On Third-Order Asymptotics for DMCs

Vincent Y. F. Tan

Institute for Infocomm Research (I²R)
National University of Singapore (NUS)

January 20, 2013
Acknowledgements

This is joint work with Marco Tomamichel

Centre for Quantum Technologies
National University of Singapore
Information theory \equiv \text{Finding fundamental limits for reliable information transmission}
Shannon abstracted away information meaning, "semantics"
• treat all data equally — bits as a "universal currency"
• crucial abstraction for modern communication and computing systems
Also relaxed computation and delay constraints to discover a fundamental limit: capacity, providing a goal-post to work toward

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

The average error probability $p_{\text{err}}(\mathcal{C})$ is

$$p_{\text{err}}(\mathcal{C}) := \Pr [\hat{M} \neq M]$$

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

![Diagram showing channel coding process](image)
A code is an triple $C = \{M, e, d\}$ where $M$ is the message set.

The average error probability $p_{\text{err}}(C)$ is

$$p_{\text{err}}(C) := \Pr[\hat{M} \neq M]$$

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

$\varepsilon$-Error Capacity is

$$M^*(W, \varepsilon) := \sup \{ m \in \mathbb{N} \mid \exists C \text{ s.t. } m = |\mathcal{M}|, p_{\text{err}}(C) \leq \varepsilon \}$$
Consider $n$ independent uses of a channel
Consider $n$ independent uses of a channel

Assume $W$ is a discrete memoryless channel
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 \((n\text{-Shot})\)

- Consider \(n\) independent uses of a channel
- Assume \(W\) is a **discrete memoryless channel**
- For vectors \(x = (x_1, \ldots, x_n) \in X^n\) and \(y := (y_1, \ldots, y_n) \in Y^n\),
  \[
  W^n(y|x) = \prod_{i=1}^{n} W(y_i|x_i)
  \]
- Blocklength \(n\), \(\varepsilon\)-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

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

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

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

\[
\log M^*(W^n, \varepsilon) \leq nC - \sqrt{nV_\varepsilon} Q^{-1}(\varepsilon) + \frac{1}{2} \log n + O(1)
\]

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

- The \( \frac{1}{2} \log n \) term is our main contribution
Main Contribution: Remarks

Our bound

$$\log M^*(W^n, \varepsilon) \leq nC - \sqrt{nV\varepsilon Q^{-1}(\varepsilon)} + \frac{1}{2} \log n + O(1)$$
Main Contribution: Remarks

- Our bound

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

- Best upper bound till date:

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

V. Strassen (1964)  
Polyanskiy-Poor-Verdú or PPV (2010)
Main Contribution: Remarks

- Our bound

\[ \log M^*(W^n, \varepsilon) \leq nC - \sqrt{nV\varepsilon}Q^{-1}(\varepsilon) + \frac{1}{2} \log n + O(1) \]

- Best upper bound till date:

\[ \log M^*(W^n, \varepsilon) \leq nC - \sqrt{nV\varepsilon}Q^{-1}(\varepsilon) + \left( \frac{|X|}{2} - \frac{1}{2} \right) \log n + O(1) \]

V. Strassen (1964)  Polyanskiy-Poor-Verdú or PPV (2010)

- Requires new converse techniques
Outline

1. Background
2. Related work
3. Main result
4. New converse
5. Proof sketch
6. Summary and open problems
Shannon’s noisy channel coding theorem and

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

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

where \( C \) is the channel capacity defined as

\[
C = C(W) = \max_P I(P, W)
\]
Background: Shannon’s Channel Coding Theorem

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

- Noisy channel coding theorem is independent of $\varepsilon \in (0, 1)$
Background: Shannon’s Channel Coding Theorem

\[ \lim_{n \to \infty} \frac{1}{n} \log M^*(W^n, \varepsilon) = C \] 

bits/channel use

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

\[ \lim_{n \to \infty} p_{err}(C) \]

0 \hspace{2cm} C \hspace{2cm} R

1

\[ \text{Phase transition at capacity} \]
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

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

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

- Phase transition at capacity

\[ \lim_{n \to \infty} p_{err}(C) \]
What happens at capacity?
Background: $\varepsilon$-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: $\varepsilon$-Dispersion

- What happens at capacity?

