Upper Bounds on the Relative Entropy and Rényi Divergence as a Function of Total Variation Distance for Finite Alphabets

Igal Sason  
Department of Electrical Engineering  
Technion – Israel Institute of Technology  
Haifa 32000, Israel  
E-mail: sason@ee.technion.ac.il

Sergio Verdú  
Department of Electrical Engineering  
Princeton University  
Princeton, New Jersey 08544, USA  
E-mail: verdu@princeton.edu

Abstract—A new upper bound on the relative entropy is derived as a function of the total variation distance for probability measures defined on a common finite alphabet. The bound improves a previously reported bound by Csiszár and Talata. It is further extended to an upper bound on the Rényi divergence of an arbitrary non-negative order (including $\infty$) as a function of the total variation distance.

Keywords: Pinsker’s inequality, relative entropy, relative information, Rényi divergence, total variation distance.

1. INTRODUCTION

Consider two probability distributions $P$ and $Q$ defined on a common measurable space $(\mathcal{A}, \mathcal{F})$. The Csiszár-Kemperman-Kullback-Pinsker inequality (a.k.a. Pinsker’s inequality) states that

$$\frac{1}{2} |P - Q|^2 \log e \leq D(P||Q)$$

where

$$D(P||Q) = E_P \left[ \log \frac{dP}{dQ} \right] = \int_{\mathcal{A}} dP(a) \log \frac{dP}{dQ}(a)$$

designates the relative entropy (a.k.a. the Kullback-Leibler divergence) from $P$ to $Q$, and

$$|P - Q| = \sup_{\mathcal{F} \in \mathcal{F}} |P(\mathcal{F}) - Q(\mathcal{F})|$$

is the total variation distance between $P$ and $Q$.

A “reverse Pinsker inequality” providing an upper bound on the relative entropy in terms of the total variation distance does not exist in general since we can find distributions that are arbitrarily close in total variation but with arbitrarily high relative entropy. Nevertheless, it is possible to introduce constraints under which such reverse Pinsker inequalities can be obtained. In the case where the probability measures $P$ and $Q$ are defined on a common discrete (i.e., finite or countable) set $\mathcal{A}$,

$$D(P||Q) = \sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{Q(a)},$$

$$|P - Q| = \sum_{a \in \mathcal{A}} |P(a) - Q(a)|,$$

One of the implications of (1) is that convergence in relative entropy implies convergence in total variation distance. The total variation distance is bounded $|P - Q| \leq 2$, whereas the relative entropy is an unbounded information measure.

Improved versions of Pinsker’s inequality were studied, e.g., in [9], [10], [14], [17], [22].

A “reverse Pinsker inequality” providing an upper bound on the relative entropy in terms of the total variation distance does not exist in general since we can find distributions that are arbitrarily close in total variation but with arbitrarily high relative entropy. Nevertheless, it is possible to introduce constraints under which such reverse Pinsker inequalities can be obtained. In the case of a finite alphabet $\mathcal{A}$, Csiszár and Talata [6, p. 1012] show that

$$D(P||Q) \leq \left( \frac{\log e}{Q_{\min}} \right) \cdot |P - Q|^2,$$

where

$$Q_{\min} \triangleq \min_{a \in \mathcal{A}} Q(a).$$

Recent applications of (6) can be found in [12, Appendix D] and [21, Lemma 7] for the analysis of the third-order asymptotics of the discrete memoryless channel with or without cost constraints.

In addition to $Q_{\min}$ in (7), the bounds in this paper involve

$$\beta_1 = \min_{a \in \mathcal{A}} \frac{Q(a)}{P(a)},$$

$$\beta_2 = \min_{a \in \mathcal{A}} \frac{P(a)}{Q(a)}$$

so, $\beta_1, \beta_2 \in [0, 1]$.

In this paper, Section 2 derives a reverse Pinsker inequality for probability measures defined on a common finite set, improving the bound in (6). The utility of this inequality is studied in Section 3, and it is extended in Section 4 to Rényi divergences of an arbitrary non-negative order.

