A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks

Marco Tomamichel*, Masahito Hayashi†*

arXiv: 1208.1478

Also discussing results of:

Second Order Asymptotics for Quantum Hypothesis Testing

Ke Li*, arXiv: 1208.1400

*CQT, National University of Singapore
†Graduate School of Mathematics, Nagoya University

Cambridge, January 2013

Outline

1. Hypothesis Testing and β^{ε}
2. Main Result: Asymptotic Expansion of β^{ε} and Smooth Entropies
3. Two Applications
   - Randomness Extraction against Quantum Side Information
   - Data Compression with Quantum Side Information
4. Finite Block Length Results
5. Main Proof Ideas
6. Conclusion and Discussion

Quantum Hypothesis Testing

- Given one out of two quantum states, either ρ or σ, we must decide which one we received.
- For a given *test POVM*, $\{Q, 1-Q\}$, $0 \leq Q \leq 1$, the error of the *first* and *second* kind are $\alpha_\rho(Q) = \mathrm{tr}(\rho(1-Q))$ and $\beta_\sigma(Q) = \mathrm{tr}(\sigma Q)$, respectively.
- We are interested in the minimal β that can be achieved if α is required to smaller than a given constant ε, i.e. the SDP

$$\beta^\varepsilon_{\rho,\sigma} := \min_{\substack{0 \leq Q \leq 1 \\ \alpha_\rho(Q) \leq \varepsilon}} \beta_\sigma(Q) = \min_{\substack{0 \leq Q \leq 1 \\ \mathrm{tr}(\rho Q) \geq 1-\varepsilon}} \mathrm{tr}(\sigma Q).$$

- Alternatively, one may consider the divergence

$$D^\varepsilon_h(\rho\|\sigma) := -\log\left(\frac{\beta^\varepsilon_{\rho,\sigma}}{1-\varepsilon}\right), \qquad 0 < \varepsilon < 1.$$

(The additive normalization $\log(1-\varepsilon)$ ensures that $\rho = \sigma \iff D^\varepsilon_h(\rho\|\sigma) = 0$.)

I.i.d. Asymptotic Expansion of D_h^ε

- Given n copies of either σ and ρ, a quantum generalization of Stein's Lemma (Hiai&Petz'91) and its strong converse (Ogawa&Nagaoka'00) imply

$$D_h^\varepsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) = nD(\rho\|\sigma) + o(n)$$

- This was recently improved (Audenaert,Mosonyi&Verstraete'12)

$$D_h^\varepsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) \leq nD(\rho\|\sigma) + O\left(\sqrt{n}\right) \quad \text{and}$$
$$D_h^\varepsilon\left(\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) \geq nD(\rho\|\sigma) - O\left(\sqrt{n}\right)$$

  by giving explicit upper and lower bounds. However, the terms proportional to $\sqrt{n}$ in the upper and lower bounds are different. (They do not have the same sign!)

- Our goal is to investigate the second order term, $O(\sqrt{n})$.

Main Result

Theorem

For two states ρ, σ with $\mathrm{supp}\{\sigma\} \supseteq \mathrm{supp}\{\rho\}$, and $0 < \varepsilon < 1$, we find[†]

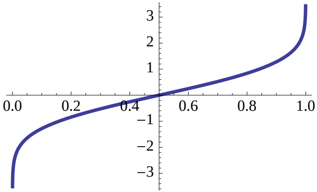
$$D_h^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \leq nD(\rho\|\sigma) + \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\varepsilon) + 2\log n + O(1), \quad \textit{and}$$

$$D_h^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \geq nD(\rho\|\sigma) + \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\varepsilon) - O(1).$$

- D and V are the mean and variance of $\log \rho - \log \sigma$ under ρ, i.e.

$$V(\rho\|\sigma) := \mathrm{tr}\Big(\rho\big(\log \rho - \log \sigma - D(\rho\|\sigma)\big)^2\Big).$$

- Φ is the cumulative normal distribution function, and $\Phi^{-1}(\varepsilon)$ is



[†] The bounds given here are due to Li. A similar result was independently derived by T&H.

