ε-Capacity, Strong Converse, and Second-Order Coding Rates for Channels with General State

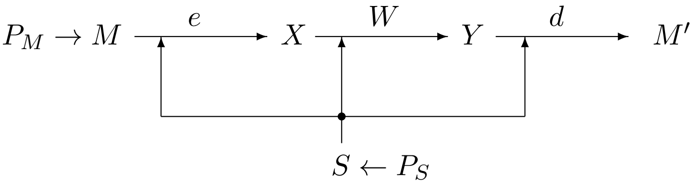
Marco Tomamichel[1] and Vincent Y. F. Tan[2,3]

[1]Centre for Quantum Technologies, National University of Singapore, Singapore
[2]Institute for Infocomm Research (I^{2}R), A*STAR
[3]Department of Electrical and Computer Engineering, National University of Singapore, Singapore

Sevilla, September 13, 2013

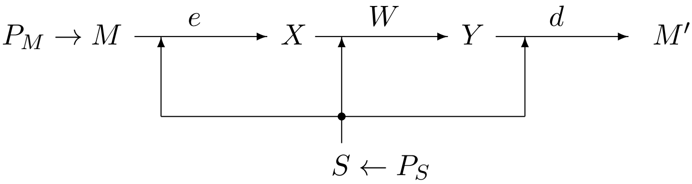
full version: `arXiv:1305.6789`

Setup and Notation



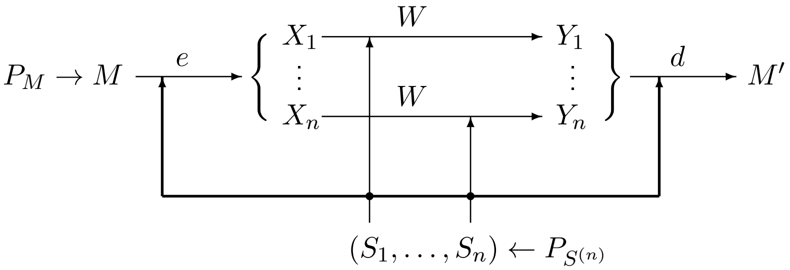
- We consider discrete channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X}, \mathcal{S})$.
- The channel state $S \leftarrow P_S \in \mathcal{P}(\mathcal{S})$ is known non-causally at the encoder and the decoder.

Setup and Notation



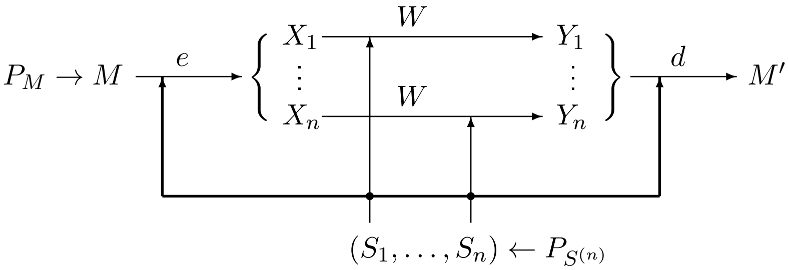
- We consider discrete channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X}, \mathcal{S})$.
- The channel state $S \leftarrow P_S \in \mathcal{P}(\mathcal{S})$ is known non-causally at the encoder and the decoder.
- We keep $\varepsilon \in (0, 1)$ fixed.
- $M^*(\varepsilon; W, P_S)$ is the maximum size of a code $\mathcal{C} = \{\mathcal{M}, e, d\}$ for W such that the **average** error probability is at most ε.

Setup and Notation



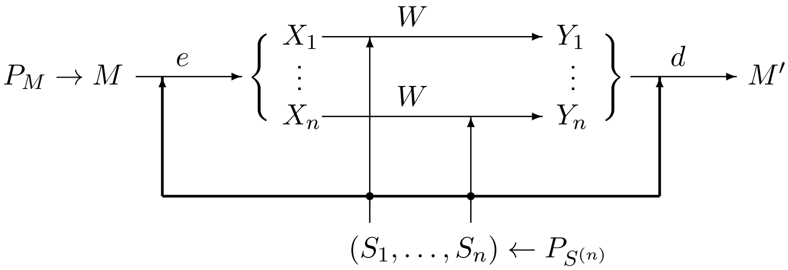
- For any $n \in \mathbb{N}$, W^n is the n-fold memoryless repetition of W.

Setup and Notation



- For any $n \in \mathbb{N}$, W^n is the n-fold memoryless repetition of W.
- $\vec{S} = \left\{ S^{(n)} := \left(S_1^{(n)}, S_2^{(n)}, \dots, S_n^{(n)} \right) \right\}$ is a general sequence of random channel states in the sense of Verdú-Han'94.