- More precisely, what happens when

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

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

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

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

where $(X, Y) \sim P^* \times W$ and $Q^*(y) = \sum_x P^*(x) W(y | x)$.
Background: \( \varepsilon \)-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 \( \varepsilon \)-dispersion is an operational quantity that is equal to

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

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

- Since CAID is unique, \( V_\varepsilon = V \)
Assume rate of the code satisfies

\[ \frac{1}{n} \log |\mathcal{M}| = C + \frac{a}{\sqrt{n}} \]
Assume rate of the code satisfies

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

Here, we have fixed \(a\), the second-order coding rate [Hayashi (2009)].
Background: $\varepsilon$-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)].
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)]
Theorem (Strassen (1964), Hayashi (2009), Polyanskiy-Poor-Verdú (2010))

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

$$
\log M^*(W^n, \varepsilon) = nC - \sqrt{nVQ^{-1}(\varepsilon)} + O(\log n)
$$
Background: \( \varepsilon \)-Dispersion

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

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

\[
\log M^*(W^n, \varepsilon) = nC - \sqrt{nVQ^{-1}(\varepsilon)} + O(\log n)
\]
Background: $\varepsilon$-Dispersion

- **Berry-Esséen theorem**: For independent $X_i$ with zero-mean and variances $\sigma_i^2$,

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

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 $\sigma_i^2$,

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

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

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

is very accurate even at moderate blocklengths of $\approx 100$. 
For a BSC with crossover probability $p = 0.11$, the normal approximation yields:
Recall that we are interested in quantifying the third-order term $\rho_n$

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

- $\rho_n = O(\log n)$ if channel is non-exotic
Recall that we are interested in quantifying the third-order term $\rho_n$

$$\rho_n = \log M^*(W^n, \varepsilon) - [nC - \sqrt{nVQ^{-1}}(\varepsilon)]$$

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

- Recall that we are interested in quantifying the third-order term $\rho_n$

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

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

- Motivation 1: $\rho_n$ may be important at very short blocklengths

- Motivation 2: Because we’re information theorists

  \textit{Wir müssen wissen – wir werden wissen (David Hilbert)}
\[ \rho_n = \log M^*(W^n, \varepsilon) - \left[ nC - \sqrt{nVQ^{-1}}(\varepsilon) \right] \]

For the BSC [PPV10]

\[ \rho_n = \frac{1}{2} \log n + O(1) \]
Related Work: Third-Order Term

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

- For the BSC [PPV10]
  \[
  \rho_n = \frac{1}{2} \log n + O(1)
  \]

- For the BEC [PPV10]
  \[
  \rho_n = O(1)
  \]
\[ \rho_n = \log M^*(W^n, \varepsilon) - \left[ nC - \sqrt{nVQ^{-1}}(\varepsilon) \right] \]

- 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]
  \[ O(1) \leq \rho_n \leq \frac{1}{2} \log n + O(1) \]
Related Work: Third-Order Term

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

- 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]
  \[ O(1) \leq \rho_n \leq \frac{1}{2} \log n + O(1) \]

- Our converse technique can be applied to the AWGN channel
Proposition (Polyanskiy (2010))

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

\[
\rho_n \geq \frac{1}{2} \log n + O(1)
\]
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
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
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
Proposition (Polyanskiy (2010))

If $W$ is weakly input-symmetric

\[ \rho_n \leq \frac{1}{2} \log n + O(1) \]
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*

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

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

*If* $W$ *is a DMC with positive* $\varepsilon$-*dispersion,*

$$\rho_n \leq \left( |X| - \frac{1}{2} \right) \log n + O(1)$$
Proposition (Strassen (1964), PPV (2010))

