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

Article · March 2015

Source: arXiv

2 authors:

Igal Sason
Technion - Israel Institute of Technology

Sergio Verdú
Princeton University

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 for arbitrary probability distributions that are defined on a common finite set. The bound is expressed in terms of the total variation distance, and it improves a previously reported bound by Csiszár and Talata. It is further extended to Rényi divergences of an arbitrary non-negative order (including ∞).**

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

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

designates the relative entropy (also known as (a.k.a.) KL 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 non-negative 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 $\frac{\log e}{2}$ on the RHS of (1) is replaced by $\frac{\log e}{408}$ (see [17, Eq. (51)]). Improved versions of Pinsker's inequality were studied in [5], [6], [7], [12], [16].

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 [17]. If $P \ll Q$, let

$$\beta_1^{-1} \triangleq \sup_{a \in \mathcal{A}} \frac{\mathrm{d}P}{\mathrm{d}Q}(a) \tag{3}$$

with the convention that $\beta_1 = 0$ if the relative information is unbounded from above. With $\beta_1 \leq 1$, as it is defined in (3), the following inequality holds (see [17, Theorem 7]):

$$\begin{aligned} \frac{1}{2}|P-Q| &\geq \left(\frac{1-\beta_1}{\log \frac{1}{\beta_1}}\right) D(P\|Q) \\ &\geq \left(\frac{\sqrt{\beta_1}}{\log e}\right) D(P\|Q). \end{aligned}$$

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

$$\begin{aligned} D(P\|Q) &= \sum_{a \in \mathcal{A}} P(a) \, \log \frac{P(a)}{Q(a)}, \\ |P-Q| &= \sum_{a \in \mathcal{A}} \bigl|P(a) - Q(a)\bigr|. \end{aligned}$$

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 $\mathcal{A}$ has led in [2, p. 1012 and Lemma 6.3] (see [18, Theorem 2.55]) to the following upper bound on the relative entropy in terms of the total variation distance:

$$D(P\|Q) \leq \left(\frac{\log e}{Q_{\min}}\right) \cdot |P-Q|^2, \tag{4}$$

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 (4), and generalizes it to Rényi divergences.

The Rényi divergence of order α from P to Q is defined as

$$\begin{aligned} D_\alpha(P||Q) \triangleq \frac{1}{\alpha - 1} \, \log\left(\sum_{x \in \mathcal{X}} P^\alpha(x) \, Q^{1-\alpha}(x)\right), \\ \forall \, \alpha \in (0,1) \cup (1, \infty). \end{aligned} \tag{5}$$

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 $D_\infty(P||Q) = \log\left(\text{ess sup}\, \frac{P}{Q}\right)$ where ess sup f denotes the essential supremum of a function f.

Pinsker's inequality was extended by Gilardoni [7] for a Rényi divergence of order $\alpha \in (0, 1]$ (see also [4, 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 (4) by Csiszár and Talata [2]. The new inequality is extended in Section 3 to Rényi divergences of an arbitrary non-negative order.

2. A New Reverse Pinsker Inequality for Distributions on a Finite Set

The present section introduces a strengthened version of inequality (4) (see Theorem 1), followed by a discussion and an example.

A. Main Result and Proof

Theorem 1. Let P and Q be distributions defined on a common finite set $\mathcal{A}$. Then

$$D(P\|Q) \leq \log\left(1 + \frac{|P - Q|^2}{2Q_{\min}}\right) \tag{6}$$

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

Remark 1. The upper bound on the relative entropy in Theorem 1 improves the bound in (4). The improvement in (6) is due to the logarithm introduced on the RHS of (6) (recall that $\log(1 + u) \leq u \log e$ for $u \geq 0$), and the introduction of the factor 2 in the denominator of the RHS of (6).