Setup and Notation



- For any $n \in \mathbb{N}$, W^n is the n-fold memoryless repetition of W.
- $\vec{S} = \left\{ S^{(n)} := \left(S_1^{(n)}, S_2^{(n)}, \ldots, S_n^{(n)} \right) \right\}$ is a general sequence of random channel states in the sense of Verdú-Han'94.
- Set $M^*(\varepsilon; n) = M^*(\varepsilon; W^n, P_{S^{(n)}})$. We are interested in the asymptotic expansion of $\log M^*(\varepsilon; n)$ for large n.

Setup and Notation

- Define $W_s(y|x) \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ as the law $W_s(y|x) := W(y|x, s)$.

Setup and Notation

- Define $W_s(y|x) \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ as the law $W_s(y|x) := W(y|x,s)$.
- For W_s, we denote its *channel capacity* as

$$\mathcal{C}_s := \sup_{P_s \in \mathcal{P}(\mathcal{X})} I(P_s; W_s) \,.$$

Setup and Notation

- Define $W_s(y|x) \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ as the law $W_s(y|x) := W(y|x,s)$.
- For W_s, we denote its *channel capacity* as

$$\mathcal{C}_s := \sup_{P_s \in \mathcal{P}(\mathcal{X})} I(P_s; W_s)\,.$$

- For our second-order results, we assume the CAID P_s^* of W_s is unique and

$$\mathcal{V}_s := V(P_s^*; W_s) > 0$$

is the *channel dispersion* of W_s.

Setup and Notation

- Define $W_s(y|x) \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ as the law $W_s(y|x) := W(y|x, s)$.
- For W_s, we denote its *channel capacity* as

$$\mathcal{C}_s := \sup_{P_s \in \mathcal{P}(\mathcal{X})} I(P_s; W_s) \,.$$

- For our second-order results, we assume the CAID P_s^* of W_s is unique and

$$\mathcal{V}_s := V(P_s^*; W_s) > 0$$

  is the *channel dispersion* of W_s.
- Φ is the cumulative (normal) Gaussian distribution.

Motivation and Prior Work

- Channels with state can be used to model many realistic scenarios, e.g. burst errors in the Gilbert-Elliott model.

Motivation and Prior Work

- Channels with state can be used to model many realistic scenarios, e.g. burst errors in the Gilbert-Elliott model.
- From a theoretical perspective, the model is rewarding because it allows to separate the effects due to randomness from the channel and the state.

Motivation and Prior Work

- Channels with state can be used to model many realistic scenarios, e.g. burst errors in the Gilbert-Elliott model.
- From a theoretical perspective, the model is rewarding because it allows to separate the effects due to randomness from the channel and the state.
- To the best of our knowledge, the problem has not been considered in this generality.
- Many special cases are discussed in the literature: see, e.g., Keshet-Steinberg-Merhav'08 and El Gamal-Kim'12.
- We point to specific prior results when we discuss particularizations of our results.

ε-Capacity and Optimistic (Converse) ε-Capacity

- A rate R is ε-achievable if there exists a sequence $\{\varepsilon_n\}$ with

$$\liminf_{n\to\infty} \frac{1}{n} \log M^*(\varepsilon_n; n) \geq R \quad \text{and} \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon \,.$$

- ε-Capacity: $C(\varepsilon) := \sup\{R \,|\, R \text{ is } \varepsilon\text{-achievable}\}$.

ε-Capacity and Optimistic (Converse) ε-Capacity

- A rate R is ε-achievable if there exists a sequence $\{\varepsilon_n\}$ with

$$\liminf_{n\to\infty} \frac{1}{n} \log M^*(\varepsilon_n; n) \geq R \quad \text{and} \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon \,.$$

- ε-Capacity: $C(\varepsilon) := \sup\{R \,|\, R \text{ is } \varepsilon\text{-achievable}\}$.
- A rate R is an ε-converse rate if for every sequence $\{\varepsilon_n\}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log M^*(\varepsilon_n; n) \geq R \implies \liminf_{n\to\infty} \varepsilon_n \geq \varepsilon \,.$$

- Optimistic (Converse) ε-Capacity:

$$C^\dagger(\varepsilon) := \inf\{R \,|\, R \text{ is an } \varepsilon\text{-converse rate}\}.$$

ε-Capacity and Optimistic (Converse) ε-Capacity

- A rate R is ε-achievable if there exists a sequence $\{\varepsilon_n\}$ with

$$\liminf_{n\to\infty} \frac{1}{n} \log M^*(\varepsilon_n; n) \geq R \quad \text{and} \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon\,.$$

- ε-Capacity: $C(\varepsilon) := \sup\{R \,|\, R \text{ is } \varepsilon\text{-achievable}\}$.
- A rate R is an ε-converse rate if for every sequence $\{\varepsilon_n\}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log M^*(\varepsilon_n; n) \geq R \implies \liminf_{n\to\infty} \varepsilon_n \geq \varepsilon\,.$$

