Simple Channel Coding Bounds

Ligong Wang
Signal and Information Processing Laboratory
ETH Zurich, Switzerland
wang@isi.ee.ethz.ch

Roger Colbeck
Institute for Theoretical Physics, and
Institute of Theoretical Computer Science
ETH Zurich, Switzerland
colbeck@phys.ethz.ch

Renato Renner
Institute for Theoretical Physics
ETH Zurich, Switzerland
renner@phys.ethz.ch

***Abstract*— New channel coding converse and achievability bounds are derived for a single use of an arbitrary channel. Both bounds are expressed using a quantity called the "smooth 0-divergence", which is a generalization of Rényi's divergence of order 0. The bounds are also studied in the limit of large block-lengths. In particular, they combine to give a general capacity formula which is equivalent to the one derived by Verdú and Han.**

I. Introduction

We consider the problem of transmitting information through a channel. A channel consists of an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are each equipped with a σ-Algebra, and the channel law which is a stochastic kernel $P_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$. We consider average error probabilities throughout this paper,[1] thus an (m, ϵ)-*code* consists of an encoder $f : \{1, \ldots, m\} \to \mathcal{X}, i \mapsto x$ and a decoder $g : \mathcal{Y} \to \{1, \ldots, m\}, y \mapsto \hat{i}$ such that the probability that $i \neq \hat{i}$ is smaller than or equal to ϵ, assuming that the message is uniformly distributed. Our aim is to derive upper and lower bounds on the largest m given $\epsilon > 0$ such that an (m, ϵ)-code exists for a given channel.

Such bounds are different from those in Shannon's original work [1] in the sense that they are nonasymptotic and do not rely on any channel structure such as memorylessness or information stability.

Previous works have demonstrated the advantages of such nonasymptotic bounds. They can lead to more general channel capacity formulas [2] as well as giving tight approximations to the maximal rate achievable for a desired error probability and a fixed block-length [3].

In this paper we prove a new converse bound and a new achievability bound. They are asymptotically tight in the sense that they combine to give a general capacity formula that is equivalent to [2, (1.4)]. We are mainly interested in proving simple bounds which offer theoretical intuitions into channel coding problems. It is not our main concern to derive bounds which outperform the existing ones in estimating the largest achievable rates in finite block-length scenarios. In fact, as will be seen in Section VI, the new achievability bound is less tight than the one in [3], though the differences are small.

[1]Note that Shannon's method of obtaining codes that have small maximum error probabilities from those that have small average error probabilities [1] can be applied to our codes. We shall not examine other such methods which might lead to tighter bounds for finite block-lengths.

Both new bounds are expressed using a quantity which we call the *smooth* 0-*divergence*, denoted as $D_0^\delta(\cdot\|\cdot)$ where δ is a positive parameter. This quantity is a generalization of Rényi's divergence of order 0 [4]. Thus, our new bounds demonstrate connections between the channel coding problem and Rényi's divergence of order 0. Various previous works [5], [6], [7] have shown connections between channel coding and Rényi's information measures of order α for $\alpha \geq \frac{1}{2}$. Also relevant is [8] where channel coding bounds were derived using the smooth min- and max-entropies introduced in [9].

As will be seen, proofs of the new bounds are simple and self-contained. The achievability bound uses random coding and suboptimal decoding, where the decoding rule can be thought of as a generalization of Shannon's joint typicality decoding rule [1]. The converse is proved by simple algebra combined with the fact that $D_0^\delta(\cdot\|\cdot)$ satisfies a Data Processing Theorem.

The quantity $D_0^\delta(\cdot\|\cdot)$ has also been defined for quantum systems [10], [11]. In [11] the present work is extended to quantum communication channels.

The remainder of this paper is arranged as follows: in Section II we introduce the quantity $D_0^\delta(\cdot\|\cdot)$; in Section III we state and prove the converse theorem; in Section IV we state and prove the achievability theorem; in Section V we analyze the bounds asymptotically for an arbitrary channel to study its capacity and ϵ-capacity; finally, in Section VI we compare numerical results obtained using our new achievability bound with some existing bounds.

