Information Theory in the Finite Blocklength Regime

Maël Le Treust
ETIS UMR 8051, CY Cergy Paris Université, ENSEA, CNRS,
6, avenue du Ponceau,
95014 Cergy-Pontoise CEDEX, France
Email: mael.le-treust@ensea.fr

I. Shannon's Entropy and Mutual Information

We consider two discrete sets $\mathcal{X}$, $\mathcal{Y}$ and we denote by

- $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ the pair of random variables,
- $(x, y) \in \mathcal{X} \times \mathcal{Y}$ a pair of realizations,
- $\mathcal{P}_X$ a probability distribution over the set $\mathcal{X}$, i.e. for all $x \in \mathcal{X}$, $\mathcal{P}_X(x) = \Pr(X = x) \in [0, 1]^{\mathbb{R}^{|\mathcal{X}|}}$ and $\sum_{x\in\mathcal{X}} \mathcal{P}_X(x) = 1$,
- $\Delta(\mathcal{X})$ the set of probability distributions $\mathcal{Q}_X$ over $\mathcal{X}$, i.e. the probability simplex,
- $\operatorname{supp} \mathcal{Q}_X = \{x \in \mathcal{X},\ \mathcal{Q}_X(x) > 0\}$ the support of the distribution $\mathcal{P}_X$,
- $\mathcal{P}_{XY} \in \Delta(\mathcal{X} \in \mathcal{Y})$ a joint distribution with marginal distributions $\mathcal{P}_X$, $\mathcal{P}_Y$, i.e. $\mathcal{P}_X(x) = \sum_y \mathcal{P}_{XY}(x, y)$, and with conditional distributions $\mathcal{P}_{X|Y} \in \Delta(\mathcal{X})^{|\mathcal{Y}|}$, $\mathcal{P}_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$, i.e. $\forall (x, y) \in \operatorname{supp} \mathcal{P}_X \times \mathcal{Y}$, $\mathcal{P}_{Y|X}(y|x) = \frac{\mathcal{P}_{XY}(x,y)}{\mathcal{P}_X(x)}$.

Definition 1 *Let $X \in \mathcal{X}$ a random variable with probability distribution $\mathcal{P}_X \in \Delta(\mathcal{X})$. The entropy is defined by*

$$H(X) = \sum_{x\in\operatorname{supp}\mathcal{P}_X} \mathcal{P}_X(x) \log_2 \frac{1}{\mathcal{P}_X(x)}. \tag{1}$$

We define the function $f : \operatorname{supp} \mathcal{P}_X \to \mathbb{R}, x \mapsto f(x) = \log_2 \frac{1}{\mathcal{P}_X(x)}$, the entropy reformulates

$$H(X) = \mathbb{E}\Big[f(X)\Big]. \tag{2}$$

Proposition 1 *We have*

- $0 \leq H(X) \leq \log_2 |\mathcal{X}|$,

- $H(X) = \log_2 |\mathcal{X}| \iff X$ *is uniformly distributed.*

Example 1 *We consider that $X \in \{0,1\}$ is drawn according to the Bernouilli distribution with parameter $p \in [0,1]$. We denote the binary entropy by $h_b(p) = p \log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$.*

Definition 2 *The conditional entropy $H(X|Y)$ and the mutual information $I(X;Y)$ are defined by*

$$H(X|Y) = \sum_{(x,y)\in \operatorname{supp} \mathcal{P}_{XY}} \mathcal{P}_{XY}(x,y) \log_2 \frac{1}{\mathcal{P}_{X|Y}(x|y)}, \tag{3}$$

$$I(X;Y) = \sum_{(x,y)\in \operatorname{supp} \mathcal{P}_{XY}} \mathcal{P}_{XY}(x,y) \log_2 \frac{\mathcal{P}_{XY}(x,y)}{\mathcal{P}_X(x)\mathcal{P}_Y(y)}. \tag{4}$$

Proposition 2

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y), \tag{5}$$

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y). \tag{6}$$

Example 2 *We consider that $(X,Y) \in \{0,1\}^2$ is drawn according to the distribution $[(1-p)/2, p/2; p/2, (1-p)/2]$, with parameter $p \in [0,1]$. Then $H(X) = H(Y) = 1$, $H(X|Y) = H(Y|X) = h_b(p)$, $H(X,Y) = 1 + h_b(p)$, $I(X;Y) = 1 - h_b(p)$.*

Definition 3 *Given a distribution $\mathcal{P}_{XY} \in \Delta(\mathcal{X} \times \mathcal{Y})$, we define the information density function*

$$i : \operatorname{supp} \mathcal{P}_{XY} \to \mathbb{R}, \tag{7}$$

$$(x,y) \mapsto i(x,y) = \log_2 \frac{\mathcal{P}_{XY}(x,y)}{\mathcal{P}_X(x)\mathcal{P}_Y(y)}. \tag{8}$$

The function is well defined since $\operatorname{supp} \mathcal{P}_{XY} \subset \operatorname{supp} \mathcal{P}_X \times \operatorname{supp} \mathcal{P}_Y$.

