A Tight Upper Bound for the Third-Order Asymptotics for Most Discrete Memoryless Channels

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

Istanbul, July 10, 2013

Notation

- We consider discrete memoryless channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.

Notation

- We consider discrete memoryless channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.
- We keep $\varepsilon \in (0, 1)$ fixed.

Notation

- We consider discrete memoryless channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.
- We keep $\varepsilon \in (0,1)$ fixed.
- $M^*(W,\varepsilon)$ is the maximum size of a code for W such that the average error probability is at most ε.

Notation

- We consider discrete memoryless channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.
- We keep $\varepsilon \in (0, 1)$ fixed.
- $M^*(W, \varepsilon)$ is the maximum size of a code for W such that the average error probability is at most ε.
- For any $n \in \mathbb{N}$, W^n is the n-fold repetition of W.

Notation

- We consider discrete memoryless channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.
- We keep $\varepsilon \in (0,1)$ fixed.
- $M^*(W,\varepsilon)$ is the maximum size of a code for W such that the average error probability is at most ε.
- For any $n \in \mathbb{N}$, W^n is the n-fold repetition of W.
- We are interested in the asymptotic expansion of $\log M^*(W^n,\varepsilon)$ for large n.

Notation

- We consider discrete memoryless channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.
- We keep $\varepsilon \in (0,1)$ fixed.
- $M^*(W,\varepsilon)$ is the maximum size of a code for W such that the average error probability is at most ε.
- For any $n \in \mathbb{N}$, W^n is the n-fold repetition of W.
- We are interested in the asymptotic expansion of $\log M^*(W^n,\varepsilon)$ for large n.
- We employ the $O(\cdot)$ and $o(\cdot)$ notation, and take note that the implied constants may depend on W and ε.

Notation

- We consider discrete memoryless channels $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.
- We keep $\varepsilon \in (0,1)$ fixed.
- $M^*(W,\varepsilon)$ is the maximum size of a code for W such that the average error probability is at most ε.
- For any $n \in \mathbb{N}$, W^n is the n-fold repetition of W.
- We are interested in the asymptotic expansion of $\log M^*(W^n,\varepsilon)$ for large n.
- We employ the $O(\cdot)$ and $o(\cdot)$ notation, and take note that the implied constants may depend on W and ε.
- Φ is the cumulative (normal) Gaussian distribution.

Known Results

- Shannon's noisy coding theorem and Wolfowitz's strong converse imply

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

where $C := \max_P I(P, W)$ is the channel capacity.

Known Results

- Shannon's noisy coding theorem and Wolfowitz's strong converse imply

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

  where $C := \max_P I(P, W)$ is the channel capacity.
- Finer asymptotic characterizations are desirable to understand the backoff from capacity for finite n.

Known Results

- Define $\Pi := \arg\max_P I(P, W)$, the set of capacity-achieving input distributions.

Known Results

- Define $\Pi := \arg\max_P I(P, W)$, the set of capacity-achieving input distributions.
- The ε-channel dispersion is

$$V_\varepsilon := \begin{cases} V_{\min} & \text{if } \varepsilon < \frac{1}{2} \\ V_{\max} & \text{if } \varepsilon \geq \frac{1}{2} \end{cases}, \quad \text{where}$$

$$V_{\min} := \min_{P \in \Pi} V(P, W) \quad \text{and} \quad V_{\max} := \max_{P \in \Pi} V(P, W) .$$

Known Results

- Define $\Pi := \arg\max_P I(P, W)$, the set of capacity-achieving input distributions.
- The ε-channel dispersion is

$$V_\varepsilon := \begin{cases} V_{\min} & \text{if } \varepsilon < \frac{1}{2} \\ V_{\max} & \text{if } \varepsilon \geq \frac{1}{2} \end{cases}, \quad \text{where}$$

$$V_{\min} := \min_{P \in \Pi} V(P, W) \quad \text{and} \quad V_{\max} := \max_{P \in \Pi} V(P, W).$$

- Strassen's Gaussian approximation states that

$$\log M^*(W^n, \varepsilon) = nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \rho_n,$$

  where $\rho_n = o(\sqrt{n})$.
- See also Hayashi, using information spectrum methods.