II. The Quantity $D_0^\delta(\cdot\|\cdot)$

In [4] Rényi defined entropies and divergences of order α for every $\alpha > 0$. We denote these $H_\alpha(\cdot)$ and $D_\alpha(\cdot\|\cdot)$ respectively. They are generalizations of Shannon's entropy $H(\cdot)$ and relative entropy $D(\cdot\|\cdot)$.

Letting α tend to zero in $D_\alpha(\cdot\|\cdot)$ yields the following definition of $D_0(\cdot\|\cdot)$.

Definition 1 (Rényi's Divergence of Order 0*):* For P and Q which are two probability measures on $(\Omega, \mathcal{F})$, $D_0(P\|Q)$ is defined as

$$D_0(P\|Q) = -\log \int_{\text{supp}(P)} \mathrm{d}Q, \tag{1}$$

where we use the convention $\log 0 = -\infty$.[2]

We generalize $D_0(\cdot\|\cdot)$ to define $D_0^\delta(\cdot\|\cdot)$ as follows.

Definition 2 (Smooth 0-Divergence): Let P and Q be two probability measures on $(\Omega, \mathcal{F})$. For $\delta > 0$, $D_0^\delta(P\|Q)$ is defined as

$$D_0^\delta(P\|Q) = \sup_{\substack{\Phi:\Omega\to[0,1] \\ \int_\Omega \Phi \,\mathrm{d}P \geq 1-\delta}} \left\{ -\log \int_\Omega \Phi \,\mathrm{d}Q \right\}. \tag{2}$$

Remark: To achieve the supremum in (2), one should choose Φ to be large (equal to 1) for large $\frac{\mathrm{d}P}{\mathrm{d}Q}$ and vice versa.

Lemma 1 (Properties of $D_0^\delta(\cdot\|\cdot)$):

1) $D_0^\delta(P\|Q)$ is monotonically nondecreasing in δ.
2) When $\delta = 0$, the supremum in (2) is achieved by choosing Φ to be 1 on supp(P) and to be 0 elsewhere, which yields $D_0^0(P\|Q) = D_0(P\|Q)$.
3) If P has no point masses, then the supremum in (2) is achieved by letting Φ take value in $\{0, 1\}$ only and
$$D_0^\delta(P\|Q) = \sup_{P':\frac{1}{2}\|P'-P\|_1 \leq \delta} D_0(P'\|Q).$$
4) (Data Processing Theorem) Let P and Q be probability measures on $(\Omega, \mathcal{F})$, and let W be a stochastic kernel from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$. For all $\delta > 0$, we have
$$D_0^\delta(P\|Q) \geq D_0^\delta(W \circ P\|W \circ Q), \tag{3}$$
where $W \circ P$ denotes the probability distribution on $(\Omega', \mathcal{F}')$ induced by P and W and similarly for $W \circ Q$.

Proof: The first three properties are immediate consequences of the definition and the remark. We therefore only prove 4).

For any $\Phi' : \Omega' \to [0, 1]$ such that

$$\int_{\Omega'} \Phi' \,\mathrm{d}(W \circ P) \geq 1 - \delta,$$

we choose $\Phi : \Omega \to \mathbb{R}$ to be

$$\Phi(\omega) = \int_{\Omega'} \Phi'(\omega')W(\mathrm{d}\omega'|\omega), \quad \omega \in \Omega.$$

then we have that $\Phi(\omega) \in [0, 1]$ for all $\omega \in \Omega$. Further,

$$\int_\Omega \Phi \,\mathrm{d}P = \int_{\Omega'} \Phi' \,\mathrm{d}(W \circ P) \geq 1 - \delta,$$
$$\int_\Omega \Phi \,\mathrm{d}Q = \int_{\Omega'} \Phi' \,\mathrm{d}(W \circ Q).$$

