On Third-Order Asymptotics for DMCs

Vincent Y. F. Tan

Institute for Infocomm Research (I^{2}R)
National University of Singapore (NUS)

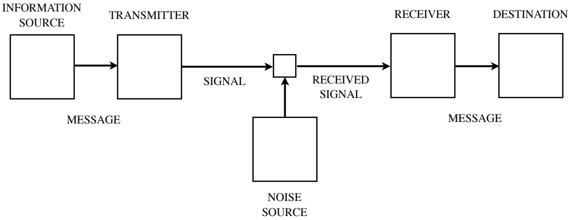
January 20, 2013

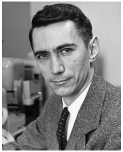
Acknowledgements

This is joint work with Marco Tomamichel

Centre for Quantum Technologies
National University of Singapore

Transmission of Information

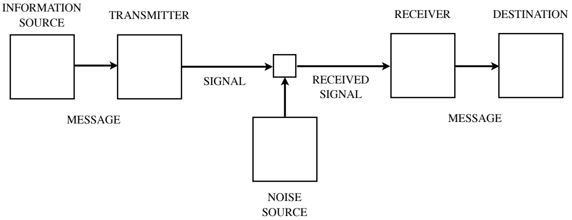


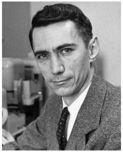


Shannon's Figure 1

- Information theory $\equiv$ Finding fundamental limits for reliable information transmission

Transmission of Information

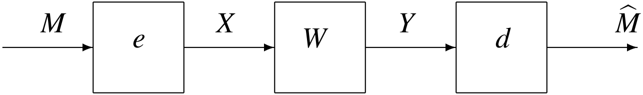




Shannon's Figure 1

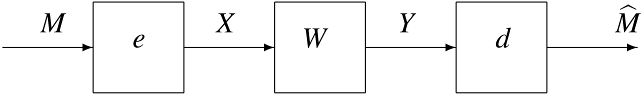
- Information theory $\equiv$ Finding fundamental limits for reliable information transmission
- Channel coding: Concerned with the maximum rate of communication in bits/channel use

Channel Coding (One-Shot)



- A code is an triple $\mathcal{C} = \{\mathcal{M}, e, d\}$ where $\mathcal{M}$ is the message set

Channel Coding (One-Shot)

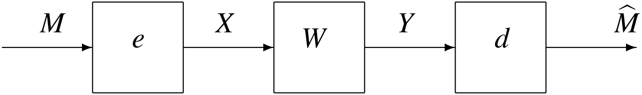


- A code is an triple $\mathcal{C} = \{\mathcal{M}, e, d\}$ where $\mathcal{M}$ is the message set
- The average error probability $p_{\mathrm{err}}(\mathcal{C})$ is

$$p_{\mathrm{err}}(\mathcal{C}) := \Pr\left[\widehat{M} \neq M\right]$$

where M is uniform on $\mathcal{M}$

Channel Coding (One-Shot)



- A code is an triple $\mathcal{C} = \{\mathcal{M}, e, d\}$ where $\mathcal{M}$ is the message set
- The average error probability $p_{\mathrm{err}}(\mathcal{C})$ is

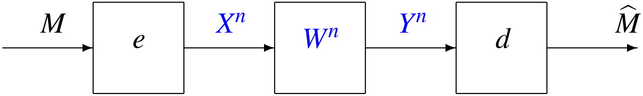
$$p_{\mathrm{err}}(\mathcal{C}) \;:=\; \Pr\left[\widehat{M} \neq M\right]$$

  where M is uniform on $\mathcal{M}$

- ε-Error Capacity is

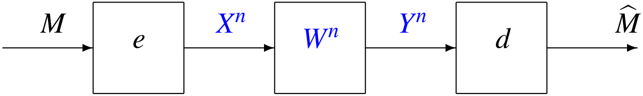
$$M^*(W,\varepsilon) := \sup\left\{m \in \mathbb{N} \,\middle|\, \exists\, \mathcal{C} \text{ s.t. } m = |\mathcal{M}|,\, p_{\mathrm{err}}(\mathcal{C}) \le \varepsilon\right\}$$

Channel Coding (n-Shot)



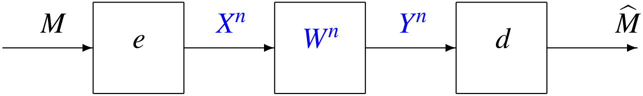
- Consider n independent uses of a channel

Channel Coding (n-Shot)



- Consider n independent uses of a channel
- Assume W is a discrete memoryless channel

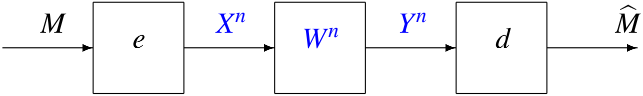
Channel Coding (n-Shot)



- Consider n independent uses of a channel
- Assume W is a discrete memoryless channel
- For vectors $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} := (y_1, \ldots, y_n) \in \mathcal{Y}^n$,

