CCIT Report #817 October 2012

An Information-Theoretic Perspective of the Poisson Approximation via the Chen-Stein Method

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

Abstract

The first part of this work considers the entropy of the sum of (possibly dependent and non-identically distributed) Bernoulli random variables. Upper bounds on the error that follows from an approximation of this entropy by the entropy of a Poisson random variable with the same mean are derived via the Chen-Stein method. The second part of this work derives new lower bounds on the total variation distance and relative entropy between the distribution of the sum of independent Bernoulli random variables and the Poisson distribution. The starting point of the derivation of the new bounds in the second part of this work is an introduction of a new lower bound on the total variation distance, whose derivation generalizes and refines the analysis by Barbour and Hall (1984), based on the Chen-Stein method for the Poisson approximation. A new lower bound on the relative entropy between these two distributions is introduced, and this lower bound is compared to a previously reported upper bound on the relative entropy by Kontoyiannis et al. (2005). The derivation of the new lower bound on the relative entropy follows from the new lower bound on the total variation distance, combined with a distribution-dependent refinement of Pinsker's inequality by Ordentlich and Weinberger (2005). Upper and lower bounds on the Bhattacharyya parameter, Chernoff information and Hellinger distance between the distribution of the sum of independent Bernoulli random variables and the Poisson distribution with the same mean are derived as well via some relations between these quantities with the total variation distance and the relative entropy. The analysis in this work combines elements of information theory with the Chen-Stein method for the Poisson approximation. The resulting bounds are easy to compute, and their applicability is exemplified.

Index Terms

Chen-Stein method, Chernoff information, entropy, error bounds, error exponents, Poisson approximation, relative entropy, total variation distance.

AMS 2000 Subject Classification: Primary 60E07, 60E15, 60G50, 94A17.

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. This approximation was treated by a limit theorem in [15, Chapter 8], and later some non-asymptotic results have considered the accuracy of this approximation. Among these old and interesting results, Le Cam's inequality [35] provides an upper bound on the total variation distance between the distribution of the sum $S_n = \sum_{i=1}^n X_i$ of n independent Bernoulli random variables $\{X_i\}_{i=1}^n$, where $X_i \sim \text{Bern}(p_i)$, and a Poisson distribution $\text{Po}(\lambda)$ with mean $\lambda = \sum_{i=1}^n p_i$. This inequality states that

$$d_{\text{TV}}\big(P_{S_n}, \text{Po}(\lambda)\big) \triangleq \frac{1}{2} \sum_{k=0}^{\infty} \left| \mathbb{P}(S_n = k) - \frac{e^{-\lambda}\lambda^k}{k!} \right| \leq \sum_{i=1}^n p_i^2$$

so if, e.g., $X_i \sim \text{Bern}\big(\frac{\lambda}{n}\big)$ for every $i \in \{1, \ldots, n\}$ (referring to the case that S_n is binomially distributed) then this upper bound is equal to $\frac{\lambda^2}{n}$, thus decaying to zero as n tends to infinity. This upper bound was later improved, e.g., by Barbour and Hall (see [4, Theorem 1]), replacing the above upper bound by $\left(\frac{1-e^{-\lambda}}{\lambda}\right) \sum_{i=1}^n p_i^2$ and therefore improving it by a factor of $\frac{1}{\lambda}$ when λ is large. This improved upper bound was also proved by Barbour and Hall to be essentially tight (see [4, Theorem 2]) with the following lower bound on the total variation distance:

$$d_{\text{TV}}\big(P_{S_n}, \text{Po}(\lambda)\big) \geq \frac{1}{32} \min\Big\{1, \frac{1}{\lambda}\Big\} \sum_{i=1}^n p_i^2$$