Proof: Theorem 1 is proved by obtaining upper and lower bounds on the χ^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:

$$\begin{aligned}
\chi^2(P, Q) &= \sum_{a \in \mathcal{A}} \frac{(P(a) - Q(a))^2}{Q(a)} \\
&= \sum_{a \in \mathcal{A}} \frac{P(a)^2}{Q(a)} - 2\sum_{a \in \mathcal{A}} P(a) + \sum_{a \in \mathcal{A}} Q(a) \\
&= \sum_{a \in \mathcal{A}} \frac{P(a)^2}{Q(a)} - 1 \\
&= \sum_{a \in \mathcal{A}} P(a) \exp\left(\log \frac{P(a)}{Q(a)}\right) - 1 \\
&\geq \exp\left(\sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{Q(a)}\right) - 1 \\
&= \exp\big(D(P\|Q)\big) - 1.
\end{aligned} \tag{7}$$

On the other hand, we have

$$\begin{aligned}
\chi^2(P, Q) &= \sum_{a \in \mathcal{A}} \frac{(P(a) - Q(a))^2}{Q(a)} \\
&\leq \frac{\sum_{a \in \mathcal{A}} \big(P(a) - Q(a)\big)^2}{\min_{a \in \mathcal{A}} Q(a)} \\
&\leq \frac{\max_{a \in \mathcal{A}} |P(a) - Q(a)| \, \sum_{a \in \mathcal{A}} \big|P(a) - Q(a)\big|}{\min_{a \in \mathcal{A}} Q(a)} \\
&= \frac{|P - Q| \; \max_{a \in \mathcal{A}} |P(a) - Q(a)|}{Q_{\min}}
\end{aligned} \tag{8}$$

and, from (2),

$$|P - Q| \geq 2 \max_{a \in \mathcal{A}} |P(a) - Q(a)| \tag{9}$$

since, for every $a \in \mathcal{A}$, the 1-element set $\{a\}$ is included in the σ-algebra $\mathcal{F}$. Combining (8) and (9) gives that

$$\chi^2(P, Q) \leq \frac{|P - Q|^2}{2Q_{\min}}. \tag{10}$$

Inequality (6) finally follows from the upper and lower bounds on the χ^2-divergence in inequalities (7) and (10), respectively. ■

B. Discussion

In the following, we discuss Theorem 1 and its proof in connection to some previously reported related results.

Remark 2. Inequality (7) improves the lower bound on the χ^2-divergence in [2, Lemma 6.3] which states that $\chi^2(P, Q) \geq D(P\|Q)$. A refined version of Jensen's inequality ([3, Theorem 1]) leads to the following strengthening of the bound in (7) (see [14, Corollary 1]):

$$\begin{aligned}
&\min_{x \in \mathcal{A}} \frac{P(x)}{Q(x)} \cdot D(Q||P) \\
&\leq \log\big(1 + \chi^2(P, Q)\big) - D(P||Q) \\
&\leq \max_{x \in \mathcal{A}} \frac{P(x)}{Q(x)} \cdot D(Q||P).
\end{aligned}$$

Remark 3. A related problem to Theorem 1 has been recently studied in [1]. Consider an arbitrary distribution 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 distributions P that are at least ε-far away from Q in total variation, i.e.,

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

From Sanov's theorem, $D^*(\varepsilon, Q)$ 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 ε (see also [10, Section 3]). Upper and lower bounds on $D^*(\varepsilon, Q)$ have been introduced in [1, Theorem 1], in terms of the balance coefficient $\beta \geq \frac{1}{2}$ that is defined as

$$\beta \triangleq \inf \left\{ x \in \big\{Q(A)\colon A \in \mathcal{F}\big\} \colon x \geq \frac{1}{2} \right\}.$$

It has been demonstrated in [1, Theorem 1] that

$$D^*(\varepsilon, Q) = C\varepsilon^2 + O(\varepsilon^3) \tag{11}$$

where

$$\frac{1}{4(2\beta-1)} \log\left(\frac{\beta}{1-\beta}\right) \leq C \leq \frac{\log e}{8\beta(1-\beta)}.$$

If the support of the distribution Q is a finite set $\mathcal{A}$, Theorem 1 implies that

$$D^*(\varepsilon, Q) \leq \log\left(1 + \frac{\varepsilon^2}{2Q_{\min}}\right) = C_1\varepsilon^2 + O(\varepsilon^4) \tag{12}$$

where

$$C_1 = \frac{\log e}{2Q_{\min}}.$$

Similarly to (11), the same quadratic scaling of $D^*(\varepsilon, Q)$ holds for small values of ε, but with different coefficients.

