Moderate deviation analysis for c-q channels
(and hypothesis testing)

Joint work with Vincent Y.F. Tan (NUS)
and Marco Tomamichel (USyd/UTS)

Christopher T. Chubb
Centre for Engineered Quantum Systems
University of Sydney

Beyond IID
Singapore, July 2017

physics.usyd.edu.au/~cchubb/
We are going to consider coding of classical-quantum channels.

For c-q channel $\mathcal{W}$, a $(n, R, \epsilon)$-code is an encoder $E$ and decoding POVM $\{D_i\}$ such that

$$
\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \text{Tr} \left[ \mathcal{W} \otimes^n \left( \otimes_{i=1}^n E_i(m) \right) D_m \right] \geq 1 - \epsilon
$$

We will be concerned with the trade-off between the block-length $n$, the rate $R$, and the error probability $\epsilon$. We define the optimal rate/error probability as

$$
R^*(\mathcal{W}; n, \epsilon) := \max \{ R \mid \exists (n, R, \epsilon)-\text{code} \},
$$

$$
\epsilon^*(\mathcal{W}; n, R) := \min \{ \epsilon \mid \exists (n, R, \epsilon)-\text{code} \}.
$$
Asymptotics

For a constant error probability $\epsilon$, the Strong Converse Theorem tells us the rate approaches a constant known as the capacity

$$\lim_{n \to \infty} R^*(\mathcal{W}; n, \epsilon) = C(\mathcal{W}).$$

Equivalently this means that the error probability must go to 0 to 1 either side of the capacity

$$\lim_{n \to \infty} \epsilon^*(\mathcal{W}; n, R) = \begin{cases} 0 & : R < C(\mathcal{W}) \\ 1 & : R > C(\mathcal{W}) \end{cases}$$

This tells us we can have either $R \to C$ OR $\epsilon \to 0$.

How fast are these convergences? Can we do both?
Asymptotics

For a constant error probability $\epsilon$, the Strong Converse Theorem tells us the rate approaches a constant known as the capacity

$$\lim_{n \to \infty} R^*(\mathcal{W}; n, \epsilon) = C(\mathcal{W}).$$

Equivalently this means that the error probability must go 0 to 1 either side of the capacity

$$\lim_{n \to \infty} \epsilon^*(\mathcal{W}; n, R) = \begin{cases} 0 & : R < C(\mathcal{W}) \\ 1 & : R > C(\mathcal{W}) \end{cases}$$

This tells us we can have either $R \to C$ OR $\epsilon \to 0$.

How fast are these convergences? Can we do both?
Small and large deviations

How fast are the convergences $R \to C$ or $\epsilon \to 0$ as $n \to \infty$?

Small deviation (Tomamichel and Tan 2015)

$$R^*(n, \epsilon) = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \frac{1}{2})$$

Large deviation (Partial progress)

$$\ln \epsilon^*(n, R) = -n \cdot E(R) + o(n) \quad R < C$$
Small and large deviations

How fast are the convergences $R \to C$ or $\epsilon \to 0$ as $n \to \infty$?

**Small deviation** (Tomamichel and Tan 2015)

$$R^*(n, \epsilon) = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \frac{1}{2})$$

**Large deviation** (Partial progress)

$$\ln \epsilon^*(n, R) = -n \cdot E(R) + o(n) \quad R < C$$
Small and large deviations

How fast are the convergences $R \to C$ or $\epsilon \to 0$ as $n \to \infty$?

Small deviation (Tomamichel and Tan 2015)

$$R^*(n, \epsilon) = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon) + o \left( \frac{1}{\sqrt{n}} \right) \quad \epsilon \in (0, \frac{1}{2})$$

Large deviation (Partial progress)

$$\ln \epsilon^*(n, R) = -n \cdot E(R) + o(n) \quad R < C$$
Small and large deviations

How fast are the convergences $R \to C$ or $\epsilon \to 0$ as $n \to \infty$?

Small deviation (Tomamichel and Tan 2015)

$$R^*(n, \epsilon) = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \frac{1}{2})$$

Large deviation (Partial progress)

$$\ln \epsilon^*(n, R) = -n \cdot E(R) + o(n) \quad R < C$$
Small and large deviations

How fast are the convergences $R \to C$ or $\epsilon \to 0$ as $n \to \infty$?

Small deviation (Tomamichel and Tan 2015)

$$R^*(n, \epsilon) = C + \sqrt{\frac{V}{n}} \phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \frac{1}{2})$$

Large deviation (Partial progress)

$$\ln \epsilon^*(n, R) = -n \cdot E(R) + o(n) \quad R < C$$
Small and large deviations

How fast are the convergences $R \to C$ or $\epsilon \to 0$ as $n \to \infty$?

**Small deviation** (Tomamichel and Tan 2015)

$$R^*(n, \epsilon) = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \frac{1}{2})$$

**Large deviation** (Partial progress)

$$\ln \epsilon^*(n, R) = -n \cdot E(R) + o(n) \quad R < C$$
Small and large deviations

How fast are the convergences $R \to C$ or $\epsilon \to 0$ as $n \to \infty$?

Small deviation (Tomamichel and Tan 2015)

$$R^*(n, \epsilon) = C + \sqrt{\frac{V}{n}} \Phi^{-1}(\epsilon) + o \left( \frac{1}{\sqrt{n}} \right) \quad \epsilon \in (0, \frac{1}{2})$$

Large deviation (Partial progress)

$$\ln \epsilon^*(n, R) = -n \cdot E(R) + o(n) \quad R < C$$
What if we want $R \rightarrow C$ AND $\epsilon \rightarrow 0$?