Thus we have

$$\sup_{\substack{\Phi:\Omega\to[0,1] \\ \int_\Omega \Phi \,\mathrm{d}P \geq 1-\delta}} \left\{ -\log \int_\Omega \Phi \,\mathrm{d}Q \right\} \geq \sup_{\substack{\Phi':\Omega'\to[0,1] \\ \int_{\Omega'} \Phi' \,\mathrm{d}(W\circ P) \geq 1-\delta}} \left\{ -\log \int_{\Omega'} \Phi' \,\mathrm{d}(W \circ Q) \right\},$$

which proves 4). ■

[2]We remark that for distributions defined on a finite alphabet, $\mathcal{X}$, the equivalent of (1) is $D_0(P\|Q) = -\log \sum_{x:P(x)>0} Q(x)$.

A relation between $D_0^\delta(P\|Q)$, $D(P\|Q)$ and the information spectrum methods [12], [13] can be seen in the next lemma. A slightly different quantum version of this theorem has been proven in [10]. We include a classical proof of it in the Appendix.

Lemma 2: Let P_n and Q_n be probability measures on $(\Omega_n, \mathcal{F}_n)$ for every $n \in \mathbb{N}$. Then

$$\lim_{\delta\downarrow 0} \varliminf_{n\to\infty} \frac{1}{n} D_0^\delta\left(P_n\|Q_n\right) = \{P_n\}\text{-}\varliminf_{n\to\infty} \frac{1}{n} \log \frac{\mathrm{d}P_n}{\mathrm{d}Q_n}. \tag{4}$$

Here $\{P_n\}$-$\varliminf$ means the $\liminf$ in probability with respect to the sequence of probability measures $\{P_n\}$, that is, for a real stochastic process $\{Z_n\}$,

$$\{P_n\}\text{-}\varliminf_{n\to\infty} Z_n \triangleq \sup\left\{a \in \mathbb{R} : \lim_{n\to\infty} P_n\left(\{Z_n < a\}\right) = 0\right\}.$$

In particular, let $P^{\times n}$ and $Q^{\times n}$ denote the product distributions of P and Q respectively on $(\Omega^{\otimes n}, \mathcal{F}^{\otimes n})$, then

$$\lim_{\delta\downarrow 0} \varliminf_{n\to\infty} \frac{1}{n} D_0^\delta\left(P^{\times n}\|Q^{\times n}\right) = D(P\|Q). \tag{5}$$

Proof: See Appendix. ■

III. The Converse

We first state and prove a lemma.

Lemma 3: Let M be uniformly distributed over $\{1, \ldots, m\}$ and let $\hat{M}$ also take value in $\{1, \ldots, m\}$. If the probability that $\hat{M} \neq M$ is at most ϵ, then

$$\log m \leq D_0^\epsilon\left(P_{M\hat{M}}\|P_M \times P_{\hat{M}}\right),$$

where $P_{M\hat{M}}$ denotes the joint distribution of M and $\hat{M}$ while P_M and $P_{\hat{M}}$ denote its marginals.

Proof: Let Φ be the indicator of the event $M = \hat{M}$, i.e.,

$$\Phi(i, \hat{i}) \triangleq \begin{cases} 1, & i = \hat{i} \\ 0, & \text{otherwise} \end{cases} \qquad i, \hat{i} \in \{1, \ldots, m\}.$$

Because, by assumption, the probability that $M \neq \hat{M}$ is not larger than ϵ, we have

$$\int_{\{1,\ldots,m\}^{\otimes 2}} \Phi \,\mathrm{d}P_{M\hat{M}} \geq 1 - \epsilon.$$

Thus, to prove the lemma, it suffices to show that

$$\log m \leq -\log \int_{\{1,\ldots,m\}^{\otimes 2}} \Phi \,\mathrm{d}(P_M \times P_{\hat{M}}). \tag{6}$$

