CCIT Report #882 March 2015

On Reverse Pinsker Inequalities

Igal Sason

Abstract

New upper bounds on the relative entropy are derived as a function of the total variation distance. One bound refines an inequality by Verdú for general probability measures. A second bound improves the tightness of an inequality by Csiszár and Talata for arbitrary probability measures that are defined on a common finite set. The latter result is further extended, for probability measures on a finite set, leading to an upper bound on the Rényi divergence of an arbitrary non-negative order (including ∞) as a function of the total variation distance. Another lower bound by Verdú on the total variation distance, expressed in terms of the distribution of the relative information, is tightened and it is attained under some conditions. The effect of these improvements is exemplified.

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

I. Introduction

Consider two probability measures 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

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

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

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

designates the total variation distance (or L_1 distance) between P and Q. One of the implications of inequality (1) is that convergence in relative entropy implies convergence in total variation distance. The total variation distance is bounded $|P - Q| \leq 2$, in contrast to the relative entropy.

Inequality (1) is a.k.a. Pinsker's inequality, although the analysis made by Pinsker [15] leads to a significantly looser bound where $\frac{\log e}{2}$ on the RHS of (1) is replaced by $\frac{\log e}{408}$ (see [25, Eq. (51)]). Improved and generalized versions of Pinsker's inequality have been studied in [7], [8], [9], [14], [18], [24].

For any $\varepsilon > 0$, there exists a pair of probability measures 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 exist in general. Nevertheless, under some conditions, such inequalities hold [4], [25], [26] (to be addressed later in this section).

If $P \ll Q$, the relative information in $a \in \mathcal{A}$ according to (P, Q) is defined to be

$$i_{P\|Q}(a) \triangleq \log \frac{\mathrm{d}P}{\mathrm{d}Q}\,(a). \tag{4}$$

I. Sason is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: sason@ee.technion.ac.il). This work has been submitted in part to the *2015 IEEE Information Theory Workshop (ITW 2015)*, Jeju Island, Korea, October 11–15, 2015. The research was supported by the Israeli Science Foundation (ISF), grant number 12/12.