If $W$ is a DMC with positive $\varepsilon$-dispersion,

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

Every code can be partitioned into no more than $(n + 1)|X|^{-1}$ constant-composition subcodes
Related Work: Converse for Third-Order Term

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

If $W$ is a DMC with positive $\varepsilon$-dispersion,

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

- 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 $\varepsilon$-dispersion,

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

- 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,

$$M^*(W^n, \varepsilon) \leq (n + 1)|\mathcal{X}|^{-1} \max_{P \in \mathcal{P}_n(\mathcal{X})} M^*_P(W^n, \varepsilon)$$
Related Work: Converse for Third-Order Term

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

If $W$ is a DMC with positive $\varepsilon$-dispersion,

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

- 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,

$$M^*(W^n, \varepsilon) \leq (n + 1)|\mathcal{X}|^{-1} \max_{P \in \mathcal{P}_n(\mathcal{X})} M^*_P(W^n, \varepsilon)$$
- 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 $\varepsilon$-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 $\varepsilon$-dispersion assumption as well.

No need for unique CAID.
Theorem (Tomamichel-Tan (2013))

If $W$ is a DMC with positive $\varepsilon$-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 $\varepsilon$-dispersion,

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

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

$$\rho_n = \frac{1}{2} \log n + O(1)$$
Theorem (Tomamichel-Tan (2013))

If $W$ is a DMC with positive $\varepsilon$-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 $\varepsilon$-dispersion assumption as well
Main Result: Tight Third-Order Term

Theorem (Tomamichel-Tan (2013))

If $W$ is a DMC with positive $\varepsilon$-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 $\varepsilon$-dispersion assumption as well.
- No need for unique CAID.
Main Result: Tight Third-Order Term

All cases are covered

$V_\varepsilon > 0$

- Yes
  - $\leq nC - \sqrt{nV_\varepsilon Q^{-1}(\varepsilon)} + \frac{1}{2} \log n + O(1)$

- No
  - not exotic or $\varepsilon < \frac{1}{2}$
    - Yes
      - $\leq nC + O(1)$
    - No
      - Yes
        - $\leq nC + \frac{1}{2} \log n + O(1)$
      - No
        - exotic and $\varepsilon = \frac{1}{2}$
          - Yes
            - $\leq nC + O(1)$
          - No
            - $\leq nC + O(n^{\frac{1}{3}})$ [PPV10]
For the regular case, \( \rho_n \leq \frac{1}{2} \log n + O(1) \)
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.
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

\[ D^\varepsilon_s(P\|Q) := \sup \left\{ R \in \mathbb{R} \mid P \left( \log \frac{P(X)}{Q(X)} \leq R \right) \leq \varepsilon \right\} \]

Proof Technique: Information Spectrum Divergence

\[ D_s^\varepsilon(P \parallel Q) := \sup \left\{ R \in \mathbb{R} \mid P \left( \log \frac{P(X)}{Q(X)} \leq R \right) \leq \varepsilon \right\} \]

"Density" of \( \log \frac{P(X)}{Q(X)} \)
Proof Technique: Information Spectrum Divergence

\[ D_s^\varepsilon(P\|Q) := \sup \left\{ R \in \mathbb{R} \mid P \left( \log \frac{P(X)}{Q(X)} \leq R \right) \leq \varepsilon \right\} \]

"Density" of \( \log \frac{P(X)}{Q(X)} \)
Proof Technique: Information Spectrum Divergence

\[ D_s^\varepsilon(P\|Q) := \sup \left\{ R \in \mathbb{R} \mid P \left( \log \frac{P(X)}{Q(X)} \leq R \right) \leq \varepsilon \right\} \]

