# Entropy Bounds for Discrete Random Variables via Coupling

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

**Abstract**

This paper derives new entropy bounds for discrete random variables via maximal coupling. It provides bounds on the difference between the entropies of two discrete random variables in terms of the local and total variation distances between their probability mass functions. These bounds address cases of finite or countable infinite alphabets. Particular cases of these bounds reproduce some known results. The use of the new entropy bounds is exemplified by relying on some bounds on the above distances via Stein's method. The improvement that is obtained by these bounds is exemplified.

**Index Terms**

Coupling, entropy, local distance, Stein's method, total variation distance.

## I. Introduction

Inequalities that relate the Shannon entropy or information divergence with the total variation distance were extensively studied during the last fifty years (see, e.g., [7]–[10], [13], [14], [16], [18]–[21], [25]–[27], [30]–[38], [42], [44]–[49]). Among the observations in these works, it is known that a sufficiently small total variation distance between a pair of discrete random variables with *a finite and fixed alphabet*, implies a small difference between their entropies. However, if the size of the alphabet is finite but it is not bounded then for an arbitrarily small $\delta > 0$ and an arbitrarily large $\mu > 0$, there exists a pair of discrete random variables such that the total variation distance between them is less than $\delta$ whereas the difference between their entropies is larger than $\mu$ (see, e.g., [21, Theorem 1] with a concrete example in its proof).

The interplay between the entropy difference of two discrete random variables and their total variation distance was studied in [7, Theorem 17.3.3] or [10, Lemma 2.7], [11, Lemma 1], [21], [34], [42, Section 2] and [49]. The bounds that are derived in this work improve some existing bounds as a result of their dependence on both the *local and total variation distances* and the alphabet sizes (the relevant distances are defined later in this section). The new bounds are derived via the use of *maximal coupling*, which is also known to be useful for the derivation of error bounds via Stein's method (see, e.g., [40, Chapter 2] and [41]). It is noted that the entropy bounds in [49] are also derived via coupling, but the approach of the analysis in this work is remarkably different (see Sections II and III). The new bounds are linked to Stein's method, and the improvement that is achieved by these bounds is exemplified.

We provide in the following the essential mathematical background that is required for the analysis in this work.

*Definition 1:* A *coupling* of a pair of two discrete random variables $(X, Y)$ is a pair of two random variables $(\hat{X}, \hat{Y})$ such that the marginal distributions of $(X, Y)$ and $(\hat{X}, \hat{Y})$ coincide, i.e., $P_X = P_{\hat{X}}$ and $P_Y = P_{\hat{Y}}$.

*Definition 2:* For a pair of random variables $(X, Y)$, a coupling $(\hat{X}, \hat{Y})$ is called a *maximal coupling* if $\mathbb{P}(\hat{X} = \hat{Y})$ is as large as possible among all the couplings of $(X, Y)$.

The following theorem is a basic result on maximal coupling that also suggests, as part of its proof, a construction for maximal coupling. We later rely on this particular construction to derive in Section III