The notation $i(X,Y)$ stands for the image of the random pair (X,Y) by the information density function $i(x,y)$. Then $i(X,Y)$ is a random variable with expected value and variance.

Definition 4 *The mutual information $I(X;Y)$, the unconditional information variance $U(X,Y)$ and the third absolute moment $T(X,Y)$ are defined by*

$$I(X;Y) = \mathbb{E}\big[i(X,Y)\big], \tag{9}$$

$$U(X;Y) = \mathbb{E}\Big[\Big|i(X,Y) - \mathbb{E}\big[i(X,Y)\big]\Big|^2\Big] = \mathbb{E}\Big[i(X,Y)^2\Big] - I(X;Y)^2, \tag{10}$$

$$T(X;Y) = \mathbb{E}\Big[\Big|i(X,Y) - \mathbb{E}\big[i(X,Y)\big]\Big|^3\Big]. \tag{11}$$

Lemma 1 (Lemma 46, pp. 24 in Polyanskiy et al. [4])

$$T(X;Y) \leq \bigg(\Big(|\mathcal{X}|^{\frac{1}{3}} + |\mathcal{Y}|^{\frac{1}{3}} \Big) \cdot \frac{3 \log_2 e}{e} + \log_2 \min \big(|\mathcal{X}|, |\mathcal{Y}| \big) \bigg)^3. \tag{12}$$

The proof of Lemma 1 is stated in [4, App. F].

Example 3 *We consider $(X, Y) \in \{0, 1\}^2$ is drawn according to the distribution $[(1-p)/2, p/2; p/2, (1-p)/2]$, with parameter $p \in [0, 1]$. We have $i(x, y) \in \{1 + \log_2(1-p), 1 + \log_2 p\}$, therefore*

$$I(X;Y) = 2 \cdot \frac{1-p}{2} \cdot (1 + \log_2(1-p)) + 2 \cdot \frac{p}{2} \cdot (1 + \log_2(p)) = 1 - h_b(p), \tag{13}$$

$$U(X;Y) = p(1-p) \cdot \Big(\log_2 \frac{1-p}{p} \Big)^2. \tag{14}$$

Definition 5 *The Kullback-Liebler (KL) divergence between $\mathcal{P}_X \in \Delta(\mathcal{X})$ and $\mathcal{Q}_X \in \Delta(\mathcal{X})$ is defined by*

$$D(\mathcal{P}_X || \mathcal{Q}_X) = \begin{cases} \sum_{x \in \operatorname{supp} \mathcal{P}_X} \mathcal{P}_X(x) \log_2 \frac{\mathcal{P}_X(x)}{\mathcal{Q}_X(x)} & \text{if} \quad \operatorname{supp} \mathcal{Q}_X \subset \operatorname{supp} \mathcal{P}_X, \\ +\infty & \text{otherwise.} \end{cases} \tag{15}$$

The divergence variance is defined by

$$V(\mathcal{P}_X || \mathcal{Q}_X) = \begin{cases} \sum_{x \in \operatorname{supp} \mathcal{P}_X} \mathcal{P}_X(x) \bigg(\log_2 \frac{\mathcal{P}_X(x)}{\mathcal{Q}_X(x)} \bigg)^2 - \Big(D(\mathcal{P}_X || \mathcal{Q}_X) \Big)^2 & \text{if} \quad \operatorname{supp} \mathcal{Q}_X \subset \operatorname{supp} \mathcal{P}_X, \\ +\infty & \text{otherwise.} \end{cases} \tag{16}$$