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

\[ D_s^\varepsilon(P^n\|Q^n) \approx n D(P\|Q) - \sqrt{n V(P\|Q)} Q - 1(\varepsilon) \]
Proof Technique: Information Spectrum Divergence

\[ D^\varepsilon_s(P \parallel Q) := \sup \left\{ R \in \mathbb{R} \mid P \left( \log \frac{P(X)}{Q(X)} \leq R \right) \leq \varepsilon \right\} \]

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

\[ D^\varepsilon_s(P^n \parallel Q^n) \approx nD(P \parallel Q) - \sqrt{nV(P \parallel Q)Q^{-1}(\varepsilon)} \]
Lemma (Tomamichel-Tan (2013))

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

$$\log M^*(W, \varepsilon) \leq \min_{Q \in \mathcal{P}(Y)} \max_{x \in X} D_{\varepsilon+\delta}^s(W(\cdot|x)\|Q) + \log \frac{1}{\delta}$$
Lemma (Tomamichel-Tan (2013))

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

\[
\log M^*(W, \varepsilon) \leq \min_{Q \in \mathcal{P}(Y)} \max_{x \in \mathcal{X}} D_{\varepsilon + \delta}^s(W(\cdot|x)\|Q) + \log \frac{1}{\delta}
\]

- When DMC is used \( n \) times,

\[
\log M^*(W^n, \varepsilon) \leq \min_{Q^{(n)} \in \mathcal{P}(Y^n)} \max_{x \in \mathcal{X}^n} D_{\varepsilon + \delta}^s(W^n(\cdot|x)\|Q^{(n)}) + \log \frac{1}{\delta}
\]
Lemma (Tomamichel-Tan (2013))

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

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

- When DMC is used $n$ times,

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

- Choose $\delta = n^{-\frac{1}{2}}$ so $\log \frac{1}{\delta} = \frac{1}{2} \log n$
Lemma (Tomamichel-Tan (2013))

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

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

- When DMC is used $n$ times,

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

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

- Since all $x$ 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 $P_x \in \mathcal{P}_n(\mathcal{X})$.
Proof Technique: Choice of Output Distribution

\[
\log M^*(W^n, \varepsilon) \leq \max_{x \in X^n} D_{s+\delta}^\varepsilon(W^n(\cdot|x)\|Q^{(n)}) + \log \frac{1}{\delta}, \quad \forall Q^{(n)} \in \mathcal{P}(Y^n)
\]

- \( Q^{(n)}(y) \): invariant to permutations of the \( n \) channel uses
Proof Technique: Choice of Output Distribution

\[
\log M^*(W^n, \varepsilon) \leq \max_{x \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|x) \| Q^{(n)}(x)) + \log \frac{1}{\delta}, \quad \forall Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n)
\]

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

\[
Q^{(n)}(y) := \frac{1}{2} \sum_{k \in \mathcal{K}} \lambda(k) Q^n_k(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} (PW)^n(y)
\]
Proof Technique: Choice of Output Distribution

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

- \(Q^{(n)}(y)\): invariant to permutations of the \(n\) channel uses
  \[
  Q^{(n)}(y) := \frac{1}{2} \sum_{k \in \mathcal{K}} \lambda(k) Q_k^n(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|}(PW)^n(y)
  \]

- First term: \(Q_k\)'s and \(\lambda(k)\)'s designed to form an \(n^{-\frac{1}{2}}\)-cover of \(\mathcal{P}(\mathcal{Y})\):
  \[
  \forall Q \in \mathcal{P}(\mathcal{Y}), \quad \exists k \in \mathcal{K} \quad \text{s.t.} \quad \|Q - Q_k\|_2 \leq n^{-\frac{1}{2}}.
  \]
Proof Technique: Choice of Output Distribution

\[ \log M^*(W^n, \varepsilon) \leq \max_{x \in \mathcal{X}^n} D_s^{\varepsilon+\delta}(W^n(\cdot|x)\|Q^{(n)}) + \log \frac{1}{\delta}, \quad \forall Q^{(n)} \in \mathcal{P}(\mathcal{Y}^n) \]