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(y_i|x_i)$$

Channel Coding (n-Shot)



- Consider n independent uses of a channel
- Assume W is a discrete memoryless channel
- For vectors $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} := (y_1, \ldots, y_n) \in \mathcal{Y}^n$,

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} W(y_i|x_i)$$

- Blocklength n, ε-Error Capacity is

$$M^*(W^n, \varepsilon)$$

Main Contribution

- Upper bound $\log M^*(W^n, \varepsilon)$ for n large (converse)

Main Contribution

- Upper bound $\log M^*(W^n, \varepsilon)$ for n large (converse)
- Concerned with the third-order term of the asymptotic expansion

Main Contribution

- Upper bound $\log M^*(W^n, \varepsilon)$ for n large (converse)
- Concerned with the third-order term of the asymptotic expansion
- Going beyond the normal approximation terms

Main Contribution

- Upper bound $\log M^*(W^n, \varepsilon)$ for n large (converse)
- Concerned with the third-order term of the asymptotic expansion
- Going beyond the normal approximation terms

Theorem (Tomamichel-Tan (2013))

For all DMCs with positive ε-dispersion V_ε,

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

where $\mathrm{Q}(a) := \int_a^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)\,\mathrm{d}x$

Main Contribution

- Upper bound $\log M^*(W^n, \varepsilon)$ for n large (converse)
- Concerned with the third-order term of the asymptotic expansion
- Going beyond the normal approximation terms

Theorem (Tomamichel-Tan (2013))

For all DMCs with positive ε-dispersion V_ε,

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\,\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

where $\mathrm{Q}(a) := \int_a^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)\,\mathrm{d}x$

- The $\frac{1}{2}\log n$ term is our main contribution

Main Contribution: Remarks

- Our bound

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

Main Contribution: Remarks

- Our bound

$$\log M^*(W^n,\varepsilon) \le nC - \sqrt{nV_\varepsilon}\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- Best upper bound till date:

$$\log M^*(W^n,\varepsilon) \le nC - \sqrt{nV_\varepsilon}\mathrm{Q}^{-1}(\varepsilon) + \left(|\mathcal{X}| - \frac{1}{2}\right)\log n + O(1)$$

V. Strassen (1964)

Polyanskiy-Poor-Verdú or PPV (2010)

Main Contribution: Remarks

- Our bound

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\,\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- Best upper bound till date:

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\,\mathrm{Q}^{-1}(\varepsilon) + \left(|\mathcal{X}| - \frac{1}{2}\right)\log n + O(1)$$

V. Strassen (1964)

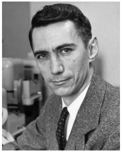
Polyanskiy-Poor-Verdú or PPV (2010)

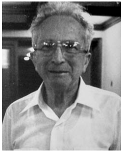
- Requires new converse techniques

Outline

Background: Shannon's Channel Coding Theorem

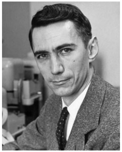
- Shannon's noisy channel coding theorem and
- Wolfowitz's strong converse state that

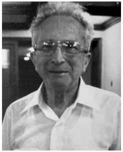




Background: Shannon's Channel Coding Theorem

- Shannon's noisy channel coding theorem and
- Wolfowitz's strong converse state that





Theorem (Shannon (1949), Wolfowitz (1959))

$$\lim_{n\to\infty} \frac{1}{n} \log M^*(W^n, \varepsilon) = C, \qquad \forall\, \varepsilon \in (0,1)$$

where C is the channel capacity defined as

$$C = C(W) = \max_P I(P, W)$$

Background: Shannon's Channel Coding Theorem

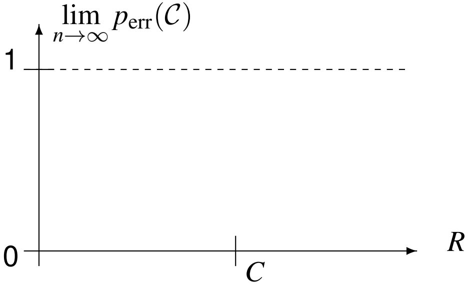
$$\lim_{n\to\infty} \frac{1}{n} \log M^*(W^n, \varepsilon) = C \quad \text{bits/channel use}$$

- Noisy channel coding theorem is independent of $\varepsilon \in (0, 1)$

Background: Shannon's Channel Coding Theorem

$$\lim_{n\to\infty} \frac{1}{n} \log M^*(W^n, \varepsilon) = C \quad \text{bits/channel use}$$

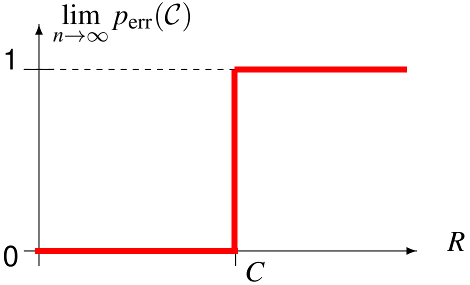
- Noisy channel coding theorem is independent of $\varepsilon \in (0, 1)$