Main Result

Theorem

For two states ρ, σ with $\mathrm{supp}\{\sigma\} \supseteq \mathrm{supp}\{\rho\}$*, and* $0 < \varepsilon < 1$*, we find*[†]

$$D_h^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \leq nD(\rho\|\sigma) + \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\varepsilon) + 2\log n + O(1), \quad \textit{and}$$

$$D_h^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) \geq nD(\rho\|\sigma) + \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\varepsilon) - O(1).$$

- We also have bounds on the constant terms, enabling us to calculate upper and lower bounds on $D_h^{\varepsilon}(\rho^{\otimes n} \| \sigma^{\otimes n})$ for finite n.
- Classically, the above holds with logarithmic terms in upper and lower bound equal to $\frac{1}{2}\log n$ (e.g. Strassen'62,Polyanskiy,Poor&Verdú'10).
- One ingredient of both proof is the Berry-Essèen theorem, which quantizes the convergence of the distribution of a sum of i.i.d. random variables to a normal distribution.
- Intuitively, our results can be seen as quantum, entropic formulation of the central limit theorem.

[†] The bounds given here are due to Li. A similar result was independently derived by T&H.

Smooth Entropies

- We also investigate the smooth min-entropy (Renner'05), where it was known (T,Colbeck&Renner'09,T'12) that, for $\varepsilon \in (0,1)$,

$$H_{\min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} \leq nH(A|B)_\rho + O\big(\sqrt{n}\big), \quad \text{and}$$
$$H_{\min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} \geq nH(A|B)_\rho - O\big(\sqrt{n}\big).$$

- We derive the following expansion

$$H_{\min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} \leq nH(A|B)_\rho + \sqrt{nV(A|B)_\rho}\,\Phi^{-1}(\varepsilon^2) + O(\log n),$$
$$H_{\min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} \geq nH(A|B)_\rho + \sqrt{nV(A|B)_\rho}\,\Phi^{-1}(\varepsilon^2) - O(\log n),$$

where $H(A|B)_\rho = D(\rho_{AB}\|1_A \otimes \rho_B)$ and $V(A|B)_\rho = V(\rho_{AB}\|1_A \otimes \rho_B)$.

- Both hypothesis testing and smooth entropies have various applications in information theory, some of which we explore next.

Randomness Extraction against Side Information

- Consider a CQ random source that outputs states $\rho_{XE} = \sum_x p_x |x\rangle\langle x| \otimes \rho_E^x$.
- Investigate the amount of randomness that can be extracted from X such that it is independent of E and the seeded randomness S.
- A protocol $\mathcal{P} : XS \to ZS$ extracts a random number Z from X, producing a state τ_{ZES} when applied to $\rho_{XE} \otimes \rho_S$.
- For any $0 \leq \varepsilon < 1$ and ρ_{XE} a CQ state, we define

$$\ell^\varepsilon(X|E) := \max\left\{\ell \in \mathbb{N} \,\middle|\, \exists \mathcal{P}, \sigma_E : |Z| = 2^\ell \wedge \tau_{ZES} \approx^\varepsilon 2^{-\ell} 1_Z \otimes \sigma_E \otimes \tau_S \right\}.$$

- This quantity can be characterized in terms of the smooth min-entropy (Renner'05). We tighten this and show

Theorem

Consider an i.i.d. source $\rho_{X^nE^n} = \rho_{XE}^{\otimes n}$ and $0 < \varepsilon < 1$. Then,

$$\ell^\varepsilon(X^n|E^n) \lesseqgtr nH(X|E) + \sqrt{nV(X|E)}\Phi^{-1}(\varepsilon^2) \pm O(\log n).$$