C. 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

$$D(P\|U) = \log|\mathcal{A}| - H(P).$$

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

$$|P-U| \leq \sqrt{\frac{2}{\log e} \cdot \big(\log|\mathcal{A}| - H(P)\big)}. \tag{13}$$

From [18, Theorem 2.51], for all distributions P and Q,

$$|P-Q| \leq 2\sqrt{1 - \exp\big(-D(P\|Q)\big)}$$

which gives the second upper bound

$$|P-U| \leq 2\sqrt{1 - \frac{1}{|\mathcal{A}|} \cdot \exp\big(H(P)\big)}. \tag{14}$$

From Theorem 1, we have

$$D(P\|U) \leq \log\left(1 + \frac{|\mathcal{A}|}{2} \cdot |P-U|^2\right)$$

which implies the following lower bound on the total variation distance:

$$|P-U| \geq \sqrt{2\left(\exp\big(-H(P)\big) - \frac{1}{|\mathcal{A}|}\right)}. \tag{15}$$

Let $H(P) = \beta \log|\mathcal{A}|$, which implies that $\beta \in [0,1]$. From (13), (14) and (15), it follows that

$$\begin{aligned} &\sqrt{2\left[\left(\frac{1}{|\mathcal{A}|}\right)^{\beta} - \frac{1}{|\mathcal{A}|}\right]} \\ &\leq |P-U| \\ &\leq \min\left\{\sqrt{2(1-\beta)\ln|\mathcal{A}|},\ 2\sqrt{1 - |\mathcal{A}|^{\beta-1}}\right\}. \end{aligned} \tag{16}$$

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 (16) improves the lower bound on the total variation distance which follows from (4):

$$|P-U| \geq \sqrt{\frac{(1-\beta)\ln|\mathcal{A}|}{|\mathcal{A}|}} \tag{17}$$

For example, for a set of size $|\mathcal{A}| = 1024$ and $\beta = 0.5$, the improvement in the new lower bound on the total variation distance is from 0.0582 to 0.2461.

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 (16) tends to $\sqrt{2}$. On the other hand, in this case, the ratio between the upper and the looser lower bound in (17) tends to $2\sqrt{\frac{|\mathcal{A}|-1}{\ln|\mathcal{A}|}}$, so it can be made arbitrarily large for a sufficiently large set $\mathcal{A}$.

3. Extension of Theorem 1 to the Rényi Divergence

The present section extends Theorem 1 to Rényi divergences of an arbitrary order $\alpha \in [0, \infty]$ (i.e., it relies on Theorem 1 to provide a generalization of the special case where $\alpha = 1$). The generalized inequality in Theorem 2 is exemplified.

A. Main Result

The following theorem provides a kind of a reverse Pinsker inequality where the Rényi divergence of an arbitrary order $\alpha \in [0, \infty]$ is upper bounded in terms of the total variation distance for probability distributions defined on a common finite set.

Theorem 2. Let P and Q be distributions on a common finite set $\mathcal{A}$. Let $\varepsilon \triangleq |P-Q|$ (recall that $\varepsilon \in [0,2]$), $\varepsilon' \triangleq \min\{1, \varepsilon\}$, 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