Known Results

- Define $\Pi := \arg\max_P I(P, W)$, the set of capacity-achieving input distributions.
- The ε-channel dispersion is

$$V_\varepsilon := \begin{cases} V_{\min} & \text{if } \varepsilon < \frac{1}{2} \\ V_{\max} & \text{if } \varepsilon \geq \frac{1}{2} \end{cases}, \quad \text{where}$$

$$V_{\min} := \min_{P \in \Pi} V(P, W) \quad \text{and} \quad V_{\max} := \max_{P \in \Pi} V(P, W) .$$

- Strassen's Gaussian approximation states that

$$\log M^*(W^n, \varepsilon) = nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \rho_n,$$

where $\rho_n = o(\sqrt{n})$.

- See also Hayashi, using information spectrum methods.
- In fact, $O(1) \leq \rho_n \leq O(\log n)$ unless $\varepsilon > \frac{1}{2}$ and the channel is exotic as pointed out by Polyanskiy-Poor-Verdú (PPV).

Known Results

- Polyanskiy showed the following further refinements.
- Lower bound (achievability): If W has positive reverse dispersion (e.g. all entries of W positive) then

$$\rho_n \geq \log \sqrt{n} + O(1).$$

(The reverse dispersion is $V(PW, W^{-1})$, where P achieves the Gaussian approximation and W^{-1} is the inverse channel.)

Known Results

- Polyanskiy showed the following further refinements.
- Lower bound (achievability): If W has positive reverse dispersion (e.g. all entries of W positive) then

$$\rho_n \geq \log \sqrt{n} + O(1).$$

  (The reverse dispersion is $V(PW, W^{-1})$, where P achieves the Gaussian approximation and W^{-1} is the inverse channel.)
- Upper bounds (converse): If W is weakly input-symmetric or we restrict to constant composition codes (also Moulin), then

$$\rho_n \leq \log \sqrt{n} + O(1).$$

Known Results

- Polyanskiy showed the following further refinements.
- Lower bound (achievability): If W has positive reverse dispersion (e.g. all entries of W positive) then

$$\rho_n \geq \log \sqrt{n} + O(1).$$

  (The reverse dispersion is $V(PW, W^{-1})$, where P achieves the Gaussian approximation and W^{-1} is the inverse channel.)
- Upper bounds (converse): If W is weakly input-symmetric or we restrict to constant composition codes (also Moulin), then

$$\rho_n \leq \log \sqrt{n} + O(1).$$

- Our contribution: The above converse bound holds **for all** DMC's (unless $\varepsilon > \frac{1}{2}$ and W is exotic).

Main Result

Theorem

For every DMC W and ε with $V_\varepsilon > 0$, the blocklength n, ε-error capacity satisfies

$$\log M^*(W^n, \varepsilon) \le nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \log\sqrt{n} + O(1).$$

If $V_\varepsilon = 0$, we have $\log M^*(W^n, \varepsilon) \le nC + O(1)$, unless the channel is exotic and $\varepsilon \ge \frac{1}{2}$.

Main Result

Theorem

For every DMC W and ε with $V_\varepsilon > 0$, the blocklength n, ε-error capacity satisfies

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

If $V_\varepsilon = 0$, we have $\log M^*(W^n, \varepsilon) \leq nC + O(1)$, unless the channel is exotic and $\varepsilon \geq \frac{1}{2}$.

- This converse is tight in the third order, except for channels with vanishing reverse dispersion, e.g. the binary erasure channel.

One-Shot Converse

- A "symbol-wise" relaxation of PPV's meta-converse to a relative entropy information spectrum.

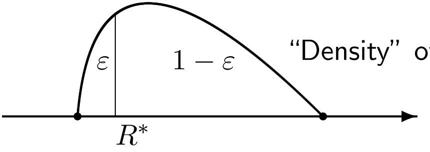
Proposition

For any $\delta \in (0, 1-\varepsilon)$, *we have*

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

- The relative entropy information spectrum is given as

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



"Density" of $\log \frac{P}{Q}$ under P

One-Shot Converse

- A "symbol-wise" relaxation of PPV's meta-converse to a relative entropy information spectrum.