**Moderate deviation** (This work, Cheng and Hsieh 2017)

For any $\{a_n\}$ such that $a_n \rightarrow 0$ and $\sqrt{n}a_n \rightarrow \infty$ we have

$$R^*(n, \epsilon_n) = C - \sqrt{2V}a_n + o(a_n) \quad \text{for} \quad \epsilon_n = e^{-na_n^2},$$

or equivalently

$$\ln \epsilon^*(n, R_n) = -\frac{na_n^2}{2V} + o(na_n^2) \quad \text{for} \quad R_n = C - a_n.$$
Moderate deviations

What if we want $R \rightarrow C \text{ AND } \epsilon \rightarrow 0$?

Moderate deviation  
(This work, Cheng and Hsieh 2017)

For any $\{a_n\}$ such that $a_n \rightarrow 0$ and $\sqrt{n}a_n \rightarrow \infty$ we have

$$R^*(n, \epsilon_n) = C - \sqrt{2V} a_n + o(a_n) \quad \text{for} \quad \epsilon_n = e^{-na_n^2},$$

or equivalently

$$\ln \epsilon^*(n, R_n) = -\frac{na_n^2}{2V} + o(na_n^2) \quad \text{for} \quad R_n = C - a_n.$$
Moderate deviations

What if we want $R \to C$ AND $\epsilon \to 0$?

For any $\{a_n\}$ such that $a_n \to 0$ and $\sqrt{n}a_n \to \infty$ we have

$$R^*(n, \epsilon_n) = C - \sqrt{2V}a_n + o(a_n) \quad \text{for} \quad \epsilon_n = e^{-na_n^2},$$

or equivalently

$$\ln \epsilon^*(n, R_n) = -\frac{na_n^2}{2V} + o(na_n^2) \quad \text{for} \quad R_n = C - a_n.$$
## Related work

<table>
<thead>
<tr>
<th></th>
<th>asymmetric binary hypothesis testing</th>
<th>channel coding</th>
<th>quantum hypothesis testing</th>
<th>classical-quantum channel coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>large dev. (&lt;)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>×</td>
</tr>
<tr>
<td>moderate dev. (&lt;)</td>
<td>✔</td>
<td>✔</td>
<td>[This talk$^2$, next talk$^3$]</td>
<td>[This talk$^2$, next talk$^3$]</td>
</tr>
<tr>
<td>small dev.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>moderate dev. (&gt;)</td>
<td>[This talk$^2$]</td>
<td>[This talk$^2$]</td>
<td>[This talk$^2$]</td>
<td>[This talk$^2$]</td>
</tr>
<tr>
<td>large dev. (&gt;)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

This talk$^2$ = Refined **small** deviation analysis  
Next talk$^3$ = Refined **large** deviation analysis

---

Concentration inequalities

Take \( \{X_i\} \) iid with \( \mathbb{E}[X_i] = 0 \) and \( \text{Var}[X_i] =: V \), and \( \bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \).

**Asymptotic (Law of large numbers)**

\[
\lim_{n \to \infty} \Pr \left[ \bar{X}_n \geq t \right] = \begin{cases} 
1 & t < 0, \\
0 & t > 0.
\end{cases}
\]

**Small deviation (Berry-Esseen)**

\[
\Pr \left[ \bar{X}_n \geq \frac{\epsilon}{\sqrt{n}} \right] = Q \left( \frac{\epsilon}{\sqrt{V}} \right) + O \left( \frac{1}{\sqrt{n}} \right) \quad \epsilon \in (0, 1)
\]

**Large deviation (Cramér)**

\[
\ln \Pr \left[ \bar{X}_n \geq t \right] = -n \cdot I(t) + o(n) \quad t \geq 0
\]

**Moderate deviation**

For any \( \{a_n\} \) such that \( a_n \to 0 \) and \( \sqrt{n}a_n \to \infty \)

\[
\ln \Pr \left[ \bar{X}_n \geq a_n \right] = -\frac{na_n^2}{2V} + o(na_n^2).
\]
Concentration inequalities

Take \( \{X_i\} \) iid with \( \mathbb{E}[X_i] = 0 \) and \( \text{Var}[X_i] =: V \), and \( \bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \).

### Asymptotic (Law of large numbers)

\[
\lim_{n \to \infty} \Pr[\bar{X}_n \geq t] = \begin{cases} 
1 & t < 0, \\
0 & t > 0.
\end{cases}
\]

### Small deviation (Berry-Esseen)

\[
\Pr[\bar{X}_n \geq \frac{\epsilon}{\sqrt{n}}] = Q\left(\frac{\epsilon}{\sqrt{V}}\right) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, 1)
\]

### Large deviation (Cramér)

\[
\ln \Pr[\bar{X}_n \geq t] = -n \cdot I(t) + o(n) \quad t \geq 0
\]

### Moderate deviation

For any \( \{a_n\} \) such that \( a_n \to 0 \) and \( \sqrt{n}a_n \to \infty \)

\[
\ln \Pr[\bar{X}_n \geq a_n] = -\frac{na_n^2}{2V} + o(na_n^2).
\]
Concentration inequalities

Take \( \{X_i\} \) iid with \( \mathbb{E}[X_i] = 0 \) and \( \text{Var}[X_i] =: V \), and \( \bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \).