$$D_\alpha(P\|Q) \le \begin{cases} \log\left(1 + \frac{\varepsilon}{2Q_{\min}}\right) & \text{if } \alpha > 2 \\ \log\left(1 + \frac{\varepsilon\,\varepsilon'}{2Q_{\min}}\right) & \text{if } \alpha \in [1,2] \\ \min\Big\{\left(\frac{\alpha}{1-\alpha}\right)\log\left(1 + \frac{\varepsilon^2}{2P_{\min}}\right), & \\ \qquad \log\left(1 + \frac{\varepsilon\,\varepsilon'}{2Q_{\min}}\right)\Big\} & \text{if } \alpha \in \left(\frac{1}{2}, 1\right) \\ \min\Big\{-2\log\left(1 - \frac{\varepsilon}{2}\right), & \\ \qquad \left(\frac{\alpha}{1-\alpha}\right)\log\left(1 + \frac{\varepsilon^2}{2P_{\min}}\right), & \\ \qquad \log\left(1 + \frac{\varepsilon\,\varepsilon'}{2Q_{\min}}\right)\Big\} & \text{if } \alpha \in \left[0, \frac{1}{2}\right] \end{cases} \tag{18}$$

Proof: The Rényi divergence of order ∞ satisfies (see, e.g., [4, Theorem 6])

$$D_\infty(P||Q) = \log\left(\text{ess sup}\,\frac{P}{Q}\right).$$

Since, by assumption, the probability distributions P and Q are defined on a common finite set $\mathcal{A}$

$$\begin{aligned} D_\infty(P||Q) &= \log\left(\max_{a\in\mathcal{A}} \frac{P(a)}{Q(a)}\right) \\ &= \log\left(1 + \max_{a\in\mathcal{A}} \frac{P(a) - Q(a)}{Q(a)}\right) \\ &\le \log\left(1 + \frac{\max_{a\in\mathcal{A}} |P(a) - Q(a)|}{\min_{a\in\mathcal{A}} Q(a)}\right) \\ &\le \log\left(1 + \frac{|P-Q|}{2Q_{\min}}\right) \end{aligned} \tag{19}$$

where the last inequality follows from (9). Since the Rényi divergence of order $\alpha \in [0, \infty]$ is monotonic non-decreasing in α (see, e.g., [4, Theorem 3]), it follows from (19) that

$$D_\alpha(P\|Q) \le D_\infty(P\|Q) \le \log\left(1 + \frac{\varepsilon}{2Q_{\min}}\right), \ \forall\, \alpha \in [0, \infty] \tag{20}$$

which proves the first line in (18) when the validity of the bound is restricted to $\alpha \in (2, \infty]$.

For proving the second line in (18), it is shown that the bound in (6) can be sharpened by replacing $D(P\|Q)$ on the LHS of (6) with the quadratic Rényi divergence $D_2(P\|Q)$ (note that $D_2(P\|Q) \ge D(P\|Q)$), leading to

$$D_2(P\|Q) \le \log\left(1 + \frac{|P-Q|^2}{2Q_{\min}}\right). \tag{21}$$

The strengthened inequality in (21), in comparison to (6), follows by replacing inequality (7) with the equality

$$\begin{aligned} \chi^2(P,Q) &= \sum_{a\in\mathcal{A}} \frac{P(a)^2}{Q(a)} - 1 \\ &\stackrel{\text{(a)}}{=} \exp\big(D_2(P\|Q)\big) - 1 \end{aligned} \tag{22}$$

where equality (a) follows from (5). Combining (10) with (22) gives inequality (21), and

$$D_\alpha(P\|Q) \le D_2(P\|Q) \le \log\left(1 + \frac{\varepsilon^2}{2Q_{\min}}\right), \quad \forall\, \alpha \in [0, 2]. \tag{23}$$

The combination of (20) with (23) gives the second line in (18) (note that $\varepsilon\varepsilon' = \min\{\varepsilon, \varepsilon^2\}$) while the validity of the bound is restricted to $\alpha \in [1, 2]$.

