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

Joint work with Vincent Y.F. Tan (NUS)
and Marco Tomamichel (USyd/UTS)
arXiv:1701.03114 (to appear in CMP)

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

Beyond IID
Singapore, July 2017

physics.usyd.edu.au/~cchubb/

THE UNIVERSITY OF SYDNEY

eQus

Classical communication over a quantum channel

We are going to consider coding of classical-quantum channels.

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

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

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

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

Asymptotics

For a constant error probability ϵ, 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 to 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?

Asymptotics

For a constant error probability ϵ, 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 to go 0 to 1 either side of the capacity

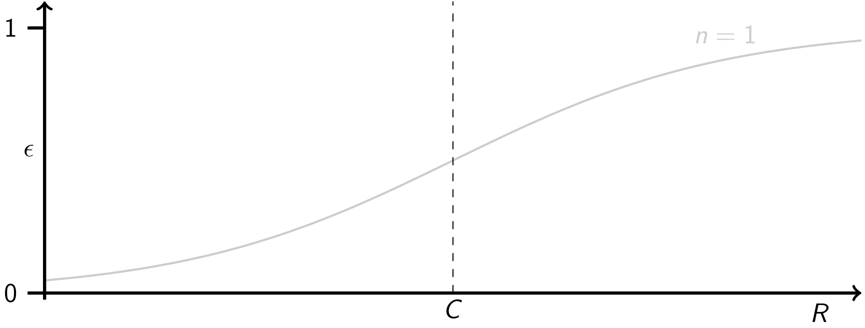
$$\lim_{n\to\infty} \epsilon^*(\mathcal{W}; n, R) = \begin{cases} 0 & : R < C(\mathcal{W}) \\ 1 & : R > C(\mathcal{W}) \end{cases}$$

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)

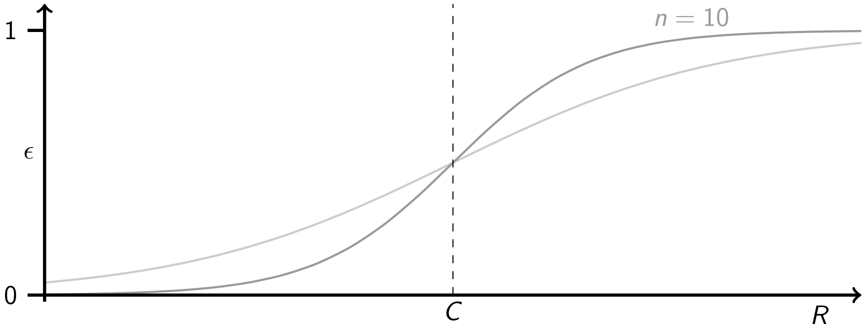
$$R^*(n,\epsilon) = C + \sqrt{\frac{V}{n}}\Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \tfrac{1}{2})$$

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)

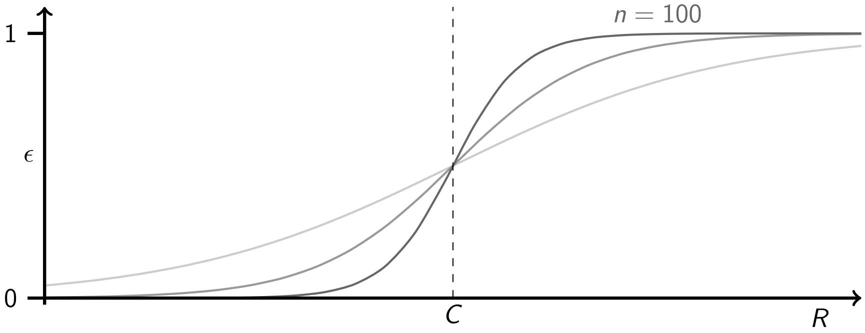
$$R^*(n,\epsilon) = C + \sqrt{\frac{V}{n}}\Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \tfrac{1}{2})$$

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)