Asymptotic (Law of large numbers)

\[
\lim_{n \to \infty} \Pr [\bar{X}_n \geq t] = \begin{cases} 
1 & t < 0, \\
0 & t > 0.
\end{cases}
\]

Small deviation (Berry-Esseen)

\[
\Pr [\bar{X}_n \geq \frac{\epsilon}{\sqrt{n}}] = Q \left( \frac{\epsilon}{\sqrt{V}} \right) + \mathcal{O} \left( \frac{1}{\sqrt{n}} \right) \quad \epsilon \in (0, 1)
\]

Large deviation (Cramér)

\[
\ln \Pr [\bar{X}_n \geq t] = -n \cdot I(t) + o(n) \quad t \geq 0
\]

Moderate deviation

For any \( \{a_n\} \) such that \( a_n \to 0 \) and \( \sqrt{n}a_n \to \infty \)

\[
\ln \Pr [\bar{X}_n \geq a_n] = -\frac{na_n^2}{2V} + o(na_n^2).
\]
Hypothesis testing

We want to test between two hypotheses, ρ and σ. For a binary POVM \( \{A, I - A\} \), we define the type-I and type-II errors as

\[
\alpha(A; \rho, \sigma) := \text{Tr}(I - A)\rho, \quad \beta(A; \rho, \sigma) := \text{Tr} A\sigma,
\]

and the \( \epsilon \)-hypothesis-testing divergence

\[
D_h^\epsilon(\rho \parallel \sigma) := -\log \min_{0 \leq A \leq I} \{\beta(A; \rho, \sigma) \mid \alpha(A; \rho, \sigma) \leq \epsilon\}.
\]

If we now consider testing between \( \rho^\otimes n \) and \( \sigma^\otimes n \), then the asymptotic behaviour is given by Quantum Stein’s Lemma.

Asymptotics (Hiai and Petz 1991, Ogawa and Nagaoka 1999)

For any \( \epsilon \in (0, 1) \)

\[
\lim_{n \to \infty} \frac{1}{n} D_h^\epsilon(\rho^\otimes n \parallel \sigma^\otimes n) = D(\rho \parallel \sigma).
\]
Hypothesis testing

We want to test between two hypotheses, $\rho$ and $\sigma$. For a binary POVM $\{A, I - A\}$, we define the type-I and type-II errors as

$$\alpha(A; \rho, \sigma) := \text{Tr}(I - A)\rho, \quad \beta(A; \rho, \sigma) := \text{Tr} A\sigma,$$

and the $\epsilon$-hypothesis-testing divergence

$$D_\ell^\epsilon(\rho\|\sigma) := -\log \min_{0 \leq A \leq I} \{\beta(A; \rho, \sigma) \mid \alpha(A; \rho, \sigma) \leq \epsilon\}.$$

If we now consider testing between $\rho^\otimes n$ and $\sigma^\otimes n$, then the asymptotic behaviour is given by Quantum Stein’s Lemma.

Asymptotics (Hiai and Petz 1991, Ogawa and Nagaoka 1999)

For any $\epsilon \in (0, 1)$

$$\lim_{n \to \infty} \frac{1}{n} D_\ell^\epsilon(\rho^\otimes n\|\sigma^\otimes n) = D(\rho\|\sigma).$$
Deviation results for hypothesis testing

**Small deviation** (Tomamichel and Hayashi 2013, Li 2014)
\[
\frac{1}{n} D_h^\epsilon (\rho^\otimes n \parallel \sigma^\otimes n) = D(\rho \parallel \sigma) + \sqrt{\frac{V(\rho \parallel \sigma)}{n}} \Phi^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right) \quad \text{for} \quad \epsilon \in (0, 1).
\]

**Large deviation** (Hayashi 2006, Nagaoka 2006)
\[
\ln \epsilon_n = -n \cdot E(R) + o(n) \quad \text{for} \quad \frac{1}{n} D_h^{\epsilon_n} (\rho^\otimes n \parallel \sigma^\otimes n) = R < D(\rho \parallel \sigma).
\]

**Moderate deviation** (This work, Cheng and Hsieh 2017)

For any \(\{a_n\}\) such that \(a_n \to 0\) and \(\sqrt{n}a_n \to \infty\) and \(\epsilon_n := e^{-na_n^2}\),
\[
\frac{1}{n} D_h^{\epsilon_n} (\rho^\otimes n \parallel \sigma^\otimes n) = D(\rho \parallel \sigma) - \sqrt{2V(\rho \parallel \sigma)a_n} + o(a_n).
\]
Deviation results for hypothesis testing

**Small deviation** (Tomamichel and Hayashi 2013, Li 2014)
\[
\frac{1}{n} D_h^\epsilon(\rho^\otimes n \parallel \sigma^\otimes n) = D(\rho \parallel \sigma) + \sqrt{\frac{V(\rho \parallel \sigma)}{n}} \Phi^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right) \quad \text{for} \quad \epsilon \in (0, 1).
\]

**Large deviation** (Hayashi 2006, Nagaoka 2006)
\[
\ln \epsilon_n = -n \cdot E(R) + o(n) \quad \text{for} \quad \frac{1}{n} D_h^{\epsilon_n}(\rho^\otimes n \parallel \sigma^\otimes n) = R < D(\rho \parallel \sigma).
\]

**Moderate deviation** (This work, Cheng and Hsieh 2017)
For any \(\{a_n\}\) such that \(a_n \to 0\) and \(\sqrt{n}a_n \to \infty\) and \(\epsilon_n := e^{-na_n^2}\),
\[
\frac{1}{n} D_h^{\epsilon_n}(\rho^\otimes n \parallel \sigma^\otimes n) = D(\rho \parallel \sigma) - \sqrt{2V(\rho \parallel \sigma)}a_n + o(a_n).
\]