For $\alpha \in (0, 1)$, $D_\alpha(P\|Q)$ satisfies the skew-symmetry property $D_\alpha(P\|Q) = \frac{\alpha}{1-\alpha} \cdot D_{1-\alpha}(Q\|P)$ (see, e.g., [4, Proposition 2]). Consequently, we have

$$\begin{aligned} D_\alpha(P\|Q) &= \left(\frac{\alpha}{1-\alpha}\right) D_{1-\alpha}(Q\|P) \\ &\le \left(\frac{\alpha}{1-\alpha}\right) D(Q\|P) \\ &\le \left(\frac{\alpha}{1-\alpha}\right) \log\left(1 + \frac{\varepsilon^2}{2P_{\min}}\right), \quad \forall\, \alpha \in (0,1) \end{aligned} \tag{24}$$

where the first inequality holds since the Rényi divergence is monotonic non-decreasing in its order, and the second inequality follows from Theorem 1. The third case in (18) follows from (20), (23) and (24) while restricting the validity of the bound to $\alpha \in \left(\frac{1}{2}, 1\right)$.

For proving the last case in (18), for $\alpha \in \left[0, \frac{1}{2}\right]$, note that from (5)

$$D_{1/2}(P\|Q) = -2\log Z(P,Q)$$

where $Z(P,Q) \triangleq \sum_{a\in\mathcal{A}} \sqrt{P(a)Q(a)}$ is the Bhattacharyya coefficient between P and Q [8]. The Bhattacharyya distance is defined as minus the logarithm of the Bhattacharyya coefficient, so it is zero if and only if $P = Q$, and it is non-negative (since $0 \le Z(P,Q) \le 1$, and $Z(P,Q) = 1$ if and only if $P = Q$). Hence, the Rényi divergence of order $\frac{1}{2}$ is twice the Bhattacharyya distance. Based on the inequality $Z(P,Q) \ge 1 - \frac{|P-Q|}{2}$ (see [14, Proposition 1 and Remark 1])

$$D_\alpha(P\|Q) \le D_{1/2}(P\|Q) \le -2\log\left(1 - \frac{\varepsilon}{2}\right), \ \forall\, \alpha \in [0, 0.5] \tag{25}$$

where $\varepsilon \triangleq |P - Q| \in [0, 2]$. Finally, the last case in (18) follows from (20), (23), (24) and (25). ■

B. Example: Rényi Divergence for Multinomial Distributions

Let $X_1, X_2, \ldots$ be independent Bernoulli random variables with $X_i \sim \text{Bernoulli}(p_i)$, and let $Y_1, Y_2, \ldots$ be independent Bernoulli random variables with $Y_i \sim \text{Bernoulli}(q_i)$ (assume w.l.o.g. that $q_i \le \frac{1}{2}$). Let U_n and V_n be the partial sums $U_n = \sum_{i=1}^n X_i$ and $V_n = \sum_{i=1}^n Y_i$, and let P_{U_n}, P_{V_n} denote their multinomial distributions. For all $\alpha \in [0, 2]$ and $n \in \mathbb{N}$, we have

$$
\begin{aligned}
&D_\alpha(P_{U_n}\|P_{V_n}) \\
&\overset{\text{(a)}}{\leq} D_\alpha(P_{X_1,\ldots,X_n}\|P_{Y_1,\ldots,Y_n}) \\
&\overset{\text{(b)}}{=} \sum_{i=1}^{n} D_\alpha(P_{X_i}\|P_{Y_i}) \\
&\overset{\text{(c)}}{\leq} \log\left(1+\frac{|P_{X_i}-P_{Y_i}|^2}{2\left(P_{Y_i}\right)_{\min}}\right) \\
&\overset{\text{(d)}}{=} \sum_{i=1}^{n} \log\left(1+2q_i\left(\frac{p_i}{q_i}-1\right)^2\right) \qquad (26)
\end{aligned}
$$

where inequality (a) follows from the data processing inequality for the Rényi divergence (see [4, Theorem 9]), equality (b) follows from the additivity property of the Rényi divergence under the independence assumption for $\{X_i\}$ and for $\{Y_i\}$ (see [4, Theorem 28]), inequality (c) follows from Theorem 2, and equality (d) holds since $|P_{X_i}-P_{Y_i}| = 2|p_i-q_i|$ for Bernoulli random variables, and $(P_{Y_i})_{\min} = \min\{q_i, 1-q_i\} = q_i$ ($q_i \leq \frac{1}{2}$). Similarly, for all $\alpha > 2$ and $n \in \mathbb{N}$,

$$D_\alpha(P_{U_n}\|P_{V_n}) \leq \sum_{i=1}^{n} \log\left(1+2\left|\frac{p_i}{q_i}-1\right|\right). \qquad (27)$$

The only difference in the derivation of (27) is in inequality (c) of (26) where the bound in the first line of (18) is used this time.