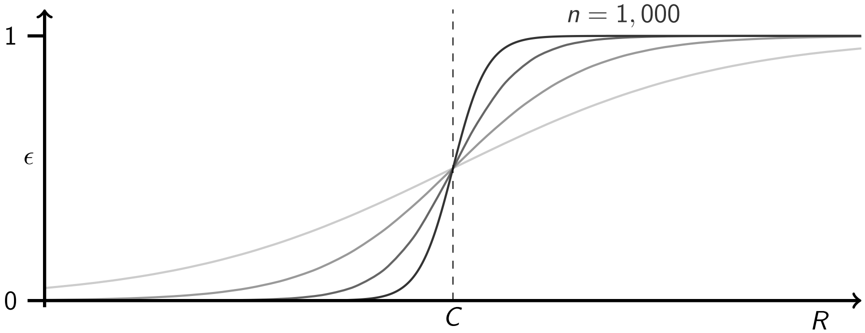
$$R^*(n,\epsilon) = C + \sqrt{\frac{V}{n}}\Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \tfrac{1}{2})$$

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)

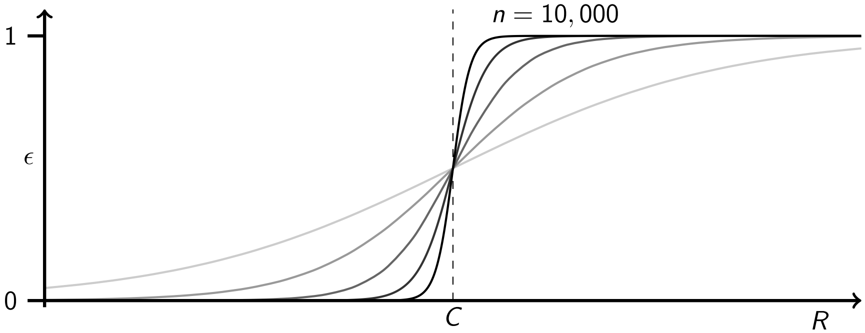
$$R^*(n,\epsilon) = C + \sqrt{\frac{V}{n}}\Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad \epsilon \in (0, \tfrac{1}{2})$$

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 $n \to \infty$?



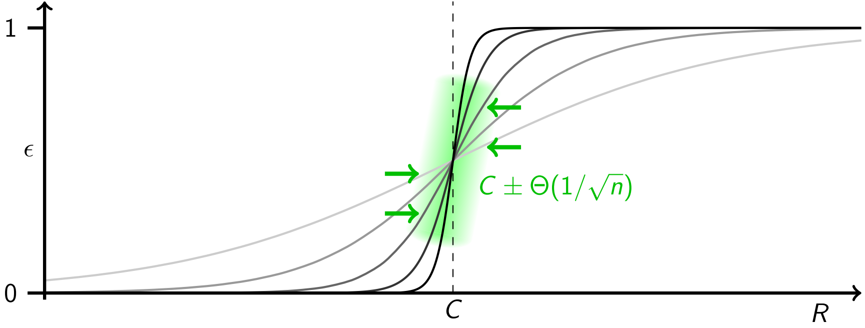
Small deviation (Tomamichel and Tan 2015)

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

Large deviation (Partial progress)

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

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



Small deviation (Tomamichel and Tan 2015)

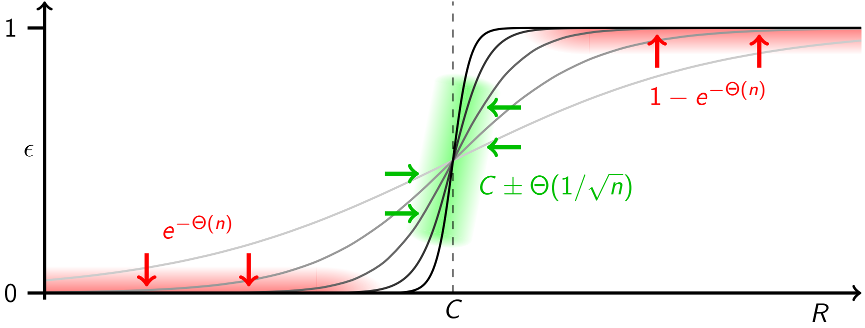
$$R^*(n,\epsilon) = C + \sqrt{\frac{V}{n}}\Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \qquad \epsilon \in (0, \tfrac{1}{2})$$