- Optimistic (Converse) ε-Capacity:

$$C^\dagger(\varepsilon) := \inf\{R \,|\, R \text{ is an } \varepsilon\text{-converse rate}\}.$$

- Strong converse: $C(0) = C^\dagger(1)$. (Verdú-Han'94)

- Introduce the sequence of random variables $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{C}^{(n)}(S_1^{(n)},\ldots,S_n^{(n)}) := \frac{1}{n}\sum_{k=1}^{n}\mathcal{C}_{S_k^{(n)}}.$$

- Introduce the sequence of random variables $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{C}^{(n)}(S_1^{(n)}, \ldots, S_n^{(n)}) := \frac{1}{n} \sum_{k=1}^{n} \mathcal{C}_{S_k^{(n)}}.$$

Theorem

For any $\varepsilon \in [0,1]$, *we have*

$$C(\varepsilon) = \sup \left\{ R \,\middle|\, \limsup_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \le R\right] \le \varepsilon \right\},$$
$$C^{\dagger}(\varepsilon) = \inf \left\{ R \,\middle|\, \liminf_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \le R\right] \ge \varepsilon \right\}.$$

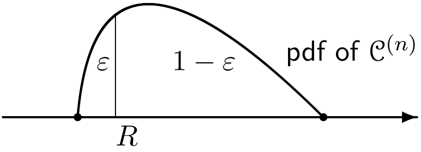
- Introduce the sequence of random variables $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{C}^{(n)}(S_1^{(n)}, \ldots, S_n^{(n)}) := \frac{1}{n}\sum_{k=1}^{n} \mathcal{C}_{S_k^{(n)}}.$$

Theorem

For any $\varepsilon \in [0,1]$, *we have*

$$C(\varepsilon) = \sup\left\{R \,\middle|\, \limsup_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \le R\right] \le \varepsilon\right\},$$
$$C^{\dagger}(\varepsilon) = \inf\left\{R \,\middle|\, \liminf_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \le R\right] \ge \varepsilon\right\}.$$



- Introduce the sequence of random variables $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{C}^{(n)}(S_1^{(n)}, \ldots, S_n^{(n)}) := \frac{1}{n} \sum_{k=1}^{n} \mathcal{C}_{S_k^{(n)}}.$$

Theorem

For any $\varepsilon \in [0,1]$, *we have*

$$C(\varepsilon) = \sup \left\{ R \,\middle|\, \limsup_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \leq R\right] \leq \varepsilon \right\},$$
$$C^{\dagger}(\varepsilon) = \inf \left\{ R \,\middle|\, \liminf_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \leq R\right] \geq \varepsilon \right\}.$$

- Depends on the randomness of the state, not the channel.
- In particular, by definition of $\mathfrak{p}\text{-}\liminf$ and $\mathfrak{p}\text{-}\limsup$,

$$C(0) = \mathfrak{p}\text{-}\liminf_{n\to\infty} \mathcal{C}^{(n)}, \quad C^{\dagger}(1) = \mathfrak{p}\text{-}\limsup_{n\to\infty} \mathcal{C}^{(n)}.$$

- Introduce the sequence of random variables $\{\mathcal{C}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{C}^{(n)}(S_1^{(n)}, \ldots, S_n^{(n)}) := \frac{1}{n}\sum_{k=1}^{n} \mathcal{C}_{S_k^{(n)}}.$$

Theorem

For any $\varepsilon \in [0,1]$, *we have*

$$C(\varepsilon) = \sup\left\{R \,\middle|\, \limsup_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \le R\right] \le \varepsilon\right\},$$
$$C^{\dagger}(\varepsilon) = \inf\left\{R \,\middle|\, \liminf_{n\to\infty} \Pr\left[\mathcal{C}^{(n)} \le R\right] \ge \varepsilon\right\}.$$

- Achievability: State-Dependent Feinstein bound — maximum probability of error.
- Converse: State-Dependent Meta-Converse:

$$M^*(\varepsilon; W, P_S) \le \max_{P\in\mathcal{P}(\mathcal{X}|\mathcal{S})} \min_{Q\in\mathcal{P}(\mathcal{Y}|\mathcal{S})} \frac{1}{\beta_{1-\varepsilon}(P_S\times P\times W \| P_S\times P\times Q)}.$$

Strong Converse

- As a corollary, the strong converse holds (for capacity C) if and only if

$$\mathfrak{p}\text{-}\liminf_{n\to\infty} \mathcal{C}^{(n)} = \mathfrak{p}\text{-}\limsup_{n\to\infty} \mathcal{C}^{(n)} = C.$$