Proposition

For any $\delta \in (0, 1-\varepsilon)$, *we have*

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

- Previous "symbol-wise" relaxations were limited to maximum error and/or constant composition codes (PPV).

One-Shot Converse

- A "symbol-wise" relaxation of PPV's meta-converse to a relative entropy information spectrum.

Proposition

For any $\delta \in (0, 1-\varepsilon)$, *we have*

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

- Previous "symbol-wise" relaxations were limited to maximum error and/or constant composition codes (PPV).
- The bound trivially specializes to channels with cost constraints, where the supremum is taken over $x \in \Gamma \subseteq \mathcal{X}$.

One-Shot Converse

- A "symbol-wise" relaxation of PPV's meta-converse to a relative entropy information spectrum.

Proposition

For any $\delta \in (0, 1-\varepsilon)$, *we have*

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

- Previous "symbol-wise" relaxations were limited to maximum error and/or constant composition codes (PPV).
- The bound trivially specializes to channels with cost constraints, where the supremum is taken over $x \in \Gamma \subseteq \mathcal{X}$.
- We will usually choose $\delta = \frac{1}{\sqrt{n}}$ so that the last term is $\log \sqrt{n}$.

Type-Counting Method

- For n repetition of the channel, we choose $\delta = \frac{1}{\sqrt{n}}$ and set $\varepsilon' := \varepsilon + \delta$. Then, the converse yields

$$\log M^*(W^n, \varepsilon) \leq \inf_{Q^{(n)}} \sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n) \| Q^{(n)}) + \log \sqrt{n},$$

Type-Counting Method

- For n repetition of the channel, we choose $\delta = \frac{1}{\sqrt{n}}$ and set $\varepsilon' := \varepsilon + \delta$. Then, the converse yields

$$\log M^*(W^n, \varepsilon) \leq \inf_{Q^{(n)}} \sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n) \| Q^{(n)}) + \log \sqrt{n},$$

- Task: Find suitable output distribution $Q^{(n)}$ such that the minimax term is bounded by the Gaussian approximation.

Type-Counting Method

- For n repetition of the channel, we choose $\delta = \frac{1}{\sqrt{n}}$ and set $\varepsilon' := \varepsilon + \delta$. Then, the converse yields

$$\log M^*(W^n, \varepsilon) \le \inf_{Q^{(n)}} \sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n) \| Q^{(n)}) + \log \sqrt{n},$$

- Task: Find suitable output distribution $Q^{(n)}$ such that the minimax term is bounded by the Gaussian approximation.
- Hayashi uses a convex combination of "output types", e.g.

$$Q^{(n)}(y^n) = \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \prod_{i=1}^{n} PW(y_i).$$

Type-Counting Method

- For n repetition of the channel, we choose $\delta = \frac{1}{\sqrt{n}}$ and set $\varepsilon' := \varepsilon + \delta$. Then, the converse yields

$$\log M^*(W^n, \varepsilon) \leq \inf_{Q^{(n)}} \sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^{(n)}) + \log\sqrt{n},$$

- Task: Find suitable output distribution $Q^{(n)}$ such that the minimax term is bounded by the Gaussian approximation.
- Hayashi uses a convex combination of "output types", e.g.

$$Q^{(n)}(y^n) = \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \prod_{i=1}^{n} PW(y_i).$$

- Then the converse relaxes to

$$\log M^*(W^n, \varepsilon) \leq \sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n)\|(P_{x^n}W)^n) + \log\sqrt{n} + \log|\mathcal{P}_n(\mathcal{X})|,$$

where P_{x^n} is the type (empirical distribution) of x^n.

Type-Counting Method

- Using the Berry-Esseen theorem, it is easy to see that

$$\begin{aligned}&\sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n)\|(P_{x^n}W)^n)\\&\qquad = \sup_{x^n} nI(P_{x^n}, W) + \sqrt{nV(P_{x^n}, W)}\Phi^{-1}(\varepsilon') + O(1)\\&\qquad \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + O(1)\end{aligned}$$

The last step assumes regularity, i.e. $V_\varepsilon > 0$, see PPV.

