\le s\le 1} \sum_i p_i^s q_i^{1-s}.$$

- The quantity $-\log Q$ is called the *Chernoff Distance* (Divergence, Bound).
- It is a measure of distinguishability between distributions.
- $-\log Q((0.5,0.5),(0.9,0.1)) = 0.0488$
- $-\log Q((0.5,0.5),(0.55,0.45)) = 0.000545$

Asymptotic Error Rate

- Asymptotic error rate hard to calculate directly: large n
- H. Chernoff (1952): Simple formula for asymptotic error rate:

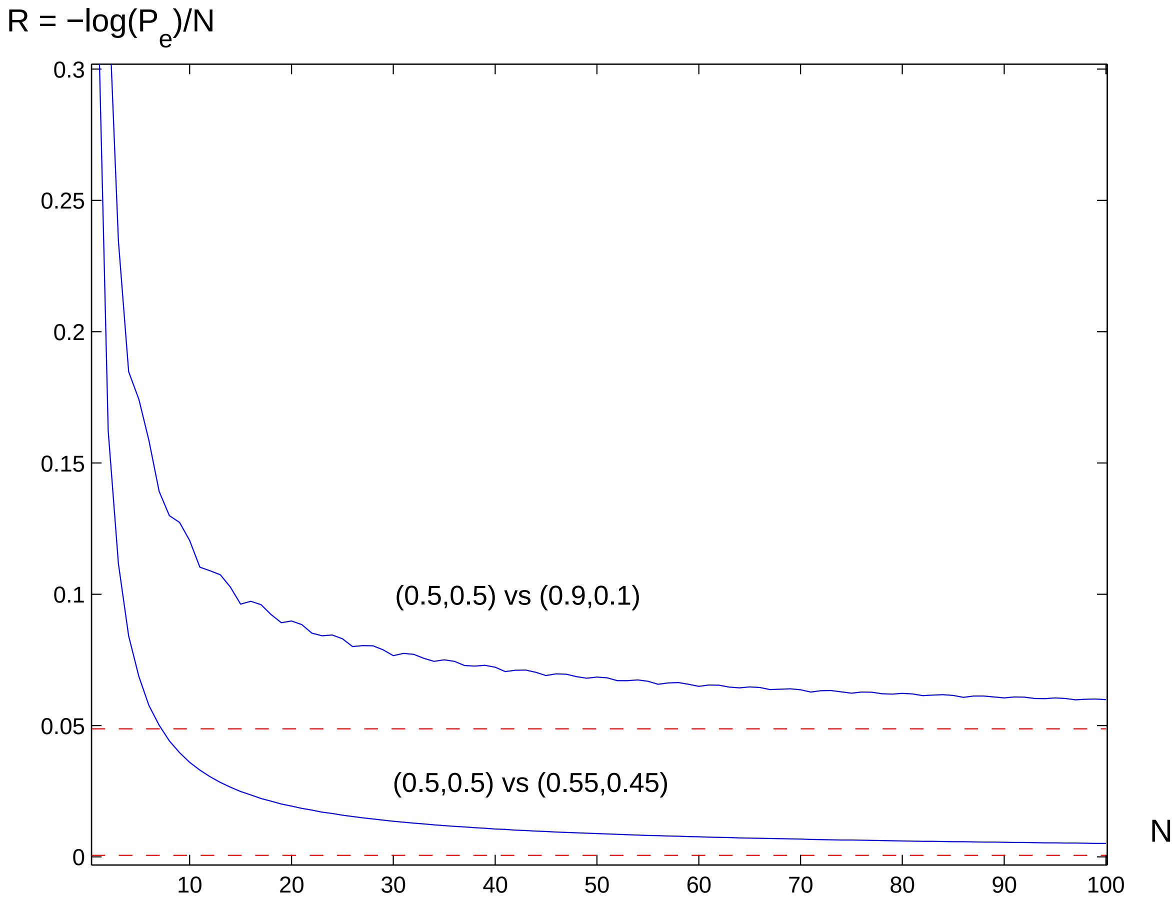
$$\lim_{n\to\infty} -\frac{1}{n}\log P_e = -\log Q(p,q),$$