Strong Converse

- As a corollary, the strong converse holds (for capacity C) if and only if

$$\mathfrak{p}\text{-}\liminf_{n\to\infty} \mathcal{C}^{(n)} = \mathfrak{p}\text{-}\limsup_{n\to\infty} \mathcal{C}^{(n)} = C.$$

- We find the following sufficient condition:

Proposition

The strong converse holds (for capacity C) if the following limits exist:

$$\lim_{n\to\infty} \mathrm{E}\left[\mathcal{C}^{(n)}\right] = C, \quad \lim_{n\to\infty} \mathrm{Var}\left[\mathcal{C}^{(n)}\right] = 0.$$

ε-Second-Order Coding Rate

- As introduced by Hayashi'09. More general than dispersion.
- We say that $r \in \mathbb{R} \cup \{\pm\infty\}$ is an *ε-achievable second-order coding rate* if there exists a sequence $\{\varepsilon_n\}$ s.t.

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}}\Big[\log M^*(\varepsilon_n; n) - nC(\varepsilon)\Big] \geq r, \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon$$

ε-Second-Order Coding Rate

- As introduced by Hayashi'09. More general than dispersion.
- We say that $r \in \mathbb{R} \cup \{\pm\infty\}$ is an *ε-achievable second-order coding rate* if there exists a sequence $\{\varepsilon_n\}$ s.t.

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \Big[\log M^*(\varepsilon_n; n) - nC(\varepsilon) \Big] \geq r, \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon$$

- If r is an ε-achievable second-order coding rate, then there exists a sequence of length-n codes with number of codewords M_n and error probability ε_n satisfying

$$\log M_n = nC(\varepsilon) + r\sqrt{n} + o\big(\sqrt{n}\big), \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon\,.$$

ε-Second-Order Coding Rate

- As introduced by Hayashi'09. More general than dispersion.
- We say that $r \in \mathbb{R} \cup \{\pm\infty\}$ is an *ε-achievable second-order coding rate* if there exists a sequence $\{\varepsilon_n\}$ s.t.

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \Big[\log M^*(\varepsilon_n; n) - nC(\varepsilon) \Big] \geq r, \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon$$

- If r is an ε-achievable second-order coding rate, then there exists a sequence of length-n codes with number of codewords M_n and error probability ε_n satisfying

$$\log M_n = nC(\varepsilon) + r\sqrt{n} + o\big(\sqrt{n}\big), \quad \limsup_{n\to\infty} \varepsilon_n \leq \varepsilon\,.$$

- Optimal ε-second order coding rate: $\Lambda(\varepsilon) :=$ $\sup\{r \,|\, r$ is an ε-achievable second-order coding rate$\}$.

- For a fixed realization of a state sequence $\vec{s} = (s_1, \ldots, s_n)$, the average error for code of length M is approximately

$$\Phi\left(\frac{\log M - \sum_{k=1}^{n} \mathcal{C}_{s_k}}{\sqrt{\sum_{k=1}^{n} \mathcal{V}_{S_k}}}\right) \quad \text{by Berry-Esseen.}$$

- For a fixed realization of a state sequence $\vec{s} = (s_1, \ldots, s_n)$, the average error for code of length M is approximately

$$\Phi\left(\frac{\log M - \sum_{k=1}^{n} \mathcal{C}_{s_k}}{\sqrt{\sum_{k=1}^{n} \mathcal{V}_{S_k}}}\right) \quad \text{by Berry-Esseen.}$$

- Introduce the sequence of random variables $\{\mathcal{V}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{V}^{(n)}(S_1^{(n)}, \ldots, S_n^{(n)}) := \frac{1}{n} \sum_{k=1}^{n} \mathcal{V}_{S_k^{(n)}}.$$

- For a fixed realization of a state sequence $\vec{s} = (s_1, \ldots, s_n)$, the average error for code of length M is approximately

$$\Phi\left(\frac{\log M - \sum_{k=1}^{n} \mathcal{C}_{s_k}}{\sqrt{\sum_{k=1}^{n} \mathcal{V}_{S_k}}}\right) \quad \text{by Berry-Esseen.}$$

- Introduce the sequence of random variables $\{\mathcal{V}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{V}^{(n)}(S_1^{(n)}, \ldots, S_n^{(n)}) := \frac{1}{n} \sum_{k=1}^{n} \mathcal{V}_{S_k^{(n)}}.$$

Theorem

For every $\varepsilon \in (0,1)$, we have

$$\Lambda(\varepsilon) = \sup\left\{ r \,\middle|\, \limsup_{n\to\infty} \mathrm{E}\left[\Phi\left(\frac{nC(\varepsilon) - n\mathcal{C}^{(n)} + r\sqrt{n}}{\sqrt{n\mathcal{V}^{(n)}}}\right)\right] \le \varepsilon \right\}.$$