Background: Shannon's Channel Coding Theorem

$$\lim_{n\to\infty} \frac{1}{n} \log M^*(W^n, \varepsilon) = C \quad \text{bits/channel use}$$

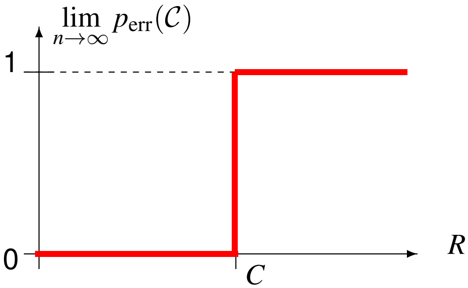
- Noisy channel coding theorem is independent of $\varepsilon \in (0, 1)$



Background: Shannon's Channel Coding Theorem

$$\lim_{n\to\infty} \frac{1}{n} \log M^*(W^n, \varepsilon) = C \quad \text{bits/channel use}$$

- Noisy channel coding theorem is independent of $\varepsilon \in (0, 1)$



- Phase transition at capacity

Background: ε-Dispersion

- What happens at capacity?

Background: ε-Dispersion

- What happens at capacity?
- More precisely, what happens when

$$\log |\mathcal{M}| \approx nC + a\sqrt{n}$$

for some $a \in \mathbb{R}$?

Background: ε-Dispersion

- What happens at capacity?
- More precisely, what happens when

$$\log |\mathcal{M}| \approx nC + a\sqrt{n}$$

for some $a \in \mathbb{R}$?

- Assume capacity-achieving input distribution (CAID) P^* is unique

Background: ε-Dispersion

- What happens at capacity?
- More precisely, what happens when

$$\log |\mathcal{M}| \approx nC + a\sqrt{n}$$

for some $a \in \mathbb{R}$?

- Assume capacity-achieving input distribution (CAID) P^* is unique
- The ε-dispersion is an operational quantity that is equal to

$$V_\varepsilon = V(P^*, W) = \mathbb{E}_{P^*}\left[\mathrm{Var}_{W(\cdot|X)}\left(\log \frac{W(\cdot|X)}{Q^*(\cdot)} \,\Big|\, X\right)\right]$$

where $(X, Y) \sim P^* \times W$ and $Q^*(y) = \sum_x P^*(x)W(y|x)$

Background: ε-Dispersion

- What happens at capacity?
- More precisely, what happens when

$$\log|\mathcal{M}| \approx nC + a\sqrt{n}$$

for some $a \in \mathbb{R}$?

- Assume capacity-achieving input distribution (CAID) P^* is unique
- The ε-dispersion is an operational quantity that is equal to

$$V_\varepsilon = V(P^*, W) = \mathbb{E}_{P^*}\left[\mathrm{Var}_{W(\cdot|X)}\left(\log \frac{W(\cdot|X)}{Q^*(\cdot)} \,\Big|\, X\right)\right]$$

where $(X, Y) \sim P^* \times W$ and $Q^*(y) = \sum_x P^*(x) W(y|x)$

- Since CAID is unique, $V_\varepsilon = V$

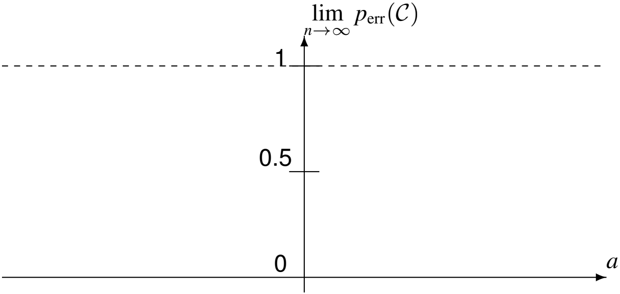
Background: ε-Dispersion

Assume rate of the code satisfies

$$\frac{1}{n}\log|\mathcal{M}| = C + \frac{a}{\sqrt{n}}$$

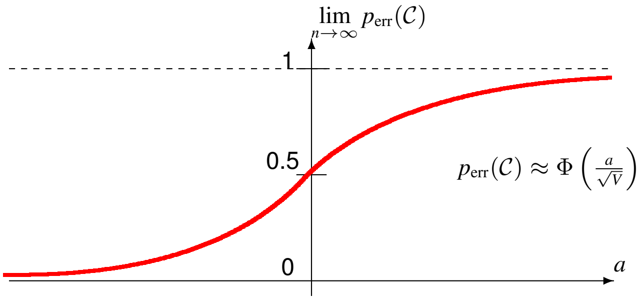
Assume rate of the code satisfies

$$\frac{1}{n}\log|\mathcal{M}| = C + \frac{a}{\sqrt{n}}$$



Assume rate of the code satisfies

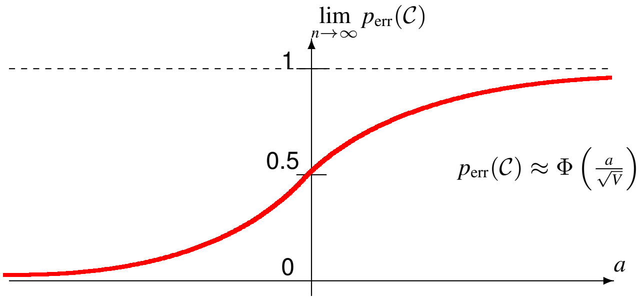
$$\frac{1}{n}\log|\mathcal{M}| = C + \frac{a}{\sqrt{n}}$$