C. T. Chubb

Moderate deviations
Reducing hyp. testing to concentration inequalities

To give a moderate deviation analysis of the HTD, we will use concentration bounds. First we see it is related to tail bounds of the Nussbaum-Szkoła distributions

\[ P^{\rho,\sigma}(a,b) := r_a |\langle \phi_a | \psi_b \rangle|^2 \quad \text{and} \quad Q^{\rho,\sigma}(a,b) := s_b |\langle \phi_a | \psi_b \rangle|^2, \]

where we have eigendecomposed our states \( \rho := \sum_a r_a |\phi_a \rangle \langle \phi_a | \) and \( \sigma := \sum_b s_b |\psi_b \rangle \langle \psi_b | \). These reproduce the first two moments of our states

\[ D(P^{\rho,\sigma} \parallel Q^{\rho,\sigma}) = D(\rho \parallel \sigma) \quad \text{and} \quad V(P^{\rho,\sigma} \parallel Q^{\rho,\sigma}) = V(\rho \parallel \sigma). \]

Specifically for iid \( Z_i = \log P^{\rho,\sigma}/Q^{\rho,\sigma} \) and \((a_i, b_i) \sim P^{\rho,\sigma}\), then

\[
\frac{1}{n} D^e_h(\rho^\otimes n \parallel \sigma^\otimes n) \geq \sup \left\{ R \mid \Pr \left[ \sum_{i=1}^{n} Z_i \leq \epsilon_n/2 \right] \right\} - O(\log 1/\epsilon_n),
\]

\[
\frac{1}{n} D^e_h(\rho^\otimes n \parallel \sigma^\otimes n) \leq \sup \left\{ R \mid \Pr \left[ \sum_{i=1}^{n} Z_i \leq 2\epsilon_n \right] \right\} + O(\log 1/\epsilon_n).
\]

\(^1\)Nussbaum and Szkoła 2009.

\(^2\)Tomamichel and Hayashi 2013
To give a moderate deviation analysis of the HTD, we will use concentration bounds. First we see it is related to tail bounds of the Nussbaum-Szkoła distributions\(^1\)

\[
P_{\rho,\sigma}(a, b) := r_a |\langle \phi_a | \psi_b \rangle|^2 \quad \text{and} \quad Q_{\rho,\sigma}(a, b) := s_b |\langle \phi_a | \psi_b \rangle|^2,
\]

where we have eigendecomposed our states \(\rho := \sum_a r_a |\phi_a\rangle\langle \phi_a| \) and \(\sigma := \sum_b s_b |\psi_b\rangle\langle \psi_b|\). These reproduce the first two moments of our states

\[
D(P_{\rho,\sigma} \| Q_{\rho,\sigma}) = D(\rho \| \sigma) \quad \text{and} \quad V(P_{\rho,\sigma} \| Q_{\rho,\sigma}) = V(\rho \| \sigma).
\]

Specifically for iid \(Z_i = \log \frac{P_{\rho,\sigma}}{Q_{\rho,\sigma}}\) and \((a_i, b_i) \sim P_{\rho,\sigma}\), then\(^2\)

\[
\frac{1}{n} D_{\epsilon_n}^{\rho_\sigma} (\rho^\otimes n \| \sigma^\otimes n) \geq \sup \left\{ R \left| \Pr \left[ \sum_{i=1}^n Z_i \leq \epsilon_n/2 \right] \right| - O(\log 1/\epsilon_n), \right\}
\]

\[
\frac{1}{n} D_{\epsilon_n}^{\rho_\sigma} (\rho^\otimes n \| \sigma^\otimes n) \leq \sup \left\{ R \left| \Pr \left[ \sum_{i=1}^n Z_i \leq 2\epsilon_n \right] \right| + O(\log 1/\epsilon_n). \right\}
\]

\(^1\)Nussbaum and Szkoła 2009.
\(^2\)Tomamichel and Hayashi 2013
Bounding the rate

For this we can use the one shot bounds

\[ R^*(1, \epsilon) \geq \sup_{P_X} D_{h}^{\epsilon/2} (\pi_{XY} \| \pi_X \otimes \pi_Y) - O(1), \]  
\[ (\text{Wang and Renner 2012}) \]

\[ R^*(1, \epsilon) \leq \inf_{\sigma} \sup_{\rho \in \text{Im}(\mathcal{W})} D_{h}^{2\epsilon} (\rho \| \sigma) + O(1), \]  
\[ (\text{Tomamichel and Tan 2015}) \]

where \( \pi_{XY} = \sum_x P_X(x) |x\rangle \langle x| \otimes \rho_Y^{(x)}. \)

This give \( n \)-shot bounds

\[ R^*(n, \epsilon_n) \geq \sup_{P_X^n} \frac{1}{n} D_{h}^{\epsilon_n/2} (\pi_{X^n Y^n} \| \pi_X^n \otimes \pi_Y^n) - O(1/n), \]

\[ R^*(n, \epsilon_n) \leq \inf_{\sigma^n} \sup_{\rho^n \in \text{Im}(\mathcal{W}^n)} \frac{1}{n} D_{h}^{2\epsilon_n} (\rho^n \| \sigma^n) + O(1/n). \]

We now want to show that a moderate deviation analysis of the rate follows from that of the hypothesis testing divergence.
Bounding the rate

For this we can use the one shot bounds

\[ R^*(1, \epsilon) \geq \sup_{P_X} D_{h}^{\epsilon/2} (\pi_{XY} \| \pi_X \otimes \pi_Y) - \mathcal{O}(1), \]  

(Wang and Renner 2012)

\[ R^*(1, \epsilon) \leq \inf_{\sigma} \sup_{\rho \in \text{Im}(\mathcal{W})} D_{h}^{2\epsilon} (\rho \| \sigma) + \mathcal{O}(1), \]  

(Tomamichel and Tan 2015)

where \( \pi_{XY} = \sum_x P_X(x) |x\rangle \langle x| \otimes \rho^{(x)}_Y \).