To justify this we write:

$$\begin{aligned} \int_{\{1,\ldots,m\}^{\otimes 2}} \Phi \,\mathrm{d}(P_M \times P_{\hat{M}}) &= \sum_{i=1}^m P_M\big(\{i\}\big) \cdot P_{\hat{M}}\big(\{i\}\big) \\ &= \sum_{i=1}^m \frac{1}{m} \cdot P_{\hat{M}}\big(\{i\}\big) \\ &= \frac{1}{m}, \end{aligned}$$

from which it follows that (6) is satisfied with equality. ■

Theorem 1 (Converse): An (m, ϵ)-code satisfies

$$\log m \leq \sup_{P_X} D_0^{\epsilon}\left(P_{XY} \| P_X \times P_Y\right), \tag{7}$$

where P_{XY} and P_Y are probability distributions on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{Y}$, respectively, induced by P_X and the channel law.

Proof: Choose P_X to be the distribution induced by the message uniformly distributed over $\{1, \ldots, m\}$, then

$$\begin{aligned}\log m &\leq D_0^{\epsilon}\left(P_{M\hat{M}} \| P_M \times P_{\hat{M}}\right)\\ &\leq D_0^{\epsilon}\left(P_{XY} \| P_X \times P_Y\right),\end{aligned}$$

where the first inequality follows by Lemma 3; the second inequality by Lemma 1 Part 4) and the fact that $M \multimap X \multimap Y \multimap \hat{M}$ forms a Markov Chain. Theorem 1 follows. ■

IV. Achievability

Theorem 2 (Achievability): For any channel, any $\epsilon > 0$ and $\epsilon' \in [0, \epsilon)$ there exists an (m, ϵ)-code satisfying

$$\log m \geq \sup_{P_X} D_0^{\epsilon'}\left(P_{XY} \| P_X \times P_Y\right) - \log \frac{1}{\epsilon - \epsilon'}, \tag{8}$$

where P_{XY} and P_Y are induced by P_X and the channel law.

The proof of Theorem 2 can be thought of as a generalization of Shannon's original achievability proof [1]. We use random coding as in [1]; for the decoder, we generalize Shannon's typicality decoder to allow, instead of the "indicator" for the jointly typical set, an arbitrary function on input-output pairs.

Proof: For any distribution P_X on $\mathcal{X}$ and any $m \in \mathbb{Z}^+$, we randomly generate a codebook of size m such that the m codewords are independent and identically distributed according to P_X. We shall show that, for any ϵ', there exists a decoding rule associated with each codebook such that the average probability of a decoding error averaged over all such codebooks satisfies

$$\Pr(\text{error}) \leq (m-1) \cdot 2^{-D_0^{\epsilon'}\left(P_{XY} \| P_X \times P_Y\right)} + \epsilon'. \tag{9}$$

Then there exists at least one codebook whose average probability of error is upper-bounded by the right hand side (RHS) of (9). That this codebook satisfies (8) follows by rearranging terms in (9).

We shall next prove (9). For a given codebook and any $\Phi : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ which satisfies

$$\int_{\mathcal{X} \times \mathcal{Y}} \Phi \, \mathrm{d}P_{XY} \geq 1 - \epsilon', \tag{10}$$

we use the following random decoding rule:[3] when y is received, select some or none of the messages such that message j is selected with probability $\Phi(f(j), y)$ independently of the other messages. If only one message is selected, output this message; otherwise declare an error.

[3]It is well-known that, for the channel model considered in this paper, the average probability of error cannot be improved by allowing random decoding rules.

To analyze the error probability, suppose i was the transmitted message. The error event is the union of $\mathcal{E}_1$ and $\mathcal{E}_2$, where $\mathcal{E}_1$ denotes the event that some message other than i is selected; $\mathcal{E}_2$ denotes the event that message i is not selected.

