Improved Lower Bounds on the Total Variation Distance for the Poisson Approximation

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

Abstract

New lower bounds on the total variation distance between the distribution of a sum of independent Bernoulli random variables and the Poisson random variable (with the same mean) are derived via the Chen-Stein method. The new bounds rely on a non-trivial modification of the analysis by Barbour and Hall (1984) which surprisingly gives a significant improvement. A use of the new lower bounds is addressed.

Keywords: Chen-Stein method, Poisson approximation, total variation distance.

I. Introduction

Convergence to the Poisson distribution, for the number of occurrences of possibly dependent events, naturally arises in various applications. Following the work of Poisson, there has been considerable interest in how well the Poisson distribution approximates the binomial distribution.

The basic idea which serves as a starting point of the so called *Chen-Stein method for the Poisson approximation* [6] is the following. Let $\{X_i\}_{i=1}^n$ be independent Bernoulli random variables with $\mathbb{E}(X_i) = p_i$. Let $W \triangleq \sum_{i=1}^n X_i$, $V_i \triangleq \sum_{j \neq i} X_j$ for every $i \in \{1, \ldots, n\}$, and $Z \sim \text{Po}(\lambda)$ with mean $\lambda \triangleq \sum_{i=1}^n p_i$. It is easy to show that

$$\mathbb{E}[\lambda f(Z+1) - Z f(Z)] = 0 \tag{1}$$

holds for an arbitrary bounded function $f : \mathbb{N}_0 \to \mathbb{R}$ where $\mathbb{N}_0 \triangleq \{0, 1, \ldots\}$. Furthermore (see, e.g., [10, Chapter 2])

$$\mathbb{E}\big[\lambda f(W+1) - W f(W)\big] = \sum_{j=1}^n p_j^2 \, \mathbb{E}\big[f(V_j+2) - f(V_j+1)\big] \tag{2}$$

which then serves to get rigorous bounds on the difference between the distributions of W and Z, by the Chen-Stein method for Poisson approximations. This method, and more generally the so called *Stein's method*, serves as a powerful tool for the derivation of rigorous bounds for various distributional approximations. Nice expositions of this method are provided, e.g., in [1], [10, Chapter 2] and [11]. Furthermore, some interesting links between the Chen-Stein method and information-theoretic functionals in the context of Poisson and compound Poisson approximations are provided in [5].

Throughout this letter, we use the term 'distribution' to refer to the discrete probability mass function of an integer-valued random variable. In the following, we introduce some known results that are related to the presentation of the new results.

Definition 1: Let P and Q be two probability measures defined on a set $\mathcal{X}$. Then, the total variation distance between P and Q is defined by

$$d_{\text{TV}}(P, Q) \triangleq \sup_{\text{Borel } A \subseteq \mathcal{X}} \big(P(A) - Q(A)\big) \tag{3}$$

where the supermum is taken w.r.t. all the Borel subsets A of $\mathcal{X}$. If $\mathcal{X}$ is a countable set then (3) is simplified to

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)| = \frac{||P - Q||_1}{2} \tag{4}$$

so the total variation distance is equal to one-half of the L_1-distance between the two probability distributions.

Among old and interesting results that are related to the Poisson approximation, Le Cam's inequality [9] provides an upper bound on the total variation distance between the distribution of the sum $W = \sum_{i=1}^n X_i$ of n independent