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 $\beta^\varepsilon$
2. Main Result: Asymptotic Expansion of $\beta^\varepsilon$ 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 $\rho$ or $\sigma$, 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 $\alpha_\rho(Q) = \text{tr}(\rho(1 - Q))$ and $\beta_\sigma(Q) = \text{tr}(\sigma Q)$, respectively.

- We are interested in the minimal $\beta$ that can be achieved if $\alpha$ is required to be smaller than a given constant $\varepsilon$, i.e. the SDP

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

- Alternatively, one may consider the divergence

$$D^\varepsilon_h(\rho\|\sigma) := -\log \left( \frac{\beta^\varepsilon_{\rho,\sigma}}{1 - \varepsilon} \right), \quad 0 < \varepsilon < 1.$$
Given $n$ copies of either $\sigma$ and $\rho$, a quantum generalization of Stein’s Lemma (Hiai&Petz’91) and its strong converse (Ogawa&Nagaoka’00) imply

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

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

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

$$D_h^\varepsilon(\rho^\otimes n \left\| \sigma^\otimes n) \geq nD(\rho \| \sigma) - O(\sqrt{n})$$

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 $\rho, \sigma$ with $\text{supp}\{\sigma\} \supseteq \text{supp}\{\rho\}$, and $0 < \varepsilon < 1$, we find†

\[
\begin{align*}
D_h^{\varepsilon}(\rho \otimes n \parallel \sigma \otimes n) &\leq nD(\rho \parallel \sigma) + \sqrt{nV(\rho \parallel \sigma)}\Phi^{-1}(\varepsilon) + 2 \log n + O(1), \quad \text{and} \\
D_h^{\varepsilon}(\rho \otimes n \parallel \sigma \otimes n) &\geq nD(\rho \parallel \sigma) + \sqrt{nV(\rho \parallel \sigma)}\Phi^{-1}(\varepsilon) - O(1).
\end{align*}
\]

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

\[
V(\rho \parallel \sigma) := \text{tr}\left(\rho \left( \log \rho - \log \sigma - D(\rho \parallel \sigma) \right)^2 \right).
\]

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

\[\begin{array}{c}
\phi \quad 0.0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1.0 \\
\text{V} \quad -3 \quad -2 \quad -1 \quad 1 \quad 2 \quad 3
\end{array}\]

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

Theorem

For two states $\rho, \sigma$ with $\text{supp}\{\sigma\} \supseteq \text{supp}\{\rho\}$, and $0 < \varepsilon < 1$, we find†

\[
D^{\varepsilon}_h(\rho^\otimes n \parallel \sigma^\otimes n) \leq n D(\rho \parallel \sigma) + \sqrt{n V(\rho \parallel \sigma)} \Phi^{-1}(\varepsilon) + 2 \log n + O(1), \quad \text{and}
\]

\[
D^{\varepsilon}_h(\rho^\otimes n \parallel \sigma^\otimes n) \geq n D(\rho \parallel \sigma) + \sqrt{n V(\rho \parallel \sigma)} \Phi^{-1}(\varepsilon) - O(1).
\]

- We also have bounds on the constant terms, enabling us to calculate upper and lower bounds on $D^{\varepsilon}_h(\rho^\otimes n \parallel \sigma^\otimes n)$ 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.
We also investigate the smooth min-entropy (Renner’05), where it was known (T,Colbeck&Renner’09,T’12) that, for $\varepsilon \in (0, 1)$,

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

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

We derive the following expansion

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

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

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
\[ \rho_{XE} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x^E. \]

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 \( \mathcal{P} : XS \rightarrow ZS \) extracts a random number \( Z \) from \( X \), producing a state \( \tau_{ZES} \) when applied to \( \rho_{XE} \otimes \rho_S \).

For any \( 0 \leq \epsilon < 1 \) and \( \rho_{XE} \) a CQ state, we define
\[ \ell^\epsilon(X|E) := \max \{ \ell \in \mathbb{N} \mid \exists \mathcal{P}, \sigma_E : |Z| = 2^\ell \wedge \tau_{ZES} \approx^\epsilon 2^{-\ell}1_Z \otimes \sigma_E \otimes \tau_S \}. \]

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 \( 0 < \epsilon < 1 \). Then,
\[ \ell^\epsilon(X^n|E^n) \lesssim nH(X|E) + \sqrt{nV(X|E)}\Phi^{-1}(\epsilon^2) \pm O(\log n). \]
Data Compression with Side Information

- Consider a CQ random source that outputs states
  \[ \rho_{XB} = \sum_x p_x |x⟩⟨x| \otimes \rho_B^x. \]
- 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 \(\rho_{XB}\) a CQ state, we define
  \[ m^\varepsilon(X|B)_\rho := \min \left\{ m \in \mathbb{N} \mid \exists \mathcal{P} : |M| = 2^m \land P[X \neq X'] \leq \varepsilon \right\}. \]
- 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,
\[ m^\varepsilon(X^n|B^n) \lesssim nH(X|B) - \sqrt{nV(X|B)\Phi^{-1}(\varepsilon)} \pm O(\log n). \]
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

\[
\rho_{XB} = \frac{1}{2} \sum |x\rangle\langle x| \otimes \left((1-p)|x\rangle\langle x| + p|1-x\rangle\langle 1-x|\right),
\]

\[
\rho_{XE} = \frac{1}{2} \sum |x\rangle\langle x| \otimes |\phi^x\rangle\langle \phi^x|,
\]

\( |\phi^x\rangle = \sqrt{p}|0\rangle + (-1)^x \sqrt{1-p}|1\rangle \).