This give \( n \)-shot bounds

\[ R^*(n, \epsilon_n) \geq \sup_{P_{X^n}} \frac{1}{n} D_{h}^{\epsilon_n/2} (\pi_{X^n Y^n} \| \pi_{X^n} \otimes \pi_{Y^n}) - \mathcal{O}(1/n), \]

\[ R^*(n, \epsilon_n) \leq \inf_{\sigma^n} \sup_{\rho^n \in \text{Im}(\mathcal{W} \otimes^n)} \frac{1}{n} D_{h}^{2\epsilon_n} (\rho^n \| \sigma^n) + \mathcal{O}(1/n). \]

We now want to show that a moderate deviation analysis of the rate follows from that of the hypothesis testing divergence.
Bounding the rate

For this we can use the one shot bounds

\[ R^*(1, \epsilon) \geq \sup_{P_X} D^{\epsilon/2}_{h}(\pi_{XY} \parallel \pi_X \otimes \pi_Y) - O(1), \]  

(Wang and Renner 2012)

\[ R^*(1, \epsilon) \leq \inf_{\sigma} \sup_{\rho \in \text{Im}(W)} D^{2\epsilon}_{h}(\rho \parallel \sigma) + O(1), \]  

(Tomamichel and Tan 2015)

where \( \pi_{XY} = \sum_x P_X(x) |x\rangle\langle x|_X \otimes \rho_Y^{(x)}. \)

This give \( n \)-shot bounds

\[ R^*(n, \epsilon_n) \geq \sup_{P_{X^n}} \frac{1}{n} D^{\epsilon_n/2}_{h}(\pi_{X^n Y^n} \parallel \pi_{X^n} \otimes \pi_{Y^n}) - O(1/n), \]

\[ R^*(n, \epsilon_n) \leq \inf_{\sigma^n} \sup_{\rho^n \in \text{Im}(W^{\otimes n})} \frac{1}{n} D^{2\epsilon_n}_{h}(\rho^n \parallel \sigma^n) + O(1/n). \]

We now want to show that a moderate deviation analysis of the rate follows from that of the hypothesis testing divergence.
Achievability

In general we have

\[ R^*(n, \epsilon_n) \gtrsim \sup_{P_{X^n}} \frac{1}{n} D_h^{\epsilon_n/2} \left( \pi_{X^n Y^n} \parallel \pi_X \otimes \pi_Y \right) \]

where \( \pi_{X^n Y^n} = \sum_{\vec{x}} P_{X^n}(\vec{x}) \langle \vec{x} | X^n \otimes \rho_{Y^n}^{(\vec{x})} \rangle \).

If we assume iid input \( P_{X^n} = (P_X)^\otimes n \) then we can apply the moderate deviation result:

\[ R^*(n, \epsilon_n) \gtrsim \sup_{P_X} \frac{1}{n} D_h^{\epsilon_n/2} \left( \pi_{X^n} \parallel \pi_X \otimes \pi_Y \right)^\otimes n \]

\[ \gtrsim \sup_{P_X} D \left( \pi_{XY} \parallel \pi_X \otimes \pi_Y \right) - \sqrt{2 V \left( \pi_{XY} \parallel \pi_X \otimes \pi_Y \right)} a_n. \]

There exists a distribution \( P_X \) such that

\[ D \left( \pi_{XY} \parallel \pi_X \otimes \pi_Y \right) = C \]

and

\[ V \left( \pi_{XY} \parallel \pi_X \otimes \pi_Y \right) = V, \]

and so substituting this in gives

\[ R^*(n, \epsilon_n) \gtrsim C - \sqrt{2V a_n}. \]

\(^3\)Tomamichel and Tan 2015.
Achievability

In general we have

\[
R^*(n, \epsilon_n) \gtrsim \sup_{P_{X^n}} \frac{1}{n} D_{\epsilon_n/2}^{\epsilon_n/2} (\pi_{X^n Y^n} \| \pi_{X^n} \otimes \pi_{Y^n})
\]

where \(\pi_{X^n Y^n} = \sum_{\vec{x}} P_{X^n}(\vec{x}) |\vec{x}\rangle \langle \vec{x}|_{X^n} \otimes \rho_{Y^n}(\vec{x})\).

If we assume iid input \(P_{X^n} = (P_X)^{\otimes n}\) then we can apply the moderate deviation result:

\[
R^*(n, \epsilon_n) \gtrsim \sup_{P_X} \frac{1}{n} D_{\epsilon_n/2}^{\epsilon_n/2} (\pi_{X^n} \otimes \pi_{Y^n})^{\otimes n}
\]

\[
\gtrsim \sup_{P_X} D (\pi_{XY} \| \pi_X \otimes \pi_Y) - \sqrt{2} V (\pi_{XY} \| \pi_X \otimes \pi_Y) a_n.
\]

There exists\(^3\) a distribution \(P_X\) such that

\[
D (\pi_{XY} \| \pi_X \otimes \pi_Y) = C \quad \text{and} \quad V (\pi_{XY} \| \pi_X \otimes \pi_Y) = V,
\]

and so substituting this in gives

\[
R^*(n, \epsilon_n) \gtrsim C - \sqrt{2} V a_n.
\]

---

\(^3\)Tomamichel and Tan 2015.
Achievability

In general we have

\[ R^*(n, \epsilon_n) \gtrsim \sup_{P_X^n} \frac{1}{n} D^{\epsilon_n/2} \left( \pi_{X^n Y^n} \mid \pi_{X^n} \otimes \pi_{Y^n} \right) \]

where \( \pi_{X^n Y^n} = \sum_{\vec{x}} P_X^n(\vec{x}) |\vec{x}\rangle \langle \vec{x}|_{X^n} \otimes \rho_{Y^n}(\vec{x}) \).