- For a fixed realization of a state sequence $\vec{s} = (s_1, \ldots, s_n)$, the average error for code of length M is approximately

$$\Phi\left(\frac{\log M - \sum_{k=1}^{n} \mathcal{C}_{s_k}}{\sqrt{\sum_{k=1}^{n} \mathcal{V}_{S_k}}}\right) \quad \text{by Berry-Esseen.}$$

- Introduce the sequence of random variables $\{\mathcal{V}^{(n)}\}_{n=1}^{\infty}$ as

$$\mathcal{V}^{(n)}(S_1^{(n)}, \ldots, S_n^{(n)}) := \frac{1}{n}\sum_{k=1}^{n} \mathcal{V}_{S_k^{(n)}}.$$

Theorem

For every $\varepsilon \in (0,1)$, *we have*

$$\Lambda(\varepsilon) = \sup\left\{ r \,\middle|\, \limsup_{n\to\infty} \mathrm{E}\left[\Phi\left(\frac{nC(\varepsilon) - n\mathcal{C}^{(n)} + r\sqrt{n}}{\sqrt{n\mathcal{V}^{(n)}}}\right)\right] \le \varepsilon \right\}.$$

- Depends on the randomness of the state, not the channel.

Example 1: Compound/Mixed Channels

- Let $S \leftarrow P_S$ and $S_i^{(n)} = S$ for all $n \geq 1, i \in [n]$.

Example 1: Compound/Mixed Channels

- Let $S \leftarrow P_S$ and $S_i^{(n)} = S$ for all $n \geq 1, i \in [n]$.
- We have $\mathcal{C}^{(n)} = \mathcal{C}_S$.

Example 1: Compound/Mixed Channels

- Let $S \leftarrow P_S$ and $S_i^{(n)} = S$ for all $n \geq 1, i \in [n]$.
- We have $\mathcal{C}^{(n)} = \mathcal{C}_S$.
- The strong converse does not hold unless $\mathrm{Var}[\mathcal{C}_S] = 0$.

Example 1: Compound/Mixed Channels

- Let $S \leftarrow P_S$ and $S_i^{(n)} = S$ for all $n \geq 1, i \in [n]$.
- We have $\mathcal{C}^{(n)} = \mathcal{C}_S$.
- The strong converse does not hold unless $\mathrm{Var}[\mathcal{C}_S] = 0$.
- The ε-Capacity is given by (cf., e.g., Han'03)

$$C(\varepsilon) = \sup\{R \,|\, \Pr[\mathcal{C}_S \leq R] \leq \varepsilon\}.$$

Example 1: Compound/Mixed Channels

- Let $S \leftarrow P_S$ and $S_i^{(n)} = S$ for all $n \geq 1, i \in [n]$.
- We have $\mathcal{C}^{(n)} = \mathcal{C}_S$.
- The strong converse does not hold unless $\mathrm{Var}[\mathcal{C}_S] = 0$.
- The ε-Capacity is given by (cf., e.g., Han'03)

$$C(\varepsilon) = \sup\{R \,|\, \Pr[\mathcal{C}_S \leq R] \leq \varepsilon\}.$$

- The second-order coding rate

$$\Lambda(\varepsilon) = \sup\left\{ r \,\middle|\, \lim_{n\to\infty} \mathop{\mathrm{E}}_{S \leftarrow P_S}\left[\Phi\left(\frac{nC(\varepsilon) - n\mathcal{C}_S + \sqrt{n}r}{\sqrt{n\mathcal{V}_S}}\right)\right] \leq \varepsilon \right\}.$$

can be calculated explicitely (Polyanskiy-Poor-Verdú'11).

Example 1: Compound/Mixed Channels

- Let $S \leftarrow P_S$ and $S_i^{(n)} = S$ for all $n \geq 1, i \in [n]$.
- We have $\mathcal{C}^{(n)} = \mathcal{C}_S$.
- The strong converse does not hold unless $\mathrm{Var}[\mathcal{C}_S] = 0$.
- The ε-Capacity is given by (cf., e.g., Han'03)

$$C(\varepsilon) = \sup\{R \,|\, \Pr[\mathcal{C}_S \leq R] \leq \varepsilon\}.$$

- The second-order coding rate

$$\Lambda(\varepsilon) = \sup\left\{ r \,\middle|\, \lim_{n\to\infty} \mathop{\mathrm{E}}_{S\leftarrow P_S}\left[\Phi\left(\frac{nC(\varepsilon) - n\mathcal{C}_S + \sqrt{n}r}{\sqrt{n\mathcal{V}_S}}\right)\right] \leq \varepsilon \right\}.$$

  can be calculated explicitely (Polyanskiy-Poor-Verdú'11).
- Let $s^* = \arg\min_{s\in\mathcal{S}} \mathcal{C}_s$ be unique. For $\varepsilon < P_S(s^*)$, we have $C(\varepsilon) = \mathcal{C}_{s^*}$ and $\Lambda(\varepsilon) = \sqrt{\mathcal{V}_{s^*}}\Phi^{-1}\big(\frac{\varepsilon}{P_S(s^*)}\big)$.