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

<table>
<thead>
<tr>
<th>Class</th>
<th>Role</th>
<th>Quantities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Optimal performance of protocol. Calculation is very difficult.</td>
<td>$m^\varepsilon(X</td>
</tr>
<tr>
<td>Class 2</td>
<td>One-shot bound for general source. SDP tractable for small systems.</td>
<td>$H^\varepsilon_h(A</td>
</tr>
<tr>
<td>Class 3</td>
<td>Quantum information spectrum.</td>
<td>$D^\varepsilon_s(\rho|\sigma)$</td>
</tr>
<tr>
<td>Class 4</td>
<td>Classical information spectrum. Approximately possible for i.i.d.</td>
<td>$D^\varepsilon_s(P_0,\rho,\sigma|P_1,\rho,\sigma)$</td>
</tr>
<tr>
<td>Class 5</td>
<td>Second order asymptotics. Calculation is easy for large $n$.</td>
<td>$nH(X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classes</th>
<th>Difference</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 → 2</td>
<td>$O(\log n)$</td>
<td>Random coding and data-processing inequalities.</td>
</tr>
<tr>
<td>2 → 4</td>
<td>$O(\log n)$</td>
<td>Relations between entropies.</td>
</tr>
<tr>
<td>4 → 5</td>
<td>$O(1)$</td>
<td>Berry-Essèen Theorem.</td>
</tr>
</tbody>
</table>

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 $\rho_{XB}$ be a CQ state and $0 < \eta \leq \epsilon < 1$. Then,

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

Theorem (One-Shot Data Compression)
Let $\rho_{XB}$ be a CQ state and $0 < \eta \leq \epsilon < 1$. Then,

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

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

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

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

- In this sense, we have $m_\varepsilon(X|B)_\rho \approx H_\varepsilon^h(X|B)_\rho$, 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.
  \[ H^\varepsilon_{h}(A|B)_\rho = \max_\sigma -D^\varepsilon_{h}(\rho_{AB}||1_A \otimes \sigma_B) \]  and  
  \[ H^\varepsilon_{\min}(A|B)_\rho = \max_\sigma -D^\varepsilon_{\max}(\rho_{AB}||1_A \otimes \sigma_B). \]

- Use relations between relative entropies:
  \[ D^\varepsilon_{h}(\rho||\sigma) \approx D^{\sqrt{1-\varepsilon}}_{\max}(\rho||\sigma) \approx D^\varepsilon_{s}(\rho||\sigma) \approx D^\varepsilon_{s}(P_0||P_1) \]

- This holds up to terms log \( \Theta \), where \( \Theta \) is at most the number of distinct eigenvalues of \( \sigma \), or \( 2[\log(\lambda_{\max}(\sigma)/\lambda_{\min}(\sigma))] \).

- In particular, \( \Theta = O(n) \) in the i.i.d. case.
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

$$P_0(x, y) := r_x |\langle v_x | u_y \rangle|^2$$

and

$$P_1(x, y) := s_y |\langle v_x | u_y \rangle|^2$$

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

$$D_\varepsilon^c(P_0 \| P_1) := \max \left\{ R \in \mathbb{R} \mid \Pr_{P_0} \left[ \log P_0 - \log P_1 \leq R \right] \leq \varepsilon \right\}.$$

So far we have not used any i.i.d. assumption except that we assumed that the eigenvalues of $\sigma$ 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

$$D_s^\varepsilon(P_0^\otimes n \parallel P_1^\otimes n) = \max \left\{ R \in \mathbb{R} \mid \Pr_{P_0} \left[ \log P_0^\otimes n - \log P_1^\otimes n \leq R \right] \leq \varepsilon \right\}$$

$$= n \cdot \max \left\{ R \in \mathbb{R} \mid \Pr_{P_0} \left[ \frac{1}{n} \sum_i Z_i \leq R \right] \leq \varepsilon \right\}$$

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 $\mathbb{E}[Z] = D(P_0 \parallel P_1)$ and variance $\mathbb{E}[(Z - \mathbb{E}[Z])^2] = V(P_0 \parallel P_1)^2$.

- This yields the expansion

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

\[ D(\varepsilon|\rho\parallel\sigma) := \sup \{ R \in \mathbb{R} \mid \limsup_{n \to \infty} \text{tr}\rho_n\{\rho_n \leq 2^{nR}\sigma_n\} \leq \varepsilon \}, \]

\[ \overline{D}(\varepsilon|\rho\parallel\sigma) := \inf \{ R \in \mathbb{R} \mid \liminf_{n \to \infty} \text{tr}\rho_n\{\rho_n \leq 2^{nR}\sigma_n\} \geq \varepsilon \} \]

\[ = \sup \{ R \in \mathbb{R} \mid \liminf_{n \to \infty} \text{tr}\rho_n\{\rho_n \leq 2^{nR}\sigma_n\} < \varepsilon \}. \]

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

\[ D(\varepsilon|\rho\parallel\sigma) = \sup_{\varepsilon'} \left\{ \liminf_{n \to \infty} \frac{1}{n} D_{\max}^{\sqrt{1-\varepsilon_n}}(\rho_n\parallel\sigma_n) \mid \limsup_{n \to \infty} \varepsilon_n \leq \varepsilon \right\}, \]

\[ \overline{D}(\varepsilon|\rho\parallel\sigma) = \sup_{\varepsilon'} \left\{ \liminf_{n \to \infty} \frac{1}{n} D_{\max}^{\sqrt{1-\varepsilon_n}}(\rho_n\parallel\sigma_n) \mid \liminf_{n \to \infty} \varepsilon_n < \varepsilon \right\}, \]

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 $\beta^\varepsilon$ 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 $\rho$ and $\sigma$ 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 $\varepsilon$ of the operational quantity, either $\ell^\varepsilon$ or $m^\varepsilon$. This implies, in particular, that the $\varepsilon$ 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.