We first bound $\Pr(\mathcal{E}_1)$ averaged over all codebooks. Fix $f(i)$ and y. The probability averaged over all codebooks of selecting a particular message other than i is given by

$$\int_{\mathcal{X}} \Phi(x, y) P_X(\mathrm{d}x).$$

Since there are $(m-1)$ such messages, we can use the union bound to obtain

$$\mathsf{E}[\Pr(\mathcal{E}_1 | f(i), y)] \leq (m-1) \cdot \int_{\mathcal{X}} \Phi(x, y) P_X(\mathrm{d}x). \tag{11}$$

Since the RHS of (11) does not depend on $f(i)$, we further have

$$\mathsf{E}[\Pr(\mathcal{E}_1 | y)] \leq (m-1) \cdot \int_{\mathcal{X}} \Phi(x, y) P_X(\mathrm{d}x).$$

Averaging this inequality over y gives

$$\begin{aligned}\mathsf{E}[\Pr(\mathcal{E}_1)] &\leq (m-1) \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} \Phi(x, y) P_X(\mathrm{d}x) \right) P_Y(\mathrm{d}y)\\ &= (m-1) \int_{\mathcal{X} \times \mathcal{Y}} \Phi \, \mathrm{d}\left(P_X \times P_Y\right).\end{aligned} \tag{12}$$

On the other hand, the probability of $\mathcal{E}_2$ averaged over all generated codebooks can be bounded as

$$\begin{aligned}\mathsf{E}[\Pr(\mathcal{E}_2)] &= \int_{\mathcal{X} \times \mathcal{Y}} (1 - \Phi) \, \mathrm{d}P_{XY}\\ &\leq \epsilon'.\end{aligned} \tag{13}$$

Combining (12) and (13) yields

$$\Pr(\text{error}) \leq (m-1) \int_{\mathcal{X} \times \mathcal{Y}} \Phi \, \mathrm{d}\left(P_X \times P_Y\right) + \epsilon'. \tag{14}$$

Finally, since (14) holds for every Φ satisfying (10), we establish (9) and thus conclude the proof of Theorem 2. ■

V. Asymptotic Analysis

In this section we use the new bounds to study the capacity of a channel whose structure can be arbitrary. Such a channel is described by stochastic kernels from $\mathcal{X}^n$ to $\mathcal{Y}^n$ for all $n \in \mathbb{Z}^+$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively. An (n, M, ϵ)-*code* on a channel consists of an encoder and a decoder such that a message of size M can be transmitted by mapping it to an element of $\mathcal{X}^n$ while the probability of error is no larger than ϵ. The capacity and the optimistic capacity [14] of a channel are defined as follows.

Definition 3 (Capacity and Optimistic Capacity): The *capacity* C of a channel is the supremum over all R for which there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$\frac{\log M_n}{n} \geq R, \quad n \in \mathbb{Z}^+ \tag{15}$$

and

$$\lim_{n \to \infty} \epsilon_n = 0.$$

The *optimistic capacity* $\overline{C}$ of a channel is the supremum over all R for which there exists a sequence of (n, M_n, ϵ_n)-codes such that (15) holds and

$$\underline{\lim}_{n\to\infty} \epsilon_n = 0.$$

Given Definition 3, the next theorem is an immediate consequence of Theorems 1 and 2.

Theorem 3 (Capacity Formulas): Any channel satisfies

$$C = \lim_{\epsilon\downarrow 0} \underline{\lim}_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} D_0^{\epsilon}\left(P_{X^nY^n} \| P_{X^n} \times P_{Y^n}\right), \quad (16)$$

$$\overline{C} = \lim_{\epsilon\downarrow 0} \overline{\lim}_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} D_0^{\epsilon}\left(P_{X^nY^n} \| P_{X^n} \times P_{Y^n}\right). \quad (17)$$

Remark: According to Lemma 2, (16) is equivalent to [2, (1.4)]. It can also be shown that (17) is equivalent to [15, Theorem 4.4].

We can also use Theorems 1 and 2 to study the ϵ-capacities which are usually defined as follows (see, for example, [2], [15]).