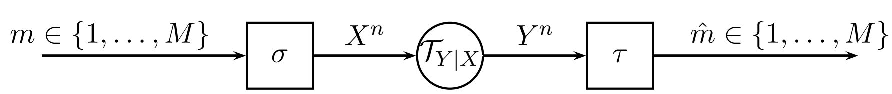
The conditional information variance is defined by

$$V(X;Y) = \mathbb{E}\Big[i(X, Y)^2 \Big] - \sum_x \mathcal{P}_X(x) \bigg(D\big(\mathcal{P}_{Y|X}(\cdot|x) \big|\big| \mathcal{P}_Y \big) \bigg)^2. \tag{17}$$

II. Channel Coding Problem

Definition 6 *A discrete channel $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ consists of two discrete sets $\mathcal{X}$, $\mathcal{Y}$ and a conditional probability distribution $\mathcal{T}_{Y|X} \in \Delta(\mathcal{Y})^{|\mathcal{X}|}$. The channel is memoryless if for all $n \in \mathbb{N}^\star = \mathbb{N} \setminus \{0\}$ and for all pair of sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$, we have*

$$\Pr(y^n | x^n) = \prod_{t=1}^{n} \mathcal{T}_{Y|X}(y_t | x_t). \tag{18}$$



Example 4 *The binary symmetric channel (BSC) with parameter $\delta \in [0, \frac{1}{2}]$.*

Definition 7 *Given $n \in \mathbb{N}^\star$ and $M \in \mathbb{N}^\star$, an (M,n)-code is a pair of functions (σ, τ) defined by*

$$\sigma : \{1, \dots, M\} \longrightarrow \mathcal{X}^n, \tag{19}$$

$$\tau : \mathcal{Y}^n \longrightarrow \{1, \dots, M\}. \tag{20}$$

We suppose that the message $m \in \{1, \dots, M\}$ is drawn uniformly at random, therefore

$$\Pr(m, x^n, y^n, \hat{m}) = \frac{1}{M}\mathbb{1}\big(x^n = \sigma(m)\big)\bigg(\prod_{t=1}^{n} \mathcal{T}_{Y|X}(y_t|x_t)\bigg)\mathbb{1}\big(\hat{m} = \tau(y^n)\big), \qquad \forall (m, x^n, y^n, \hat{m}). \tag{21}$$

The maximal error probability is defined by

$$\mathcal{P}_{\textsf{max}} = \max_{m \in \{1,\dots,M\}} \Pr(\hat{m} \neq m | m). \tag{22}$$

The average error probability is defined by

$$\mathcal{P}_{\textsf{ave}} = \frac{1}{M} \sum_{m \in \{1,\dots,M\}} \Pr(\hat{m} \neq m | m). \tag{23}$$

Definition 8 *Given $n \in \mathbb{N}^\star$ and $\varepsilon > 0$, the maximal code sizes are defined by*

$$M^\star_{\textsf{max}}(n, \varepsilon) = \max_{\substack{\exists (M,n)\text{-code}, \\ \mathcal{P}_{\textsf{max}} \leq \varepsilon}} M, \tag{24}$$

$$M^\star_{\textsf{ave}}(n, \varepsilon) = \max_{\substack{\exists (M,n)\text{-code}, \\ \mathcal{P}_{\textsf{vea}} \leq \varepsilon}} M. \tag{25}$$

For all $n \in \mathbb{N}^\star$ and $\varepsilon > 0$, we have $M^\star_{\textsf{max}}(n, \varepsilon) \leq M^\star_{\textsf{ave}}(n, \varepsilon)$.

Proposition 3 *Given $n \in \mathbb{N}^\star$ and $\varepsilon' > \varepsilon > 0$,*

$$M^\star_{\textsf{max}}(n, \varepsilon') \geq M^\star_{\textsf{ave}}(n, \varepsilon) \cdot \Big(1 - \frac{\varepsilon}{\varepsilon'}\Big). \tag{26}$$

Example 5 *We remove half of the codewords of $M^\star_{\textsf{ave}}(n, \varepsilon)$ with higher error probability $\Pr(\hat{m} \neq m | m)$. Since the codewords are equiprobable, all the remaining codewords have an error probability less than $2\varepsilon > 0$, otherwise $\mathcal{P}_{\textsf{ave}} > \varepsilon$. Therefore $M^\star_{\textsf{max}}(n, 2\varepsilon) \geq \frac{1}{2} M^\star_{\textsf{ave}}(n, \varepsilon)$.*

Definition 9 *The capacity of the discrete channel $(\mathcal{X}, \mathcal{Y}, \mathcal{T}_{Y|X})$ is defined by*

$$C = \max_{\mathcal{P}_X \in \Delta(\mathcal{X})} I(X; Y), \tag{27}$$

where

$$I(X;Y) = \sum_{x,y} \mathcal{P}_X(x) \mathcal{T}_{Y|X}(y|x) \log_2 \frac{\mathcal{T}_{Y|X}(y|x)}{\sum_{x'} \mathcal{P}_X(x') \mathcal{T}_{Y|X}(y|x')}. \tag{28}$$

Example 6 *The capacity of the binary symmetric channel (BSC) with parameter* $p \in [0, \frac{1}{2}]$ *is* $C = 1 - h_b(p)$.