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

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

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

\[ \forall Q \in \mathcal{P}(\mathcal{Y}), \quad \exists k \in \mathcal{K} \quad \text{s.t.} \quad \|Q - Q_k\|_2 \leq n^{-\frac{1}{2}}. \]

- **Second term**: Mixture over output distributions induced by input types [Hayashi (2009)]
Proof Technique: Choice of Output Distribution

\[
Q^{(n)}(y) := \frac{1}{2} \sum_{k \in K} \lambda(k) Q_k^n(y) + \frac{1}{2} \sum_{P \in P_n(X')} \frac{1}{|P_n(X')|} (PW)^n(y)
\]
Proof Technique: Choice of Output Distribution

\[ Q^{(n)}(y) := \frac{1}{2} \sum_{k \in K} \lambda(k) Q^n_k(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} (PW)^n(y) \]
Proof Technique: Choice of Output Distribution

\[ Q^{(n)}(y) := \frac{1}{2} \sum_{k \in \mathcal{K}} \lambda(k) Q_k^n(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(X')} \frac{1}{|\mathcal{P}_n(X')|} (PW)^n(y) \]
Proof Technique: Choice of Output Distribution

\[ Q^{(n)}(y) := \frac{1}{2} \sum_{k \in \mathcal{K}} \lambda(k) Q_k^n(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} (PW)^n(y) \]
Proof Technique: Choice of Output Distribution

\[ Q^{(n)}(y) := \frac{1}{2} \sum_{k \in \mathcal{K}} \lambda(k) Q_k^{n}(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X}')} \frac{1}{|\mathcal{P}_n(\mathcal{X}')|} (PW)^n(y) \]
Proof Technique: Summary

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

This construction ensures that for every type \( P_x \) near the CAID is well-approximated by by a \( Q_{k(x)} \).
Proof Technique: Summary

\[ Q^{(n)}(y) := \frac{1}{2} \sum_{k \in \mathcal{K}} \lambda(k) Q^n_k(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(x')} \frac{1}{|\mathcal{P}_n(x')|} (PW)^n(y) \]

- This construction ensures that for every type \( P_x \) near the CAID is well-approximated by by a \( Q_k(x) \)

- Well in the sense that the loss is

\[ -\log \lambda(k) = O(1) \]

for every \( x \) such that \( P_x \) is near the CAID
Proof Technique: Summary

\[ Q^{(n)}(y) := \frac{1}{2} \sum_{k \in \mathcal{K}} \lambda(k) Q_k^n(y) + \frac{1}{2} \sum_{P \in \mathcal{P}_n(X)} \frac{1}{|P_n(X)|} (PW)^n(y) \]

- This construction ensures that for every type \( P_x \) near the CAID is well-approximated by by a \( Q_k(x) \)
- Well in the sense that the loss is

\[ -\log \lambda(k) = O(1) \]

for every \( x \) such that \( P_x \) is near the CAID

- For types \( P_x \) far from the CAID, use the second part and

\[ I(P_x, W) \leq C' < C \]
We showed that for DMCs with positive $\varepsilon$-dispersion,

$$\log M^*(W^n, \varepsilon) \leq nC - \sqrt{nV_\varepsilon Q^{-1}(\varepsilon)} + \frac{1}{2} \log n + O(1)$$
Summary and Food for Thought

- We showed that for DMCs with positive $\varepsilon$-dispersion,

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

- How important is the assumption of discreteness?
We showed that for DMCs with positive $\varepsilon$-dispersion,

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

How important is the assumption of discreteness?

Does our uniform quantization technique extend to lossy source coding? [Ingber-Kochman (2010), Kostina-Verdú (2012)]
We showed that for DMCs with positive $\varepsilon$-dispersion,

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

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)]?

$$\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \geq c \right) = \Theta \left( \frac{\exp(-nI(c))}{\sqrt{n}} \right)$$
We showed that for DMCs with positive $\varepsilon$-dispersion,

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

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)]?

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

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