Improved Lower Bounds on the Total Variation Distance and Relative Entropy 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. Corresponding lower bounds on the relative entropy are derived, based on the lower bounds on the total variation distance and an existing distribution-dependent refinement of Pinsker’s inequality. Two uses of these bounds are finally outlined. The full version for this shortened paper is available at http://arxiv.org/abs/1206.6811.

Index Terms—Chen-Stein method, Poisson approximation, relative entropy, 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 [2] is the following. Let \( \{X_i\}_{i=1}^n \) be independent Bernoulli random variables with \( \mathbb{P}(X_i) = p_i \). Let \( W = \sum_{i=1}^n X_i \), \( V_i = \sum_{j \neq i} X_j \) for every \( i \in \{1, \ldots, n\} \), and \( Z \sim \text{Po}(\lambda) \) with mean \( \lambda = \sum_{i=1}^n p_i \). It is rather easy to show that the equality

\[
E[\lambda f(Z + 1) - Z f(Z)] = 0
\]

holds for an arbitrary bounded function \( f : \mathbb{N}_0 \rightarrow \mathbb{R} \). Furthermore, one can show that (see, e.g., [16] Chapter 2)

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

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 [2], [16] Chapter 2 and [17]. 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 [6].

Throughout this paper, 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

\[
d_{TV}(P, Q) = \sup_{A \subseteq \mathcal{X}} |P(A) - Q(A)|
\]

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

\[
d_{TV}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)| = \frac{||P - Q||_1}{2}
\]

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 [14] 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 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 \( d_{TV}(P_W, P(\lambda)) \leq \sum_{i=1}^n p_i^2 \) so if, e.g., \( X_i \sim \text{Bern}(\frac{1}{2}) \) for every \( i \in \{1, \ldots, n\} \) (referring to the case that \( W \) is binomially distributed) then this upper bound is equal to \( \frac{\lambda^2}{2} \), thus decaying to zero as \( n \) tends to infinity. The following theorem combines [4] Theorems 1 and 2, and its proof relies on the Chen-Stein method:

Theorem 1: Let \( W = \sum_{i=1}^n X_i \) be a sum of \( n \) independent Bernoulli random variables with \( \mathbb{E}(X_i) = p_i \) for \( i \in \{1, \ldots, n} \), and \( \mathbb{E}(W) = \lambda \). Then, the total variation distance between the probability distribution of \( W \) and the Poisson distribution with mean \( \lambda \) satisfies

\[
\frac{1}{32} \left( 1 + \frac{1}{\lambda} \right) \sum_{i=1}^n p_i^2 \leq d_{TV}(P_W, P(\lambda)) \leq \left( 1 - e^{-\lambda} \right) \sum_{i=1}^n p_i^2
\]

where \( a \wedge b \triangleq \min\{a, b\} \) for every \( a, b \in \mathbb{R} \).

As a consequence of Theorem 1 it follows that the ratio between the upper and lower bounds in (5) is not larger than 32, irrespectively of the values of \( \{p_i\} \). The factor \( \frac{1}{32} \) in the lower bound was claimed to be improvable to \( \frac{1}{17} \) with no explicit proof (see [5] Remark 3.2.2). This shows that, for independent Bernoulli random variables, these bounds are essentially tight. Furthermore, note that the upper bound in
improves Le Cam’s inequality; for large values of \( \lambda \), this improvement is by approximately a factor of \( \frac{1}{\lambda} \).

This paper presents 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). The starting point of the derivation of these new bounds generalizes and improves the analysis by Barbour and Hall [4], 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. [13]. The presentation in this conference paper relies on the analysis in the full paper version [18].

II. IMPROVED LOWER BOUNDS ON THE TOTAL VARIATION DISTANCE

In the following, we introduce an improved lower bound on the total variation distance and then provide a loosened version of this bound that is expressed in closed form.

