An Improved Reverse Pinsker Inequality for Probability Distributions on a Finite Set

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

Abstract—A new upper bound on the relative entropy is derived as a function of the total variation distance for arbitrary probability distributions that are defined on a common finite set. The bound improves a previously reported bound by Csiszár and Talata. It is further extended, for probability distributions on a finite set, to a derivation of an upper bound on the Rényi divergence of an arbitrary non-negative order (including ∞) as a function of the total variation distance. An extended version of this paper, which also includes reverse Pinsker inequalities for general probability measures and further results and discussions, is available at http://arxiv.org/abs/1503.07118.

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

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 states that

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

where

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

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

$$|P - Q| = 2 \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \tag{2}$$

designates the total variation distance between $P$ and $Q$. 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 un-bounded information measure.

Inequality (1) is a.k.a. Pinsker’s inequality, although the analysis made by Pinsker [11] leads to a significantly looser bound where $\log e$ on the RHS of (1) is replaced by $\log \frac{\log e}{2}$ (see [16] Eq. (51)). Improved versions of Pinsker’s inequality were studied in, e.g., [7], [8], [12].

For any $\varepsilon > 0$, there exists a pair of probability distributions $P$ and $Q$ such that $|P - Q| \leq \varepsilon$ while $D(P||Q) = \infty$. Consequently, a reverse Pinsker inequality which provides an upper bound on the relative entropy in terms of the total variation distance does not hold. Nevertheless, under some conditions, such an inequality can be derived [16]. If $P \ll Q$, let

$$\beta_1^{-1} \triangleq \sup_{a \in \mathcal{A}} \frac{dP}{dQ}(a) \tag{3}$$

with the convention that $\beta_1 = 0$ if the relative information $i_P||Q \triangleq \log \frac{dP}{dQ}$ is unbounded from above. With $\beta_1 \leq 1$, as in [3], the following inequality holds (see [16] Theorem 7):

$$\frac{1}{2} |P - Q| \geq \left( \frac{1 - \beta_1}{\log e} \right) D(P||Q). \tag{4}$$

From (4), if the relative information is bounded from above, a reverse Pinsker inequality holds. This inequality has been recently used in the context of the optimal quantization of probability measures when the distortion is either characterized by the total variation distance or the relative entropy between the approximating and the original probability measures [2] Proposition 4. A refined version of inequality [4] has been introduced, as part of this work, in [14] Theorem 1.

In the special case where two probability mass functions $P$ and $Q$ are defined on a common discrete (i.e., finite or countable) set $\mathcal{A}$,

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

$$|P - Q| = \sum_{a \in \mathcal{A}} |P(a) - Q(a)| \triangleq |P - Q|_1.$$

Throughout this paper, we restrict our attention to probability mass functions $P$ and $Q$ defined on a finite set, and use the term probability distributions for $P$ and $Q$.

A restriction to probability distributions on a finite set $\mathcal{A}$ led in [4] p. 1012 and Lemma 6.3) to the following bound on the relative entropy in terms of the total variation distance:

$$D(P||Q) \leq \left( \frac{\log e}{Q_{\min}} \right) |P - Q|^2, \tag{5}$$

where $Q_{\min} \triangleq \min_{a \in \mathcal{A}} Q(a)$, suggesting a kind of a reverse Pinsker inequality for probability distributions on a finite set. A recent application of this bound has been exemplified in [9] Appendix D) and [15] Lemma 7) for the analysis of the third-order asymptotics of the discrete memoryless channel with or without cost constraints. The present paper improves the bound in [5], and generalizes it to Rényi divergences.
The Rényi divergence of order \( \alpha \) from \( P \) to \( Q \) is defined as
\[
D_{\alpha}(P||Q) \triangleq \frac{1}{\alpha - 1} \log \left( \sum_{a \in \mathcal{A}} P^\alpha(a) Q^{1-\alpha}(a) \right),
\]
\[\forall \alpha \in (0, 1) \cup (1, \infty). \tag{6}\]
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 \( D(P||Q) = \lim_{\alpha \to 1^-} D_\alpha(P||Q) \)).