2. A NEW REVERSE PINSDKER INEQUALITY FOR DISTRIBUTIONS ON A FINITE SET

The present section introduces a strengthened version of (6), followed by some remarks and an example.
A. Main Result and Proof

**Theorem 1.** Let $P$ and $Q$ be probability measures defined on a common finite set $\mathcal{A}$, and assume that $Q$ is strictly positive on $\mathcal{A}$. Then, the following inequality holds:

$$D(P\|Q) \leq \log \left(1 + \frac{P - Q^2}{2Q_{\min}}\right) - \frac{\beta_2 \log e}{2} \cdot |P - Q|^2$$

(10)

where $Q_{\min}$ and $\beta_2$ are given in (7) and (9), respectively.

**Proof.** Theorem 1 is proved by obtaining upper and lower bounds on the $\chi^2$-divergence from $P$ to $Q$

$$\chi^2(P\|Q) \triangleq \sum_{a \in \mathcal{A}} \frac{(P(a) - Q(a))^2}{Q(a)}.$$  

(12)

A lower bound follows by invoking Jensen’s inequality

$$\chi^2(P\|Q) = \sum_{a \in \mathcal{A}} \frac{P(a)^2}{Q(a)} - 1$$

(13)

$$= \sum_{a \in \mathcal{A}} P(a) \exp \left(\log \frac{P(a)}{Q(a)}\right) - 1$$

(14)

$$\geq \exp \left(\sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{Q(a)}\right) - 1$$

(15)

$$= \exp(D(P\|Q)) - 1.$$  

(16)

Alternatively, (16) can be obtained by combining the equality

$$\chi^2(P\|Q) = \exp(D_2(P\|Q)) - 1$$

(17)

with the monotonicity of the Rényi divergence $D_\alpha(P\|Q)$ in $\alpha$, which implies that $D_2(P\|Q) \geq D(P\|Q)$.

A refined version of (16) is derived in the following. The starting point is a refined version of Jensen’s inequality in (20, Lemma 1), generalizing a result from [7, Theorem 1], which leads to (see [20, Theorem 7])

$$\min_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \cdot D(Q\|P)$$

$$\leq \log(1 + \chi^2(P\|Q)) - D(P\|Q)$$

(18)

$$\leq \max_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \cdot D(Q\|P).$$

(19)

From (19) and the definition of $\beta_2$ in (9), we have

$$\chi^2(P\|Q) \geq \exp\left(D(P\|Q) + \beta_2 D(Q\|P)\right) - 1$$

(20)

$$\geq \exp\left(D(P\|Q) + \frac{\beta_2 \log e}{2} \cdot |P - Q|^2\right) - 1$$

(21)

where (20) follows from (18) and the definition of $\beta_2$ in (9), and (21) follows from Pinsker’s inequality (1). Note that the lower bound in (21) refines the lower bound in (16) since $\beta_2 \in [0, 1]$. An upper bound on $\chi^2(P\|Q)$ is derived as follows:

$$\chi^2(P\|Q) = \sum_{a \in \mathcal{A}} \frac{(P(a) - Q(a))^2}{Q(a)}$$

$$\leq \sum_{a \in \mathcal{A}} \frac{(P(a) - Q(a))^2}{Q_{\min}}$$

$$= \frac{|P - Q|}{Q_{\min}} \cdot \max_{a \in \mathcal{A}} |P(a) - Q(a)|$$

(22)

and, from (3),

$$|P - Q| \geq 2 \max_{a \in \mathcal{A}} |P(a) - Q(a)|.$$  

(24)

Combining (23) and (24) yields

$$\chi^2(P\|Q) \leq \frac{|P - Q|^2}{2Q_{\min}}.$$  

(25)

Finally, (10) follows by combining the upper and lower bounds on the $\chi^2$-divergence in (21) and (25).

□

**Remark 1.** It is easy to check that Theorem 1 strengthens the bound by Csiszár and Talata in (6) by at least a factor of 2 since upper bounding the logarithm in (10) gives

$$D(P\|Q) \leq \frac{(1 - \beta_2 Q_{\min}) \log e}{2Q_{\min}} \cdot |P - Q|^2.$$  

(26)

In the finite-alphabet case, we can obtain another upper bound on $D(P\|Q)$ as a function of the $\ell_2$ norm $\|P - Q\|_2$: $D(P\|Q) \leq \log \left(1 + \frac{\|P - Q\|_2^2}{Q_{\min}}\right) - \beta_2 \log e \cdot \frac{\|P - Q\|_2^2}{2}$

(27)

which follows by combining (21), (22), and $\|P - Q\|_2 \leq |P - Q|$. Using the inequality $\log(1 + x) \leq x \log e$ for $x \geq 0$ in the right side of (27), and also loosening this bound by ignoring the term $\beta_2 \log e \cdot \|P - Q\|_2$, we recover the bound

$$D(P\|Q) \leq \frac{|P - Q|^2}{Q_{\min}} \log e$$

(28)

which appears in the proof of Property 4 of [21, Lemma 7], and also used in [12, (174)].

