CCIT Report #859 May 2014

Bounds on f-Divergences and Related Distances

Igal Sason

Abstract

Derivation of tight bounds on f-divergences and related distances is of interest in information theory and statistics. This paper improves some existing bounds on f-divergences. In some cases, an alternative approach leads to a simplified proof of an existing bound. Following bounds on the chi-squared divergence, an improved version of a reversed Pinsker's inequality is derived for an arbitrary pair of probability distributions on a finite set. Following bounds on the relative entropy and Jeffreys' divergence, a tightened inequality for lossless source coding is derived and considered. Finally, a new inequality relating f-divergences is derived and studied.

Index Terms – Bhattacharyya distance, Chernoff information, chi-squared divergence, f-divergence, Hellinger distance, Jeffreys' divergence, lossless source coding, relative entropy (Kullback-Leibler divergence), total variation distance.

I. Introduction

Divergence measures are widely used in information theory, machine learning, statistics, and other theoretical and applied branches of mathematics (see, e.g., [3], [12], [15], [34], [40] and [44]). The class of f-divergences, introduced independently in [1], [8] and [32], forms an important class of divergence measures which includes the relative entropy (a.k.a. information divergence or the Kullback-Leibler divergence), its dual and symmetrized divergences, the total variation distance, squared Hellinger distance, chi-squared divergence, etc. Properties of f-divergences, including relations to statistical tests and estimators, were extensively studied in [30].

In the following, some related papers that are most relevant to the scope of this work are briefly reviewed. Pinsker's inequality (a.k.a the Csiszár-Kemperman-Kullback-Pinsker inequality) and Vajda's inequality [44] have been derived during the sixties to provide lower bounds on the relative entropy in terms of the total variation distance. Following these bounds, Fedotov *et al.* [21] derived an exact parametrization of the infimum of the relative entropy with respect to all possible pairs of probability distributions with a given total variation distance. The derivation of the parametrization in [21] relies on the data processing theorem for the relative entropy, leading to a maximization problem of the convex conjugate function of the relative entropy between two-element probability distributions; this approach leads to closed-form solutions that are used to identify a possible form for the required parametrization. As an extension to this problem, Harremoës and Vajda studied in [26] the joint range of pairs of f-divergences, characterizing all the possible points in $[0, \infty]^2$ that are achievable by a given pair of f-divergences. It was shown that this region is convex where each point is a convex combination of two achievable points that are obtained by a pair of probability distributions over two elements; hence, every such an achievable point is obtained by a pair of probability distributions over at most 4 elements.

In [23], Gilardoni studied the problem of minimizing an arbitrary *symmetric* f-divergence for a given total variation distance, and provided a closed-form solution of this optimization problem. Furthermore, an alternative parametrization of the infimum of the relative entropy for a given total variation distance was derived in [23]. In a follow-up paper by the same author [24], Pinsker's and Vajda's type inequalities were studied for symmetric f-divergences, and the issue of obtaining lower bounds on f-divergences for a given total variation distance was further studied. One of the main results in [24] was a derivation of an improved and simple closed-form lower bound on the relative entropy in terms of the total variation distance, as well as a simple and reasonably tight closed-form upper bound on the infimum of the relative entropy for a given total variation distance.

Sharp inequalities for f-divergences were recently studied in [25] as a problem of maximizing or minimizing an arbitrary f-divergence between two probability measures subject to a finite number of inequality constraints on other f-divergences. The main result stated in [25] is that such infinite-dimensional optimization problems are equivalent to optimization problems over finite-dimensional spaces where the latter are numerically solvable.

I. Sason is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: sason@ee.technion.ac.il).