Data Compression with Side Information

- Consider a CQ random source that outputs states $\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \rho_B^x$.
- Find the minimum encoding length for data reconciliation of X if quantum side information B is available.
- A protocol $\mathcal{P}$ encodes X into M and then produces an estimate X' of X from B and M.
- For any $0 \leq \varepsilon < 1$ and ρ_{XB} a CQ state, we define

$$m^{\varepsilon}(X|B)_{\rho} := \min \left\{ m \in \mathbb{N} \,\middle|\, \exists \mathcal{P} : |M| = 2^m \wedge P[X \neq X'] \leq \varepsilon \right\}.$$

- This quantity can be characterized using hypothesis testing (H&Nagaoka'04). We tighten this and show

Theorem

Consider an i.i.d. source $\rho_{X^nB^n} = \rho_{XB}^{\otimes n}$ and $0 < \varepsilon < 1$. Then,

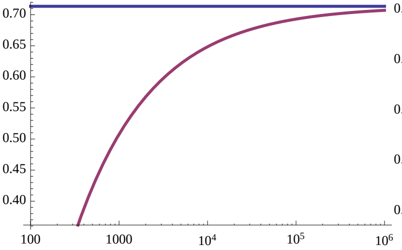
$$m^{\varepsilon}(X^n|B^n) \lesseqgtr nH(X|B) - \sqrt{nV(X|B)}\Phi^{-1}(\varepsilon) \pm O(\log n).$$

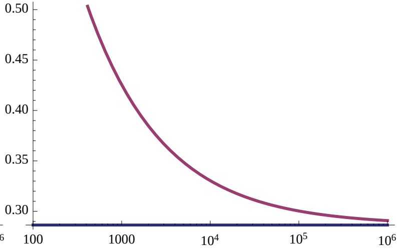
Example of Second Order Asymptotics

- Consider transmission of $|0\rangle$, $|1\rangle$ through a Pauli channel to B (phase and bit flip independent) with environment E. This yields the states

$$\rho_{XB} = \frac{1}{2}\sum |x\rangle\langle x| \otimes \big((1-p)|x\rangle\langle x| + p|1-x\rangle\langle 1-x|\big),$$

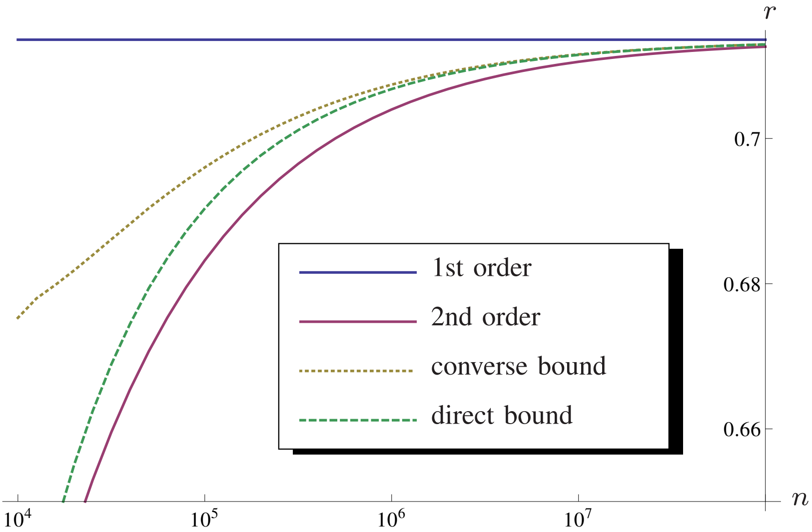
$$\rho_{XE} = \frac{1}{2}\sum |x\rangle\langle x| \otimes |\phi^x\rangle\langle\phi^x|, \qquad |\phi^x\rangle = \sqrt{p}|0\rangle + (-1)^x\sqrt{1-p}|1\rangle.$$





- Plot of first and second order asymptotic approximation of $\frac{1}{n}\ell^{\varepsilon}(X|E)$ and $\frac{1}{n}m^{\varepsilon}(X|B)$ for $p = 0.05$ and $\varepsilon = 10^{-6}$.