Let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a non-negative sequence such that

$$(1-\varepsilon_n)q_n \leq p_n \leq (1+\varepsilon_n)q_n, \quad \forall\, n \in \mathbb{N}$$

and assume that

$$\sum_{n=1}^{\infty} \varepsilon_n^2 < \infty.$$

Then, from (26), it follows that $D_\alpha(P_{U_n}\|P_{V_n}) \leq K_1$ for all $\alpha \in [0,2]$ and $n \in \mathbb{N}$ where

$$K_1 \triangleq \sum_{n=1}^{\infty} \log\left(1+2\varepsilon_n^2\right) < \infty.$$

Furthermore, if $\sum_{n=1}^{\infty} \varepsilon_n < \infty$, it follows from (27) that $D_\alpha(P_{U_n}\|P_{V_n}) \leq K_2$ for all $\alpha > 2$ and $n \in \mathbb{N}$ where

$$K_2 \triangleq \sum_{n=1}^{\infty} \log\left(1+2\varepsilon_n\right) < \infty.$$

Note that although $D_\alpha(P_{X_i}\|P_{Y_i})$ in equality (b) of (26) is equal to the binary Rényi divergence

$$d_\alpha(p_i\|q_i) \triangleq \begin{cases} \dfrac{\log\left(p_i^\alpha q_i^{1-\alpha} + (1-p_i)^\alpha (1-q_i)^{1-\alpha}\right)}{\alpha-1} \\ \quad \text{if } \alpha \in (0,1)\cup(1,\infty), \\ p_i \, \log\left(\frac{p_i}{q_i}\right) + (1-p_i) \, \log\left(\frac{1-p_i}{1-q_i}\right) \\ \quad \text{if } \alpha = 1. \end{cases}$$

the reason for the use of the upper bounds in step (c) of (26) and (27) is to state sufficient conditions, in terms of $\{\varepsilon_n\}_{n=1}^{\infty}$, for the boundedness of the Rényi divergence $D_\alpha(P_{U_n}\|P_{V_n})$.

ACKNOWLEDGMENT

Sergio Verdú is acknowledged for providing a draft of [18], and Vincent Tan is acknowledged for pointing out [15] and a simplified proof of (9). 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] 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.

[3] S. S. Dragomir, "Bounds for the normalized Jensen functional," *Bulletin of the Australian Mathematical Society*, vol. 74, no. 3, pp. 471–478, 2006.

[4] 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.

[5] A. A. Fedotov, P. Harremoës and F. Topsøe, "Refinements of Pinsker's inequality," *IEEE Trans. on Information Theory*, vol. 49, no. 6, pp. 1491–1498, June 2003.

[6] G. L. Gilardoni, "On the minimum f-divergence for given total variation," *Comptes Rendus Mathematique*, vol. 343, no. 11–12, pp. 763–766, 2006.

[7] 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.

[8] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. on Communication Technology*, vol. 15, no. 1, pp. 52–60, February 1967.

[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-Fransisco: 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, "Tight bounds on symmetric divergence measures and a new inequality relating f-divergences," accepted to the *IEEE 2015 Information Theory Workshop*, Jerusalem, Israel, April 26–May 1, 2015. [Online]. Available at http://arxiv.org/abs/1502.06428.

[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] I. Vajda, "Note on discrimination information and variation," *IEEE Trans. on Information Theory*, vol. 16, no. 6, pp. 771–773, Nov. 1970.

[17] S. Verdú, "Total variation distance and the distribution of the relative information," *Proceedings of the 2014 Information Theory and Applications (ITA) Workshop*, pp. 499–501, San-Diego, California, USA, February 2014.

[18] S. Verdú, *Information Theory*, in preparation.