Large deviation (Partial progress)

$$\ln \epsilon^*(n,R) = -n \cdot E(R) + o(n) \qquad R < C$$

Small and large deviations

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



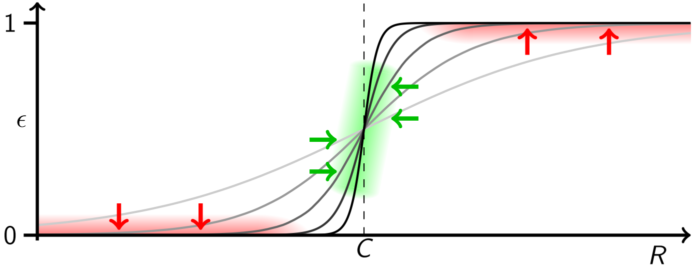
Small deviation (Tomamichel and Tan 2015)

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

Large deviation (Partial progress)

$$\ln \epsilon^*(n,R) = -n \cdot E(R) + o(n) \qquad R < C$$

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



Moderate deviation (This work, Cheng and Hsieh 2017)

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

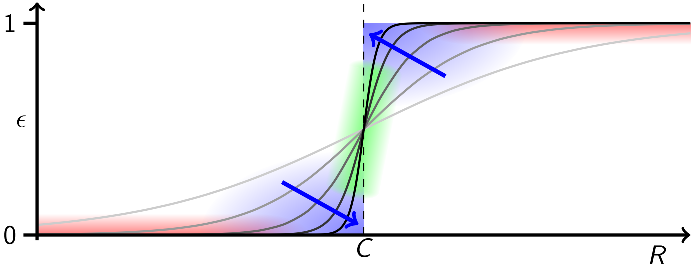
$$R^*(n,\epsilon_n) = C - \sqrt{2V}a_n + o(a_n) \quad \text{for} \quad \epsilon_n = e^{-na_n^2},$$

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 $\epsilon \to 0$?



Moderate deviation (This work, Cheng and Hsieh 2017)

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

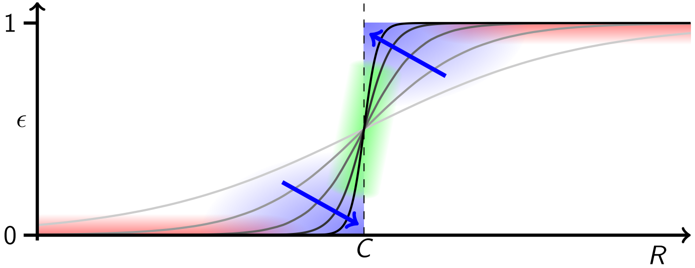
$$R^*(n, \epsilon_n) = C - \sqrt{2V}a_n + o(a_n) \quad \text{for} \quad \epsilon_n = e^{-na_n^2},$$

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 $\epsilon \to 0$?



Moderate deviation (This work, Cheng and Hsieh 2017)

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

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

or equivalently

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

Related work

	asymmetric binary hypothesis testing	channel coding	quantum hypothesis testing	classical-quantum channel coding
large dev. ($<$)	✓	✓	✓	×
moderate dev. ($<$)	✓	✓	[This talk[2], next talk[3]]	[This talk[2], next talk[3]]
small dev.	✓	✓	✓	✓
moderate dev. ($>$)	[This talk[2]]	[This talk[2]]	[This talk[2]]	[This talk[2]]
large dev. ($>$)	✓	✓	✓	✓

This talk[2] = Refined **small** deviation anlaysis

Next talk[3] = Refined **large** deviation anlaysis

[2] Chubb, Tan, and Tomamichel (arXiv:1701.03114).
[3] Cheng and Hsieh (arXiv:1701.03195).