Type-Counting Method

- Using the Berry-Esseen theorem, it is easy to see that

$$\begin{aligned}\sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n)\|(P_{x^n}W)^n) &= \sup_{x^n} nI(P_{x^n},W) + \sqrt{nV(P_{x^n},W)}\Phi^{-1}(\varepsilon') + O(1) \\ &\leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + O(1)\end{aligned}$$

  The last step assumes regularity, i.e. $V_\varepsilon > 0$, see PPV.
- We also have $\log|\mathcal{P}_n(\mathcal{X})| \leq (|\mathcal{X}|-1)\log n + o(1)$.

Type-Counting Method

- Using the Berry-Esseen theorem, it is easy to see that

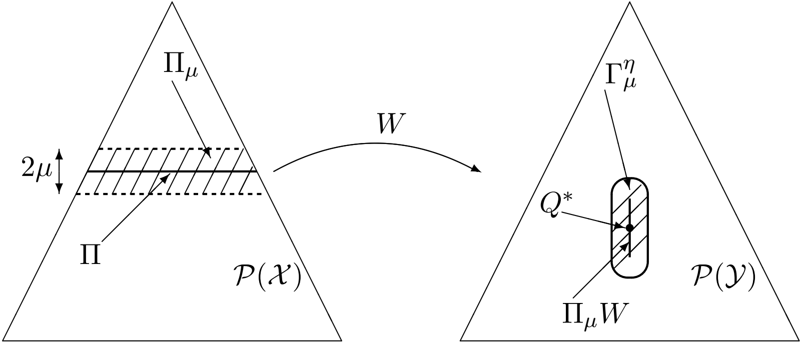
$$\sup_{x^n} D_s^{\varepsilon'}(W^n(\cdot|x^n)\|(P_{x^n}W)^n)$$
$$= \sup_{x^n} nI(P_{x^n}, W) + \sqrt{nV(P_{x^n}, W)}\Phi^{-1}(\varepsilon') + O(1)$$
$$\leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + O(1)$$

  The last step assumes regularity, i.e. $V_\varepsilon > 0$, see PPV.
- We also have $\log|\mathcal{P}_n(\mathcal{X})| \leq (|\mathcal{X}|-1)\log n + o(1)$.
- Combining these upper bounds, we find

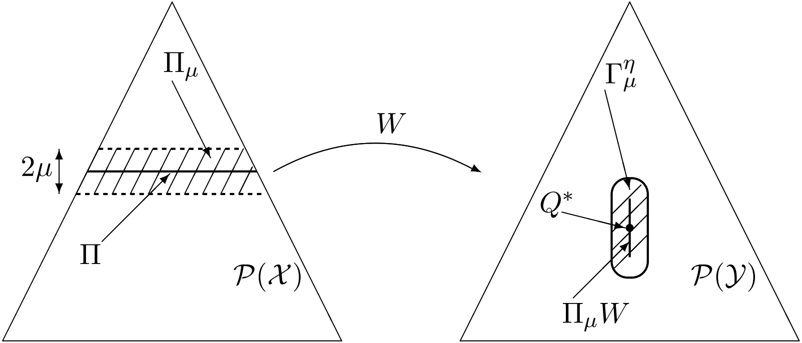
$$\log M^*(W^n, \varepsilon) \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + \log\sqrt{n} + \big(|\mathcal{X}|-1\big)\log n + O(1).$$

- This recovers the converse bounds by PPV (and Strassen) for general DMCs.