Theorem 2: In the setting of Theorem 1, the total variation distance between the probability distribution of \( W \) and the Poisson distribution with mean \( \lambda \) satisfies the inequality

\[
K_1(\lambda) \sum_{i=1}^{n} p_i^2 \leq d_{TV}(P_W, \text{Po}(\lambda)) \leq \left( \frac{1 - e^{-\lambda}}{\lambda} \right) \sum_{i=1}^{n} p_i^2 \tag{6}
\]

where

\[
K_1(\lambda) \triangleq \sup_{\alpha_1, \alpha_2 \in \mathbb{R}, \alpha_2 \leq \lambda + \frac{2}{\lambda}} \left( \frac{1 - h_\lambda(\alpha_1, \alpha_2, \theta)}{2 \, g_\lambda(\alpha_1, \alpha_2, \theta)} \right) \tag{7}
\]

and

\[
h_\lambda(\alpha_1, \alpha_2, \theta) \triangleq \frac{3\lambda + (2 - \alpha_2 + \lambda)^3 - (1 - \alpha_2 + \lambda)^3}{\theta \lambda} + \frac{\left| \alpha_1 - \alpha_2 \right| (2\lambda + |3 - 2\alpha_2|)}{\theta \lambda} \exp \left( -\frac{\left(1 - \alpha_2\right) \lambda}{\theta \lambda} \right) \tag{8}
\]

\[
x_+ \triangleq \max\{x, 0\}, \quad x_+^2 \triangleq (x_+)^2, \quad \forall x \in \mathbb{R}
\]

\[
g_\lambda(\alpha_1, \alpha_2, \theta) \triangleq \max \left\{ \left| \left( 1 + \sqrt{\frac{2}{\theta \lambda e}} \cdot |\alpha_1 - \alpha_2| \right) \lambda + \max\{x(u)\} \right|, \left| \left( 2e^{-\frac{2}{\theta \lambda e}} + \sqrt{\frac{2}{\theta \lambda e}} \cdot |\alpha_1 - \alpha_2| \right) \lambda - \min\{x(u)\} \right| \right\} \tag{10}
\]

\[
x(u) \triangleq (c_0 + c_1 u + c_2 u^2) \exp(-u^2), \quad \forall u \in \mathbb{R} \tag{11}
\]

\[
\{u_i\} \triangleq \left\{ u \in \mathbb{R} : 2c_2 u^3 + 2c_1 u^2 - 2(c_2 - c_0) u - c_1 = 0 \right\} \tag{12}
\]

\[
c_0 \triangleq (\alpha_2 - \alpha_1)(\lambda - \alpha_2) \tag{13}
\]

\[
c_1 \triangleq \sqrt{\theta \lambda} (\lambda + \alpha_1 - 2\alpha_2) \tag{14}
\]

\[
c_2 \triangleq -\theta \lambda. \tag{15}
\]

Proof: See [18, Section 4.F]. The derivation relies on the Chen-Stein method for the Poisson approximation, and it improves (significantly) the constant in the lower bound by Barbour and Hall [4, Theorem 2]. The proof in [18] is self-contained.

Remark 1: The upper and lower bounds on the total variation distance in (6) scale like \( \sum_{i=1}^{n} p_i^2 \), similarly to the known bounds in Theorem 1. The ratio of the upper and lower bounds in Theorem 1 tends to 32.00 when either \( \lambda \) tends to zero or infinity. It was obtained numerically that the ratio of the upper bound and new lower bound on the total variation distance is reduced to 1.69 when \( \lambda \to 0 \), it is 10.54 when \( \lambda \to \infty \), and it is no more than 12.91 for all \( \lambda \in (0, \infty) \).

Remark 2: [9, Theorem 1.2] provides an asymptotic result for the total variation distance between the distribution of the sum \( W \) of \( n \) independent Bernoulli random variables with \( \mathbb{E}(X_i) = p_i \) and the Poisson distribution with mean \( \lambda = \sum_{i=1}^{n} p_i \). It shows that when \( \sum_{i=1}^{n} p_i \to \infty \) and \( \max_{1 \leq i \leq n} p_i \to 0 \) as \( n \to \infty \) then

\[
d_{TV}(P_W, \text{Po}(\lambda)) \sim \frac{1}{\sqrt{2\pi e\lambda}} \sum_{i=1}^{n} p_i^2. \tag{16}
\]

This implies that the ratio of the upper bound on the total variation distance in [4, Theorem 1] (see Theorems 1 here) and this asymptotic expression is equal to \( \frac{10.54}{\sqrt{\pi e \lambda}} \approx 2.55 \). It therefore follows from Remark 1 that in the limit where \( \lambda \to 0 \), the new lower bound on the total variation in (6) is smaller than the exact value by no more than 1.69, and for \( \lambda \gg 1 \), it is smaller than the exact asymptotic result by a factor of 2.55.

