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 ∞) 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 $(\mathcal{A}, \mathscr{F})$. The Csiszár-Kemperman-Kullback-Pinsker inequality (a.k.a. Pinsker's inequality) states that

$$\tfrac{1}{2}\,|P-Q|^2\,\log e \le D(P\|Q) \tag{1}$$

where

$$D(P\|Q) = \mathbb{E}_P\left[\log \frac{\mathrm{d}P}{\mathrm{d}Q}\right] = \int_{\mathcal{A}} \mathrm{d}P(a)\,\log \frac{\mathrm{d}P}{\mathrm{d}Q}\,(a) \tag{2}$$

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

$$|P-Q| = 2 \sup_{\mathcal{F}\in\mathscr{F}} |P(\mathcal{F}) - Q(\mathcal{F})| \tag{3}$$

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 $\mathcal{A}$,

$$D(P\|Q) = \sum_{a\in\mathcal{A}} P(a)\,\log \frac{P(a)}{Q(a)}, \tag{4}$$

$$|P-Q| = \sum_{a\in\mathcal{A}} \bigl|P(a) - Q(a)\bigr|. \tag{5}$$

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

$$D(P\|Q) \le \left(\frac{\log e}{Q_{\min}}\right) \cdot |P-Q|^2, \tag{6}$$

where

$$Q_{\min} \triangleq \min_{a\in\mathcal{A}} Q(a). \tag{7}$$

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

$$\beta_1 = \min_{a\in\mathcal{A}} \frac{Q(a)}{P(a)}, \tag{8}$$

$$\beta_2 = \min_{a\in\mathcal{A}} \frac{P(a)}{Q(a)} \tag{9}$$

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 Pinsker 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:

$$D(P\|Q) \leq \log\left(1 + \frac{|P-Q|^2}{2Q_{\min}}\right) - \frac{\beta_2 \log e}{2} \cdot |P-Q|^2 \tag{10}$$

$$\leq \log\left(1 + \frac{|P-Q|^2}{2Q_{\min}}\right) \tag{11}$$

where $Q_{\min}$ and β_2 are given in (7) and (9), respectively.

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

$$\chi^2(P\|Q) \triangleq \sum_{a\in\mathcal{A}} \frac{(P(a)-Q(a))^2}{Q(a)}. \tag{12}$$

A lower bound follows by invoking Jensen's inequality

$$\chi^2(P\|Q) = \sum_{a\in\mathcal{A}} \frac{P(a)^2}{Q(a)} - 1 \tag{13}$$

$$= \sum_{a\in\mathcal{A}} P(a) \exp\left(\log \frac{P(a)}{Q(a)}\right) - 1 \tag{14}$$

$$\geq \exp\left(\sum_{a\in\mathcal{A}} P(a) \log \frac{P(a)}{Q(a)}\right) - 1 \tag{15}$$

$$= \exp\bigl(D(P\|Q)\bigr) - 1. \tag{16}$$

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

$$\chi^2(P\|Q) = \exp\bigl(D_2(P\|Q)\bigr) - 1 \tag{17}$$

with the monotonicity of the Rényi divergence $D_\alpha(P\|Q)$ in α, 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])

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

$$\leq \log\bigl(1 + \chi^2(P\|Q)\bigr) - D(P\|Q) \tag{18}$$

$$\leq \max_{a\in\mathcal{A}} \frac{P(a)}{Q(a)} \cdot D(Q\|P). \tag{19}$$

From (19) and the definition of β_2 in (9), we have

$$\chi^2(P\|Q)$$

$$\geq \exp\Bigl(D(P\|Q) + \beta_2\, D(Q\|P)\Bigr) - 1 \tag{20}$$

$$\geq \exp\left(D(P\|Q) + \frac{\beta_2 \log e}{2} \cdot |P-Q|^2\right) - 1 \tag{21}$$

where (20) follows from (18) and the definition of β_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:

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

$$\leq \frac{\sum_{a\in\mathcal{A}} \bigl(P(a)-Q(a)\bigr)^2}{Q_{\min}} \tag{22}$$

$$= \frac{|P-Q|}{Q_{\min}} \cdot \max_{a\in\mathcal{A}} |P(a)-Q(a)| \tag{23}$$

and, from (3),

$$|P-Q| \geq 2 \max_{a\in\mathcal{A}} |P(a)-Q(a)|. \tag{24}$$

Combining (23) and (24) yields

$$\chi^2(P\|Q) \leq \frac{|P-Q|^2}{2Q_{\min}}. \tag{25}$$

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

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

$$D(P\|Q) \leq \frac{(1 - \beta_2\, Q_{\min}) \log e}{2Q_{\min}} \cdot |P-Q|^2. \tag{26}$$