where Q is defined as

$$Q(p,q) = \inf_{0\leq s\leq 1} \sum_i p_i^s q_i^{1-s}.$$

- The quantity $-\log Q$ is called the *Chernoff Distance* (Divergence, Bound).
- It is a measure of distinguishability between distributions.
- $-\log Q((0.5,0.5),(0.9,0.1)) = 0.0488$
- $-\log Q((0.5,0.5),(0.55,0.45)) = 0.000545$

Asymptotic Error Rate



The Quantum Chernoff Bound

- Can we "*quantise*" this?
- Question 1: What is the optimal symmetric hypothesis test for discriminating between two *quantum states* ρ and σ?
- Quantum measurement theory by Helstrom and Holevo from 70's:
  - Hypothesis H_0: n draws yield state $\rho^{\otimes n}$
  - Hypothesis H_1: n draws yield state $\sigma^{\otimes n}$
  - ML decision rule $\mapsto$ "optimal measurement"
- Question 2: What is the error exponent?
- Would yield a distinguishability measure for quantum states
- Answered last year.

Optimal Measurement

- Quantum version of n throws = $\rho^{\otimes n}$ vs $\sigma^{\otimes n} \in \mathcal{H}^{\otimes n}$.
- Measurement = POVM $\{E_0, E_1\}$ on $\mathcal{H}^{\otimes n}$, with $0 \leq E_0, E_1 \leq \mathbb{1}$ and $E_0+E_1 = \mathbb{1}$.
- Decide on H_0 if outcome is '0' (E_0), otherwise H_1.
- Type-I error: $\alpha_n = \mathrm{Tr}[E_0\, \sigma^{\otimes n}]$, Type-II error: $\beta_n = \mathrm{Tr}[E_1\, \rho^{\otimes n}]$.
- Total error: $P_{e,n} = (\alpha_n + \beta_n)/2$ (assuming equal priors).
- Optimal measurement: minimise P_e over all E_0, E_1

$$\begin{aligned} P_{e,\min,n} &= \min_{0\leq E_1\leq \mathbb{1}} \mathrm{Tr}[(\mathbb{1} - E_1)\, \sigma^{\otimes n} + E_1\, \rho^{\otimes n}]/2 \\ &= (1 - \max_{0\leq E_1\leq \mathbb{1}} \mathrm{Tr}[E_1\, (\sigma^{\otimes n} - \rho^{\otimes n})])/2. \end{aligned}$$

- Solution is based on the *positive part* of an operator/matrix.

The Positive Part

- The positive part H_+ of a Hermitian matrix H is obtained by setting its negative eigenvalues equal to 0.
- In terms of the matrix absolute value: $H_+ = (H + |H|)/2$.
- If P is the projector on (the support of) H_+, we can write $H_+ = PH$.
- For all Hermitian H, one has $H_+ \geq H$, and $H_+ \geq 0$.
- Variational expression for $\mathrm{Tr}\, H_+$: $\mathrm{Tr}\, H_+ = \max_Q \mathrm{Tr}\, QH$, where the maximisation is over all Hermitian projectors Q, and the optimum is achieved in $Q = P$, the projector on H_+.
- Variant: maximise $\mathrm{Tr}\, QH$ over all positive contractions Q ($0 \leq Q \leq \mathbb{1}$). Same answer.