Background: ε-Dispersion

Assume rate of the code satisfies

$$\frac{1}{n}\log|\mathcal{M}| = C + \frac{a}{\sqrt{n}}$$



Here, we have fixed a, the second-order coding rate [Hayashi (2009)]

Background: ε-Dispersion

Theorem (Strassen (1964), Hayashi (2009), Polyanskiy-Poor-Verdú (2010))

For every $\varepsilon \in (0,1)$, and if $V_\varepsilon > 0$, we have

$$\log M^*(W^n, \varepsilon) = nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon) + O(\log n)$$

Background: ε-Dispersion

Theorem (Strassen (1964), Hayashi (2009), Polyanskiy-Poor-Verdú (2010))

For every $\varepsilon \in (0,1)$, *and if* $V_\varepsilon > 0$, *we have*

$$\log M^*(W^n, \varepsilon) = nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon) + O(\log n)$$

V. Strassen (1964)

M. Hayashi (2009)

Polyanskiy-Poor-Verdú (2010)

Background: ε-Dispersion

- Berry-Esséen theorem: For independent X_i with zero-mean and variances σ_i^2,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \geq a\right) = \mathrm{Q}\left(\frac{a}{\bar{\sigma}}\right) \pm \frac{6\,B}{\sqrt{n}}$$

where $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$ and B is related to the third moment

Background: ε-Dispersion

- Berry-Esséen theorem: For independent X_i with zero-mean and variances σ_i^2,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \geq a\right) = \mathrm{Q}\left(\frac{a}{\bar{\sigma}}\right) \pm \frac{6\,B}{\sqrt{n}}$$

  where $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$ and B is related to the third moment

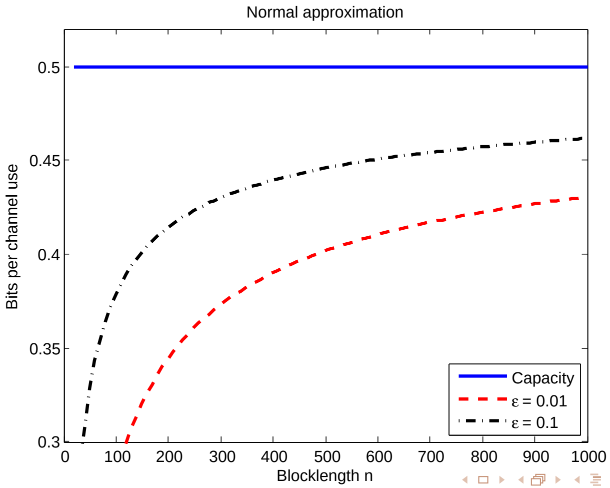
- PPV showed that the normal approximation

$$\log M^*(W^n, \varepsilon) \approx nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)$$

  is very accurate even at moderate blocklengths of ≈ 100

Background: ε-Dispersion for the BSC

For a BSC with crossover probability $p = 0.11$, the normal approximation yields:



Related Work: Third-Order Term

- Recall that we are interested in quantifying the third-order term ρ_n

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)\right]$$

- $\rho_n = O(\log n)$ if channel is non-exotic

Related Work: Third-Order Term

- Recall that we are interested in quantifying the third-order term ρ_n

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)\right]$$

- $\rho_n = O(\log n)$ if channel is non-exotic
- Motivation 1: ρ_n may be important at very short blocklengths

Related Work: Third-Order Term

- Recall that we are interested in quantifying the third-order term ρ_n

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)\right]$$

- $\rho_n = O(\log n)$ if channel is non-exotic
- Motivation 1: ρ_n may be important at very short blocklengths
- Motivation 2: Because we're information theorists

Wir müssen wissen – wir werden wissen (David Hilbert)

Related Work: Third-Order Term

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)\right]$$

- For the BSC [PPV10]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

Related Work: Third-Order Term

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)\right]$$

- For the BSC [PPV10]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

- For the BEC [PPV10]

$$\rho_n = O(1)$$

Related Work: Third-Order Term

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)\right]$$

- For the BSC [PPV10]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

- For the BEC [PPV10]

$$\rho_n = O(1)$$

- For the AWGN under maximum-power constraints [PPV10]

$$O(1) \le \rho_n \le \frac{1}{2}\log n + O(1)$$

Related Work: Third-Order Term

$$\rho_n = \log M^*(W^n, \varepsilon) - \left[nC - \sqrt{nV}\mathrm{Q}^{-1}(\varepsilon)\right]$$