The extreme cases of \( \alpha = 0, \infty \) are defined as follows:
- If \( \alpha = 0 \) then \( D_0(P||Q) = -\log Q(\text{Support}(P)) \) where \( \text{Support}(P) = \{ x \in \mathcal{X} : P(x) > 0 \} \) denotes the support of \( P \).
- If \( \alpha = +\infty \) then \( D_\infty(P||Q) = \log \left( \text{ess sup}_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \right) \) where \( \text{ess sup} P \) denotes the essential supremum of a function \( f \).

Pinsker’s inequality was extended by Gilardoni [3] for a Rényi divergence of order \( \alpha \in (0, 1] \) (see also [6] Theorem 30]), and it gets the form
\[
D_\alpha(P||Q) \geq \frac{\alpha \log e}{2} \cdot |P - Q|^2.
\]
An improved bound, providing the best lower bound on the Rényi divergence of order \( \alpha > 0 \) in terms of the total variation distance, has been recently introduced in [13] Section 2.

Motivated by these findings, our analysis extends the upper bound on the relative entropy to provide an upper bound on the Rényi divergence of orders \( \alpha \in (0, \infty) \) in terms of the total variation distance for distributions defined on a common finite set.

In this paper, Section 2 derives a reverse Pinsker inequality for probability distributions on a finite set. This inequality improves inequality [5] by Csiszár and Talata [4]. The new inequality is extended in Section 3 to Rényi divergences of an arbitrary non-negative order. Section 4 exemplifies the utility of a reverse Pinsker inequality in the context of typical sequences.

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

The present section introduces a strengthened version of inequality [5], 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:
\[
D(P||Q) \leq \log \left( 1 + \frac{|P - Q|^2}{2Q_{\min}} \right) - \frac{\beta_2 \log e}{2} \cdot |P - Q|^2 \tag{7}\]
where
\[
Q_{\min} \triangleq \min_{a \in \mathcal{A}} Q(a) > 0, \quad \beta_2 \triangleq \min_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \in [0, 1]. \tag{8}\]

**Remark 1.** The upper bound on the relative entropy in Theorem 1 improves the bound in [5]. The improvement in (7) is demonstrated as follows: let \( V \triangleq |P - Q| \), then the RHS of (7) satisfies
\[
\log \left( 1 + \frac{V^2}{2Q_{\min}} \right) - \frac{\beta_2 \log e}{2} \cdot V^2
\leq \log \left( 1 + \frac{V^2}{2Q_{\min}} \right)
\leq \frac{V^2 \log e}{2Q_{\min}}
\leq \frac{V^2 \log e}{Q_{\min}}.
\]
Hence, the upper bound on \( D(P||Q) \) in Theorem 1 can be loosened to (5).

**Proof:** Theorem 1 is proved by obtaining upper and lower bounds on the \( \chi^2 \)-divergence from \( P \) to \( Q \)
\[
\chi^2(P, Q) \triangleq \sum_{a \in \mathcal{A}} \frac{(P(a) - Q(a))^2}{Q(a)}.
\]
A lower bound follows by invoking Jensen’s inequality (or, alternatively, by combining [6] Eqs. (6) and (7)):
\[
\chi^2(P, Q) = \sum_{a \in \mathcal{A}} \frac{P(a)^2}{Q(a)} - 1
\geq \sum_{a \in \mathcal{A}} P(a) \exp \left( \log \frac{P(a)}{Q(a)} \right) - 1
\geq \exp \left( \sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{Q(a)} \right) - 1
= \exp(D(P||Q)) - 1. \tag{9}\]
A refined version of (9) is derived in the following. The starting point of its derivation relies on a refined version of Jensen’s inequality from [5] Theorem 1], which enables to get the inequality
\[
\min_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \cdot D(Q||P)
\leq \log(1 + \chi^2(P, Q)) - D(P||Q)
\leq \max_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \cdot D(Q||P). \tag{10}\]
Inequality (10) is proved in [14] Appendix. From the LHS of (10) and the definition of \( \beta_2 \) in (8), we have
\[
\chi^2(P, Q) \geq \exp \left( D(P||Q) + \beta_2 D(Q||P) \right) - 1
\geq \exp \left( D(P||Q) + \frac{\beta_2 \log e}{2} \cdot |P - Q|^2 \right) - 1 \tag{11}\]
where the last inequality relies on Pinsker’s lower bound on \( D(Q||P) \). Inequality (11) refines the lower bound in (9) since \( \beta_2 \in [0, 1] \), and they coincide in the worst case where \( \beta_2 = 0 \).
An upper bound on $\chi^2(P, Q)$ is derived as follows:

