Refined Bounds on the Empirical Distribution of Good Channel Codes via Concentration Inequalities

Maxim Raginsky and **Igal Sason**

ISIT 2013, Istanbul, Turkey

Capacity-Achieving Channel Codes

The set-up

- DMC $T : \mathcal{X} \to \mathcal{Y}$ with capacity

$$C = C(T) = \max_{P_X} I(X;Y)$$

- (n, M)-code: $\mathcal{C} = (f, g)$ with encoder $f : \{1, \ldots, M\} \to \mathcal{X}^n$ and decoder $g : \mathcal{Y}^n \to \{1, \ldots, M\}$

Capacity-achieving codes:

A sequence $\{\mathcal{C}_n\}_{n=1}^{\infty}$, where each $\mathcal{C}_n$ is an (n, M_n)-code, is capacity-achieving if

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = C.$$

Capacity-Achieving Channel Codes

Capacity-achieving input and output distributions:

$$P_X^* \in \arg\max_{P_X} I(X;Y) \qquad \text{(may not be unique)}$$

$$P_X^* \xrightarrow{\quad T \quad} P_Y^* \qquad \text{(always unique)}$$

Theorem **(Shamai–Verdú, 1997)** Let $\{\mathcal{C}_n\}$ be any capacity-achieving code sequence with vanishing error probability. Then

$$\lim_{n\to\infty} \frac{1}{n} D\left(P_{Y^n}^{(\mathcal{C}_n)} \middle\| P_{Y^n}^*\right) = 0,$$

where $P_{Y^n}^{(\mathcal{C}_n)}$ is the output distribution induced by the code $\mathcal{C}_n$ when the messages in $\{1, \ldots, M_n\}$ are equiprobable.

Capacity-Achieving Channel Codes

$$\lim_{n\to\infty} \frac{1}{n} D(P_{Y^n} \| P^*_{Y^n}) = 0$$

Main message: channel output sequences induced by good code "resemble" i.i.d. sequences drawn from the CAOD P^*_Y

Useful implications: estimate performance characteristics of good channel codes by their expectations w.r.t. $P^*_{Y^n} = (P^*_Y)^n$

- often much easier to compute explicitly
- bound estimation accuracy using large-deviation theory (e.g., Sanov's theorem)

Question: what about good codes with nonvanishing error probability?

Codes with Nonvanishing Error Probability

Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability" (2012)

1. Let $\mathcal{C} = (f, g)$ be any (n, M, ε)-code for T:

$$\max_{1 \le j \le M} \mathbb{P}\left(g(Y^n) \neq j \middle| f(X^n) = j\right) \le \varepsilon.$$

   Then $D\left(P_{Y^n}^{(\mathcal{C})} \middle\| P_{Y^n}^*\right) \le nC - \log M + o(n)$.*

2. If $\{\mathcal{C}_n\}_{n=1}^{\infty}$ is a capacity-achieving sequence, where each $\mathcal{C}_n$ is an (n, M_n, ε)-code for some fixed $\varepsilon > 0$, then

$$\lim_{n \to \infty} \frac{1}{n} D\left(P_{Y^n}^{(\mathcal{C}_n)} \middle\| P_{Y^n}^*\right) = 0.$$

* In some cases, the $o(n)$ term can be improved to $O(\sqrt{n})$.

Codes with Nonvanishing Error Probability

$$D(P_{Y^n} \| P^*_{Y^n}) \le nC - \log M + o(n)$$

The same message: channel output sequences induced by good codes “resemble” i.i.d. sequences drawn from P^*_Y

Main technical tool: concentration of measure

Our contribution: sharpening of the Polyanskiy–Verdú bounds by identifying explicit expressions for the $o(n)$ term

Preliminaries on Concentration of Measure

Let $Z_1, \ldots, Z_n \in \mathcal{Z}$ be independent random variables. We seek tight bounds on the deviation probabilities

$$\mathbb{P}\left(f(Z^n) \geq r\right) \qquad \text{for } r > 0$$

where $f : \mathcal{Z}^n \to \mathbb{R}$ is some function with $\mathbb{E}[f(Z^n)] = 0$.

Subgaussian tails:

$$\log \mathbb{E}[e^{tf(Z^n)}] \leq \kappa t^2/2, \ \forall t > 0$$
$$\implies \quad \mathbb{P}\left(f(Z^n) \geq r\right) \leq \exp\left(-\frac{r^2}{2\kappa}\right), \ \forall r > 0$$

Preliminaries on Concentration of Measure

Suppose that $\mathcal{Z}^n$ is a metric space with metric $d(\cdot, \cdot)$.