Optimal Measurement

- To Do: find $\max_{0\le E_1\le \mathbb{1}} \mathrm{Tr}[E_1\,(\sigma^{\otimes n}-\rho^{\otimes n})]$.
- Maximisation over positive contractions E_1!
- Optimal E_1 is therefore the projector on $(\sigma^{\otimes n}-\rho^{\otimes n})_+$.
- Optimal value:

$$\begin{aligned}\mathrm{Tr}[(\sigma^{\otimes n}-\rho^{\otimes n})_+] &= (\mathrm{Tr}[\sigma^{\otimes n}-\rho^{\otimes n}]+\mathrm{Tr}[|\sigma^{\otimes n}-\rho^{\otimes n}|])/2 \\ &= ||\sigma^{\otimes n}-\rho^{\otimes n}||_1/2.\end{aligned}$$

- Total error probability of the optimal measurement scheme is thus

$$P_{e,\min,n}=(1-T(\rho^{\otimes n},\sigma^{\otimes n}))/2, \quad T(\rho,\sigma):=||\rho-\sigma||_1/2.$$

The Quantum Chernoff Bound

- Again, P_e goes down exponentially with n, with asymptotical rate

$$\lim_{n\to\infty} -\frac{1}{n}\log(1 - T(\rho^{\otimes n}, \sigma^{\otimes n}))$$

- Can we find a closed-form expression in the sense of Chernoff?
- Long-standing open problem.
- Ogawa and Hayashi (2004): three candidate expressions, based on the quantities

$$\begin{aligned}
\psi_1(s) &= \min\{\mathrm{Tr}[\rho\sigma^{s/2}\rho^{-s}\sigma^{s/2}], \mathrm{Tr}[\sigma\rho^{(1-s)/2}\sigma^{-(1-s)}\rho^{(1-s)/2}]\} \\
\psi_2(s) &= \mathrm{Tr}[\rho^s\sigma^{1-s}] \\
\psi_3(s) &= \mathrm{Tr}[\exp((1-s)\log\rho + s\log\sigma)],
\end{aligned}$$

each of which reduces to $\sum_k p_k^s q_k^{1-s}$ for commuting ρ and σ.

Candidate #2 is an upper bound

- Nussbaum and Szkola ('06) proved that candidate #2,

$$-\log \min_{0\leq s\leq 1} \mathrm{Tr}[\rho^s \sigma^{1-s}],$$

is an upper bound to the error rate.

- Proof is based on a very special mapping of pairs of d-dim. states to pairs of d^2-dim. probability vectors:

$$\rho = U\Lambda U^*, \sigma = VMV^* \mapsto p = \mathrm{vec}(\Lambda W), q = \mathrm{vec}(WM),$$

where W is an entrywise positive matrix s.t. $\mathrm{Tr}[\rho\sigma] = \sum_{i,j}(\Lambda W M)_{i,j}$.

- Can this bound be achieved? Is it also a lower bound?
- If so, this solves the problem completely!

Is the bound of candidate #2 achievable?

- Let us define the quantity $Q(\rho, \sigma) := \min_{0 \le s \le 1} \mathrm{Tr}[\rho^s \sigma^{1-s}]$.
- We have