Concentration inequalities

Take $\{X_i\}$ iid with $\mathbb{E}[X_i] = 0$ and $\mathrm{Var}[X_i] =: V$, 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)

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

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 $\mathrm{Var}[X_i] =: V$, 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)

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

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 $\mathrm{Var}[X_i] =: V$, 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)

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

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 $\sqrt{n}a_n \to \infty$

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

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

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

and the ϵ-hypothesis-testing divergence

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

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 $\epsilon \in (0, 1)$

$$\lim_{n\to\infty} \frac{1}{n} D_h^\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma).$$

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

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

and the ϵ-hypothesis-testing divergence

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

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 $\epsilon \in (0,1)$

$$\lim_{n\to\infty} \frac{1}{n} D_h^\epsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma).$$

Deviation results for hypothesis testing

Small deviation (Tomamichel and Hayashi 2013, Li 2014)

$$\frac{1}{n}D_h^{\epsilon}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}}\Phi^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for} \quad \epsilon \in (0,1).$$

Large deviation (Hayashi 2006, Nagaoka 2006)

$$\ln \epsilon_n = -n \cdot E(R) + o(n) \quad \text{for} \quad \frac{1}{n}D_h^{\epsilon_n}(\rho^{\otimes n}\|\sigma^{\otimes n}) = R < D(\rho\|\sigma).$$

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 $\epsilon_n := e^{-na_n^2}$,

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

Deviation results for hypothesis testing

Small deviation (Tomamichel and Hayashi 2013, Li 2014)

$$\frac{1}{n} D_h^{\epsilon}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}} \Phi^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{for} \quad \epsilon \in (0,1).$$

Large deviation (Hayashi 2006, Nagaoka 2006)

$$\ln \epsilon_n = -n \cdot E(R) + o(n) \quad \text{for} \quad \frac{1}{n} D_h^{\epsilon_n}(\rho^{\otimes n} \| \sigma^{\otimes n}) = R < D(\rho\|\sigma).$$

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 $\epsilon_n := e^{-n a_n^2}$,

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

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[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

$$D\left(P^{\rho,\sigma}\middle\|Q^{\rho,\sigma}\right) = D(\rho\|\sigma) \qquad \text{and} \qquad V\left(P^{\rho,\sigma}\middle\|Q^{\rho,\sigma}\right) = V(\rho\|\sigma).$$

Specifically for iid $Z_i = \log P^{\rho,\sigma}/Q^{\rho,\sigma}$ and $(a_i,b_i) \sim P^{\rho,\sigma}$, then[2]

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

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

[1] Nussbaum and Szkoła 2009.

[2] Tomamichel and Hayashi 2013

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[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

$$D\left(P^{\rho,\sigma} \| Q^{\rho,\sigma}\right) = D(\rho\|\sigma) \qquad \text{and} \qquad V\left(P^{\rho,\sigma} \| Q^{\rho,\sigma}\right) = V(\rho\|\sigma).$$

Specifically for iid $Z_i = \log P^{\rho,\sigma}/Q^{\rho,\sigma}$ and $(a_i, b_i) \sim P^{\rho,\sigma}$, then[2]

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

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

[1] Nussbaum and Szkoła 2009.
[2] Tomamichel and Hayashi 2013

Bounding the rate

For this we can use the one shot bounds

$$R^*(1,\epsilon) \geq \sup_{P_X} D_h^{\epsilon/2}(\pi_{XY} \| \pi_X \otimes \pi_Y) - \mathcal{O}(1), \qquad \text{(Wang and Renner 2012)}$$

$$R^*(1,\epsilon) \leq \inf_{\sigma} \sup_{\rho \in \mathrm{Im}(\mathcal{W})} D_h^{2\epsilon}(\rho \| \sigma) + \mathcal{O}(1), \qquad \text{(Tomamichel and Tan 2015)}$$