Remark 3: The cardinality of the set \( \{u_i\} \in (12) \) can be shown to be 3 (see [18, Section 4]).

Remark 4: The optimization that is required for the computation of \( K_1 \) in (2) w.r.t. the three parameters \( \alpha_1, \alpha_2 \in \mathbb{R} \) and \( \theta \in \mathbb{R}^+ \) is performed numerically. The numerical procedure for the computation of \( K_1 \) is presented in [18, Section 4].

In the following, we introduce a looser lower bound on the total variation distance as compared to the lower bound in Theorem 2 but its advantage is that it is expressed in closed-form. Both lower bounds improve (significantly) the lower bound in [4, Theorem 2]. The following lower bound follows from Theorem 2 by the special choice of \( \alpha_1 = \alpha_2 = \lambda \) that is included in the optimization set for \( K_1 \) on the right-hand side of (7). Following this sub-optimal choice, the lower bound in
the next corollary is obtained by a derivation of a closed-form expression for the third free parameter \( \theta \in \mathbb{R}^+ \) (in fact, this was our first step towards the derivation of an improved lower bound on the total variation distance).

**Corollary 1:** Under the assumptions in Theorem 2 then

\[
\tilde{K}_1(\lambda) \leq d_{TV}(P_W, \text{Po}(\lambda)) \leq \sum_{i=1}^n p_i^2 \left( 1 - e^{-\lambda} \right)
\]

where

\[
\tilde{K}_1(\lambda) \triangleq \frac{e^{-\lambda} - 1}{2\lambda} \cdot \frac{1}{\theta + 2e^{-1/2}} - \frac{1}{\lambda} \cdot \sqrt{3(\lambda + 7)[(3 + 2e^{-1/2})\lambda + 7]}.
\]

**Proof:** See [18, Section 4].

**Remark 5:** The lower bound on the total variation distance on the left-hand side of (17) improves uniformly the lower bound in [4, Theorem 2] (i.e., the left-hand side of Eq. (5) here). This improvement is shown in Fig. 1.

![Fig. 1](image)

**Fig. 1.** The figure presents curves that correspond to ratios of upper and lower bounds on the total variation distance between the sum of independent Bernoulli random variables and the Poisson distribution with the same mean \( \lambda \). The upper bound on the total variation distance for all these three curves is the bound by Barbour and Hall (see [4, Theorem 1]) or Theorem 2, here. The lower bounds that the three curves refer to them are the following: the curve at the bottom (i.e., the one which provides the lowest ratio for a fixed \( \lambda \)) is the improved lower bound on the total variation distance that is introduced in Theorem 2. The curve slightly above it for small values of \( \lambda \) corresponds to the original lower bound when \( \alpha_1 \) and \( \alpha_2 \) in (6) are set to be equal (i.e., \( \alpha_1 = \alpha_2 \triangleq \alpha \) is their common value), so that the optimization of \( K_1 \) for this curve is reduced to a two-parameter maximization of \( K_1 \) over the two free parameters \( \alpha \in \mathbb{R} \) and \( \theta \in \mathbb{R}^+ \). Finally, the curve at the top of this figure corresponds to the further loosening of this lower bound where \( \alpha \) is set to be equal to \( \lambda \); this leads to a single-parameter maximization of \( K_1 \) over the parameter \( \theta \in \mathbb{R}^+ \) whose optimization leads to the closed-form expression of the lower bound in Corollary 1. For comparison, in order to assess the enhanced tightness of the new lower bounds, note that the ratio of the upper and lower bounds on the total variation distance from [4, Theorems 1 and 2] (or Theorem 2 here) is roughly equal to 32 for all values of \( \lambda \).