Theorem 1 (Shannon 1948 [1])

$$\textit{Achievability result} \qquad \lim_{\varepsilon \to 0} \lim_{n \to +\infty} \frac{\log_2 M^\star_{\mathit{max}}(n, \varepsilon)}{n} \geq C, \tag{29}$$

$$\textit{Converse result} \qquad \forall n \in \mathbb{N}^\star \quad \lim_{\varepsilon \to 0} \frac{\log_2 M^\star_{\mathit{ave}}(n, \varepsilon)}{n} \leq C. \tag{30}$$

The proof is provided in [2, Chap.7].

Corollary 1

$$\lim_{\varepsilon \to 0} \lim_{n \to +\infty} \frac{\log_2 M^\star_{\mathit{max}}(n, \varepsilon)}{n} = \lim_{\varepsilon \to 0} \lim_{n \to +\infty} \frac{\log_2 M^\star_{\mathit{ave}}(n, \varepsilon)}{n} = C. \tag{31}$$

Remark 1 *When we interchange the limits in* (31)*, the problem is called* zero-error channel coding problem*. It corresponds to an open problem in Graph Theory, see [3].*

$$\lim_{n \to +\infty} \lim_{\varepsilon \to 0} \frac{\log_2 M^\star_{\mathit{max}}(n, \varepsilon)}{n} = \lim_{n \to +\infty} \frac{\log_2 M^\star_{\mathit{max}}(n, 0)}{n} = \lim_{n \to +\infty} \frac{\log_2 M^\star_{\mathit{ave}}(n, 0)}{n} = \lim_{n \to +\infty} \lim_{\varepsilon \to 0} \frac{\log_2 M^\star_{\mathit{ave}}(n, \varepsilon)}{n}. \tag{32}$$

For example, noisy type-writer with 7 elements.

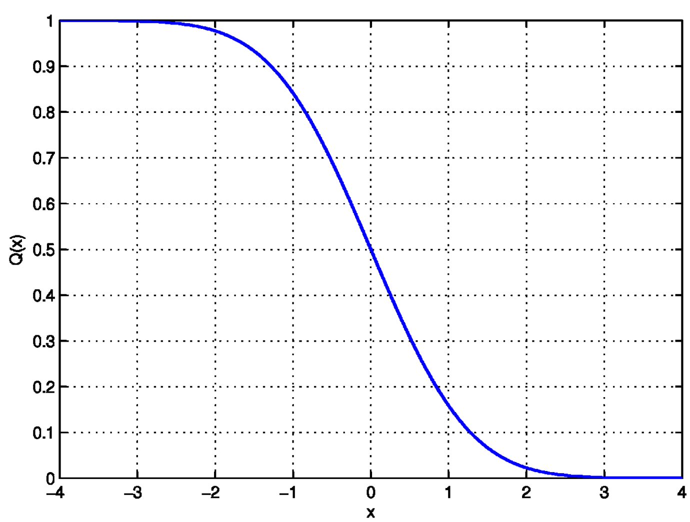


Fig. 1. Plot of the Q-function. Figure available at https://en.wikipedia.org/wiki/Q-function

III. Achievability Result in the Finite Blocklength Regime

Definition 10 *The Q-function is defined by*

$$Q : \mathbb{R} \to [0,1], \tag{33}$$

$$x \mapsto Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{t^2}{2}\Big) dt. \tag{34}$$

It is the tail distribution function $Q(x) = Pr(X > x)$ of the standard normal distribution $X \sim \mathcal{N}(0,1)$. The inverse Q-function is defined by

$$Q^{-1} : [0,1] \to \mathbb{R}, \tag{35}$$

$$y \mapsto Q^{-1}(y) = x \ s.t. \ Q^{-1} \circ Q = id_{\mathbb{R}}. \tag{36}$$

Theorem 2 (Theorem 45, pp. 24 in Polyanskiy et al. 2010 [4]) *For any $\mathcal{P}_X \in \Delta(\mathcal{X})$, for all pair $(n, \varepsilon) \in \mathbb{N}^\star \times]0, +\infty[$ that satisfies*

$$\varepsilon \cdot \sqrt{n} \geq \frac{2}{\sqrt{U(X;Y)}} \left(\frac{\ln 2}{\sqrt{2\pi}} + \frac{15 \cdot T(X;Y)}{U(X;Y)} \right). \tag{37}$$