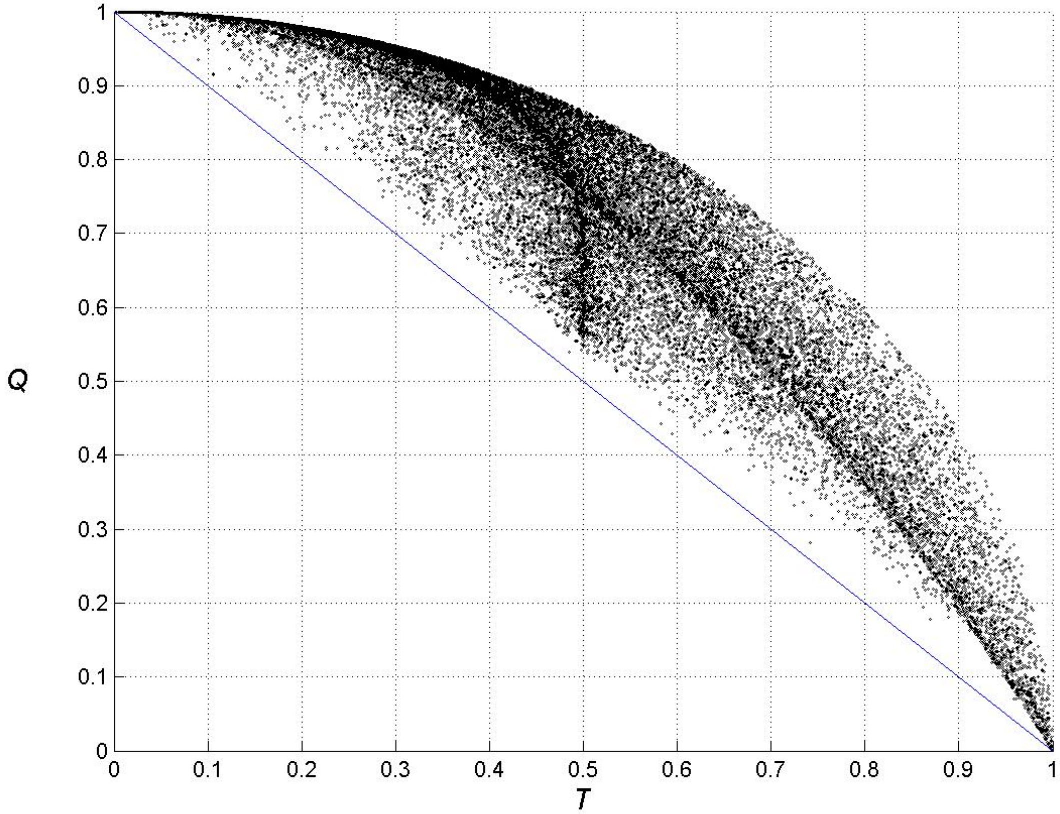
$$-\log Q(\rho, \sigma) \ge \lim_{n \to \infty} -\frac{1}{n} \log P_e = \lim_{n \to \infty} -\frac{1}{n} \log(1 - T(\rho^{\otimes n}, \sigma^{\otimes n})).$$

- Now we want to know whether

$$-\log Q(\rho, \sigma) \le \lim_{n \to \infty} -\frac{1}{n} \log(1 - T(\rho^{\otimes n}, \sigma^{\otimes n})).$$

- Let's try a simple numerical experiment to get a feel for the problem: plot $Q(\rho, \sigma)$ vs $T(\rho, \sigma)$, for various d.

Matlab scatter plot of Q vs T



Is the bound #2 achieved?

- Not only do we get a feel for the problem, we actually get the solution!
- These numerics suggest that, in any dimension:

$$Q(\rho, \sigma) \geq 1 - T(\rho, \sigma).$$

- Thus, in particular,

$$Q(\rho^{\otimes n}, \sigma^{\otimes n}) \geq 1 - T(\rho^{\otimes n}, \sigma^{\otimes n}).$$

- Now, Q is multiplicative w.r.t. tensor powers:

$$\log Q(\rho^{\otimes n}, \sigma^{\otimes n}) = n \log Q(\rho, \sigma)$$

- That would imply achievability!

$$-\log Q(\rho, \sigma) \leq \lim_{n\to\infty} -\frac{1}{n}(1 - T(\rho^{\otimes n}, \sigma^{\otimes n})) = \lim_{n\to\infty} -\frac{1}{n} \log P_e.$$

Main Theorem

- We are thus led to conjecture the validity of the following statement, which is the generalisation of the inequality $Q + T \geq 1$ to non-normalised positive operators:

For all positive operators $a, b \geq 0$*, and for all* $s \in [0, 1]$ *one has:*

$$\mathrm{Tr}[a^s b^{1-s}] \geq \mathrm{Tr}\left[(a+b) - |a-b|\right]/2.$$

- Amazing features: tensor powers not explicitly appearing, no limiting process needed
- Nice matrix analysis problem!
- We get a truly “quantum” solution!

Overview of what is to come

- A 'heuristic' walk through the proof, highlighting its main ingredients:
  - a few tricks from matrix analysis
  - tons of luck
- Implications of the Theorem.
- Properties of Q and $-\log(Q)$.
- Further applications of the techniques we have used.