L_1 **Wasserstein distance:** for any $\mu, \nu \in \mathcal{P}(\mathcal{Z}^n)$,

$$W_1(\mu, \nu) \triangleq \inf_{Z^n \sim \mu, \bar{Z}^n \sim \nu} \mathbb{E}[d(Z^n, \bar{Z}^n)]$$

L_1 **transportation cost inequalities (Marton)**: $\mu \in \mathcal{P}(\mathcal{Z}^n)$ satisfies a $\mathrm{T}_1(c)$ inequality if

$$W_1(\mu, \nu) \leq \sqrt{2cD(\nu \| \mu)}, \qquad \forall \nu \ll \mu$$

$T_1(c)$ implies concentration!

Preliminaries on Concentration of Measure

L_1 **transportation cost inequalities** **(Marton)**: $\mu \in \mathcal{P}(\mathcal{Z}^n)$ satisfies a $\mathrm{T}_1(c)$ inequality if

$$W_1(\mu, \nu) \le \sqrt{2cD(\nu \| \mu)}, \qquad \forall \nu \ll \mu$$

Theorem (Bobkov–Götze, 1999) A probability measure $\mu \in \mathcal{P}(\mathcal{Z}^n)$ satisfies $\mathrm{T}_1(c)$ if and only if

$$\log \mathbb{E}_\mu[e^{tf(Z^n)}] \le ct^2/2$$

for all f with $\mathbb{E}_\mu[f(Z^n)] = 0$ and

$$\|f\|_{\mathrm{Lip}} \triangleq \sup_{z^n \neq \bar{z}^n} \frac{|f(z^n) - f(\bar{z}^n)|}{d(z^n, \bar{z}^n)} \le 1$$

Preliminaries on Concentration of Measure

Endow $\mathcal{Z}^n$ with the weighted Hamming metric

$$d(z^n, \bar{z}^n) = \sum_{i=1}^{n} c_i \mathbf{1}_{\{z_i \neq \bar{z}_i\}}, \text{ for some fixed } c_1, \ldots, c_n > 0.$$

Marton's coupling argument: Any product measure $\mu = \mu_1 \otimes \ldots \otimes \mu_n \in \mathcal{P}(\mathcal{Z}^n)$ satisfies $\mathrm{T}_1(c)$ (relative to d) with

$$c = \frac{1}{4} \sum_{i=1}^{n} c_i^2$$

By Bobkov–Götze, this is equivalent to the subgaussian property

$$\log \mathbb{E}_\mu \left[e^{tf(Z^n)} \right] \leq \frac{t^2}{8} \sum_{i=1}^{n} c_i^2$$

for any $f : \mathcal{Z}^n \to \mathbb{R}$ with $\mathbb{E}_\mu f = 0$ and $\|f\|_{\mathrm{Lip}} \leq 1$ (another way to derive McDiarmid's inequality)

Relative Entropy at the Output of a Code

Consider a DMC $T : \mathcal{X} \to \mathcal{Y}$ with $T(\cdot|\cdot) > 0$, and let

$$c(T) = 2 \max_{x \in \mathcal{X}} \max_{y, y' \in \mathcal{Y}} \left| \ln \frac{T(y|x)}{T(y'|x)} \right|$$

Theorem. Any (n, M, ε)-code $\mathcal{C}$ for T, where $\varepsilon \in (0, 1/2)$, satisfies

$$D\Big(P_{Y^n}^{(\mathcal{C})} \Big\| P_{Y^n}^{*}\Big) \le nC - \log M + \log \frac{1}{\varepsilon} + c(T) \sqrt{\frac{n}{2} \log \frac{1}{1 - 2\varepsilon}}$$

Remark: Polyanskiy and Verdú show that

$$D\Big(P_{Y^n}^{(\mathcal{C})} \Big\| P_{Y^n}^{*}\Big) \le nC - \log M + a\sqrt{n}$$

for some constant $a = a(\varepsilon)$

Proof Idea: I

Fix $x^n \in \mathcal{X}^n$ and study concentration of the function

$$h_{x^n}(y^n) = \log \frac{\mathrm{d}P_{Y^n|X^n=x^n}}{\mathrm{d}P^{(\mathcal{C})}_{Y^n}}(y^n)$$

around its expectation w.r.t. $P_{Y^n|X^n=x^n}$:

$$\mathbb{E}[h_{x^n}(Y^n)|X^n = x^n] = D\left(P_{Y^n|X^n=x^n} \middle\| P^{(\mathcal{C})}_{Y^n}\right)$$

Step 1: Because $T(\cdot|\cdot) > 0$, the function $h_{x^n}(y^n)$ is 1-Lipschitz w.r.t. scaled Hamming metric

$$d(y^n, \bar{y}^n) = c(T) \sum_{i=1}^{n} \mathbf{1}_{\{y_i \neq \bar{y}_i\}}$$

Proof Idea: II

Step 1: Because $T(\cdot|\cdot) > 0$, the function $h_{x^n}(y^n)$ is 1-Lipschitz w.r.t. scaled Hamming metric

$$d(y^n, \bar{y}^n) = c(T) \sum_{i=1}^{n} \mathbf{1}_{\{y_i \neq \bar{y}_i\}}$$

Step 2: Any product probability measure μ on $(\mathcal{Y}^n, d)$ satisfies

$$\log \mathbb{E}_\mu \left[e^{tf(Y^n)} \right] \le \frac{nc(T)^2 t^2}{8}$$

for any f with $\mathbb{E}_\mu f = 0$ and $\|F\|_{\mathrm{Lip}} \le 1$.