Approximately Capacity-Achieving Types

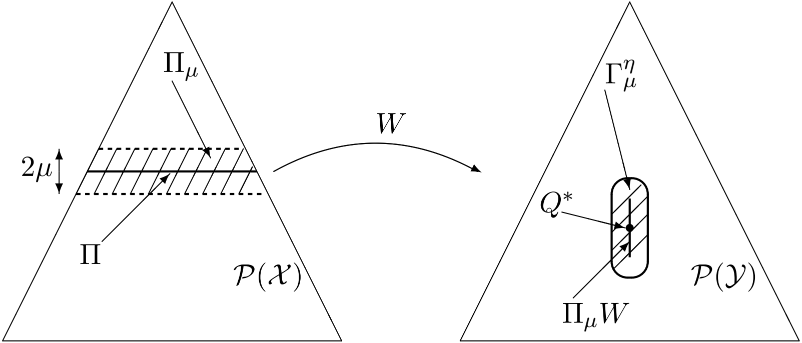


Approximately Capacity-Achieving Types



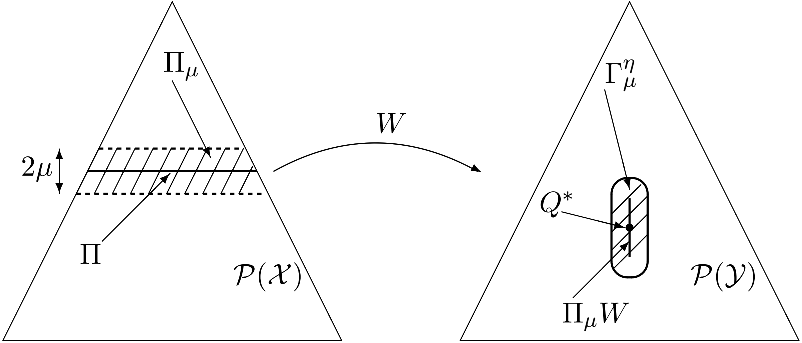
- For types $P_{x^n} \notin \Pi_\mu$, the mutual information $I(P_{x^n}, W)$ is bounded away from C.

Approximately Capacity-Achieving Types



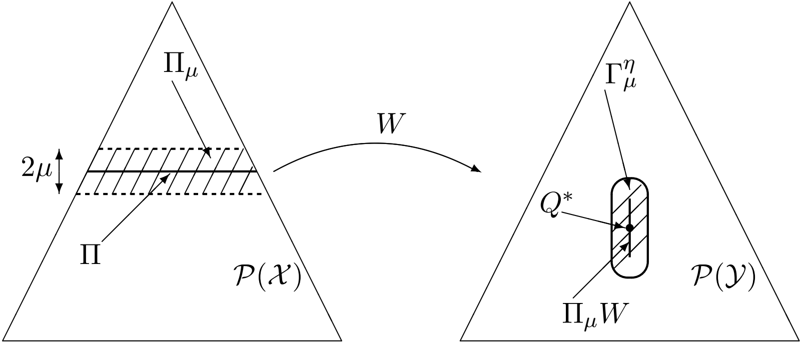
- For types $P_{x^n} \notin \Pi_\mu$, the mutual information $I(P_{x^n}, W)$ is bounded away from C.
- Thus, $D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^{(n)}) \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + O(1)$.

Approximately Capacity-Achieving Types



- For types $P_{x^n} \notin \Pi_\mu$, the mutual information $I(P_{x^n}, W)$ is bounded away from C.
- Thus, $D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^{(n)}) \le nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + O(1)$.
- But we need a different technique for $P_{x^n} \in \Pi_\mu$!

Approximately Capacity-Achieving Types



- For types $P_{x^n} \notin \Pi_\mu$, the mutual information $I(P_{x^n}, W)$ is bounded away from C.
- Thus, $D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^{(n)}) \leq nC + \sqrt{nV_\varepsilon}\Phi^{-1}(\varepsilon) + O(1)$.
- But we need a different technique for $P_{x^n} \in \Pi_\mu$!
- Our approach: Span an ϵ-net over Γ_μ^η.