Example of Finite Block Length Bounds



Different Layers of Approximation

Class	Role	Quantities
Class 1	Optimal performance of protocol. Calculation is very difficult.	$m^{\varepsilon}(X\|B)_\rho$ $\ell^{\varepsilon}(X\|B)_\rho$,
Class 2	One-shot bound for general source. SDP tractable for small systems.	$H_h^{\varepsilon}(A\|B)_\rho$, $H_{\min}^{\varepsilon}(A\|B)_\rho$
Class 3	Quantum information spectrum.	$D_s^{\varepsilon}(\rho\\|\sigma)$
Class 4	Classical information spectrum. Approximately possible for i.i.d.	$D_s^{\varepsilon}(P_{0,\rho,\sigma}\\|P_{1,\rho,\sigma})$
Class 5	Second order asymptotics. Calculation is easy for large n.	$nH(X\|B)+$ $\sqrt{n}\,s(X\|B)\Phi^{-1}(\varepsilon)$

Classes	Difference	Method
$1 \to 2$	$O(\log n)$	Random coding and data-processing inequalities.
$2 \to 4$	$O(\log n)$	Relations between entropies.
$4 \to 5$	$O(1)$	Berry-Essèen Theorem.

Let $\varepsilon \in (0,1)$ and consider n i.i.d. repetitions of tasks for large n.

Step 1: One-Shot bounds I

Theorem (One-Shot Randomness Extraction)

Let ρ_{XB} be a CQ state and $0 < \eta \leq \varepsilon < 1$. Then,

$$H_{\min}^{\varepsilon-\eta}(X|B)_\rho - \log \frac{1}{\eta^4} - 3 \leq \ell^\varepsilon(X|B)_\rho \leq H_{\min}^{\varepsilon}(X|B)_\rho.$$

Theorem (One-Shot Data Compression)

Let ρ_{XB} be a CQ state and $0 < \eta \leq \varepsilon < 1$. Then,

$$H_h^{\varepsilon}(X|B)_\rho \leq m^\varepsilon(X|B)_\rho \leq H_h^{\varepsilon-\eta}(X|B)_\rho + \log \frac{\varepsilon}{\eta^2} + 3.$$

- Converse bounds keep ε intact.
- Achievability up to η, where η can be chosen arbitrarily small. The idea is to choose $\eta \sim 1/\sqrt{n}$ for the i.i.d. analysis.

Step 1: One-Shot bounds II

Theorem

Let ρ_{XB} be a CQ state and $0 < \eta \leq \varepsilon < 1$. Then,

$$H_{\min}^{\varepsilon-\eta}(X|B)_\rho - \log\frac{1}{\eta^4} - 3 \leq \ell^\varepsilon(X|B)_\rho \leq H_{\min}^{\varepsilon}(X|B)_\rho.$$

- In this sense, we have $\ell^\varepsilon(X|B)_\rho \approx H_{\min}^{\varepsilon}(X|B)_\rho$, i.e. the smooth min-entropy characterizes randomness extraction.
- The smooth min-entropy can be calculated efficiently (using an SDP) for small systems.
- The quantity $\ell^\varepsilon(X|B)_\rho$ has some natural properties, i.e.
  - For any function: $\ell^\varepsilon(f(X)|B)_\rho \leq \ell^\varepsilon(X|B)_\rho$.
  - For any quantum channel: $\ell^\varepsilon(X|\mathcal{E}(B))_\rho \geq \ell^\varepsilon(X|B)_\rho$.
- These properties are mimicked by the smooth min-entropy.
- The converse follows solely from the first of these monotonicity properties.
- Achievability uses two-universal hashing.