\[
\chi^2(P, Q) = \sum_{a \in \mathcal{A}} \frac{(P(a) - Q(a))^2}{Q(a)} \\
\leq \frac{\sum_{a \in \mathcal{A}} (P(a) - Q(a))^2}{\min_{a \in \mathcal{A}} Q(a)} \\
\leq \frac{\max_{a \in \mathcal{A}} |P(a) - Q(a)| \sum_{a \in \mathcal{A}} |P(a) - Q(a)|}{\min_{a \in \mathcal{A}} Q(a)} \\
= \frac{|P - Q| \max_{a \in \mathcal{A}} |P(a) - Q(a)|}{Q_{\min}} \\
\tag{12}
\]

and, from (2),

\[
|P - Q| \geq 2 \max_{a \in \mathcal{A}} |P(a) - Q(a)| \\
\tag{13}
\]

since, for every $a \in \mathcal{A}$, the 1-element set $\{a\}$ is included in the $\sigma$-algebra $\mathcal{F}$. Combining (12) and (13) gives that

\[
\chi^2(P, Q) \leq \frac{|P - Q|^2}{2Q_{\min}}. \\
\tag{14}
\]

Inequality (14) finally follows from the bounds on the $\chi^2$-divergence in (11) and (14).

**Corollary 1.** Under the same setting as in Theorem 1 we have

\[
D(P\|Q) \leq \log \left(1 + \frac{|P - Q|^2}{2Q_{\min}}\right). \\
\tag{15}
\]

*Proof:* This inequality follows from (7) and since $\beta_2 \geq 0$.

**Remark 2.** The combination of (9) with the second line of (12), without further loosening the upper bound on the $\chi^2$-divergence as is done in the third line of (12) and inequality (13), gives the following tighter upper bound on the relative entropy in terms of the Euclidean norm $|P - Q|_2$:

\[
D(P\|Q) \leq \log \left(1 + \frac{|P - Q|_2^2}{2Q_{\min}}\right). \\
\tag{16}
\]

This improves the upper bound on the relative entropy in the proofs of Property 4 of [15, Lemma 7] and [9, Appendix D]:

\[
D(P\|Q) \leq \frac{|P - Q|_2^2 \log e}{Q_{\min}}. \\
\]

Furthermore, avoiding the use of Jensen’s inequality in (9), gives the equality

\[
\chi^2(P, Q) = \exp(D_2(P\|Q)) - 1 \\
\tag{17}
\]

whose combination with the second line of (12) gives

\[
D_2(P\|Q) \leq \log \left(1 + \frac{|P - Q|_2^2}{Q_{\min}}\right). \\
\tag{18}
\]

Inequality (18) improves the tightness of the bound in (16). Note that (18) is satisfied with equality when $Q$ is an equiprobable distribution over a finite set.

**Remark 3.** Inequality (2) improves the lower bound on the $\chi^2$-divergence in [4, Lemma 6.3] which states that $\chi^2(P, Q) \geq D(P\|Q)$.