In the finite-alphabet case, we can obtain another upper bound on $D(P\|Q)$ as a function of the ℓ_2 norm $\|P-Q\|_2$:

$$D(P\|Q) \leq \log\left(1 + \frac{\|P-Q\|_2^2}{Q_{\min}}\right) - \frac{\beta_2 \log e}{2} \cdot \|P-Q\|_2^2 \tag{27}$$

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 $\frac{\beta_2 \log e}{2} \cdot \|P-Q\|_2^2$, we recover the bound

$$D(P\|Q) \leq \frac{\|P-Q\|_2^2 \log e}{Q_{\min}} \tag{28}$$

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 χ^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 the weakend form in (11). When the variational distance is much smaller than the minimal eigenvalue (see [1, Eq. (57)]), the latter bounds have a quadratic scaling in this distance, similarly 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

$$L_\delta(Q) = \min_{P \notin \mathcal{T}_\delta(Q)} D(P\|Q) \tag{29}$$

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

$$\mathcal{T}_\delta(Q) = \Big\{P \colon \forall\, a \in \mathcal{A},\;\; |P(a) - Q(a)| \le \delta\, Q(a)\Big\} \tag{30}$$

Note that $(a_1, \ldots, a_n)$ is strongly δ-typical according to Q if its empirical distribution belongs to $\mathcal{T}_\delta(Q)$. 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 δ-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

$$\beta_Q = \inf_{A \in \mathscr{F} \colon Q(A) \ge \frac{1}{2}} Q(A). \tag{31}$$

The function $\phi\colon (0, \frac{1}{2}] \to [\frac{1}{2}\log e, \infty)$ is given by

$$\phi(p) = \begin{cases} \frac{1}{4(1-2p)} \log\left(\frac{1-p}{p}\right), & p \in \left(0, \frac{1}{2}\right), \\ \frac{1}{2}\log e, & p = \frac{1}{2}. \end{cases} \tag{32}$$

Theorem 2. If $Q_{\min} > 0$, then

$$\begin{aligned} \phi(1-\beta_Q)\, Q_{\min}^2\, \delta^2 &\le L_\delta(Q) & (33)\\ &\le \log\left(1 + 2Q_{\min}\,\delta^2\right) & (34) \end{aligned}$$

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

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

$$\phi(1-\beta_Q)\; |P - Q|^2 \le D(P\|Q). \tag{35}$$

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

$$|P - Q| > \delta Q_{\min} \tag{36}$$

together with (35) yields (33).

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

$$\min_{P \notin \mathcal{T}_\delta(Q)} |P - Q| = 2\delta Q_{\min}. \tag{37}$$

□

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

$$\frac{1}{Q_{\min}} \cdot \frac{\log e}{2\,\phi(1-\beta_Q)} \cdot \frac{\log\left(1 + 2Q_{\min}\,\delta^2\right)}{\frac{1}{2}\log e\; Q_{\min}\,\delta^2} \le \frac{4}{Q_{\min}} \tag{38}$$

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 δ^2 for $\delta \approx 0$.

B. Distance from Equiprobable

If P is a distribution on a finite set $\mathcal{A}$, $H(P)$ gauges the "distance" from U, the equiprobable distribution, since

$$H(P) = \log|\mathcal{A}| - D(P\|U). \tag{39}$$

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

$$|P - U| \le \sqrt{\frac{2}{\log e} \cdot \big(\log|\mathcal{A}| - H(P)\big)}, \tag{40}$$

$$|P - U| \le 2\sqrt{1 - \frac{1}{|\mathcal{A}|} \cdot \exp\big(H(P)\big)}, \tag{41}$$

$$|P - U| \ge \sqrt{2\left(\exp\big(-H(P)\big) - \frac{1}{|\mathcal{A}|}\right)}, \tag{42}$$

respectively.