- For the BSC [PPV10]

$$\rho_n = \frac{1}{2}\log n + O(1)$$

- For the BEC [PPV10]

$$\rho_n = O(1)$$

- For the AWGN under maximum-power constraints [PPV10]

$$O(1) \le \rho_n \le \frac{1}{2}\log n + O(1)$$

- Our converse technique can be applied to the AWGN channel

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ *are positive and* $C > 0$*. Then,*

$$\rho_n \geq \frac{1}{2} \log n + O(1)$$

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ are positive and $C > 0$. Then,

$$\rho_n \geq \frac{1}{2}\log n + O(1)$$

- This is an achievability result

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ *are positive and* $C > 0$. *Then,*

$$\rho_n \geq \frac{1}{2} \log n + O(1)$$

- This is an achievability result
- BEC doesn't satisfy assumptions

Related Work: Achievability for Third-Order Term

Proposition (Polyanskiy (2010))

Assume that all elements of $\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ *are positive and* $C > 0$. *Then,*

$$\rho_n \geq \frac{1}{2} \log n + O(1)$$

- This is an achievability result
- BEC doesn't satisfy assumptions
- We will not try to improve on it

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \le \frac{1}{2}\log n + O(1)$$

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- This is a converse result

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \le \frac{1}{2}\log n + O(1)$$

- This is a converse result
- Gallager-symmetric channels are weakly input-symmetric

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2}\log n + O(1)$$

- This is a converse result
- Gallager-symmetric channels are weakly input-symmetric
- The set of weakly input-symmetric channels is very thin

Related Work: Converse for Third-Order Term

Proposition (Polyanskiy (2010))

If W is weakly input-symmetric

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- This is a converse result
- Gallager-symmetric channels are weakly input-symmetric
- The set of weakly input-symmetric channels is very thin
- We dispense of this symmetry assumption

Related Work: Converse for Third-Order Term

Proposition (Strassen (1964), PPV (2010))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \left(|\mathcal{X}| - \frac{1}{2} \right) \log n + O(1)$$

Related Work: Converse for Third-Order Term

Proposition (Strassen (1964), PPV (2010))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \left(|\mathcal{X}| - \frac{1}{2} \right) \log n + O(1)$$

- Every code can be partitioned into no more than $(n+1)^{|\mathcal{X}|-1}$ constant-composition subcodes

Related Work: Converse for Third-Order Term

Proposition (Strassen (1964), PPV (2010))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \left(|\mathcal{X}| - \frac{1}{2} \right) \log n + O(1)$$

- Every code can be partitioned into no more than $(n+1)^{|\mathcal{X}|-1}$ constant-composition subcodes
- $M_P^*(W^n, \varepsilon)$: Max size of a constant-composition code with type P

Related Work: Converse for Third-Order Term

Proposition (Strassen (1964), PPV (2010))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \left(|\mathcal{X}| - \frac{1}{2} \right) \log n + O(1)$$

- Every code can be partitioned into no more than $(n+1)^{|\mathcal{X}|-1}$ constant-composition subcodes
- $M_P^*(W^n, \varepsilon)$: Max size of a constant-composition code with type P
- As such,

$$M^*(W^n, \varepsilon) \le (n+1)^{|\mathcal{X}|-1} \max_{P \in \mathcal{P}_n(\mathcal{X})} M_P^*(W^n, \varepsilon)$$

Related Work: Converse for Third-Order Term

Proposition (Strassen (1964), PPV (2010))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \left(|\mathcal{X}| - \frac{1}{2}\right) \log n + O(1)$$

- Every code can be partitioned into no more than $(n+1)^{|\mathcal{X}|-1}$ constant-composition subcodes
- $M_P^*(W^n, \varepsilon)$: Max size of a constant-composition code with type P
- As such,

$$M^*(W^n, \varepsilon) \le (n+1)^{|\mathcal{X}|-1} \max_{P \in \mathcal{P}_n(\mathcal{X})} M_P^*(W^n, \varepsilon)$$

- This is where the dependence on $|\mathcal{X}|$ comes in

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved without further assumptions

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved without further assumptions
- For BSC

$$\rho_n = \frac{1}{2} \log n + O(1)$$

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \leq \frac{1}{2} \log n + O(1)$$

- The $\frac{1}{2}$ cannot be improved without further assumptions
- For BSC
$$\rho_n = \frac{1}{2} \log n + O(1)$$
- We can dispense of the positive ε-dispersion assumption as well

Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If W is a DMC with positive ε-dispersion,

$$\rho_n \le \frac{1}{2}\log n + O(1)$$

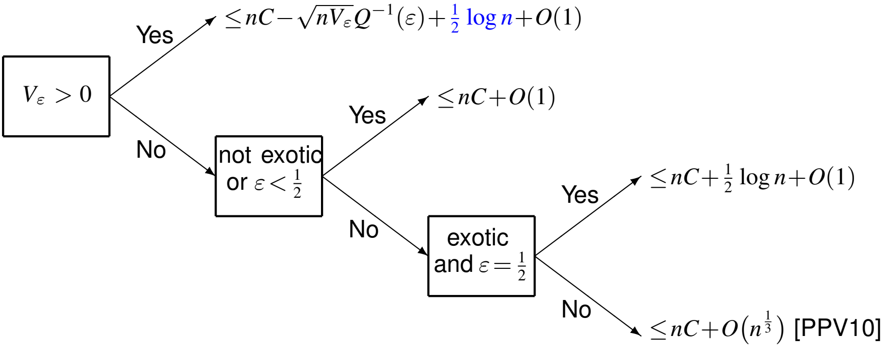
- The $\frac{1}{2}$ cannot be improved without further assumptions
- For BSC

$$\rho_n = \frac{1}{2}\log n + O(1)$$

- We can dispense of the positive ε-dispersion assumption as well
- No need for unique CAID

Main Result: Tight Third-Order Term

All cases are covered



Proof Technique for Tight Third-Order Term

- For the regular case, $\rho_n \leq \frac{1}{2} \log n + O(1)$

Proof Technique for Tight Third-Order Term

- For the regular case, $\rho_n \leq \frac{1}{2}\log n + O(1)$
- The type-counting trick and upper bounds on $M_P^*(W^n, \varepsilon)$ are not sufficiently tight

Proof Technique for Tight Third-Order Term

- For the regular case, $\rho_n \leq \frac{1}{2}\log n + O(1)$
- The type-counting trick and upper bounds on $M_P^*(W^n, \varepsilon)$ are not sufficiently tight
- We need a new converse bound for general DMCs

Proof Technique for Tight Third-Order Term

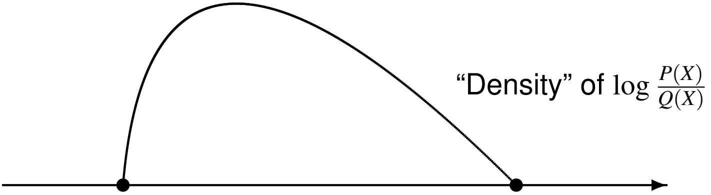
- For the regular case, $\rho_n \leq \frac{1}{2}\log n + O(1)$
- The type-counting trick and upper bounds on $M_P^*(W^n, \varepsilon)$ are not sufficiently tight
- We need a new converse bound for general DMCs
- Information spectrum divergence

$$D_s^\varepsilon(P\|Q) := \sup\left\{R \in \mathbb{R} \,\middle|\, P\left(\log\frac{P(X)}{Q(X)} \leq R\right) \leq \varepsilon\right\}$$

"Information Spectrum Methods in Information Theory"
by T. S. Han (2003)

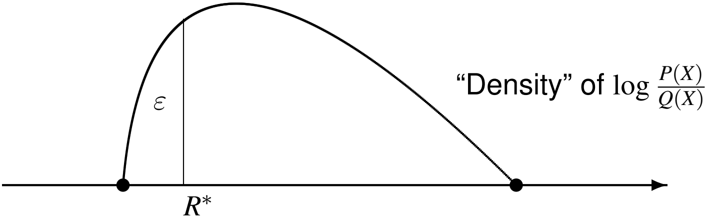
Proof Technique: Information Spectrum Divergence

$$D_s^{\varepsilon}(P\|Q) := \sup\left\{ R \in \mathbb{R} \,\middle|\, P\left(\log \frac{P(X)}{Q(X)} \le R\right) \le \varepsilon \right\}$$



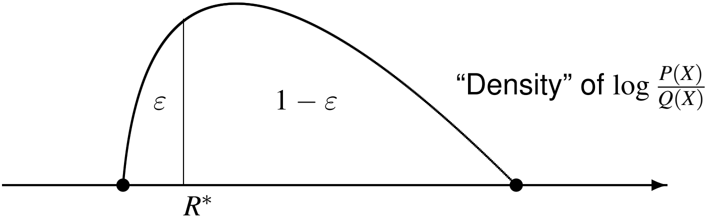
Proof Technique: Information Spectrum Divergence

$$D_s^{\varepsilon}(P\|Q) := \sup\left\{ R \in \mathbb{R} \,\middle|\, P\left(\log \frac{P(X)}{Q(X)} \le R\right) \le \varepsilon \right\}$$



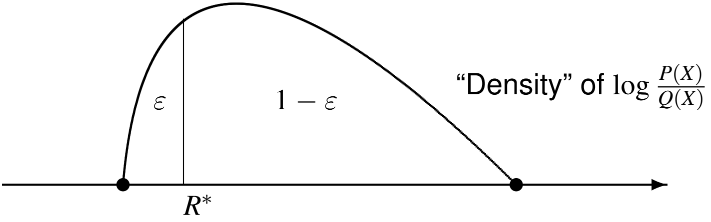
Proof Technique: Information Spectrum Divergence

$$D_s^\varepsilon(P\|Q) := \sup \left\{ R \in \mathbb{R} \,\middle|\, P\left(\log \frac{P(X)}{Q(X)} \le R \right) \le \varepsilon \right\}$$