**Remark 2.** The lower bounds on the $\chi^2$-divergence in (16) and (21) improve the one in [6, Lemma 6.3] which states that $D(P\|Q) \leq \chi^2(P\|Q) \log e$.

**Remark 3.** Reverse Pinsker inequalities have been also derived in quantum information theory ([1], [2]), providing upper bounds on the relative entropy of two quantum states as a function of the trace norm distance when the minimal eigenvalues of the states are positive (c.f. [1, Theorem 6] and [2, Theorem 1]). These type of bounds are akin to (11); they are also inversely proportional to the minimal eigenvalue, similarly to the dependence of (11) in $Q_{\min}$.
3. Applications of Theorem 1

A. The Exponential Decay of the Probability for a Non-Typical Sequence

To exemplify the utility of Theorem 1, we bound the function

\[ L_\delta(Q) = \min_{P \notin T_\delta(Q)} D(P||Q) \]  

(29)

where we have denoted the subset of probability measures on \((A, \mathcal{F})\) which are \(\delta\)-close to \(Q\) as

\[ T_\delta(Q) = \left\{ P; \forall a \in A, \ |P(a) - Q(a)| \leq \delta \ Q(a) \right\} \]  

(30)

Note that \((a_1, \ldots, a_n)\) is strongly \(\delta\)-typical according to \(Q\) if its empirical distribution belongs to \(T_\delta(Q)\). According to Sanov’s theorem (e.g. [5, Theorem 11.4.1]), if the random variables are independent distributed according to \(Q\), then the probability that \((Y_1, \ldots, Y_n)\), is not \(\delta\)-typical vanishes exponentially with exponent \(L_\delta(Q)\).

To state the next result, we invoke the following notions from [14]. Given a probability measure \(Q\), its balance coefficient is given by

\[ \beta_Q = \inf_{A \in \mathcal{F}, \ Q(A) \geq \frac{1}{2}} Q(A). \]  

(31)

The function \(\phi: (0, \frac{1}{2}) \rightarrow \left[ \frac{1}{2} \log e, \infty \right)\) is given by

\[ \phi(p) = \begin{cases} \frac{1}{p \log e} \log \left( \frac{1-p}{p} \right), & p \in \left(0, \frac{1}{2}\right), \\ \frac{1}{2} \log e, & p = \frac{1}{2}. \end{cases} \]  

(32)

Theorem 2. If \(Q_{\text{min}} > 0\), then

\[ \phi(1 - \beta_Q) Q_{\text{min}}^2 \delta^2 \leq L_\delta(Q) \leq \log \left( 1 + 2Q_{\text{min}} \delta^2 \right) \]  

(33)

(34)

where (34) holds if \(\delta \leq Q_{\text{min}}^{-1} - 1\).

Proof. Ordentlich and Weinberger [14, Section 4] show the refinement of Pinsker’s inequality:

\[ \phi(1 - \beta_Q) |P - Q|^2 \leq D(P||Q). \]  

(35)

Note that if \(Q_{\text{min}} > 0\) then \(\beta_Q \leq 1 - Q_{\text{min}} < 1\), and therefore \(\phi(1 - \beta_Q)\) is well defined and finite. If \(P \notin T_\delta(Q)\) the simple bound

\[ |P - Q| > \delta Q_{\text{min}} \]  

(36)

together with (35) yields (33).