If we assume iid input \( P_X^n = (P_X)^n \) then we can apply the moderate deviation result:

\[ R^*(n, \epsilon_n) \gtrsim \sup_{P_X} \frac{1}{n} D^{\epsilon_n/2} \left( \pi_{XY} \mid \pi_X \otimes \pi_Y \otimes^n \right) \]

\[ \gtrsim \sup_{P_X} D \left( \pi_{XY} \mid \pi_X \otimes \pi_Y \right) - \sqrt{2V \left( \pi_{XY} \parallel \pi_X \otimes \pi_Y \right)a_n} \].

There exists\(^3\) a distribution \( P_X \) such that

\[ D \left( \pi_{XY} \parallel \pi_X \otimes \pi_Y \right) = C \quad \text{and} \quad V \left( \pi_{XY} \parallel \pi_X \otimes \pi_Y \right) = V, \]

and so substituting this in gives

\[ R^*(n, \epsilon_n) \gtrsim C - \sqrt{2V} a_n. \]

---

\(^3\)Tomamichel and Tan 2015.
Optimality

We start with

\[ R^*(n, \epsilon_n) \lesssim \inf_{\sigma^n} \left( \sup_{\rho^n \in \text{Im}(W \otimes^n)} \frac{1}{n} D_h^{2\epsilon_n}(\rho^n \| \sigma^n) \right). \]

As \( W \) is c-q we have that \( \rho^n := \otimes_{i=1}^n \rho_i \), so

\[ R^*(n, \epsilon_n) \lesssim \inf_{\sigma^n} \left( \sup_{\{\rho_i\} \subset \text{Im}(W)} \frac{1}{n} D_h^{2\epsilon_n} \left( \bigotimes_{i=1}^n \rho_i \bigg\| \sigma^n \right) \right). \]

We need to find a choice of \( \sigma^n \) such that the above is appropriately bounded

\[ \frac{1}{n} D_h^{2\epsilon_n} \left( \bigotimes_{i=1}^n \rho_i \bigg\| \sigma^n \right) \lesssim C - \sqrt{2V} a_n \]

for any \( \{\rho_i\} \subset \text{Im}(W) \).
Optimality

We start with

$$R^*(n, \epsilon_n) \lesi \inf_{\sigma^n} \sup_{\rho^n \in \text{Im}(\mathcal{W}^\otimes n)} \frac{1}{n} D_h^{2\epsilon_n}(\rho^n \| \sigma^n).$$

As $\mathcal{W}$ is c-q we have that $\rho^n := \bigotimes_{i=1}^n \rho_i$, so

$$R^*(n, \epsilon_n) \lesi \inf_{\sigma^n} \sup_{\{\rho_i\} \subset \text{Im}(\mathcal{W})} \frac{1}{n} D_h^{2\epsilon_n} \left( \bigotimes_{i=1}^n \rho_i \biggm\| \sigma^n \right).$$

We need to find a choice of $\sigma^n$ such that the above is appropriately bounded

$$\frac{1}{n} D_h^{2\epsilon_n} \left( \bigotimes_{i=1}^n \rho_i \biggm\| \sigma^n \right) \lesi C - \sqrt{2V} a_n$$

for any $\{\rho_i\} \subset \text{Im}(\mathcal{W})$. 
Optimality

We start with

\[ R^*(n, \epsilon_n) \lesssim \inf_{\sigma^n} \sup_{\rho^n \in \text{Im}(\mathcal{W} \otimes n)} \frac{1}{n} D_{h}^{2\epsilon_n} (\rho^n \| \sigma^n). \]

As \( \mathcal{W} \) is c-q we have that \( \rho^n := \bigotimes_{i=1}^{n} \rho_i \), so

\[ R^*(n, \epsilon_n) \lesssim \inf_{\sigma^n} \sup_{\{\rho_i\} \subset \text{Im}(\mathcal{W})} \frac{1}{n} D_{h}^{2\epsilon_n} \left( \bigotimes_{i=1}^{n} \rho_i \bigg\| \sigma^n \right). \]

We need to find a choice of \( \sigma^n \) such that the above is appropriately bounded

\[ \frac{1}{n} D_{h}^{2\epsilon_n} \left( \bigotimes_{i=1}^{n} \rho_i \bigg\| \sigma^n \right) \lesssim C - \sqrt{2V} a_n \]

for any \( \{\rho_i\} \subset \text{Im}(\mathcal{W}). \)
To find a $\sigma^n$, we first need to split our sequences into ‘high’ and ‘low’ sequences

High: $\frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \bar{\rho}_n) > C - \eta$

Low: $\frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \bar{\rho}_n) \leq C - \eta$

where $\bar{\rho}_n := \frac{1}{n} \sum_{j=1}^{n} \rho_j$.