We have

$$\frac{\log_2 M^\star_{\textsf{ave}}(n,\varepsilon)}{n} \geq I(X;Y) - \sqrt{\frac{U(X;Y)}{n}} \cdot Q^{-1}\left(\varepsilon - \frac{2}{\sqrt{n \cdot U(X;Y)}} \left(\frac{\ln 2}{\sqrt{2\pi}} + \frac{15 T(X;Y)}{U(X;Y)} \right) \right). \tag{38}$$

The proof of Theorem 2 is provided in Sec. IV and relies on the Berry-Esseen's Theorem.

Theorem 3 (Berry-Esseen) *Let Z_k, $k \in \{1, \ldots, n\}$ i.i.d with*

$$\mu = \mathbb{E}[Z_k], \quad \sigma^2 = \mathbb{E}\big[|Z_k - \mu|^2\big], \quad t = \mathbb{E}\big[|Z_k - \mu|^3\big] < +\infty. \tag{39}$$

We have

$$\sup_{x \in \mathbb{R}} \left| Pr\Big[\sum_{k=1}^n \frac{Z_k - \mu}{\sqrt{n} \cdot \sigma} \geq x \Big] - Q(x) \right| \leq \frac{6t}{\sqrt{n} \cdot \sigma^3}, \tag{40}$$

The proof of Theorem 3 is available in [5, Chap. XVI.5]. The convergence speed of the cumulative distribution function (CDF) of the sum of i.i.d. random variables is at least on the order of $\frac{1}{\sqrt{n}}$.

Remark 2 *The random variable $\sum_{k=1}^n Z_k$ has mean and variance*

$$\mathbb{E}\Big[\sum_{k=1}^n Z_k \Big] = n \cdot \mu, \tag{41}$$

$$\mathbb{E}\Big[\big(\sum_{k=1}^n (Z_k - \mu) \big)^2 \Big] = \Big(\sqrt{n} \cdot \sigma \Big)^2. \tag{42}$$

Therefore the random variable $\sum_{k=1}^n \frac{Z_k - \mu}{\sqrt{n} \cdot \sigma}$ has expected value 0 and variance 1.

Remark 3 *We also have for all* $\lambda \in \mathbb{R}$

$$\sup_{\lambda \in \mathbb{R}} \left| Pr\left[\sum_{k=1}^{n} \frac{Z_k - \mu}{\sqrt{n} \cdot \sigma} \leq -\lambda \right] - Q(\lambda) \right| \leq \frac{6t}{\sqrt{n} \cdot \sigma^3}. \tag{43}$$

Moreover, for all $\lambda \in \mathbb{R}$ *and for all* $\delta > 0$*, we have*

$$Pr\left[\lambda \leq \sum_{k=1}^{n} \frac{Z_k - \mu}{\sqrt{n} \cdot \sigma} \leq \lambda + \delta \right] \leq Q(\lambda) - Q(\lambda + \delta) + \frac{12t}{\sqrt{n} \cdot \sigma^3}. \tag{44}$$

Theorem 4 (Theorem 17 in Polyanskiy et al. 2010 [4]; Dependence Testing (DT) Bound) *For any distribution* $\mathcal{P}_X$*, for all* $\gamma > 0$*,* $n \in \mathbb{N}^\star$ *we define* $\varepsilon > 0$ *by*

$$\varepsilon = \mathbb{E}\left[\exp\left(- \left(i(X^n; Y^n) - \ln \gamma \right)^+ \right) \right], \tag{45}$$

with the notation $(x)^+ = \max(0, x)$*. Then* $M^\star_{\textsf{ave}}(n, \varepsilon) \geq 2\gamma + 1$*.*

The proof of Theorem 4 is stated in [4, pp. 7].