Proof, Step 1

- The LHS and RHS look very different, but can be brought closer together by expressing $(a+b)-|a-b|$ in terms of the *positive part* $(a-b)_+$.
- The statement of the Theorem is equivalent to

$$\begin{aligned}
\mathrm{Tr}[a-a^s b^{1-s}] &\le \mathrm{Tr}[a-((a+b)-|a-b|)/2] \\
&= \mathrm{Tr}[((a-b)+|a-b|)/2] \\
&= \mathrm{Tr}[(a-b)_+] \\
&= \mathrm{Tr}[Q(a-b)],
\end{aligned}$$

with Q the projector on $(a-b)_+$.

- Other formulation (used in proof of Hoeffding bound):

$$\mathrm{Tr}[a^s b^{1-s}] \ge \mathrm{Tr}[Qb+(\mathbb{1}-Q)a].$$

Proof, Step 1

- We can do sth like that in LHS too:

$$\begin{aligned}\mathrm{Tr}[a - a^s b^{1-s}] = \mathrm{Tr}[a^s(a^{1-s} - b^{1-s})] &\leq \mathrm{Tr}[a^s(a^{1-s} - b^{1-s})_+] \\ &= \mathrm{Tr}[a^s P(a^{1-s} - b^{1-s})] \\ &= \mathrm{Tr}[P(a - b^{1-s}a^s)],\end{aligned}$$

  where P is the projector on $(a^{1-s} - b^{1-s})_+$.

- Thee Theorem would follow if, for that P,

$$\mathrm{Tr}[P(a - b^{1-s}a^s)] \leq \mathrm{Tr}\, P(a - b).$$

- After simplification: $\mathrm{Tr}[Pb^{1-s}(a^s - b^s)] \geq 0.$
- Much nicer form, but also much stronger (Don't try this at home!)
- Still..., what to do with all those matrix powers?

Getting rid of one matrix power

- Step 2: absorb one of the powers via appropriate substitution.
- We certainly don't want a power in the definition of projector P, so let's use

$$A = a^{1-s}, \quad B = b^{1-s}, \quad t = s/(1-s).$$

- This yields a t between 0 and 1 only when $0 \leq s \leq 1/2$.

  The case $1/2 \leq s \leq 1$ can be treated after the substitution $s \to 1-s$.
- The Theorem is thus implied by the statement ("Lemma"):

$$\mathrm{Tr}[PB(A^t - B^t)] \geq 0,$$

  for $A, B \geq 0$, and $0 \leq t \leq 1$, and P the projector on $(A-B)_+$.
- What about the remaining power?

Getting rid of the second matrix power

- Inspired by Loewner's theory of operator monotones...
- Step 3: Represent matrix power A^t using integral (V.56).
- For scalars $a \geq 0$ and $0 \leq t \leq 1$

$$a^t = \frac{\sin(t\pi)}{\pi} \int_0^{+\infty} dx\ x^{t-1}\ \frac{a}{a+x}.$$

- This can be extended to positive operators:

$$A^t = \frac{\sin(t\pi)}{\pi} \int_0^{+\infty} dx\ x^{t-1}\ A(A + x\mathbb{1})^{-1}.$$

- Potential benefit: statements about the integral might follow from statements about the integrand, which is a simpler quantity.

Getting rid of the matrix powers

- Applying the integral representation to A^t and B^t, we get

$$\mathrm{Tr}[PB(A^t - B^t)] = \frac{\sin(t\pi)}{\pi} \int_0^{+\infty} dx\ x^{t-1}\,\mathrm{Tr}[PB(A(A+x)^{-1} - B(B+x)^{-1})].$$

- If the integrand is positive for all $x > 0$, the whole integral is positive.
- The Theorem follows if indeed we have

$$\mathrm{Tr}[PB(A(A+x)^{-1} - B(B+x)^{-1})] \geq 0.$$

- Again a stronger statement!
- But this not nice enough yet: products and difference.
- I want all products.