For the high sequences we will need a second-order (moderate deviations) bound, but for low first-order (Stein’s lemma) will suffice.
The average of a high sequence is close\(^4\) to the divergence centre \(\sigma^*\)

\[
\frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \rho_n) \approx C \quad \implies \quad \bar{\rho}_n \approx \sigma^* := \arg \min_{\sigma} \max_{\rho \in \text{Im}(\mathcal{W})} D(\rho \| \sigma)
\]

Moreover, the channel dispersion can be characterised as

\[
V(\mathcal{W}) = \inf_{\{\rho_i\} \subseteq \text{Im}(\mathcal{W})} \left\{ \frac{1}{n} \sum_{i=1}^{n} V(\rho_i \| \sigma^*) \left| \frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \bar{\rho}_n) = C \right. \right\}.
\]

If we let \(\sigma^n := (\sigma^*) \otimes^n\), then by continuity arguments

\[
\frac{1}{n} D_{h}^{2\epsilon_n} \left( \bigotimes_{i=1}^{n} \rho_i \left\| (\sigma^*) \otimes^n \right) \right) \lesssim \frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \sigma^*) - \sqrt{\frac{2}{n} \sum_{i=1}^{n} V(\rho_i \| \sigma^*) \alpha_n} \lesssim C - \sqrt{2 V \alpha_n}
\]

\(^4\)Tomamichel and Tan 2015
High sequences

The average of a high sequence is close\(^4\) to the divergence centre \(\sigma^*\)

\[
\frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \bar{\rho}_n) \approx C \quad \implies \quad \bar{\rho}_n \approx \sigma^* := \arg \min_{\sigma} \max_{\rho \in \text{Im}(\mathcal{W})} D(\rho \| \sigma)
\]

Moreover, the channel dispersion can be characterised as

\[
V(\mathcal{W}) = \inf_{\{\rho_i\} \subseteq \text{Im}(\mathcal{W})} \left\{ \frac{1}{n} \sum_{i=1}^{n} V(\rho_i \| \sigma^*) \left| \frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \bar{\rho}_n) = C \right. \right\}.
\]

If we let \(\sigma^n := (\sigma^*)^{\otimes n}\), then by continuity arguments

\[
\frac{1}{n} D^2_{h_{\epsilon_n}} \left( \bigotimes_{i=1}^{n} \rho_i \bigg\| (\sigma^*)^{\otimes n} \right) \lesssim \frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \sigma^*) - \sqrt{\frac{2}{n} \sum_{i=1}^{n} V(\rho_i \| \sigma^*) a_n} \lesssim C - \sqrt{2V a_n}
\]

\(^4\text{Tomamichel and Tan 2015}\)
High sequences

The average of a high sequence is close\(^4\) to the divergence centre \(\sigma^*\)

\[
\frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \bar{\rho}_n) \approx C \quad \implies \quad \bar{\rho}_n \approx \sigma^* := \arg \min_{\sigma} \max_{\rho \in \text{Im}(\mathcal{W})} D(\rho \| \sigma)
\]

Moreover, the channel dispersion can be characterised as

\[
V(\mathcal{W}) = \inf_{\{\rho_i\} \subseteq \text{Im}(\mathcal{W})} \left\{ \frac{1}{n} \sum_{i=1}^{n} V(\rho_i \| \sigma^*) \left| \frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \bar{\rho}_n) = C \right. \right\}.
\]

If we let \(\sigma^n := (\sigma^*)^\otimes n\), then by continuity arguments

\[
\frac{1}{n} D_{2\epsilon_n}^{\mathcal{W}} \left( \bigotimes_{i=1}^{n} \rho_i \quad (\sigma^*)^\otimes n \right) \lesssim \frac{1}{n} \sum_{i=1}^{n} D(\rho_i \| \sigma^*) - \sqrt{\frac{2}{n} \sum_{i=1}^{n} V(\rho_i \| \sigma^*) a_n} \lesssim C - \sqrt{2V a_n}
\]

\(^4\)Tomamichel and Tan 2015
Low sequences

For low sequences we have no control over the variance term.

Consider a covering $^5\mathcal{N}$ such that for every $\rho$ there exists a $\tau \in \mathcal{N}$ such that $D(\rho\|\tau) \leq \eta/2$. We now define our $\sigma^n$ as

$$\sigma^n = \frac{1}{|\mathcal{N}|} \sum_{\tau \in \mathcal{N}} \tau \otimes^n.$$

If we now let $\tau_n \in \mathcal{N}$ be the specific element of the covering which is closest to $\tilde{\rho}_n$, then we can use $D_h^c(\rho\|\mu\sigma + (1 - \mu)\sigma') \leq D_h^c(\rho\|\sigma) - \log \mu$ as well as (non-uniform) Stein’s lemma

$$\frac{1}{n} D_h^{2\epsilon n} \left( \bigotimes_{i=1}^n \rho_i \bigg|\bigg| \sigma^n \right) \leq \frac{1}{n} D_h^{2\epsilon n} \left( \bigotimes_{i=1}^n \rho_i \bigg|\bigg| \tau_n \otimes^n \right) + O(1/n)$$

$$\leq \frac{1}{n} \sum_{i=1}^n D(\rho_i\|\tau_n) + o(1)$$

$$= \frac{1}{n} \sum_{i=1}^n D(\rho_i\|\tilde{\rho}_n) + D(\tilde{\rho}_n\|\tau_n) + o(1)$$

$$\leq C - \eta/2 + o(1)$$

---

Low sequences

For low sequences we have no control over the variance term.