**Remark 4.** The upper bound on the relative entropy in (7) involves the parameter $\beta_2 \in [0, 1]$ as defined in (8). A non-trivial lower bound on $\beta_2$ can be used in conjunction with (7) for improving the upper bound in Corollary 1. We derive in the following a lower bound on $\beta_2$ for a given probability measure $Q$ and a given total variation distance $|P - Q|$, which can be used in conjunction with (7), to get an upper bound on the relative entropy $D(P\|Q)$. We have

\[
\beta_2 = \min_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \geq \frac{P_{\min}}{Q_{\max}} \\
\]

where

\[
P_{\min} = \min_{a \in \mathcal{A}} P(a), \quad Q_{\max} = \max_{a \in \mathcal{A}} Q(a). \\
\]

Note that, if $|P - Q| < Q_{\min}$ then

\[
P_{\min} \geq Q_{\min} - |P - Q| > 0. \\
\]

Let $(x)^+ = \max\{x, 0\}$, then

\[
\beta_2 \geq \frac{(Q_{\min} - |P - Q|)^+}{Q_{\max}}. \\
\tag{19}
\]

**Remark 5.** A related problem to Theorem 1 has been recently studied in [1]. Consider an arbitrary probability measure $Q$, and an arbitrary $\varepsilon \in [0, 2]$. The problem studied in [1] is the characterization of $D^*(\varepsilon, Q)$, defined to be the infimum of $D(P\|Q)$ over all probability measures $P$ that are at least $\varepsilon$-far away from $Q$ in total variation, i.e.,

\[
D^*(\varepsilon, Q) = \inf_{P : |P - Q|_2 \geq \varepsilon} D(P\|Q), \quad \varepsilon \in [0, 2]. \\
\]

Note that $D(P\|Q) < \infty$ yields that $\text{Supp}(P) \subseteq \text{Supp}(Q)$. From Sanov’s theorem (see [3, Theorem 11.4.1]), $D^*(\varepsilon, Q)$ is equal to the asymptotic exponential decay of the probability that the total variation distance between the empirical distribution of a sequence of i.i.d. random variables and the true distribution $(Q)$ is more than a specified value $\varepsilon$. This problem is further discussed in [14, Section 3-B] in light of Theorem 1.

**B. Example: Total Variation Distance From the Equiprobable Distribution**

Let $\mathcal{A}$ be a finite set, and let $U$ be the equiprobable probability measure on $\mathcal{A}$ (i.e., $U(a) = \frac{1}{|\mathcal{A}|}$ for every $a \in \mathcal{A}$). The relative entropy of an arbitrary distribution $P$ on $\mathcal{A}$ with respect to the equiprobable distribution satisfies

\[
D(P\|U) = \log |\mathcal{A}| - H(P). \\
\tag{20}
\]

From Pinsker’s inequality (1), the following upper bound on the total variation distance holds:

\[
|P - U| \leq \sqrt{\frac{2}{\log e} \cdot (\log |\mathcal{A}| - H(P))}. \\
\tag{21}
\]
From [17] Theorem 2.51, for all probability measures $P$ and $Q$, 
\[ |P - Q| \leq 2 \sqrt{1 - \exp(-D(P||Q))} \]
which gives the second upper bound 
\[ |P - U| \leq 2 \sqrt{1 - \frac{1}{|A|} \cdot \exp(H(P))} . \] (22)

From Theorem 1 and (19), we have 
\[ D(P||U) \leq \log \left( 1 + \frac{|A|}{2} \cdot |P - U|^2 \right) \]
\[ - \left( \frac{|A|}{2} \log e \right) \left( \frac{1}{|A|} - |P - U| \right)^+ |P - U|^2 . \]
A loosening of the latter bound by removing its second non-negative term on the RHS of this inequality, in conjunction with (20), leads to the following closed-form expression for the lower bound on the total variation distance:
\[ |P - U| \geq \sqrt{2 \left( \exp(-H(P)) - \frac{1}{|A|} \right)} . \] (23)

Let $H(P) = \beta \log |A|$, so $\beta \in [0, 1]$. From (21), (22) and (24), it follows that 
\[ \sqrt{2 \left[ \left( \frac{1}{|A|} \right)^\beta - \frac{1}{|A|} \right]} \]
\[ \leq |P - U| \]
\[ \leq \min \left\{ \sqrt{2(1 - \beta) \log |A|}, 2\sqrt{1 - |A|^\beta - 1} \right\} . \] (24)