Example 2: State is i.i.d.

- Suppose $S^{(n)} = (S_1, \ldots, S_n)$ is i.i.d. with $\pi \in \mathcal{P}(\mathcal{S})$ for all n.

Example 2: State is i.i.d.

- Suppose $S^{(n)} = (S_1, \ldots, S_n)$ is i.i.d. with $\pi \in \mathcal{P}(\mathcal{S})$ for all n.
- Strong converse holds since

$$\mathrm{Var}\left[\mathcal{C}^{(n)}\right] = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathrm{Cov}[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}] = \frac{1}{n} \underset{S \leftarrow \pi}{\mathrm{Var}}[\mathcal{C}_S].$$

Example 2: State is i.i.d.

- Suppose $S^{(n)} = (S_1, \ldots, S_n)$ is i.i.d. with $\pi \in \mathcal{P}(\mathcal{S})$ for all n.
- Strong converse holds since

$$\mathrm{Var}\left[\mathcal{C}^{(n)}\right] = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathrm{Cov}[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}] = \frac{1}{n} \underset{S \leftarrow \pi}{\mathrm{Var}}[\mathcal{C}_S].$$

- In fact (Wolfowitz'78)

$$C = \mathrm{E}\left[\mathcal{C}^{(n)}\right] = \underset{S \leftarrow \pi}{\mathrm{E}}[\mathcal{C}_S] = \max_{P \in \mathcal{P}(\mathcal{X}|\mathcal{S})} I(P, W|\pi).$$

Example 2: State is i.i.d.

- Suppose $S^{(n)} = (S_1, \ldots, S_n)$ is i.i.d. with $\pi \in \mathcal{P}(\mathcal{S})$ for all n.
- Strong converse holds since

$$\mathrm{Var}\left[\mathcal{C}^{(n)}\right] = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathrm{Cov}[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}] = \frac{1}{n} \underset{S \leftarrow \pi}{\mathrm{Var}}[\mathcal{C}_S].$$

- In fact (Wolfowitz'78)

$$C = \mathrm{E}\left[\mathcal{C}^{(n)}\right] = \underset{S \leftarrow \pi}{\mathrm{E}}[\mathcal{C}_S] = \max_{P \in \mathcal{P}(\mathcal{X}|\mathcal{S})} I(P, W|\pi).$$

- The second-order coding rate is

$$\Lambda(\varepsilon) = \sqrt{V_1 + V_2} \cdot \Phi^{-1}(\varepsilon), \quad \text{where}$$
$$V_1 := \underset{S \leftarrow \pi}{\mathrm{E}}[\mathcal{V}_S], \quad V_2 := \underset{S \leftarrow \pi}{\mathrm{Var}}[\mathcal{C}_S].$$

Example 2: State is i.i.d.

- Suppose $S^{(n)} = (S_1, \ldots, S_n)$ is i.i.d. with $\pi \in \mathcal{P}(\mathcal{S})$ for all n.
- Strong converse holds since

$$\mathrm{Var}\left[\mathcal{C}^{(n)}\right] = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathrm{Cov}[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}] = \frac{1}{n} \underset{S \leftarrow \pi}{\mathrm{Var}}[\mathcal{C}_S].$$

- In fact (Wolfowitz'78)

$$C = \mathrm{E}\left[\mathcal{C}^{(n)}\right] = \underset{S \leftarrow \pi}{\mathrm{E}}[\mathcal{C}_S] = \max_{P \in \mathcal{P}(\mathcal{X}|\mathcal{S})} I(P, W|\pi).$$

- The second-order coding rate is

$$\Lambda(\varepsilon) = \sqrt{V_1 + V_2} \cdot \Phi^{-1}(\varepsilon)\,, \quad \text{where}$$
$$V_1 := \underset{S \leftarrow \pi}{\mathrm{E}}[\mathcal{V}_S], \quad V_2 := \underset{S \leftarrow \pi}{\mathrm{Var}}[\mathcal{C}_S]\,.$$

- Similar to Ingber-Feder'10, but converse more challenging due to channel information at the *encoder*.