Integral Representation of a Difference

- Step 4: A difference can be expressed as an integral of a derivative:

$$f(a) - f(b) = f(b + (a - b)) - f(b) = \int_0^1 dt\, \frac{d}{dt} f(b + (a - b)t)$$

- Here: apply this to the expression $A(A + x)^{-1} - B(B + x)^{-1}$.
  Let $\Delta = A - B$. Then

$$A(A + x)^{-1} - B(B + x)^{-1} = \int_0^1 dt\, \frac{d}{dt}(B + t\Delta)(B + t\Delta + x)^{-1}.$$

- Potential benefit: statement might again follow from statement about integrand.
- One may be able to calculate the derivative explicitly.
- Not a stronger statement: has to hold for the derivative anyway (A close to B).

Getting rid of the difference

- We can indeed calculate the derivative:

$$\frac{d}{dt}(B+t\Delta)(B+t\Delta+x)^{-1} = x\,(B+t\Delta+x)^{-1}\,\Delta\,(B+t\Delta+x)^{-1}.$$

  Therefore,

$$\begin{aligned}&\mathrm{Tr}[PB(A(A+x)^{-1}-B(B+x)^{-1})]\\ &= x\int_0^1 dt\ \mathrm{Tr}[PB(B+t\Delta+x)^{-1}\Delta(B+t\Delta+x)^{-1}].\end{aligned}$$

- Again, if the integrand is positive for $0 \le t \le 1$, the whole integral is.
- Absorbing t in Δ we need to show, with P the projector on Δ_+:

$$\mathrm{Tr}[PB\,V\,\Delta\,V] \ge 0, \qquad \text{where } V := (B+\Delta+x)^{-1} \ge 0.$$

Final Steps

- Since $B = V^{-1} - x - \Delta$, we have $BV\Delta V = \Delta(V - V\Delta V) - xV\Delta V$.
- $B \geq 0$ implies $VBV = V - V\Delta V - xV^2 \geq 0$,
  thus $V - V\Delta V \geq xV^2$, and

$$\begin{aligned} \mathrm{Tr}[PBV\Delta V] &= \mathrm{Tr}[\Delta_+(V - V\Delta V)] - x\,\mathrm{Tr}[PV\Delta V] \\ &\geq x(\mathrm{Tr}[\Delta_+V^2] - \mathrm{Tr}[PV\Delta V]), \end{aligned}$$

  since $P\Delta = \Delta_+ \geq 0$.
- Because $\mathbb{1} \geq P \geq 0$, $\Delta_+ \geq 0$, and $\Delta_+ \geq \Delta$,

$$\mathrm{Tr}[\Delta_+V^2] = \mathrm{Tr}[V\Delta_+V] \geq \mathrm{Tr}[P(V\Delta_+V)] \geq \mathrm{Tr}[P(V\Delta V)].$$

- Conclusion: $\mathrm{Tr}[PBV\Delta V] \geq 0$. $\square$

Importance of this result

- Having defined a quantum version of Chernoff's quantity

$$Q(\rho, \sigma) = \min_{0 \le s \le 1} Q_s, \qquad Q_s := \mathrm{Tr}\, \rho^s \sigma^{1-s}$$

  "we" have proven that the asymptotic error rate in symmetric hypothesis testing is given by $-\log Q$.
  We can thus rightfully call $-\log Q$ the *Quantum Chernoff Bound*.

- The quantities $-\log Q_s$ are known as the Renyi relative entropies.
  Since $-\log Q = \max_s(-\log Q_s)$, this gives the Renyi relative entropies a full operational meaning.

- The QCB has properties that make it an excellent distinguishability measure.

Coming up next

- We discuss some properties of Q...
- ...and show that Q and $-\log Q$ are excellent distinguishability measures, lacking many undesirable features of other measures.