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

This give n-shot bounds

$$R^*(n,\epsilon_n) \geq \sup_{P_{X^n}} \frac{1}{n} D_h^{\epsilon_n/2}(\pi_{X^nY^n} \| \pi_{X^n} \otimes \pi_{Y^n}) - \mathcal{O}(1/n),$$

$$R^*(n,\epsilon_n) \leq \inf_{\sigma^n} \sup_{\rho^n \in \mathrm{Im}(\mathcal{W}^{\otimes n})} \frac{1}{n} D_h^{2\epsilon_n}(\rho^n \| \sigma^n) + \mathcal{O}(1/n).$$

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

$$R^*(1,\epsilon) \geq \sup_{P_X} D_h^{\epsilon/2}(\pi_{XY} \| \pi_X \otimes \pi_Y) - \mathcal{O}(1), \qquad \text{(Wang and Renner 2012)}$$

$$R^*(1,\epsilon) \leq \inf_{\sigma} \sup_{\rho \in \mathrm{Im}(\mathcal{W})} D_h^{2\epsilon}(\rho \| \sigma) + \mathcal{O}(1), \qquad \text{(Tomamichel and Tan 2015)}$$

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

This give n-shot bounds

$$R^*(n,\epsilon_n) \geq \sup_{P_{X^n}} \frac{1}{n} D_h^{\epsilon_n/2}(\pi_{X^nY^n} \| \pi_{X^n} \otimes \pi_{Y^n}) - \mathcal{O}(1/n),$$

$$R^*(n,\epsilon_n) \leq \inf_{\sigma^n} \sup_{\rho^n \in \mathrm{Im}(\mathcal{W}^{\otimes n})} \frac{1}{n} D_h^{2\epsilon_n}(\rho^n \| \sigma^n) + \mathcal{O}(1/n).$$

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

$$R^*(1,\epsilon) \geq \sup_{P_X} D_h^{\epsilon/2}(\pi_{XY} \| \pi_X \otimes \pi_Y) - \mathcal{O}(1), \qquad \text{(Wang and Renner 2012)}$$

$$R^*(1,\epsilon) \leq \inf_{\sigma} \sup_{\rho \in \mathrm{Im}(\mathcal{W})} D_h^{2\epsilon}(\rho \| \sigma) + \mathcal{O}(1), \qquad \text{(Tomamichel and Tan 2015)}$$

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

This give n-shot bounds

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

$$R^*(n,\epsilon_n) \leq \inf_{\sigma^n} \sup_{\rho^n \in \mathrm{Im}(\mathcal{W}^{\otimes n})} \frac{1}{n} D_h^{2\epsilon_n}(\rho^n \| \sigma^n) + \mathcal{O}(1/n).$$

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

Achievability

In general we have

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

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 $P_{X^n} = (P_X)^{\times n}$ then we can apply the moderate deviation result:

$$\begin{aligned} R^*(n, \epsilon_n) &\gtrsim \sup_{P_X} \frac{1}{n} D_h^{\epsilon_n/2} \left(\pi_{XY}^{\otimes n} \| (\pi_X \otimes \pi_Y)^{\otimes n}\right) \\ &\gtrsim \sup_{P_X} D\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) - \sqrt{2V\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right)} a_n. \end{aligned}$$

There exists[3] a distribution P_X such that

$$D\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) = C \qquad \text{and} \qquad V\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) = V,$$

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

$$R^*(n,\epsilon_n) \gtrsim \sup_{P_{X^n}} \frac{1}{n} D_h^{\epsilon_n/2}(\pi_{X^nY^n} \| \pi_{X^n} \otimes \pi_{Y^n})$$

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

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

$$\begin{aligned} R^*(n,\epsilon_n) &\gtrsim \sup_{P_X} \frac{1}{n} D_h^{\epsilon_n/2}\left(\pi_{XY}^{\otimes n} \middle\| (\pi_X \otimes \pi_Y)^{\otimes n}\right) \\ &\gtrsim \sup_{P_X} D\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) - \sqrt{2V\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right)}\, a_n. \end{aligned}$$