Example 2: State is i.i.d. — Proof sketch

- We want to evaluate

$$\Lambda(\varepsilon) = \sup\left\{ r \,\middle|\, \limsup_{n\to\infty} \mathrm{E}\left[\Phi\left(\frac{nC(\varepsilon)-n\mathcal{C}^{(n)}+\sqrt{n}r}{\sqrt{n\mathcal{V}^{(n)}}}\right)\right] \le \varepsilon \right\}$$

Example 2: State is i.i.d. — Proof sketch

- We want to evaluate

$$\Lambda(\varepsilon) = \sup\left\{ r \;\middle|\; \limsup_{n\to\infty} \mathrm{E}\left[\Phi\left(\frac{nC(\varepsilon) - n\mathcal{C}^{(n)} + \sqrt{n}r}{\sqrt{n\mathcal{V}^{(n)}}}\right)\right] \le \varepsilon \right\}$$

- Note that $\mathcal{C}^{(n)}$ and $\mathcal{V}^{(n)}$ are approximately Gaussian, with

$$\mathrm{E}[\mathcal{C}^{(n)}] = C,\ \mathrm{E}[\mathcal{V}^{(n)}] = V_1,\ \mathrm{Var}[\mathcal{C}^{(n)}] = \frac{V_2}{n},\ \mathrm{Var}[\mathcal{V}^{(n)}] \sim \frac{1}{n}.$$

Example 2: State is i.i.d. — Proof sketch

- We want to evaluate

$$\Lambda(\varepsilon) = \sup\left\{ r \,\middle|\, \limsup_{n\to\infty} \mathrm{E}\left[\Phi\left(\frac{nC(\varepsilon) - n\mathcal{C}^{(n)} + \sqrt{n}r}{\sqrt{n\mathcal{V}^{(n)}}}\right)\right] \leq \varepsilon \right\}$$

- Note that $\mathcal{C}^{(n)}$ and $\mathcal{V}^{(n)}$ are approximately Gaussian, with

$$\mathrm{E}[\mathcal{C}^{(n)}] = C,\ \mathrm{E}[\mathcal{V}^{(n)}] = V_1,\ \mathrm{Var}[\mathcal{C}^{(n)}] = \frac{V_2}{n},\ \mathrm{Var}[\mathcal{V}^{(n)}] \sim \frac{1}{n}.$$

- Employing a variant of Berry-Esseen, we find the following result which might be of independent interest.

Lemma

The following holds uniformly in x:

$$\mathrm{E}\left[\Phi\left(\sqrt{n}\,\frac{x - \mathcal{C}^{(n)}}{\sqrt{\mathcal{V}^{(n)}}}\right)\right] - \Phi\left(\sqrt{n}\frac{x - C}{\sqrt{V_1 + V_2}}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

Example 3: State is Markov

- $\vec{S} = \{S^n = (S_1, \ldots, S_n)\}_{n=1}^{\infty}$ evolves according to a time-homogenous, irreducible and ergodic Markov chain.

Example 3: State is Markov

- $\vec{S} = \{S^n = (S_1, \ldots, S_n)\}_{n=1}^{\infty}$ evolves according to a time-homogenous, irreducible and ergodic Markov chain.
- Such a chain admits a stationary state $\pi \in \mathcal{P}(\mathcal{S})$.

Example 3: State is Markov

- $\vec{S} = \{S^n = (S_1, \ldots, S_n)\}_{n=1}^{\infty}$ evolves according to a time-homogenous, irreducible and ergodic Markov chain.
- Such a chain admits a stationary state $\pi \in \mathcal{P}(\mathcal{S})$.
- This channel satisfies a strong converse since

$$\lim_{n\to\infty} \mathrm{E}[\mathcal{C}^{(n)}] = \mathop{\mathrm{E}}_{S\leftarrow\pi}[\mathcal{C}_S] = C, \qquad \text{and}$$

$$\mathrm{Var}[\mathcal{C}^{(n)}] = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathrm{Cov}\left[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}\right],$$

where $\mathrm{Cov}\left[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}\right] \le a e^{-b|k-l|}$ for some $a, b > 0$.

Example 3: State is Markov

- $\vec{S} = \{S^n = (S_1, \dots, S_n)\}_{n=1}^{\infty}$ evolves according to a time-homogenous, irreducible and ergodic Markov chain.
- Such a chain admits a stationary state $\pi \in \mathcal{P}(\mathcal{S})$.
- This channel satisfies a strong converse since

$$\lim_{n\to\infty} \mathrm{E}[\mathcal{C}^{(n)}] = \mathop{\mathrm{E}}_{S\leftarrow\pi}[\mathcal{C}_S] = C, \qquad \text{and}$$

$$\mathrm{Var}[\mathcal{C}^{(n)}] = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} \mathrm{Cov}\left[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}\right],$$

where $\mathrm{Cov}\left[\mathcal{C}_{S_k}, \mathcal{C}_{S_l}\right] \le a e^{-b|k-l|}$ for some $a, b > 0$.
- In fact, we see that

$$\lim_{n\to\infty} n\,\mathrm{Var}[\mathcal{C}^{(n)}] = \underbrace{\mathop{\mathrm{Var}}_{S\leftarrow\pi}[\mathcal{C}_S]}_{V_2} + \underbrace{2\sum_{k=1}^{\infty} \mathop{\mathrm{Cov}}_{S_1\leftarrow\pi}\left[\mathcal{C}_{S_1}, \mathcal{C}_{S_{1+k}}\right]}_{=:V_3} .$$

Example 3: State is Markov — Gilbert-Elliott

- The second-order coding rate depends on the mixing properties of the chain through V_3.