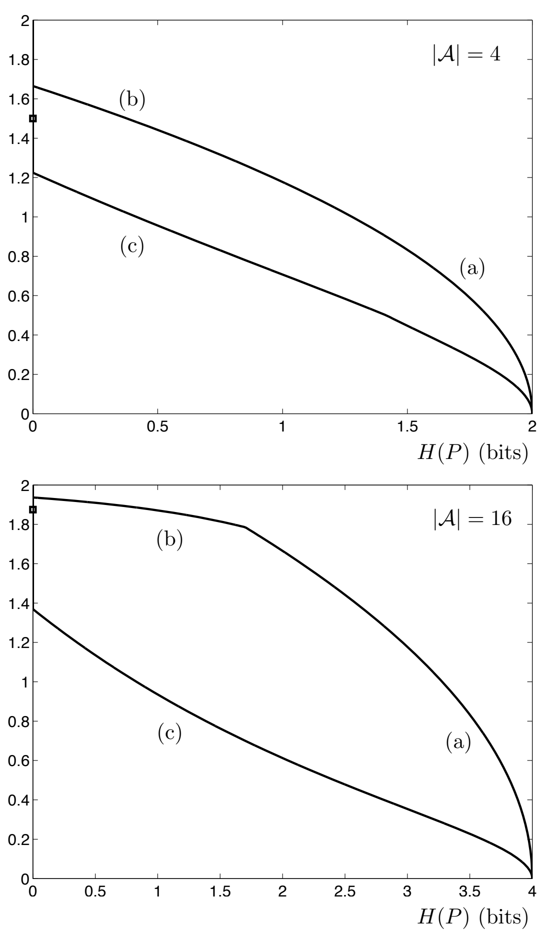

Fig. 1. Bounds on $|P - U|$ as a function of $H(P)$ for $|\mathcal{A}| = 4$, and $|\mathcal{A}| = 16$. The point $(H(P), |P-U|) = (0, 2(1 - |\mathcal{A}|^{-1}))$ is depicted on the y-axis. In the curves of the two plots, the bounds (a), (b) and (c) refer, respectively, to (40), (41) and (42).

The bounds in (40)–(42) are illustrated for $|\mathcal{A}| = 4, 16$ in Figure 1. For $H(P) = 0$, $|P - U| = 2(1 - |\mathcal{A}|^{-1})$ 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).

Results on the more general problem of finding bounds on $|H(P) - H(Q)|$ based on $|P - Q|$ can be found in [5, Theorem 17.3.3], [11], [16], [18], [26, Section 1.7] and [27].

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

$$D_\alpha(P||Q) \triangleq \frac{1}{\alpha - 1} \, \log \left(\sum_{a \in \mathcal{A}} P^\alpha(a) \, Q^{1-\alpha}(a) \right). \tag{43}$$

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 $D(P\|Q) = \lim_{\alpha \uparrow 1} D_\alpha(P\|Q)$). 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

$$D_\infty(P||Q) = \log \left(\sup_{a \in \mathcal{A}} \frac{P(a)}{Q(a)} \right).$$

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

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

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 $P_{\min}$ and $Q_{\min}$, respectively. Let β_1 and β_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

$$D_\alpha(P\|Q) \leq \begin{cases} f_1, & \alpha \in (2, \infty] \\ f_2, & \alpha \in [1, 2] \\ \min\left\{f_2, f_3, f_4\right\}, & \alpha \in \left(\frac{1}{2}, 1\right) \\ \min\left\{2 \log\left(\frac{1}{1-\delta}\right), f_2, f_3, f_4\right\}, & \alpha \in \left[0, \frac{1}{2}\right] \end{cases} \tag{44}$$

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

$$f_1(\alpha, \beta_1, \delta) \triangleq \begin{cases} \frac{1}{\alpha - 1} \, \log\left(1 + \frac{\delta(\beta_1^{1-\alpha} - 1)}{1 - \beta_1}\right) & \alpha \in [0, 1) \cup (1, \infty) \\ \frac{\delta}{1 - \beta_1} \, \log \frac{1}{\beta_1}, & \alpha = 1, \\ \log \frac{1}{\beta_1}, & \alpha = \infty \end{cases} \tag{45}$$

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

$$f_2(\alpha, \beta_1, Q_{\min}, \delta) \triangleq \min\left\{ f_1(\alpha, \beta_1, \delta), \; \log\left(1 + \frac{2\delta^2}{Q_{\min}}\right) \right\} \tag{46}$$

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

$$f_3(\alpha, P_{\min}, \beta_1, \delta) \triangleq \left(\frac{\alpha}{1 - \alpha}\right) \left[\log\left(1 + \frac{2\delta^2}{P_{\min}}\right) - 2\beta_1\delta^2 \, \log e \right], \tag{47}$$

$$\begin{aligned} f_4(\beta_2, Q_{\min}, \delta) \triangleq \min \Bigg\{ & \log\left(1 + \frac{2\delta^2}{Q_{\min}}\right) - 2\beta_2\delta^2 \, \log e, \\ & \log\left(1 + \frac{\min\{\delta, 2\delta^2\}}{Q_{\min}}\right) \Bigg\}. \end{aligned} \tag{48}$$

Proof. See [20, Section 7.C]. □

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

$$D_\alpha(P\|Q) \leq \log\left(1 + \frac{|P - Q|}{2Q_{\min}}\right) \tag{49}$$

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 $Q_{\min} > 0$. 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.