Definition 4 (ϵ-Capacity and Optimistic ϵ-Capacity): The *ϵ-capacity* C_ϵ of a channel is the supremum over all R such that, for every large enough n, there exists an (n, M_n, ϵ)-code satisfying

$$\frac{\log M_n}{n} \geq R.$$

The *optimistic ϵ-capacity* $\overline{C}_\epsilon$ of a channel is the supremum over all R for which there exist (n, M_n, ϵ)-codes for infinitely many ns satisfying

$$\frac{\log M_n}{n} \geq R.$$

The following bounds on the ϵ-capacity and optimistic ϵ-capacity of a channel are immediate consequences of Theorems 1 and 2. They can be shown to be equivalent to those in [2, Theorem 6], [16, Theorem 7] and [15, Theorem 4.3]. As in those previous results, the bounds for C_ϵ ($\overline{C}_\epsilon$) coincide except possibly at the points of discontinuity of C_ϵ ($\overline{C}_\epsilon$).

Theorem 4 (Bounds on ϵ-Capacities): For any channel and any $\epsilon \in (0,1)$, the ϵ-capacity of the channel satisfies

$$C_\epsilon \leq \underline{\lim}_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} D_0^{\epsilon}\left(P_{X^nY^n} \| P_{X^n} \times P_{Y^n}\right),$$

$$C_\epsilon \geq \lim_{\epsilon'\uparrow\epsilon} \underline{\lim}_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} D_0^{\epsilon'}\left(P_{X^nY^n} \| P_{X^n} \times P_{Y^n}\right);$$

and the optimistic ϵ-capacity of the channel satisfies

$$\overline{C}_\epsilon \leq \overline{\lim}_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} D_0^{\epsilon}\left(P_{X^nY^n} \| P_{X^n} \times P_{Y^n}\right),$$

$$\overline{C}_\epsilon \geq \lim_{\epsilon'\uparrow\epsilon} \overline{\lim}_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} D_0^{\epsilon'}\left(P_{X^nY^n} \| P_{X^n} \times P_{Y^n}\right).$$

VI. Numerical Comparison with Existing Bounds for the BSC

In this section we compare the new achievability bound obtained in this paper with the bounds by Gallager [5] and Polyanskiy et al. [3]. We consider the memoryless binary symmetric channel (BSC) with crossover probability 0.11. Thus, for n channel uses, the input and output alphabets are both $\{0,1\}^n$ and the channel law is given by

$$P_{Y^n|X^n}(y^n|x^n) = 0.11^{|y^n - x^n|} 0.89^{n - |y^n - x^n|},$$

where $|\cdot|$ denotes the Hamming weight of a binary vector. The average block-error rate is chosen to be 10^{-3}.

In the calculations of all three achievability bounds we choose P_{X^n} to be uniform on $\{0,1\}^n$. For comparison we include the plot of the converse used in [3]. Our new converse bound involves optimization over input distributions and is thus difficult to compute. In fact, in this example it is less tight compared to the one in [3] since for the uniform input distribution $D_0^{0.001}(P_{X^nY^n} \| P_{X^n} \times P_{Y^n})$ coincides with the latter.

Comparison of the curves is shown in Figure 1. For the example we consider, the new achievability is always less tight than the one in [3], though the difference is small. It outperforms Gallager's bound for large block-lengths.

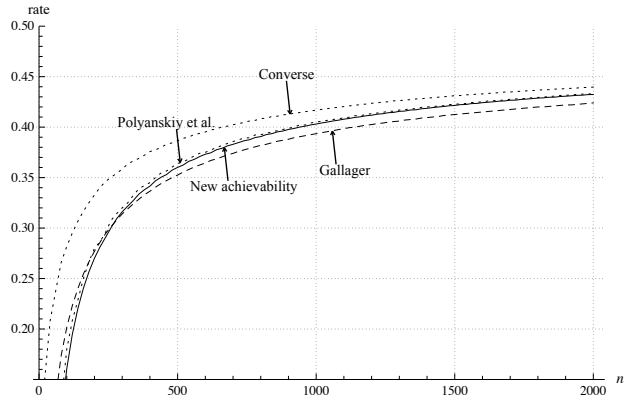

Fig. 1. Comparison of the new achievability bound with Gallager [5] and Polyanskiy et al. [3] for the BSC with crossover probability 0.11 and average block-error rate 10^{-3}. The converse is the one used in [3].