IV. Proof of Theorem 2

We select a distribution $\mathcal{P}_X \in \Delta(\mathcal{X})$ and we denote $\mathcal{P}_{XY} = \mathcal{P}_X \mathcal{T}_{Y|X} \in \Delta(\mathcal{X} \times \mathcal{Y})$. For $n \in \mathbb{N}^\star$, the pair of sequences (X^n, Y^n) is drawn i.i.d. according to $\mathcal{P}_{XY}$. We consider the sequence of i.i.d. random variables Z_k, $k \in \{1, \dots, n\}$ defined by

$$Z_k = i(X_k, Y_k) = \log_2 \frac{P_{XY}(X_k, Y_k)}{P_X(X_k) P_Y(Y_k)}. \tag{46}$$

We have $i(X^n, Y^n) = \sum_{k=1}^{n} Z_k$ and moreover

$$\forall k \in \{1, \dots, n\} \qquad \mathbb{E}[Z_k] = I(X; Y) = \mu, \tag{47}$$

$$\mathbb{E}[|Z_k - \mu|^2] = U(X; Y) = \sigma^2, \tag{48}$$

$$\mathbb{E}[|Z_k - \mu|^3] = T(X; Y) = t. \tag{49}$$

By Lemma 1, $t < +\infty$ and we apply Theorem 3 (Berry-Esseen), to the sequence of random variables $-Z_k$, $k \in \{1, \dots, n\}$. For any $\lambda \in \mathbb{R}$, we have

$$\left| \Pr\left[\sum_{k=1}^{n} \frac{-Z_k + \mu}{\sqrt{n} \cdot \sigma} \geq \lambda \right] - Q(\lambda) \right| \leq \frac{6t}{\sqrt{n} \cdot \sigma^3}. \tag{50}$$

By hypothesis $\varepsilon \cdot \sqrt{n} \geq \frac{2}{\sigma}\left(\frac{\ln 2}{\sqrt{2\pi}} + \frac{15t}{\sigma^2} \right)$. We introduce the parameters

$$\lambda = Q^{-1}\left(\varepsilon - \frac{2}{\sqrt{n} \cdot \sigma}\left(\frac{\ln 2}{\sqrt{2\pi}} + \frac{15t}{\sigma^2} \right) \right), \tag{51}$$

$$\ln \gamma = n \cdot I(X; Y) - \lambda \sqrt{n \cdot U(X; Y)}, \tag{52}$$

and (50) reformulates

$$\Pr\bigg[\sum_{k=1}^{n}\frac{-Z_k+\mu}{\sqrt{n}\cdot\sigma}\geq\lambda\bigg]\leq Q(\lambda)+\frac{6t}{\sqrt{n}\cdot\sigma^3} \tag{53}$$

$$\iff\quad \Pr\bigg[\sum_{k=1}^{n}\frac{Z_k-\mu}{\sqrt{n}\cdot\sigma}\leq-\lambda\bigg]\leq Q(\lambda)+\frac{6t}{\sqrt{n}\cdot\sigma^3} \tag{54}$$

$$\iff\quad \Pr\bigg[i(X^n,Y^n)\leq n\cdot I(X;Y)-\lambda\sqrt{n\cdot U(X;Y)}\bigg]\leq Q(\lambda)+\frac{6t}{\sqrt{n}\cdot\sigma^3} \tag{55}$$

$$\iff\quad \Pr\bigg[i(X^n,Y^n)\leq\ln\gamma\bigg]\leq\varepsilon-\frac{2}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}}+\frac{12t}{\sigma^2}\bigg). \tag{56}$$

Lemma 2 (Lemma 47 in Polyanskiy et al. 2010 [4]) *For $n\in\mathbb{N}^\star$, let Z_k, $k\in\{1,\dots,n\}$ i.i.d. discrete random variables with $\sigma^2=\mathbb{E}[|X_k-\mu|^2]$ and $t=\mathbb{E}[|X_k-\mu|^3]<+\infty$. Then for all $A\in\mathbb{R}$*

$$\mathbb{E}\bigg[\exp\bigg(-\sum_{k=1}^{n}Z_k\bigg)\cdot\mathbb{1}\bigg(\sum_{k=1}^{n}Z_k>A\bigg)\bigg]\leq\exp\big(-A\big)\cdot\frac{2}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}}+\frac{12t}{\sigma^2}\bigg). \tag{57}$$

The proof of Lemma 2 is in Sec. IV-A. By replacing $A=\ln\gamma$ in (57), we have

$$\gamma\cdot\mathbb{E}\bigg[\exp\Big(-i(X^n;Y^n)\Big)\cdot\mathbb{1}\Big(i(X^n;Y^n)>\ln\gamma\Big)\bigg]\leq\frac{2}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}}+\frac{12t}{\sigma^2}\bigg). \tag{58}$$