The upper bound (34) follows from (11) and the fact that if \(\delta \leq Q_{\text{min}}^{-1} - 1\), then

\[ \min_{P \notin T_\delta(Q)} |P - Q| = 2\delta Q_{\text{min}}. \]  

(37)

If \(\delta \leq Q_{\text{min}}^{-1} - 1\), the ratio between the upper and lower bounds in (34), satisfies

\[ \frac{1}{Q_{\text{min}}} \cdot \frac{\log e}{2 \phi(1 - \beta_Q)} \cdot \frac{\log (1 + 2Q_{\text{min}} \delta^2)}{\frac{1}{2} \log e Q_{\text{min}} \delta^2} \leq \frac{4}{Q_{\text{min}}} \]  

(38)

where (38) follows from the fact that its second and third factors are less than or equal to 1 and 4, respectively. Note that the bounds in (33) and (34) scale like \(\delta^2\) for \(\delta \approx 0\).

B. Distance from Equiprobable

If \(P\) is a distribution on a finite set \(A\), \(H(P)\) gauges the “distance” from \(U\), the equiprobable distribution, since

\[ H(P) = \log |A| - D(P||U). \]  

(39)

Thus, it is of interest to explore the relationship between \(H(P)\) and \(|P - U|\). Particularizing (1), [4, (2.2)] (see also [24, pp. 30–31]), and (11) we obtain

\[ |P - U| \leq \sqrt{\frac{2}{\log e} \cdot (\log |A| - H(P))}, \]  

(40)

\[ |P - U| \leq \sqrt{2 \left( 1 - \frac{1}{|A|} \right) \cdot \exp(H(P))}, \]  

(41)

\[ |P - U| \geq \sqrt{2 \left( \exp(-H(P)) - \frac{1}{|A|} \right)}, \]  

(42)

respectively.

Fig. 1. Bounds on \(|P - U|\) as a function of \(H(P)\) for \(|A| = 4\), and \(|A| = 16\). The point \((H(P), |P - U|) = (0, 2(1 - |A|^{-1}))\) is depicted on the \(y\)-axis. In the curves of the two plots, the bounds (a), (b) and (c) refer, respectively, to (40), (41) and (42).
The bounds in (40)–(42) are illustrated for $|A| = 4, 16$ in Figure 1. For $H(P) = 0$, $|P - U| = 2(1 - |A|^{-1})$ is shown for reference in Figure 1, as the cardinality of the alphabet increases, the gap between $|P - U|$ and its upper bound is reduced (and this gap decays asymptotically to zero).

4. EXTENSION OF THEOREM 1 TO RÉNYI DIVERGENCES

Definition 1. The Rényi divergence of order $\alpha \in [0, \infty]$ from $P$ to $Q$ is defined for $\alpha \in (0, 1) \cup (1, \infty)$ as

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha - 1} \log \left( \sum_{a \in A} P^\alpha(a) Q^{1-\alpha}(a) \right).$$

Recall that $D_1(P\|Q) \triangleq D(P\|Q)$ is defined to be the analytic extension of $D_{\alpha}(P\|Q)$ at $\alpha = 1$ (if $D(P\|Q) < \infty$, L'Hôpital's rule gives that $D(P\|Q) = \lim_{\alpha \downarrow 1} D_\alpha(P\|Q)$).

The extreme cases of $\alpha = 0, \infty$ are defined as follows:

- If $\alpha = 0$ then $D_0(P\|Q) = -\log Q(\text{Support}(P))$,
- If $\alpha = +\infty$ then $D_\infty(P\|Q) = \log \left( \sup_{a \in A} \frac{P(a)}{Q(a)} \right)$.

 Pinsker's inequality was extended by Gilardoni [10] for a Rényi divergence of order $\alpha \in (0, 1]$ (see also [8, Theorem 30]), and it gets the form

$$\frac{\alpha}{2} |P - Q|^2 \log e \leq D_\alpha(P\|Q).$$

A tight lower bound on the Rényi divergence of order $\alpha > 0$ as a function of the total variation distance is given in [19], which is consistent with Vajda's tight lower bound for $f$-divergences in [23, Theorem 3].

Motivated by these findings, we extend the upper bound on the relative entropy in Theorem 1 to Rényi divergences of an arbitrary order.