Proof Technique: Information Spectrum Divergence

$$D_s^\varepsilon(P\|Q) := \sup\left\{R \in \mathbb{R} \,\middle|\, P\left(\log\frac{P(X)}{Q(X)} \le R\right) \le \varepsilon\right\}$$



If X^n is i.i.d. P, the central limit theorem yields

$$D_s^\varepsilon(P^n\|Q^n) \approx nD(P\|Q) - \sqrt{nV(P\|Q)}\mathrm{Q}^{-1}(\varepsilon)$$

Proof Technique: The New Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

Proof Technique: The New Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

- When DMC is used n times,

$$\log M^*(W^n,\varepsilon) \le \min_{Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)} \max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) + \log \frac{1}{\delta}$$

Proof Technique: The New Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

- When DMC is used n times,

$$\log M^*(W^n,\varepsilon) \le \min_{Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)} \max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) + \log \frac{1}{\delta}$$

- Choose $\delta = n^{-\frac{1}{2}}$ so $\log \frac{1}{\delta} = \frac{1}{2} \log n$

Proof Technique: The New Converse Bound

Lemma (Tomamichel-Tan (2013))

For every channel W, every $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$, we have

$$\log M^*(W,\varepsilon) \le \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D_s^{\varepsilon+\delta}(W(\cdot|x) \| Q) + \log \frac{1}{\delta}$$

- When DMC is used n times,

$$\log M^*(W^n,\varepsilon) \le \min_{Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)} \max_{\mathbf{x} \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x}) \| Q^{(n)}) + \log \frac{1}{\delta}$$

- Choose $\delta = n^{-\frac{1}{2}}$ so $\log \frac{1}{\delta} = \frac{1}{2} \log n$
- Since all $\mathbf{x}$ within a type class result in the same $D_s^{\varepsilon+\delta}$ (if $Q^{(n)}$ is permutation invariant), it's really a $\max$ over types $P_{\mathbf{x}} \in \mathcal{P}_n(\mathcal{X})$

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \le \max_{\mathbf{x}\in\mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x})\|Q^{(n)}) + \log\frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \le \max_{\mathbf{x}\in\mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x})\|Q^{(n)}) + \log\frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \le \max_{\mathbf{x}\in\mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x})\|Q^{(n)}) + \log\frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

- First term: $Q_{\mathbf{k}}$'s and $\lambda(\mathbf{k})$'s designed to form an $n^{-\frac{1}{2}}$-cover of $\mathcal{P}(\mathcal{Y})$:

$$\forall\, Q \in \mathcal{P}(\mathcal{Y}), \quad \exists\, \mathbf{k}\in\mathcal{K} \quad \text{s.t.} \quad \|Q - Q_{\mathbf{k}}\|_2 \le n^{-\frac{1}{2}}.$$

Proof Technique: Choice of Output Distribution

$$\log M^*(W^n, \varepsilon) \le \max_{\mathbf{x}\in\mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|\mathbf{x})\|Q^{(n)}) + \log\frac{1}{\delta}, \qquad \forall\, Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)$$

- $Q^{(n)}(\mathbf{y})$: invariant to permutations of the n channel uses

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

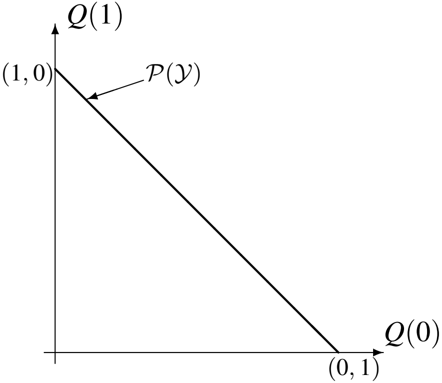
- First term: $Q_{\mathbf{k}}$'s and $\lambda(\mathbf{k})$'s designed to form an $n^{-\frac{1}{2}}$-cover of $\mathcal{P}(\mathcal{Y})$:

$$\forall\, Q \in \mathcal{P}(\mathcal{Y}), \quad \exists\, \mathbf{k}\in\mathcal{K} \quad \text{s.t.} \quad \|Q - Q_{\mathbf{k}}\|_2 \le n^{-\frac{1}{2}}.$$

- Second term: Mixture over output distributions induced by input types [Hayashi (2009)]

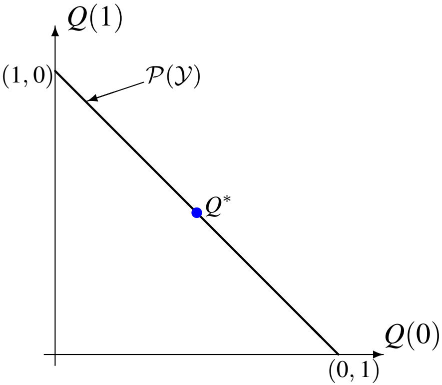
Proof Technique: Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