An ϵ-Net of Output Distributions

- We choose the following output distribution:

$$Q^{(n)}(y^n) := \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \prod_{i=1}^{n} PW(y_i) + \frac{1}{2} \sum_{\mathbf{k} \in \mathcal{K}} \frac{\exp\left(-\gamma \|\mathbf{k}\|_2^2\right)}{F(n)} \prod_{i=1}^{n} Q_{\mathbf{k}}(y_i)$$

where $F(n)$ is used for normalization, $\gamma > 0$, and

$$Q_{\mathbf{k}}(y) := Q^*(y) + \frac{k_y}{\sqrt{n}}$$

An ϵ-Net of Output Distributions

- We choose the following output distribution:

$$Q^{(n)}(y^n) := \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \prod_{i=1}^{n} PW(y_i) + \frac{1}{2} \sum_{\mathbf{k} \in \mathcal{K}} \frac{\exp\left(-\gamma \|\mathbf{k}\|_2^2\right)}{F(n)} \prod_{i=1}^{n} Q_{\mathbf{k}}(y_i)$$

where $F(n)$ is used for normalization, $\gamma > 0$, and

$$Q_{\mathbf{k}}(y) := Q^*(y) + \frac{k_y}{\sqrt{n}}$$

- The net get denser as n increases.

An ϵ-Net of Output Distributions

- We choose the following output distribution:

$$Q^{(n)}(y^n) := \frac{1}{2} \sum_{P \in \mathcal{P}_n(\mathcal{X})} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \prod_{i=1}^{n} PW(y_i) + \frac{1}{2} \sum_{\mathbf{k} \in \mathcal{K}} \frac{\exp\left(-\gamma \|\mathbf{k}\|_2^2\right)}{F(n)} \prod_{i=1}^{n} Q_{\mathbf{k}}(y_i)$$

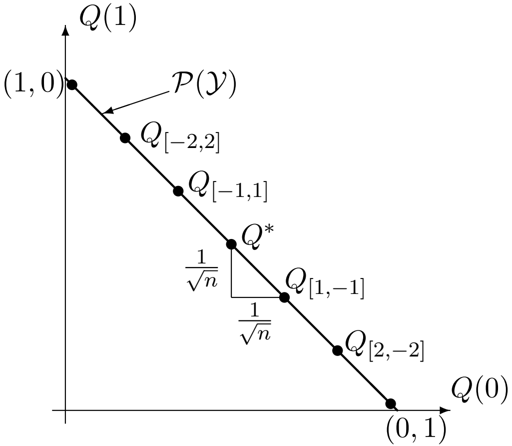
where $F(n)$ is used for normalization, $\gamma > 0$, and

$$Q_{\mathbf{k}}(y) := Q^*(y) + \frac{k_y}{\sqrt{n}}$$

- The net get denser as n increases.
- Due to the exponential damping, $F(n)$ can be bounded by a constant independent of n.

An ϵ-Net of Output Distributions

- For every $P \in \mathcal{P}(\mathcal{X})$, $\exists\,\mathbf{k}$ with $\|Q_{\mathbf{k}} - PW\|_2 \le O(\frac{1}{\sqrt{n}})$.



Almost-Capacity-Achieving Types

- For each $P_{x^n} \in \Pi_\mu$, fix $\mathbf{k}$ such that $\|Q_{\mathbf{k}} - P_{x^n} W\|_2 \leq O(\frac{1}{\sqrt{n}})$.

Almost-Capacity-Achieving Types

- For each $P_{x^n} \in \Pi_\mu$, fix $\mathbf{k}$ such that $\|Q_{\mathbf{k}} - P_{x^n} W\|_2 \leq O(\frac{1}{\sqrt{n}})$.
- Then, using various continuity properties

$$
\begin{aligned}
&D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^{(n)}) \\
&\quad \leq D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q_{\mathbf{k}}^n) + \gamma\|\mathbf{k}\|_2^2 + O(1) \\
&\quad \leq nI(P_{x^n}, W) + \sqrt{nV(P_{x^n}, W)}\Phi^{-1}(\varepsilon) + \gamma\|\mathbf{k}\|_2^2 + O(1).
\end{aligned}
$$

Almost-Capacity-Achieving Types

- For each $P_{x^n} \in \Pi_\mu$, fix $\mathbf{k}$ such that $\|Q_{\mathbf{k}} - P_{x^n}W\|_2 \leq O(\frac{1}{\sqrt{n}})$.
- Then, using various continuity properties

$$
\begin{aligned}
&D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^{(n)}) \\
&\quad \leq D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^n_{\mathbf{k}}) + \gamma\|\mathbf{k}\|_2^2 + O(1) \\
&\quad \leq nI(P_{x^n}, W) + \sqrt{nV(P_{x^n}, W)}\Phi^{-1}(\varepsilon) + \gamma\|\mathbf{k}\|_2^2 + O(1).
\end{aligned}
$$