There exists[3] a distribution P_X such that

$$D\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) = C \qquad \text{and} \qquad V\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) = V,$$

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

$$R^*(n,\epsilon_n) \gtrsim \sup_{P_{X^n}} \frac{1}{n} D_h^{\epsilon_n/2}(\pi_{X^nY^n} \| \pi_{X^n} \otimes \pi_{Y^n})$$

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

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

$$\begin{aligned} R^*(n,\epsilon_n) &\gtrsim \sup_{P_X} \frac{1}{n} D_h^{\epsilon_n/2}\left(\pi_{XY}^{\otimes n} \| (\pi_X \otimes \pi_Y)^{\otimes n}\right) \\ &\gtrsim \sup_{P_X} D\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) - \sqrt{2V\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right)}\, a_n. \end{aligned}$$

There exists[3] a distribution P_X such that

$$D\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) = C \qquad \text{and} \qquad V\left(\pi_{XY} \| \pi_X \otimes \pi_Y\right) = V,$$

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

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

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

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

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

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

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

Optimality

We start with

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

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

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

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

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

for <u>any</u> $\{\rho_i\} \subset \mathrm{Im}(\mathcal{W})$.

Optimality

We start with

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

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

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

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

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

for <u>any</u> $\{\rho_i\} \subset \mathrm{Im}(\mathcal{W})$.

‘High’ and ‘low’ sequences

To find a σ^n, we first need to split our sequences into ‘high’ and ‘low’ sequences

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

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

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.

High sequences

The average of a high sequence is close[4] to the divergence centre σ^*

$$\frac{1}{n}\sum_{i=1}^{n} D\left(\rho_i \| \bar{\rho}_n\right) \approx C \qquad \Longrightarrow \qquad \bar{\rho}_n \approx \sigma^* := \arg\min_{\sigma} \max_{\rho\in\mathrm{Im}(\mathcal{W})} D(\rho\|\sigma)$$

Moreover, the channel dispersion can be characterised as

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

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

$$\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^{n} \rho_i \middle\| (\sigma^*)^{\otimes n} \right) \lesssim \frac{1}{n}\sum_{i=1}^{n} D\left(\rho_i\|\sigma^*\right) - \sqrt{\frac{2}{n}\sum_{i=1}^{n} V(\rho_i\|\sigma^*)}a_n \lesssim C - \sqrt{2V}a_n$$

[4] Tomamichel and Tan 2015

High sequences

The average of a high sequence is close[4] to the divergence centre σ^*

$$\frac{1}{n}\sum_{i=1}^{n} D\left(\rho_i \| \bar{\rho}_n\right) \approx C \qquad \implies \qquad \bar{\rho}_n \approx \sigma^* := \arg\min_{\sigma} \max_{\rho\in\mathrm{Im}(\mathcal{W})} D(\rho\|\sigma)$$

Moreover, the channel dispersion can be characterised as

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

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

$$\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^{n} \rho_i \middle\| (\sigma^*)^{\otimes n}\right) \lesssim \frac{1}{n}\sum_{i=1}^{n} D\left(\rho_i\|\sigma^*\right) - \sqrt{\frac{2}{n}\sum_{i=1}^{n} V(\rho_i\|\sigma^*)}a_n \lesssim C - \sqrt{2V}a_n$$

[4] Tomamichel and Tan 2015

High sequences

The average of a high sequence is close[4] to the divergence centre σ^*

$$\frac{1}{n}\sum_{i=1}^{n} D\left(\rho_i \| \bar{\rho}_n\right) \approx C \qquad \Longrightarrow \qquad \bar{\rho}_n \approx \sigma^* := \arg\min_{\sigma} \max_{\rho\in\mathrm{Im}(\mathcal{W})} D(\rho\|\sigma)$$