As expected, if $\beta = 1$, both upper and lower bounds are equal to zero (since $D(P||U) = 0$). The lower bound on the LHS of (24) improves the lower bound on the total variation distance which follows from (5):
\[ |P - U| \geq \sqrt{\frac{(1 - \beta) \ln |A|}{|A|}} . \] (25)

For example, for a set of size $|A| = 1024$ and $\beta = 0.5$, the improvement in the new lower bound on the total variation distance is from 0.0582 to 0.2464.

Note that if $\beta \to 0$ (i.e., $P$ is far in relative entropy from the equiprobable distribution), and the set $\mathcal{A}$ stays fixed, the ratio between the upper and lower bounds in (24) tends to $\sqrt{2}$. On the other hand, in this case, the ratio between the upper and the looser lower bound in (25) tends to
\[ 2 \sqrt{\frac{|A| - 1}{\ln |A|}} , \]
which can be made arbitrarily large for a sufficiently large set $\mathcal{A}$.

3. Extension of Theorem 1 to Rényi Divergences

**Theorem 2.** Let $P$ and $Q$ be probability measures on a common finite set $\mathcal{A}$, and assume that $P, Q$ are strictly positive. Let $\epsilon \triangleq |P - Q|$ (recall that $\epsilon \in [0, 2]$), $\epsilon' \triangleq \min\{1, \epsilon\}$, and
\[ P_{\min} \triangleq \min_{a \in \mathcal{A}} P(a) , \quad Q_{\min} \triangleq \min_{a \in \mathcal{A}} Q(a) . \]
Then, the Rényi divergence of order $\alpha \in [0, \infty)$ satisfies
\[ D_\alpha(P||Q) \leq \begin{cases} \log \left( 1 + \frac{\epsilon}{2Q_{\min}} \right) , & \text{if } \alpha \in (2, \infty] \\ \log \left( 1 + \frac{\epsilon'}{2Q_{\min}} \right) , & \text{if } \alpha \in [1, 2] \\ \min \{ f_1, f_2 \} , & \text{if } \alpha \in \left( \frac{1}{2}, 1 \right) \\ \min \left\{ -2 \log \left( \frac{1}{2 \epsilon} \right), f_1, f_2 \right\} , & \text{if } \alpha \in \left[ 0, \frac{1}{2} \right] \end{cases} , \]
where, for $\alpha \in [0, 1)$,
\[ f_1 \triangleq \frac{\alpha}{1 - \alpha} \left[ \log \left( 1 + \frac{\epsilon^2}{2Q_{\min}^2} \right) - \left( \frac{P_{\min} \log e}{2} \right) \epsilon^2 \right] , \]
\[ f_2 \triangleq \log \left( 1 + \frac{\epsilon'}{2Q_{\min}} \right) - \left( \frac{P_{\min} \log e}{2} \right) \epsilon^2 . \]

**Proof:** See [14] Section 4.

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

Let $U^N = (U_1, \ldots, U_N)$ be a sequence of i.i.d. symbols that are emitted by a memoryless and stationary source with distribution $Q$ and a finite alphabet $\mathcal{A}$. Let $|\mathcal{A}| = r < \infty$ denote the cardinality of the source alphabet, and assume that all symbols are emitted with positive probability (i.e., $Q_{\min} \triangleq \min_{a \in \mathcal{A}} Q(a) > 0$). The empirical probability distribution of the emitted sequence $\hat{P}_{U^N}$ is given by
\[ \hat{P}_{U^N}(a) \triangleq \frac{1}{N} \sum_{k=1}^{N} 1\{U_k = a\} , \quad \forall a \in \mathcal{A} . \]