Step 1: One-Shot bounds III

Theorem

Let ρ_{XB} be a CQ state and $0 < \eta \leq \varepsilon < 1$. Then,

$$H_h^{\varepsilon}(X|B)_\rho \leq m^{\varepsilon}(X|B)_\rho \leq H_h^{\varepsilon-\eta}(X|B)_\rho + \log \frac{\varepsilon}{\eta^2} + 3.$$

- In this sense, we have $m^{\varepsilon}(X|B)_\rho \approx H_h^{\varepsilon}(X|B)_\rho$, i.e. source coding is characterized by a conditional hypothesis testing entropy.
- The hypothesis testing entropy can be calculated efficiently (using an SDP) for small systems.
- The quantity $m^{\varepsilon}(X|B)_\rho$ has some natural properties, i.e.
  - For any message: $m^{\varepsilon}(X|BM)_\rho \geq m^{\varepsilon}(X|B)_\rho - \log|M|$.
  - For any quantum channel: $m^{\varepsilon}(X|\mathcal{E}(B))_\rho \geq m^{\varepsilon}(X|B)_\rho$.
- These properties are mimicked by H_h.
- The converse follows solely from these properties.
- Achievability uses two-universal hashing (corresponds to random coding) and pretty good measurements.

Step 2: Relation to Information Spectrum I

- While the one-shot entropies can be computed for small systems, they are generally intractable for large systems.
- Hence, we want to approximate them further.

This is done as follows.

- We write everything in terms of relative entropies.
  $H_h^\varepsilon(A|B)_\rho = \max_\sigma -D_h^\varepsilon(\rho_{AB} \| 1_A \otimes \sigma_B)$ and
  $H_{\min}^\varepsilon(A|B)_\rho = \max_\sigma -D_{\max}^\varepsilon(\rho_{AB} \| 1_A \otimes \sigma_B)$.
- Use relations between relative entropies:

$$D_h^\varepsilon(\rho\|\sigma) \approx D_{\max}^{\sqrt{1-\varepsilon}}(\rho\|\sigma) \approx D_s^\varepsilon(\rho\|\sigma) \approx D_s^\varepsilon(P_0\|P_1)$$

- This holds up to terms $\log \Theta$, where Θ is at most the number of distinct eigenvalues of σ, or $2\lceil \log(\lambda_{\max}(\sigma)/\lambda_{\min}(\sigma)) \rceil$.
- In particular, $\Theta = O(n)$ in the i.i.d. case.

Step 2: Relation to Information Spectrum II

- We focus on the last quantity, the classical information spectrum.
- Decompose $\rho = \sum_x r_x |v_x\rangle\langle v_x|$ and $\sigma = \sum_y s_y |u_y\rangle\langle u_y|$.
- Nussbaum&Szkoła'09 introduced the distributions

$$P_0(x,y) := r_x |\langle v_x | u_y \rangle|^2 \quad \text{and} \quad P_1(x,y) := s_y |\langle v_x | u_y \rangle|^2$$

- They satisfy $D(P_0\|P_1) = D(\rho\|\sigma)$ and $V(P_0\|P_1) = V(\rho\|\sigma)$, i.e. the first two moments agree with the quantum analogue.
- This reduces the problem to analyzing the classical quantity

$$D_s^\varepsilon(P_0\|P_1) := \max\left\{ R \in \mathbb{R} \,\middle|\, \Pr_{P_0}\left[\log P_0 - \log P_1 \le R\right] \le \varepsilon \right\}.$$

- So far we have not used any i.i.d. assumption except that we assumed that the eigenvalues of σ behave nicely.

Step 3: Asymptotic Expansion

- Assume i.i.d. states $\rho_{XB}^{\otimes n}$. The corresponding distributions, $P_0^{\otimes n}$ and $P_1^{\otimes n}$, are i.i.d. as well.
- We write $P_0^{\otimes n}(x^n, y^n) = \prod_i P_0^{[i]}(x_i, y_i)$, $Z_i = \log P_0^{[i]} - \log P_1^{[i]}$, and

$$D_s^\varepsilon(P_0^{\otimes n} \| P_1^{\otimes n}) = \max\left\{ R \in \mathbb{R} \,\middle|\, \Pr_{P_0}\left[\log P_0^{\otimes n} - \log P_1^{\otimes n} \le R \right] \le \varepsilon \right\}$$
$$= n \cdot \max\left\{ R \in \mathbb{R} \,\middle|\, \Pr_{P_0}\left[\frac{1}{n} \sum_i Z_i \le R \right] \le \varepsilon \right\}$$