Moreover, the channel dispersion can be characterised as

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

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

$$\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^{n}\rho_i \middle\| (\sigma^*)^{\otimes n}\right) \lesssim \frac{1}{n}\sum_{i=1}^{n} D\left(\rho_i\|\sigma^*\right) - \sqrt{\frac{2}{n}\sum_{i=1}^{n} V(\rho_i\|\sigma^*)}a_n \lesssim C - \sqrt{2V}a_n$$

[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 ρ there exists a $\tau \in \mathcal{N}$ such that $D(\rho\|\tau) \leq \eta/2$. We now define our σ^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

$$\begin{aligned}
\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \sigma^n\right) &\leq \frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \tau_n^{\otimes n}\right) + \mathcal{O}(1/n) \\
&\leq \frac{1}{n}\sum_{i=1}^n D(\rho_i\|\tau_n) + o(1) \\
&= \frac{1}{n}\sum_{i=1}^n D(\rho_i\|\bar{\rho}_n) + D(\bar{\rho}_n\|\tau_n) + o(1) \\
&\leq C - \eta/2 + o(1)
\end{aligned}$$

[5] Tomamichel and Tan 2015, Hayden, Leung, Shor, and Winter 2004.

Low sequences

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

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

$$\begin{aligned}
\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \sigma^n\right) &\leq \frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \tau_n^{\otimes n}\right) + \mathcal{O}(1/n) \\
&\leq \frac{1}{n}\sum_{i=1}^n D(\rho_i\|\tau_n) + o(1) \\
&= \frac{1}{n}\sum_{i=1}^n D(\rho_i\|\bar{\rho}_n) + D(\bar{\rho}_n\|\tau_n) + o(1) \\
&\leq C - \eta/2 + o(1)
\end{aligned}$$

[5] Tomamichel and Tan 2015, Hayden, Leung, Shor, and Winter 2004.

Low sequences

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

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

$$\begin{aligned}
\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \sigma^n\right) &\leq \frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \tau_n^{\otimes n}\right) + \mathcal{O}(1/n) \\
&\leq \frac{1}{n}\sum_{i=1}^n D(\rho_i\|\tau_n) + o(1) \\
&= \frac{1}{n}\sum_{i=1}^n D(\rho_i\|\bar{\rho}_n) + D(\bar{\rho}_n\|\tau_n) + o(1) \\
&\leq C - \eta/2 + o(1)
\end{aligned}$$

[5] Tomamichel and Tan 2015, Hayden, Leung, Shor, and Winter 2004.

Arbitrary sequences

We know that

$$\text{High}: \quad \frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| (\sigma^*)^{\otimes n}\right) \leq C - \sqrt{2V} a_n + o(a_n),$$

$$\text{Low}: \quad \frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \frac{1}{\mathcal{N}} \sum_{\tau \in \mathcal{N}} \tau^{\otimes n}\right) \leq C - \eta/2 + o(1).$$

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

$$\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \sigma^n\right) \leq C - \sqrt{2V} a_n + o(a_n),$$

for arbitrary $\{\rho_i\}$.

Arbitrary sequences

We know that

$$\text{High}: \quad \frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| (\sigma^*)^{\otimes n}\right) \leq C - \sqrt{2V} a_n + o(a_n),$$

$$\text{Low}: \quad \frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \frac{1}{\mathcal{N}} \sum_{\tau \in \mathcal{N}} \tau^{\otimes n}\right) \leq C - \eta/2 + o(1).$$

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

$$\frac{1}{n} D_h^{2\epsilon_n}\left(\bigotimes_{i=1}^n \rho_i \middle\| \sigma^n\right) \leq C - \sqrt{2V} a_n + o(a_n),$$

for arbitrary $\{\rho_i\}$.

Conclusion and further work

- We have give 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 $\mathcal{O}(a_n^2 + \log n)$.
- What about other channels (entanglement-breaking) or other capacities (quantum, entanglement-assisted)?

ArXiv: 1701.03114 (to appear in Comm. Math. Phys.)