Consider a covering\(^5\) \(\mathcal{N}\) such that for every \(\rho\) there exists a \(\tau \in \mathcal{N}\) such that
\[
D(\rho\|\tau) \leq \eta/2.
\]
We now define our \(\sigma^n\) as
\[
\sigma^n = \frac{1}{|\mathcal{N}|} \sum_{\tau \in \mathcal{N}} \tau \otimes^n.
\]

If we now let \(\tau_n \in \mathcal{N}\) be the specific element of the covering which is closest to \(\bar{\rho}_n\), then we can use
\[
D^c_h(\rho\|\mu\sigma + (1 - \mu)\sigma') \leq D^c_h(\rho\|\sigma) - \log \mu
\]
as well as (non-uniform) Stein’s lemma
\[
\frac{1}{n} D^2_{\epsilon n} \left( \bigotimes_{i=1}^n \rho_i \biggm\| \sigma^n \right) \leq \frac{1}{n} D^2_{\epsilon n} \left( \bigotimes_{i=1}^n \rho_i \biggm\| \tau_n \otimes^n \right) + O(1/n)
\]
\[
\leq \frac{1}{n} \sum_{i=1}^n D(\rho_i\|\tau_n) + o(1)
\]
\[
= \frac{1}{n} \sum_{i=1}^n D(\rho_i\|\bar{\rho}_n) + D(\bar{\rho}_n\|\tau_n) + o(1)
\]
\[
\leq C - \eta/2 + o(1)
\]

Low sequences

For low sequences we have no control over the variance term.

Consider a covering\(^5\) \(\mathcal{N}\) such that for every \(\rho\) there exists a \(\tau \in \mathcal{N}\) such that \(D(\rho \| \tau) \leq \eta/2\). We now define our \(\sigma^n\) as

\[
\sigma^n = \frac{1}{|\mathcal{N}|} \sum_{\tau \in \mathcal{N}} \tau \otimes^n.
\]

If we now let \(\tau_n \in \mathcal{N}\) be the specific element of the covering which is closest to \(\bar{\rho}_n\), then we can use \(D_h^\epsilon(\rho \| \mu \sigma + (1 - \mu)\sigma') \leq D_h^\epsilon(\rho \| \sigma) - \log \mu\) as well as (non-uniform) Stein’s lemma

\[
\frac{1}{n} D_h^{2\epsilon n} \left( \bigotimes_{i=1}^n \rho_i \bigg\| \sigma^n \right) \leq \frac{1}{n} D_h^{2\epsilon n} \left( \bigotimes_{i=1}^n \rho_i \bigg\| \tau_n \otimes^n \right) + O(1/n)
\]

\[
\leq \frac{1}{n} \sum_{i=1}^n D(\rho_i \| \tau_n) + o(1)
\]

\[
= \frac{1}{n} \sum_{i=1}^n D(\rho_i \| \bar{\rho}_n) + D(\bar{\rho}_n \| \tau_n) + o(1)
\]

\[
\leq C - \eta/2 + o(1)
\]

Arbitrary sequences

We know that

\[ \frac{1}{n} D^2_{\epsilon n} \left( \bigotimes_{i=1}^{n} \rho_i \right) \left\| (\sigma^*)^\otimes n \right\| \leq C - \sqrt{2} V a_n + o(a_n), \]

High :

\[ \frac{1}{n} D^2_{\epsilon n} \left( \bigotimes_{i=1}^{n} \rho_i \right) \leq C - \sqrt{2} V a_n + o(a_n), \]

Low :

\[ \frac{1}{n} D^2_{\epsilon n} \left( \bigotimes_{i=1}^{n} \rho_i \right) \leq C - \eta/2 + o(1). \]

If we now take

\[ \sigma^n := \frac{1}{2} (\sigma^*)^\otimes n + \frac{1}{2} \frac{1}{\mathcal{N}} \sum_{\tau \in \mathcal{N}} \tau^\otimes n, \]

then

\[ \frac{1}{n} D^2_{\epsilon n} \left( \bigotimes_{i=1}^{n} \rho_i \right) \sigma^n \leq C - \sqrt{2} V a_n + o(a_n), \]

for arbitrary \{\rho_i\}. 
Arbitrary sequences

We know that

\[
\begin{align*}
\text{High :} & \quad \frac{1}{n} D_h^{2\epsilon_n} \left( \prod_{i=1}^n \rho_i \right) || (\sigma^*)^{\otimes n} \right) \leq C - \sqrt{2V} a_n + o(a_n), \\
\text{Low :} & \quad \frac{1}{n} D_h^{2\epsilon_n} \left( \prod_{i=1}^n \rho_i \right) \left( \frac{1}{N} \sum_{\tau \in \mathcal{N}} \tau^{\otimes n} \right) \leq C - \eta/2 + o(1).
\end{align*}
\]

If we now take

\[
\sigma^n := \frac{1}{2} (\sigma^*)^{\otimes n} + \frac{1}{2} \frac{1}{N} \sum_{\tau \in \mathcal{N}} \tau^{\otimes n},
\]

then

\[
\frac{1}{n} D_h^{2\epsilon_n} \left( \prod_{i=1}^n \rho_i \right) \sigma^n \right) \leq C - \sqrt{2V} a_n + o(a_n),
\]

for arbitrary \( \{\rho_i\} \).
Conclusion and further work

- We have given a moderate deviation analysis for the performance of c-q channels, and hypothesis testing of product states, specifically for $\epsilon_n := \exp(-na_n^2)$

$$R(\mathcal{W}; n, \epsilon_n) = C(\mathcal{W}) - \sqrt{2V(\mathcal{W})}a_n + o(a_n),$$

$$\frac{1}{n}D_{h}^{\epsilon_n}(\rho\|\sigma) = D(\rho\|\sigma) - \sqrt{2V(\rho\|\sigma)}a_n + o(a_n).$$

- Our proof covers the strong converse and $V = 0$ cases which had not been considered in the classical literature.
- This proof naturally extends to image-additive channels (separable encodings) and arbitrary input alphabets.

- Can we improve the $o(a_n)$ error terms? It seems they might actually be $O(a_n^2 + \log n)$.
- What about other channels (entanglement-breaking) or other capacities (quantum, entanglement-assisted)?