For an arbitrary $\delta > 0$, let the $\delta$-typical set be defined as 
\[ T_Q(\delta) \triangleq \left\{ u^N \in \mathcal{A}^N : |\hat{P}_{U^N}(a) - Q(a)| < \delta Q(a) , \forall a \in \mathcal{A} \right\} , \]
i.e., the empirical distribution of every symbol in an $N$-length $\delta$-typical sequence deviates from the true distribution of this symbol by a fraction of less than $\delta$. Consequently, the complementary of (26) is given by
\[ T_Q(\delta^c) \triangleq \left\{ u^N \in \mathcal{A}^N : \exists a \in \mathcal{A} \left| \hat{P}_{U^N}(a) - Q(a) \right| \geq \delta Q(a) \right\} . \]

From Sanov’s theorem (see [5] Theorem 11.1.4.1)), the asymptotic exponential decay of the probability that a sequence $U^N$ is not $\delta$-typical, for a specified $\delta > 0$, is given by
\[ \lim_{N \to \infty} \frac{1}{N} \log Q^N(T_Q(\delta^c)) = \min_{P \in \mathcal{P}_Q} D(P||Q) \] (27)
where
\[ P_Q \triangleq \left\{ P \text{ is a probability measure on } (A, \mathcal{F}) : \exists a \in A, \; |P(a) - Q(a)| \geq \delta Q(a) \right\}. \] (28)

We obtain the following explicit upper and lower bounds on the exponential decay rate on the RHS of (27). The emphasis is on the upper bound, which is based on Theorem 1 and we first introduce the lower bound for completeness. The derivation of the lower bound is similar to the analysis in [10, Section 4]; note, however, that there is a difference between the \( \delta \)-typicality in [10, Eq. (19)] and the way it is defined in (26). The probability-dependent refinement of Pinsker’s inequality (see [10, Theorem 2.1]) states that
\[ D(P\|Q) \geq \varphi(\pi_Q) |P - Q|^2 \] (29)

where
\[ \pi_Q \triangleq \max_{A \in \mathcal{F}} \min \{Q(A), 1 - Q(A)\} \leq \frac{1}{2} \] (30)
and
\[ \varphi(p) = \begin{cases} \frac{1}{4(1-2p)} \log \left( \frac{1-p}{p} \right), & \text{if } p \in [0, \frac{1}{2}), \\ \log e \frac{2}{p}, & \text{if } p = \frac{1}{2} \end{cases} \] (31)
is a monotonic decreasing and continuous function. Hence, \( \varphi(\pi_Q) \geq \frac{\log e}{2} \), and (29) forms a probability-dependent refinement of Pinsker’s inequality [10]. From (28) and (29),
\[ \min_{P \in P_Q} D(P\|Q) \geq \varphi(\pi_Q) \min_{P \in P_Q} |P - Q|^2 \]
\[ = \varphi(\pi_Q) \left( \min_{a \in A} Q(a) \right)^2 \]
\[ = \varphi(\pi_Q) Q_{\text{min}}^2 \delta^2 \triangleq E_L \] (32)
\[ \geq \left( Q_{\text{min}} \log e \frac{2}{\delta^2} \right) \delta^2 \] (33)
where the transition from (32) to (33) follows from the global lower bound on \( \varphi(\pi_Q) \).

We derive in the following an upper bound on the asymptotic exponential decay rate in (27):
\[ \min_{P \in P_Q} D(P\|Q) \overset{a}{\leq} \min_{P \in P_Q} \left\{ \log \left( 1 + \frac{|P - Q|^2}{2Q_{\text{min}}} \right) \right\} \]
\[ = \log \left( 1 + \frac{\min_{P \in P_Q} |P - Q|^2}{2Q_{\text{min}}} \right) \overset{b}{=} \log \left( 1 + \frac{\min_{a \in A} (\delta Q(a))^2}{2Q_{\text{min}}} \right) \]
\[ = \log \left( 1 + \frac{Q_{\text{min}} \delta^2}{2} \right) \triangleq E_U \] (34)
where inequality (a) follows from (15), and equality (b) follows from (28).