Appendix

In this appendix we prove Lemma 2. We first show that

$$\lim_{\delta\downarrow 0} \underline{\lim}_{n\to\infty} \frac{1}{n} D_0^{\delta}\left(P_n \| Q_n\right) \geq \{P_n\}\text{-}\underline{\lim}_{n\to\infty} \frac{1}{n} \log \frac{\mathrm{d}P_n}{\mathrm{d}Q_n}. \quad (18)$$

To this end, consider any a satisfying

$$0 < a < \{P_n\}\text{-}\underline{\lim}_{n\to\infty} \frac{1}{n} \log \frac{\mathrm{d}P_n}{\mathrm{d}Q_n}. \quad (19)$$

Let $\mathcal{A}_n(a) \in \mathcal{F}_n$, $n \in \mathbb{N}$, be the union of all measurable sets on which

$$\frac{1}{n} \log \frac{\mathrm{d}P_n}{\mathrm{d}Q_n} \geq a. \quad (20)$$

Let $\Phi_n : \Omega_n \to [0,1]$, $n \in \mathbb{N}$, equal 1 on $\mathcal{A}_n(a)$ and equal 0 elsewhere, then by (19) we have

$$\lim_{n\to\infty} \int_{\Omega_n} \Phi_n \, \mathrm{d}P_n = \lim_{n\to\infty} P_n\left(\mathcal{A}_n(a)\right) = 1. \quad (21)$$

Thus we have

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\varliminf_{n\to\infty}\frac{1}{n}D_0^{\delta}\left(P_n\|Q_n\right)\\
&\qquad\geq \varliminf_{n\to\infty}\left(-\frac{1}{n}\log\int_{\Omega_n}\Phi_n\,\mathrm{d}Q_n\right)\\
&\qquad\geq \varliminf_{n\to\infty}\left(-\frac{1}{n}\log\int_{\Omega_n}\Phi_n\,\mathrm{d}P_n\cdot 2^{-na}\right)\\
&\qquad= \lim_{n\to\infty}\left(-\frac{1}{n}\log\left(2^{-na}\right)\right)\\
&\qquad= a, \qquad (22)
\end{aligned}$$

where the first inequality follows because, according to (21), for any $\delta > 0$, $\int_{\Omega_n}\Phi_n\,\mathrm{d}P_n = P_n\left(\mathcal{A}_n(a)\right) \geq 1-\delta$ for large enough n; the second inequality by (20) and the fact that Φ_n is zero outside $\mathcal{A}_n(a)$; the next equality by (21). Since (22) holds for every a satisfying (19), we obtain (18).

We next show the other direction, namely, we show that

$$\lim_{\delta\downarrow 0}\varliminf_{n\to\infty}\frac{1}{n}D_0^{\delta}\left(P_n\|Q_n\right) \leq \{P_n\}\text{-}\varliminf_{n\to\infty}\frac{1}{n}\log\frac{\mathrm{d}P_n}{\mathrm{d}Q_n}. \qquad (23)$$

To this end, consider any

$$b > \{P_n\} - \varliminf_{n\to\infty}\frac{1}{n}\log\frac{\mathrm{d}P_n}{\mathrm{d}Q_n}. \qquad (24)$$

Let $\mathcal{A}'_n(b)$, $n\in\mathbb{N}$, be the union of all measurable sets on which

$$\frac{1}{n}\log\frac{\mathrm{d}P_n}{\mathrm{d}Q_n} \leq b. \qquad (25)$$

By (24) we have that there exists some $c\in(0,1]$ such that

$$\varlimsup_{n\to\infty} P_n\left(\mathcal{A}'_n(b)\right) = c. \qquad (26)$$

For every $\delta\in(0,c)$, consider any sequence of $\Phi_n:\Omega_n\to[0,1]$ satisfying

$$\int_{\Omega_n}\Phi_n\,\mathrm{d}P_n \geq 1-\delta, \quad n\in\mathbb{N}. \qquad (27)$$