Proof: tensorization of the Csiszár–Kullback–Pinsker inequality, followed by appeal to Bobkov–Götze.

Proof Idea: III

$$h_{x^n}(y^n) = \log \frac{\mathrm{d}P_{Y^n|X^n=x^n}}{\mathrm{d}P_{Y^n}^{(\mathcal{C})}}(y^n)$$

$$\mathbb{E}[h_{x^n}(Y^n)|X^n = x^n] = D\left(P_{Y^n|X^n=x^n} \middle\| P_{Y^n}^{(\mathcal{C})}\right)$$

Step 3: For any x^n, $\mu = P_{Y^n|X^n=x^n}$ is a product measure, so

$$\mathbb{P}\left(h_{x^n}(Y^n) \geq D\left(P_{Y^n|X^n=x^n} \middle\| P_{Y^n}^{(\mathcal{C})}\right) + r\right) \leq \exp\left(-\frac{2r^2}{nc(T)^2}\right)$$

Use this with $r = c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}}$:

$$\mathbb{P}\left(h_{x^n}(Y^n) \geq D\left(P_{Y^n|X^n=x^n} \middle\| P_{Y^n}^{(\mathcal{C})}\right) + c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}}\right) \leq 1 - 2\varepsilon$$

Remark: Polyanskiy–Verdú show $\mathrm{Var}[h_{x^n}(Y^n)|X^n = x^n] = O(n)$.

Proof Idea: IV

Recall:

$$\mathbb{P}\left(h_{x^n}(Y^n) \geq D\left(P_{Y^n|X^n=x^n}\middle\|P_{Y^n}^{(\mathcal{C})}\right) + c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}}\right) \leq 1-2\varepsilon$$

Step 4: Same as Polyanskiy–Verdú, appeal to Augustin's strong converse to get

$$\log M \leq \log\frac{1}{\varepsilon} + D\left(P_{Y^n|X^n}\middle\|P_{Y^n}^{(\mathcal{C})}\middle|P_{X^n}^{(\mathcal{C})}\right) + c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}}$$

$$\begin{aligned} D\left(P_{Y^n}^{(\mathcal{C})}\middle\|P_{Y^n}^{*}\right) &= D\left(P_{Y^n|X^n}\middle\|P_{Y^n}^{*}\middle|P_{X^n}^{(\mathcal{C})}\right) - D\left(P_{Y^n|X^n}\middle\|P_{Y^n}^{(\mathcal{C})}\middle|P_{X^n}^{(\mathcal{C})}\right) \\ &\leq nC - \log M + \log\frac{1}{\varepsilon} + c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}} \qquad \blacksquare \end{aligned}$$

Relative Entropy at the Output of a Code

Theorem. Let $T : \mathcal{X} \to \mathcal{Y}$ be a DMC with $C > 0$. Then, for any $0 < \varepsilon < 1$, any (n, M, ε)-code $\mathcal{C}$ for T satisfies

$$D\big(P_{Y^n}^{(\mathcal{C})} \big\| P_{Y^n}^{*}\big) \leq nC - \log M$$
$$+ \sqrt{2n}\,(\log n)^{3/2} \left(1 + \sqrt{\frac{1}{\log n}\,\log\left(\frac{1}{1-\varepsilon}\right)}\right) \left(1 + \frac{\log|\mathcal{Y}|}{\log n}\right)$$
$$+ 3\log n + \log\big(2|\mathcal{X}||\mathcal{Y}|^2\big).$$

Remark: Polyanskiy and Verdú show that

$$D\big(P_{Y^n}^{(\mathcal{C})} \big\| P_{Y^n}^{*}\big) \leq nC - \log M + b\sqrt{n}\log^{3/2} n$$

for some constant $b > 0$.

Concentration of Lipschitz Functions

Theorem. Let $T : \mathcal{X} \to \mathcal{Y}$ be a DMC with $c(T) < \infty$. Let $d : \mathcal{Y}^n \times \mathcal{Y}^n \to \mathbb{R}_+$ be a metric, and suppose that $P_{Y^n|X^n=x^n}$, $x^n \in \mathcal{X}^n$, as well as $P^*_{Y^n}$, satisfy $\mathrm{T}_1(c)$ for some $c > 0$.