The ratio between the upper and lower bounds on the asymptotic exponent in (27) is
\[ \frac{E_U}{E_L} = \frac{1}{Q_{\text{min}}} \log e \frac{2}{\varphi(\pi_Q)} \log \left( 1 + \frac{Q_{\text{min}} \delta^2}{2} \right) \in [1, \frac{1}{Q_{\text{min}}} \] which follows from the fact that the second and third multiplicands on the RHS are both less than or equal to 1. Note that both bounds in (32) and (33) scale like \( \delta^2 \) for \( \delta \approx 0 \).

ACKNOWLEDGMENT
Sergio Verdú is gratefully acknowledged for his earlier results in [16] that motivated this work, and for providing a draft of [17]. Vincent Tan is acknowledged for pointing out [15] and suggesting a simplified proof of (13). A discussion with Georg Böcherer and Bernhard Geiger on their paper [2] has been stimulating along the writing of this manuscript. This work has been supported by the Israeli Science Foundation (ISF), grant number 12/12.

REFERENCES
[1] D. Berend, P. Harremoës and A. Kontorovich, “Minimum KL-divergence on complements of \( L_1 \) balls,” IEEE Trans. on Information Theory, vol. 60, no. 6, pp. 3172–3177, June 2014.
[2] G. Böcherer and B. C. Geiger, “Optimal quantization for distribution synthesis,” March 2015. [Online]. Available: http://arxiv.org/abs/1307.6843.
[3] T. M. Cover and J. A. Thomas, Elements of Information Theory, second edition, John Wiley & Sons, 2006.
[4] I. Csiszár and Z. Talata, “Context tree estimation for not necessarily finite memory processes, via BIC and MDL,” IEEE Trans. on Information Theory, vol. 52, no. 3, pp. 1007–1016, March 2006.
[5] S. S. Dragomir, “Bounds for the normalized Jensen functional,” Bulletin of the Australian Mathematical Society, vol. 74, no. 3, pp. 471–478, 2006.
[6] T. van Erven and P. Harremoës, “Rényi divergence and Kullback-Leibler divergence,” IEEE Trans. on Information Theory, vol. 60, no. 7, pp. 3797–3820, July 2014.
[7] A. A. Fedotov, P. Harremoës and F. Topsoe, “Refinements of Pinsker’s inequality,” IEEE Trans. on Information Theory, vol. 49, no. 6, pp. 1491–1498, June 2003.
[8] G. L. Gilardoni, “On Pinsker’s and Vajda’s type inequalities for Csiszár’s f-divergences,” IEEE Trans. on Information Theory, vol. 56, no. 11, pp. 5377–5386, November 2010.
[9] V. Kostina and S. Verdú, “Channels with cost constraints: strong converse and dispersion,” to appear in the IEEE Trans. on Information Theory, vol. 61, no. 5, May 2015.
[10] E. Ordentlich and M. J. Weinberger, “A distribution dependent refinement of Pinsker’s inequality,” IEEE Trans. on Information Theory, vol. 51, no. 5, pp. 1836–1840, May 2005.
[11] M. S. Pinsker, Information and Information Stability of Random Variables and Random Processes, San-Francisco: Holden-Day, 1964, originally published in Russian in 1960.
[12] M. D. Reid and R. C. Williamson, “Information, divergence and risk for binary experiments,” Journal of Machine Learning Research, vol. 12, no. 3, pp. 731–817, March 2011.
[13] I. Sason, “On the Rényi divergence and the joint range of relative entropies,” March 2015. [Online]. Available at http://arxiv.org/abs/1501.03616
[14] I. Sason, “On reverse Pinsker inequalities,” March 2015. [Online]. Available at http://arxiv.org/abs/1503.07118
[15] M. Tomamichel and V. Y. F. Tan, “A tight upper bound for the third-order asymptotics for most discrete memoryless channels,” IEEE Trans. on Information Theory, vol. 59, no. 11, pp. 7041–7051, November 2013.
[16] S. Verdú, “Total variation distance and the distribution of the relative information,” Proceedings of the Information Theory and Applications Workshop, pp. 499–501, San-Diego, California, USA, February 2014.
[17] S. Verdú, Information Theory, in preparation.