Theorem 3. Assume that $P, Q$ are strictly positive with minimum masses denoted by $P_{\text{min}}$ and $Q_{\text{min}}$, respectively. Let $\beta_1$ and $\beta_2$ be given in (8) and (9), respectively, and abbreviate $\delta \triangleq \frac{1}{2} |P - Q| \in [0, 1]$. Then, the Rényi divergence of order $\alpha \in [0, \infty]$ satisfies

$$D_\alpha(P\|Q) \leq \begin{cases} 
    f_1, & \alpha \in (2, \infty] \\
    f_2, & \alpha \in [1, 2] \\
    \min \{f_2, f_3, f_4\}, & \alpha = \frac{1}{2}, 1 \\
    \min \left\{2 \log \left( \frac{1}{1-\delta} \right), f_2, f_3, f_4 \right\}, & \alpha \in [0, \frac{1}{2}]
\end{cases}$$

where, for $\alpha \in [0, \infty]$,

$$f_1(\alpha, \beta_1, \delta) \triangleq \begin{cases}
\frac{1}{\alpha - 1} \log \left( 1 + \frac{\delta (\beta_1^\alpha - 1)}{1-\beta_1} \right), & \alpha \in [0, 1) \cup (1, \infty) \\
\frac{\delta}{1-\beta_1} \log \frac{1}{\beta_1}, & \alpha = 1, \\
\log \frac{1}{\beta_1}, & \alpha = \infty
\end{cases}$$

and, for $\alpha \in [0, 2]$,

$$f_2(\alpha, \beta_1, Q_{\text{min}}, \delta) \triangleq \min \left\{ f_1(\alpha, \beta_1, \delta), \log \left( 1 + \frac{2 \delta^2}{Q_{\text{min}}} \right) \right\}$$

and, for $\alpha \in [0, 1)$, $f_3$ and $f_4$ are given by

$$f_3(\alpha, P_{\text{min}}, \beta_1, \delta) \triangleq \left( \frac{\alpha}{1-\alpha} \right) \left[ \log \left( 1 + \frac{2 \delta^2}{P_{\text{min}}} \right) - 2 \beta_1 \delta^2 \log e \right],$$

$$f_4(\beta_2, Q_{\text{min}}, \delta) \triangleq \min \left\{ \log \left( 1 + \frac{2 \delta^2}{Q_{\text{min}}} \right) - 2 \beta_2 \delta^2 \log e, \right\} \log \left( 1 + \frac{\min \{\delta, 2 \delta^2\}}{Q_{\text{min}}} \right).$$

Proof. See [20, Section 7.C].

Remark 4. A simple bound, albeit looser than the one in Theorem 3 is

$$D_\alpha(P\|Q) \leq \log \left( 1 + \frac{|P - Q|}{2Q_{\text{min}}} \right)$$

which is asymptotically tight as $\alpha \to \infty$ in the case of a binary alphabet with equiprobable $Q$.

Example 1. Figure 2 illustrates the bound in (45), which is valid for all $\alpha \in [0, \infty]$ (see [20, Theorem 23]), and the upper bounds of Theorem 3 in the case of binary alphabets.

5. SUMMARY

We derive in this paper some “reverse Pinsker inequalities” for probability measures $P \ll Q$ defined on a common finite set, which provide lower bounds on the total variation distance $P - Q$ as a function of the relative entropy $D(P\|Q)$ under the assumption of a bounded relative information or $Q_{\text{min}} > 0$. More general results for an arbitrary alphabet are available in [20, Section 5].

In [20], we study bounds among various $f$-divergences, dealing with arbitrary alphabets and deriving bounds on the ratios of various distance measures. New expressions of the Rényi divergence in terms of the relative information spectrum are derived, leading to upper and lower bounds on the Rényi divergence in terms of the variational distance.
Fig. 2. The Rényi divergence $D_\alpha(P\|Q)$ for $P$ and $Q$ which are defined on a binary alphabet with $P(0) = Q(1) = 0.65$, compared to (a) its upper bound in (44), and (b) its upper bound in (45) (see [20, Theorem 23]). The two bounds coincide here when $\alpha \in (1, 1.291) \cup (2, \infty)$.

ACKNOWLEDGMENT

The work of I. Sason has been supported by the Israeli Science Foundation (ISF) under Grant 12/12, and the work of S. Verdú has been supported by the US National Science Foundation under Grant CCF-1016625, and in part by the Center for Science of Information, an NSF Science and Technology Center under Grant CCF-0939370.

REFERENCES