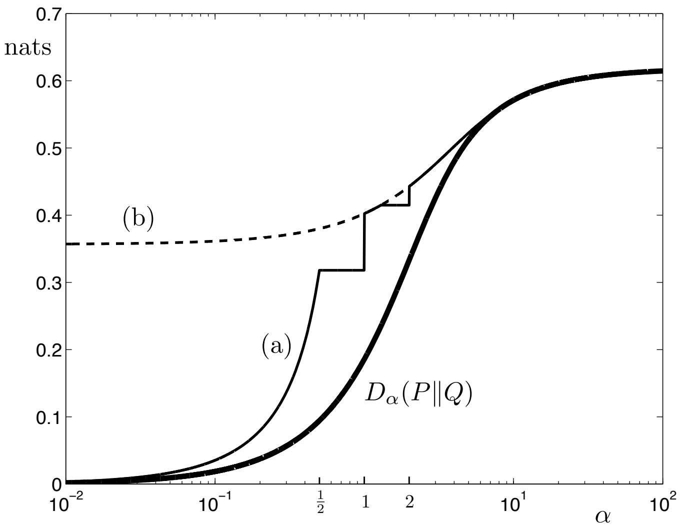

Fig. 2. The Rényi divergence $D_\alpha(P\|Q)$ for P and Q which are defined on a binary alphabet with $P(0) = Q(1) = 0.65$, compared to (a) its upper bound in (44), and (b) its upper bound in (45) (see [20, Theorem 23]). The two bounds coincide here when $\alpha \in (1, 1.291) \cup (2, \infty)$.

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

[1] K. M. R. Audenaert and J. Eisert, "Continuity bounds on the quantum relative entropy," *Journal of Mathematical Physics*, vol. 46, paper 102104, October 2005.

[2] K. M. R. Audenaert and J. Eisert, "Continuity bounds on the quantum relative entropy - II," *Journal of Mathematical Physics*, vol. 52, paper 112201, November 2011.

[3] G. Böcherer and B. C. Geiger, "Optimal quantization for distribution synthesis," March 2015. Available at http://arxiv.org/abs/1307.6843.

[4] J. Bretagnolle and C. Huber, "Estimation des densités: risque minimax," *Probability Theory and Related Fields*, vol. 47, no. 2, pp. 119–137, 1979.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, second edition, John Wiley & Sons, 2006.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] S. W. Ho and R. W. Yeung, "The interplay between entropy and variational distance," *IEEE Trans. on Information Theory*, vol. 56, no. 12, pp. 5906–5929, December 2010.

[12] 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.

[13] M. Krajči, C. F. Liu, L. Mikeš and S. M. Moser, "Performance analysis of Fano coding," *Proceedings of the IEEE 2015 International Symposium on Information Theory*, Hong Kong, June 14–19, 2015.

[14] 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.

[15] M. S. Pinsker, *Information and Information Stability of Random Variables and Random Processes*, San-Fransisco: Holden-Day, 1964, originally published in Russian in 1960.

[16] V. V. Prelov and E. C. van der Meulen, "Mutual information, variation, and Fano's inequality," *Problems of Information Transmission*, vol. 44, no. 3, pp. 185–197, September 2008.

[17] 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.

[18] I. Sason, "Entropy bounds for discrete random variables via maximal coupling," *IEEE Trans. on Information Theory*, vol. 59, no. 11, pp. 7118–7131, November 2013.

[19] I. Sason, "On the Rényi divergence and the joint range of relative entropies," *Proceedings of the 2015 IEEE International Symposium on Information Theory*, pp. 1610–1614, Hong Kong, June 14–19, 2015.

[20] I. Sason and S. Verdú, "Bounds among f-divergences," submitted to the *IEEE Trans. on Information Theory*, July 2015. [Online]. Available at http://arxiv.org/abs/1508.00335.

[21] 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.

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

[23] I. Vajda, "On f-divergence and singularity of probability measures," *Periodica Mathematica Hungarica*, vol. 2, no. 1–4, pp. 223–234, 1972.

[24] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, 1998.

[25] 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.

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

[27] Z. Zhang, "Estimating mutual information via Kolmogorov distance," *IEEE Trans. on Information Theory*, vol. 53, no. 9, pp. 3280–3282, September 2007.