Equations (56) and (58) imply

$$\mathbb{E}\bigg[\exp\bigg(-\Big(i(X^n;Y^n)-\ln\gamma\Big)^+\bigg)\bigg] \tag{59}$$

$$=\Pr\bigg[i(X^n,Y^n)\leq\ln\gamma\bigg]+\gamma\cdot\mathbb{E}\bigg[\exp\Big(-i(X^n;Y^n)\Big)\cdot\mathbb{1}\Big(i(X^n;Y^n)>\ln\gamma\Big)\bigg] \tag{60}$$

$$\leq\varepsilon-\frac{2}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}}+\frac{12t}{\sigma^2}\bigg)+\frac{2}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}}+\frac{12t}{\sigma^2}\bigg)=\varepsilon. \tag{61}$$

Theorem 4 concludes the proof of Theorem 2.

$$\frac{\log_2 M^\star_{\mathsf{ave}}(n,\varepsilon)}{n}\geq\frac{\log_2 2\gamma+1}{n} \tag{62}$$

$$\geq\frac{\log_2\gamma}{n} \tag{63}$$

$$=I(X;Y)-\lambda\sqrt{\frac{U(X;Y)}{n}} \tag{64}$$

$$=I(X;Y)-\sqrt{\frac{U(X;Y)}{n}}\cdot Q^{-1}\bigg(\varepsilon-\frac{2}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}}+\frac{15t}{\sigma^2}\bigg)\bigg). \tag{65}$$

A. Proof of Lemma 2

By Lemma 1, $t < +\infty$ and we apply Theorem 3 (Berry-Esseen), for all $\lambda \in \mathbb{R}$ and for all $\delta > 0$, we have

$$\begin{aligned}
\Pr\bigg[\lambda \leq \sum_{k=1}^{n} \frac{Z_k - \mu}{\sqrt{n}\cdot\sigma} \leq \lambda + \delta\bigg] =&\Pr\bigg[\sum_{k=1}^{n} \frac{Z_k - \mu}{\sqrt{n}\cdot\sigma} \geq \lambda\bigg] - \Pr\bigg[\sum_{k=1}^{n} \frac{Z_k - \mu}{\sqrt{n}\cdot\sigma} \geq \lambda + \delta\bigg] && (66)\\
\leq& Q(\lambda) - Q(\lambda+\delta) + \frac{1}{\sqrt{n}}\frac{12t}{\sigma^3} && (67)\\
=& \int_{\lambda}^{\lambda+\delta} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{t^2}{2}\Big)dt + \frac{1}{\sqrt{n}}\frac{12t}{\sigma^3} && (68)\\
\leq& \frac{\delta}{\sqrt{2\pi}} + \frac{1}{\sqrt{n}}\frac{12t}{\sigma^3}. && (69)
\end{aligned}$$

For all $A \in \mathbb{R}$ we have

$$\begin{aligned}
&\mathbb{E}\bigg[\exp\bigg(-\sum_{k=1}^{n} Z_k\bigg)\cdot \mathbb{1}\bigg(\sum_{k=1}^{n} Z_k > A\bigg)\bigg] && (70)\\
\leq& \sum_{l=0}^{+\infty} \exp\big(-(A + l\ln 2)\big)\cdot \Pr\bigg[A + l\ln 2 \leq \sum_{k=1}^{n} Z_k \leq A + (l+1)\ln 2\bigg] && (71)\\
=& \sum_{l=0}^{+\infty} \exp\big(-(A + l\ln 2)\big)\cdot \Pr\bigg[\frac{A + l\ln 2 - \mu}{\sqrt{n}\cdot\sigma} \leq \sum_{k=1}^{n} \frac{Z_k - \mu}{\sqrt{n}\cdot\sigma} \leq \frac{A + l\ln 2 - \mu}{\sqrt{n}\cdot\sigma} + \frac{\ln 2}{\sqrt{n}\cdot\sigma}\bigg] && (72)\\
\leq& \sum_{l=0}^{+\infty} \exp\big(-(A + l\ln 2)\big)\cdot \frac{1}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}} + \frac{12t}{\sigma^2}\bigg) && (73)\\
=& \exp\big(-A\big)\cdot \frac{1}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}} + \frac{12t}{\sigma^2}\bigg)\cdot \sum_{l=0}^{+\infty} 2^{-l} && (74)\\
=& \exp\big(-A\big)\cdot \frac{2}{\sqrt{n}\cdot\sigma}\bigg(\frac{\ln 2}{\sqrt{2\pi}} + \frac{12t}{\sigma^2}\bigg). && (75)
\end{aligned}$$

This concludes the proof of Lemma 2.