- We also find

$$
\begin{aligned}
\|\mathbf{k}\|_2 &= \sqrt{n}\|Q_{\mathbf{k}} - Q^*\|_2 \\
&\leq \sqrt{n}\|P_{x^n}W - Q^*\|_2 + O(1) \\
&\leq \sqrt{n}\|W\|_2\|P_{x^n} - \Pi\|_2 + O(1)
\end{aligned}
$$

Almost-Capacity-Achieving Types

- For each $P_{x^n} \in \Pi_\mu$, fix $\mathbf{k}$ such that $\|Q_{\mathbf{k}} - P_{x^n} W\|_2 \le O(\frac{1}{\sqrt{n}})$.
- Then, using various continuity properties

$$
\begin{aligned}
& D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q^{(n)}) \\
& \quad \le D_s^{\varepsilon'}(W^n(\cdot|x^n)\|Q_{\mathbf{k}}^n) + \gamma\|\mathbf{k}\|_2^2 + O(1) \\
& \quad \le nI(P_{x^n}, W) + \sqrt{nV(P_{x^n}, W)}\Phi^{-1}(\varepsilon) + \gamma\|\mathbf{k}\|_2^2 + O(1).
\end{aligned}
$$

- We also find

$$
\begin{aligned}
\|\mathbf{k}\|_2 &= \sqrt{n}\|Q_{\mathbf{k}} - Q^*\|_2 \\
&\le \sqrt{n}\|P_{x^n} W - Q^*\|_2 + O(1) \\
&\le \sqrt{n}\|W\|_2\|P_{x^n} - \Pi\|_2 + O(1)
\end{aligned}
$$

- This term can be compensated for since we have, for $x^n \in \Pi_\mu$,

$$
I(P_{x^n}, W) \le C - \alpha\|P_{x^n} - \Pi\|_2^2
$$

for some $\alpha > 0$.

Conclusion

- We prove a converse bound to show that the logarithmic contribution to the coding rate does not depend on the alphabet size.

Conclusion

- We prove a converse bound to show that the logarithmic contribution to the coding rate does not depend on the alphabet size.
- Combining our results with PPV, we find ($\varepsilon < \frac{1}{2}$):

dispersion	reverse dispersion	bounds
> 0	> 0	$\rho_n = \log \sqrt{n} + O(1)$
$= 0$	≥ 0	$\rho_n = O(1)$
> 0	$= 0$	$O(1) \leq \rho_n \leq \log \sqrt{n} + O(1)$

Conclusion

- We prove a converse bound to show that the logarithmic contribution to the coding rate does not depend on the alphabet size.
- Combining our results with PPV, we find ($\varepsilon < \frac{1}{2}$):

dispersion	reverse dispersion	bounds
> 0	> 0	$\rho_n = \log \sqrt{n} + O(1)$
$= 0$	≥ 0	$\rho_n = O(1)$
> 0	$= 0$	$O(1) \leq \rho_n \leq \log \sqrt{n} + O(1)$

- We know that $\rho_n = O(1)$ for the binary erasure channel (see, e.g., PPV).

Conclusion

- We prove a converse bound to show that the logarithmic contribution to the coding rate does not depend on the alphabet size.
- Combining our results with PPV, we find ($\varepsilon < \frac{1}{2}$):

dispersion	reverse dispersion	bounds
> 0	> 0	$\rho_n = \log\sqrt{n} + O(1)$
$= 0$	≥ 0	$\rho_n = O(1)$
> 0	$= 0$	$O(1) \leq \rho_n \leq \log\sqrt{n} + O(1)$

- We know that $\rho_n = O(1)$ for the binary erasure channel (see, e.g., PPV).
- It remains to find the third order term for other channels with vanishing reverse dispersion (and non-vanishing dispersion).