Combining (26) and (27) yields

$$\varlimsup_{n\to\infty}\int_{\mathcal{A}'_n(b)}\Phi_n\,\mathrm{d}P_n \geq c-\delta. \qquad (28)$$

On the other hand, from (25) it follows that

$$\int_{\mathcal{A}'_n(b)}\Phi_n\,\mathrm{d}Q_n \geq \int_{\mathcal{A}'_n(b)}\Phi_n\,\mathrm{d}P_n\cdot 2^{-nb}. \qquad (29)$$

Combining (28) and (29) yields

$$\varliminf_{n\to\infty}\left(-\frac{1}{n}\log\int_{\mathcal{A}'_n(b)}\Phi_n\,\mathrm{d}Q_n\right) \leq b.$$

Thus we obtain that for every $\delta\in(0,c)$ and every sequence $\Phi_n:\Omega_n\to[0,1]$ satisfying (27),

$$\begin{aligned}
&\varliminf_{n\to\infty}\left(-\frac{1}{n}\log\int_{\Omega_n}\Phi_n\,\mathrm{d}Q_n\right)\\
&\qquad\leq \varliminf_{n\to\infty}\left(-\frac{1}{n}\log\int_{\mathcal{A}'_n(b)}\Phi_n\,\mathrm{d}Q_n\right)\\
&\qquad\leq b.
\end{aligned}$$

This implies that, for every $\delta\in(0,c)$,

$$\varliminf_{n\to\infty}\frac{1}{n}D_0^{\delta}\left(P_n\|Q_n\right) \leq b. \qquad (30)$$

Inequality (30) still holds when we take the limit $\delta\downarrow 0$. Since this is true for every b satisfying (24), we establish (23).

Combining (18) and (23) proves (4).

Finally, (5) follows from (4) because, by the law of large numbers,

$$\frac{1}{n}\log\frac{\mathrm{d}(P^{\times n})}{\mathrm{d}(Q^{\times n})} \to \mathsf{E}\left[\log\frac{\mathrm{d}P}{\mathrm{d}Q}\right] = D(P\|Q) \qquad (31)$$

as $n\to\infty$ $P^{\times n}$-almost surely. This completes the proof of Lemma 2.

Acknowledgment

RR acknowledges support from the Swiss National Science Foundation (grant No. 200021-119868).

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[2] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "New channel coding achievability bounds," in *Proc. IEEE Int. Symposium on Inf. Theory*, Toronto, Canada, July 6–11, 2008.

[4] A. Rényi, "On measures of entropy and information," in *4th Berkeley Conference on Mathematical Statistics and Probability*, University of California, Berkeley, CA, USA, 1961.

[5] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no. 1, pp. 3–19, Jan. 1965.

[6] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory*, I. Csiszár and P. Elias, Eds., 1977, pp. 41–52.

[7] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 26–34, Jan. 1995.

[8] R. Renner, S. Wolf, and J. Wullschleger, "The single-serving channel capacity," in *Proc. IEEE Int. Symposium on Inf. Theory*, Seattle, Washington, USA, July 9–14, 2006.

[9] R. Renner and S. Wolf, "Smooth Rényi entropy and applications," in *Proc. IEEE Int. Symposium on Inf. Theory*, Chicago, Illinois, USA, June 27 – July 2, 2004.

[10] N. Datta, "Min- and max-relative entropies and a new entanglement monotone," 2008. [Online]. Available: http://arxiv.org/abs/0803.2770

[11] L. Wang and R. Renner, "One-shot classical capacity of quantum channels," presented at *The Twelfth Workshop on Quantum Inform. Processing (QIP)*, poster session.

[12] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[13] T. S. Han, *Information Spectrum Methods in Information Theory*. Springer Verlag, 2003.

[14] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.

[15] P.-N. Chen and F. Alajaji, "Optimistic Shannon coding theorems for arbitrary single-user systems," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2623–2629, Nov. 1999.

[16] Y. Steinberg, "New converses in the theory of identification via channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 984–998, May 1998.