  where the Z_i are i.i.d. distributed.
- The central limit theorem states that the distribution of $\frac{1}{n}\sum_i Z_i$ converges to a Gaussian distribution with mean $\mathbb{E}[Z] = D(P_0\|P_1)$ and variance $\mathbb{E}\big[(Z - \mathbb{E}[Z])^2\big] = V(P_0\|P_1)^2$.
- This yields the expansion

$$D_s^\varepsilon(P_0^{\otimes n} \| P_1^{\otimes n}) = nD(P_0\|P_1) + \sqrt{n}V(P_0\|P_1)\Phi^{-1}(\varepsilon) + O(1).$$

Relation to Asymptotic Information Spectrum

- Our results also strengthens the relation of one-shot entropies to the sup/inf-information spectrum (Datta&Renner'09).
- These are generally defined as (Nagaoka&Hayashi'07)

$$\begin{aligned}\underline{D}(\varepsilon|\vec{\rho}\|\vec{\sigma}) &:= \sup\left\{R \in \mathbb{R} \,\middle|\, \limsup_{n\to\infty} \mathrm{tr}\rho_n\{\rho_n \le 2^{nR}\sigma_n\} \le \varepsilon\right\},\\ \overline{D}(\varepsilon|\vec{\rho}\|\vec{\sigma}) &:= \inf\left\{R \in \mathbb{R} \,\middle|\, \liminf_{n\to\infty} \mathrm{tr}\rho_n\{\rho_n \le 2^{nR}\sigma_n\} \ge \varepsilon\right\}\\ &= \sup\left\{R \in \mathbb{R} \,\middle|\, \liminf_{n\to\infty} \mathrm{tr}\rho_n\{\rho_n \le 2^{nR}\sigma_n\} < \varepsilon\right\}.\end{aligned}$$

- We find the following relations, which are valid for all $\varepsilon \in [0,1]$.

$$\begin{aligned}\underline{D}(\varepsilon|\vec{\rho}\|\vec{\sigma}) &= \sup_{\vec{\epsilon}}\left\{\liminf_{n\to\infty} \frac{1}{n} D_{\max}^{\sqrt{1-\epsilon_n}}(\rho_n\|\sigma_n) \,\middle|\, \limsup_{n\to\infty} \epsilon_n \le \varepsilon\right\},\\ \overline{D}(\varepsilon|\vec{\rho}\|\vec{\sigma}) &= \sup_{\vec{\epsilon}}\left\{\liminf_{n\to\infty} \frac{1}{n} D_{\max}^{\sqrt{1-\epsilon_n}}(\rho_n\|\sigma_n) \,\middle|\, \liminf_{n\to\infty} \epsilon_n < \varepsilon\right\},\end{aligned}$$

- The original relations by Datta&Renner only consider $\varepsilon \in \{0,1\}$.

Conclusion, Open Questions and Discussion

- We find the second order asymptotics for quantum hypothesis testing and give bounds on β^ε for finite n.
- We use one-shot entropies and techniques developed for hypothesis testing and the information spectrum method to find a second order expansion and finite block length bounds for operational quantities.
- There is a difference of $2\log n$ between the current upper and lower bounds on $D_h^\varepsilon(\rho\|\sigma)$. Is this fundamental, i.e. do there exist ρ and σ for which these bounds are tight? Or can this be further improved? Note that classically, the upper and lower bounds only differ in the constant term.
- The bounds depend on the parameter ε of the operational quantity, either ℓ^ε or m^ε. This implies, in particular, that the ε in the respective one-shot entropies has clear operational meaning.
- Provocative question to those of us using one-shot entropies in information theory: What does it mean to characterize a quantity? When is a result considered tight?

Thank you for your attention.