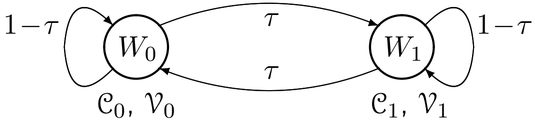
$$\Lambda(\varepsilon) = \sqrt{V_1 + V_2 + V_3}\,\Phi^{-1}(\varepsilon).$$

Example 3: State is Markov — Gilbert-Elliott

- The second-order coding rate depends on the mixing properties of the chain through V_3.

$$\Lambda(\varepsilon) = \sqrt{V_1 + V_2 + V_3}\,\Phi^{-1}(\varepsilon).$$

- E.g. states governed by a Gilbert-Eliott type Markov chain:

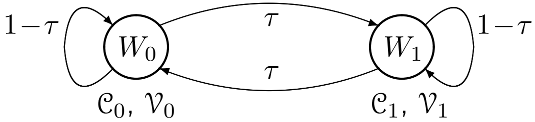


Example 3: State is Markov — Gilbert-Elliott

- The second-order coding rate depends on the mixing properties of the chain through V_3.

$$\Lambda(\varepsilon) = \sqrt{V_1 + V_2 + V_3}\, \Phi^{-1}(\varepsilon).$$

- E.g. states governed by a Gilbert-Eliott type Markov chain:



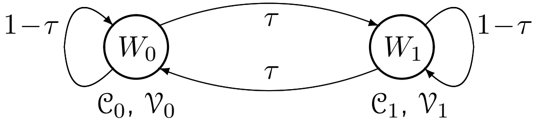
- We find $V_1 = \frac{1}{2}\mathcal{V}_0 + \frac{1}{2}\mathcal{V}_1$ and $V_2 + V_3 = \frac{1-\tau}{4\tau}(\mathcal{C}_0 - \mathcal{C}_1)^2$.

Example 3: State is Markov — Gilbert-Elliott

- The second-order coding rate depends on the mixing properties of the chain through V_3.

$$\Lambda(\varepsilon) = \sqrt{V_1 + V_2 + V_3}\,\Phi^{-1}(\varepsilon).$$

- E.g. states governed by a Gilbert-Eliott type Markov chain:



- We find $V_1 = \frac{1}{2}\mathcal{V}_0 + \frac{1}{2}\mathcal{V}_1$ and $V_2 + V_3 = \frac{1-\tau}{4\tau}(\mathcal{C}_0 - \mathcal{C}_1)^2$.
- Further specializing to W_0, W_1 BSCs, this recovers the converse result by Polyanskiy-Poor-Verdú'11.
- Additionally, we see that non-causal state information at the encoder does not help.

Conclusion

- We find the ε-capacity and conditions for the strong converse for general channels.

Conclusion

- We find the ε-capacity and conditions for the strong converse for general channels.
- We find the second-order coding rate under mild regularity assumptions on the channel.

Conclusion

- We find the ε-capacity and conditions for the strong converse for general channels.
- We find the second-order coding rate under mild regularity assumptions on the channel.
- We specialize the results to the case where the state is governed by an i.i.d. law or a Markov process.

Conclusion

- We find the ε-capacity and conditions for the strong converse for general channels.
- We find the second-order coding rate under mild regularity assumptions on the channel.
- We specialize the results to the case where the state is governed by an i.i.d. law or a Markov process.
- If the CAIDs for all W_s are equal, the (second-order) converse bounds can be achieved even if the channel state is unknown to the encoder, e.g. for the Gilbert-Elliott channel studied in Polyanskiy-Poor-Verdú'11.

Conclusion

- We find the ε-capacity and conditions for the strong converse for general channels.
- We find the second-order coding rate under mild regularity assumptions on the channel.
- We specialize the results to the case where the state is governed by an i.i.d. law or a Markov process.
- If the CAIDs for all W_s are equal, the (second-order) converse bounds can be achieved even if the channel state is unknown to the encoder, e.g. for the Gilbert-Elliott channel studied in Polyanskiy-Poor-Verdú'11.
- It would be interesting to see (but requires new techniques) if the result can be extended to general alphabets.