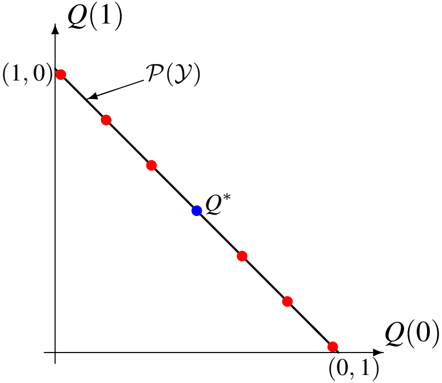
Proof Technique: Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



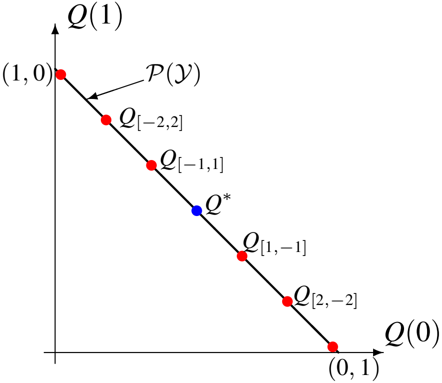
Proof Technique: Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q^n_{\mathbf{k}}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



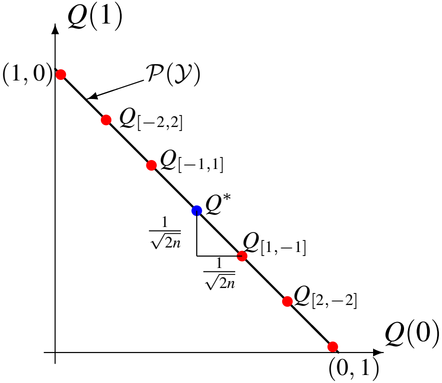
Proof Technique: Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q^n_{\mathbf{k}}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



Proof Technique: Choice of Output Distribution

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^n(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$



Proof Technique: Summary

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

- This construction ensures that for every type $P_{\mathbf{x}}$ near the CAID is well-approximated by by a $Q_{\mathbf{k}(\mathbf{x})}$

Proof Technique: Summary

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

- This construction ensures that for every type $P_{\mathbf{x}}$ near the CAID is well-approximated by by a $Q_{\mathbf{k}(\mathbf{x})}$
- Well in the sense that the loss is

$$-\log\lambda(\mathbf{k}) = O(1)$$

for every $\mathbf{x}$ such that $P_{\mathbf{x}}$ is near the CAID

Proof Technique: Summary

$$Q^{(n)}(\mathbf{y}) := \frac{1}{2}\sum_{\mathbf{k}\in\mathcal{K}} \lambda(\mathbf{k}) Q_{\mathbf{k}}^{n}(\mathbf{y}) + \frac{1}{2}\sum_{P\in\mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(\mathbf{y})$$

- This construction ensures that for every type $P_{\mathbf{x}}$ near the CAID is well-approximated by by a $Q_{\mathbf{k}(\mathbf{x})}$
- Well in the sense that the loss is

$$-\log\lambda(\mathbf{k}) = O(1)$$

  for every $\mathbf{x}$ such that $P_{\mathbf{x}}$ is near the CAID
- For types $P_{\mathbf{x}}$ far from the CAID, use the second part and

$$I(P_{\mathbf{x}}, W) \le C' < C$$

Summary and Food for Thought

- We showed that for DMCs with positive ε-dispersion,

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\,\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

Summary and Food for Thought

- We showed that for DMCs with positive ε-dispersion,

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- How important is the assumption of discreteness?

Summary and Food for Thought

- We showed that for DMCs with positive ε-dispersion,

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- How important is the assumption of discreteness?
- Does our uniform quantization technique extend to lossy source coding? [Ingber-Kochman (2010), Kostina-Verdú (2012)]

Summary and Food for Thought

- We showed that for DMCs with positive ε-dispersion,

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon} \mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- How important is the assumption of discreteness?
- Does our uniform quantization technique extend to lossy source coding? [Ingber-Kochman (2010), Kostina-Verdú (2012)]
- Alternate proof using Bahadur-Ranga Rao [Moulin (2012)]?

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \ge c \right) = \Theta\left(\frac{\exp(-nI(c))}{\sqrt{n}} \right)$$

Summary and Food for Thought

- We showed that for DMCs with positive ε-dispersion,

$$\log M^*(W^n, \varepsilon) \le nC - \sqrt{nV_\varepsilon}\mathrm{Q}^{-1}(\varepsilon) + \frac{1}{2}\log n + O(1)$$

- How important is the assumption of discreteness?
- Does our uniform quantization technique extend to lossy source coding? [Ingber-Kochman (2010), Kostina-Verdú (2012)]
- Alternate proof using Bahadur-Ranga Rao [Moulin (2012)]?

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \ge c\right) = \Theta\left(\frac{\exp(-nI(c))}{\sqrt{n}}\right)$$

- This result has been used to refine the sphere-packing bound [Altug-Wagner (2012)]