Coming up next

- We discuss some properties of Q...
- ...and show that Q and $-\log Q$ are excellent distinguishability measures, lacking many undesirable features of other measures.
- The following 3 pages are to be inserted at the end of Chapter 13 of Bengtsson and Zyczkowski.

Properties of QCB

Inverted measure. — The maximum value Q can attain is 1, and this is reached when $\rho = \sigma$. The minimal value is 0, and this is only attained for pairs of orthogonal states, i.e. states such that $\rho\sigma = 0$. If you don't like the log in $-\log Q$, use $1 - Q$.

Convexity in s. — The function to be minimised in Q is $s \mapsto \mathrm{Tr}[\rho^s \sigma^{1-s}]$ which is convex in $s \in [0, 1]$. That means that the minimisation has only one local minimum. This makes numerical and analytical calculations very efficient.

Joint concavity. — $Q(\rho, \sigma)$ is jointly concave in (ρ, σ), by Lieb concavity.

Monotonicity under CPT maps. — For all CPT maps Φ, $Q(\Phi(\rho), \Phi(\sigma)) \geq Q(\rho, \sigma)$.

Properties of QCB

Relation to Trace Norm Distance: $T(\rho, \sigma) := ||\rho - \sigma||_1/2$
We can show $0 \leq 1 - Q \leq T \leq \sqrt{1 - Q^2}$.
The lower bound implies that Q is continuous: states that are close in trace norm distance are also close in $1 - Q$ distance.

Relation to Uhlmann Fidelity: $F(\rho, \sigma) := ||\rho^{1/2}\sigma^{1/2}||_1$
F is an upper bound to Q. Indeed: $Q \leq Q_{s=1/2} = \mathrm{Tr}\, \rho^{1/2}\sigma^{1/2} \leq F$.
If the states are pure, then equality holds.

Relation to Overlap:
If one of the states is pure, Q is equal to the overlap $\mathrm{Tr}\, \rho\sigma$.
Indeed, if $\rho = |\psi\rangle\langle\psi|$ is pure, the optimum s is 0.

Properties of QCB

Relation to Relative Entropy: $S(\rho||\sigma) := \mathrm{Tr}\,\rho(\log\rho - \log\sigma)$
When dealing with pure states, the relative entropy is pretty useless: $S = 0$ only when the states are the same, otherwise it is $+\infty$. In contrast, $-\log Q$ is infinite only when the states have disjoint support, e.g. for orthogonal pure states.

Interpretation of optimal s: "Quantum Hellinger arc"
Define, for s between 0 and 1, the (non self-adjoint) operator

$$\tau_s := \frac{\rho^s \sigma^{1-s}}{\mathrm{Tr}\,\rho^s \sigma^{1-s}}.$$

Optimal s in Q is achieved for τ_s the metric midpoint between ρ and σ:

$$S(\tau_s||\rho) = S(\tau_s||\sigma).$$

Asymmetric Hypothesis Testing

- For distinguishing coins, type-I and type-II errors are treated equally.
- But what if the 'costs' of the two types of error are different, or even incommensurate?
- "What colour did my wife want again for the living room?"
  - H_0: Beige $\qquad$ H_1: Hot Pink
  - Type-II error: repaint more likely
- Medical Diagnosis:
  - H_0: Ordinary Flu $\qquad$ H_1: Bird Flu
  - Type-I error: expensive and annoying
  - Type-II error: might be lethal

Quantum Hoeffding Bound

- Several ways for dealing with asymmetry: Stein's Lemma, Hoeffding bound.
- Let α_R and β_R be the asymptotic rates of α and β.
- Quantum Hoeffding bound: under the constraint $\beta_R \geq r$, a_R is at most $e(r)$, the *error-exponent function*

$$e(r) = \max_{0 \leq s \leq 1} \frac{-rs - \log Q_s(\rho, \sigma)}{1-s}.$$

- Proof of optimality: Nagaoka, using the Nussbaum-Szkola mapping.
- Proof of achievability: Hayashi, with the inequality used for Quantum Chernoff.