Then, for any $\varepsilon \in (0, 1/2)$, any (n, M, ε)-code $\mathcal{C}$ for T, and any function $f : \mathcal{Y}^n \to \mathbb{R}$ we have

$$P^{(\mathcal{C})}_{Y^n}\Big(|f(Y^n) - \mathbb{E}[f(Y^{*n})]| \geq r \Big)$$
$$\leq \frac{4}{\varepsilon} \, \exp\left(nC - \ln M + a\sqrt{n} - \frac{r^2}{8c\|f\|^2_{\mathrm{Lip}}} \right), \; \forall\, r \geq 0$$

where $Y^{*n} \sim P^*_{Y^n}$, and $a \triangleq c(T)\sqrt{\frac{1}{2} \ln \frac{1}{1-2\varepsilon}}$.

Proof Idea

Step 1: For each $x^n \in \mathcal{X}^n$, let $\phi(x^n) \triangleq \mathbb{E}[f(Y^n)|X^n = x^n]$. Then, by Bobkov–Götze,

$$\mathbb{P}\Big(|f(Y^n) - \phi(x^n)| \geq r \Big| X^n = x^n \Big) \leq 2 \exp\left(-\frac{r^2}{2c\|f\|_{\mathrm{Lip}}^2} \right)$$

Step 2: By restricting to a subcode $\mathcal{C}'$ with codewords $x^n \in \mathcal{X}^n$ satisfying $\phi(x^n) \geq \mathbb{E}[f(Y^{*n})] + r$, we can show that

$$r \leq \|f\|_{\mathrm{Lip}} \sqrt{2c\left(nC - \log M' + a\sqrt{n} + \log\frac{1}{\varepsilon}\right)},$$

with $M' = M P_{X^n}^{(\mathcal{C})}\Big(\phi(X^n) \geq \mathbb{E}[f(Y^{*n})] + r\Big)$. Solve to get

$$P_{X^n}^{(\mathcal{C})}\Big(|\phi(X^n) - \mathbb{E}[f(Y^{*n})]| \geq r \Big) \leq 2e^{nC - \log M + a\sqrt{n} + \log\frac{1}{\varepsilon} - \frac{r^2}{2c\|f\|_{\mathrm{Lip}}^2}}$$

Step 3: Apply union bound. ■

Empirical Averages at the Code Output

- Equip $\mathcal{Y}^n$ with the Hamming metric

$$d(y^n, \bar{y}^n) = \sum_{i=1}^{n} \mathbf{1}_{\{y_i \neq \bar{y}_i\}}$$

- Consider functions of the form

$$f(y^n) = \frac{1}{n} \sum_{i=1}^{n} f_i(y_i),$$

  where $|f_i(y_i) - f_i(\bar{y}_i)| \leq L\mathbf{1}_{\{y_i \neq \bar{y}_i\}}$ for all $i, y_i, \bar{y}_i$. Then $\|f\|_{\mathrm{Lip}} \leq L/n$.
- Since $P_{Y^n|X^n=x^n}$ for all x^n and $P^*_{Y^n}$ are product measures on $\mathcal{Y}^n$, they all satisfy $\mathrm{T}_1(n/4)$ (by tensorization)
- Therefore, for any (n, M, ε)-code and any such f we have

$$P^{(\mathcal{C})}_{Y^n}\big(\,|f(Y^n) - \mathbb{E}[f(Y^{*n})]| \geq r\big) \leq \frac{4}{\varepsilon}\,\exp\left(nC - \log M + a\sqrt{n} - \frac{nr^2}{2L^2}\right)$$

Operational Significance

- A bound like

$$C(T) - \log M \geq \frac{1}{n} D\big(P_{Y^n}^{(\mathcal{C})} \big\| P_{Y^n}^{*}\big) - \frac{a(T, \varepsilon)}{\sqrt{n}}$$

  quantifies trade-offs between minimal blocklength required for achieving a certain gap (in rate) to capacity with a fixed block error probability ε, and normalized divergence between output distribution induced by the code and the (unique) CAOD of the channel
- We have identified the precise dependence of $a(T, \varepsilon)$ on the channel T and on the block error probability ε
- These results are similar to a lower bound on rate loss w.r.t. fully random block codes (whose average distance spectrum is binomially distributed) in terms of normalized divergence between the distance spectrum of a specific code and the binomial distribution (Shamai–Sason, 2002).

Concentration of measure = powerful tool for studying nonasymptotic behavior of stochastic objects in information theory!

For more information, see M. Raginsky and I. Sason, "Concentration of Measure Inequalities in Information Theory, Communications and Coding," `arXiv:1212.4663`

That’s All, Folks!