V. Example: The Binary Symmetric Channel

Theorem 5 (Theorem 52, pp. 264 in Polyanskiy et al. 2010 [4]) *We consider the Binary Symmetric Channel (BSC) with parameter $p \in [0,1] \setminus \{0, \frac{1}{2}, 1\}$ and uniform input distribution $\mathcal{P}_X$. For all $\varepsilon > 0$, there exists $\bar{n} \in \mathbb{N}^\star$, for all $n \geq \bar{n}$ we have*

$$\frac{\log_2 M^\star_{\mathit{max}}(n,\varepsilon)}{n} \geq 1 - h_b(p) - \sqrt{\frac{p(1-p)}{n}}\log_2\frac{1-p}{p}\cdot Q^{-1}(\epsilon) + \frac{1}{2}\frac{\log_2 n}{n} + O\Big(\frac{1}{n}\Big), \tag{76}$$

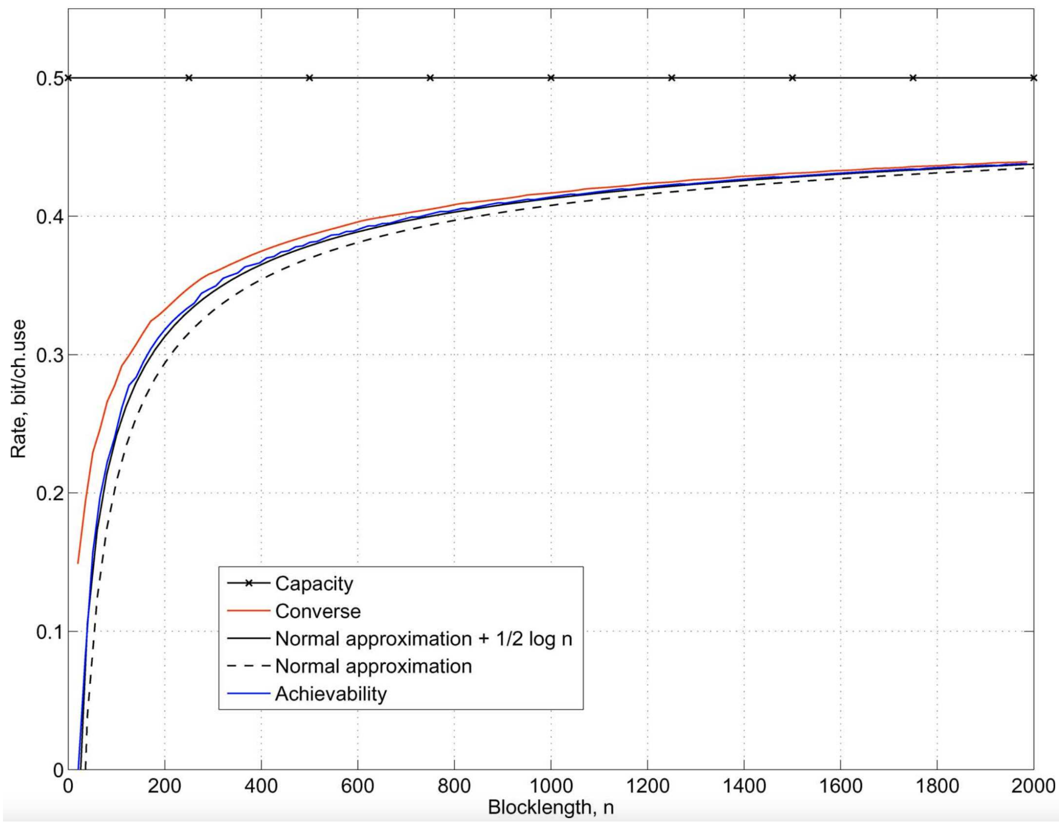


Fig. 2. [4, Fig. 8]. Rate $\frac{\log_2 M^{\star}_{\mathsf{max}}(n,\varepsilon)}{n}$ and blocklength n tradeoff for the BSC with crossover probability $p = 0.11$ and error probability $\varepsilon = 10^{-3}$.

References

[1] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: 2nd. Ed., Wiley-Interscience, 2006.

[3] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, pp. 8–19, 1956.

[4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, p. 2307, 2010.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*. Wiley, New York, 1971.