### III. Improved Lower Bounds on the Relative Entropy

The following theorem relies on the new lower bound on the total variation distance in Theorem 2 and the distribution-dependent refinement of Pinsker’s inequality in [13]. Their combination serves to derive a new lower bound on the relative entropy between the distribution of a sum of independent Bernoulli random variables and a Poisson distribution with the same mean. The following upper bound on the relative entropy was introduced in [13, Theorem 1]. Together with the new lower bound on the relative entropy, it leads to the following statement:

**Theorem 3:** In the setting of Theorem 1, the relative entropy between the probability distribution of \( W \) and the Poisson distribution with mean \( \lambda = \mathbb{E}(W) \) satisfies the inequality:

\[
K_2(\lambda) \left( \sum_{i=1}^n p_i^2 \right)^2 \leq D\left(W \mid \mid \text{Po}(\lambda)\right) \leq \frac{1}{\lambda} \sum_{i=1}^n p_i^2
\]

with \( K_1 \) from (7), and

\[
m(\lambda) \triangleq \left\{ \begin{array}{ll} \log \left( \frac{1}{\alpha_1} \right) & \text{if } \lambda \in (0, \log 2) \\ 2 & \text{if } \lambda \geq \log 2. \end{array} \right.
\]

**Proof:** See [18, Section 4].

**Remark 6:** The combination of the original lower bound on the total variation distance from [4, Theorem 2] (see (5)) with Pinsker’s inequality gives the following lower bound on the relative entropy:

\[
D\left(W \mid \mid \text{Po}(\lambda)\right) \geq \frac{1}{512} \left( 1 - \frac{1}{\lambda^2} \right) \left( \sum_{i=1}^n p_i^2 \right)^2.
\]

In light of Remark 1 it is possible to quantify the improvement that is obtained by the new lower bound of Theorem 3 in comparison to the looser lower bound in (23). The improvement of the new lower bound on the relative entropy is by a factor of 179.7 \( \log \left( \frac{1}{\alpha_1} \right) \) for \( \lambda \approx 0 \), a factor of 9.22 for \( \lambda \rightarrow \infty \), and at least by a factor of 6.14 for all \( \lambda \in (0, \infty) \). The above conclusions are supported by Figure 2 that refers to the special case of the relative entropy between the binomial and Poisson distributions.

**Remark 7:** In [10, Example 6], it is shown that if \( \mathbb{E}(X) \leq \lambda \) then \( D(P_X \mid \mid \text{Po}(\lambda)) \geq \frac{1}{\lambda^2} \left( \mathbb{E}(X) - \lambda \right)^2 \). Since \( \mathbb{E}(W) = \lambda \) then this lower bound on the relative entropy is not informative for the relative entropy \( D(P_W \mid \mid \text{Po}(\lambda)) \) where \( \mathbb{E}(W) = \lambda \). Theorem 3 and the loosened bound in (23) are, on the other hand, informative in the studied case.

We were notified in [12] about the existence of another recently derived lower bound on the relative entropy \( D(P_X \mid \mid \text{Po}(\lambda)) \) in terms of the variance of a random variable \( X \) with values in \( \mathbb{N}_0 \) (this lower bound appears in a
is also supported by the new lower bounds in Theorems 3 and Eq. (23) since the common term in these lower bounds is equal to $\lambda < \log(2)$. The inequality, applied to the Poisson distribution, affects the lower bound for this figure, it shows that the probability-dependent refinement of Pinsker’s inequality will be fully covered in our next work. In the setting $X \sim \text{Bernoulli}(p_i)$, with $p_i \triangleq \frac{1}{n}$ is fixed for all $i \in \{1, \ldots, n\}$. This scaling of the relative entropy is supported by the upper bound on the relative entropy by Kontoyiannis et al. (see [13, Theorem 1]) that is equal to $\frac{\lambda}{2n} \log(2) = \frac{\lambda}{2n} + O(\frac{1}{n})$. It is also supported by the new lower bounds in Theorems 2 and Eq. (23) since the common term in these lower bounds is equal to $(\sum_{i=1}^{n} p_i^2)^{-\frac{1}{2}} = \frac{1}{\sqrt{n}}$, so a multiplication of these lower bounds on the relative entropy by $n^2$ gives an expression that only depends on $\lambda$. It follows from [11, Theorem 1] (see also [1, p. 2302]) that $D(\text{Bin}(n, \frac{1}{2}) || \text{Pois}(\lambda)) = \frac{\lambda}{2n} + O(\frac{1}{n})$ (so, the exact value is asymptotically equal to one-quarter of the upper bound). This figure shows the upper and lower bounds, as well as the exact asymptotic result, in order to study the tightness of the existing upper bound and the new lower bounds. By comparing the dotted and dashed lines, this figure also shows the significant impact of the refinement of the lower bound on the total variation distance by Barbour and Hall (see [11, Theorem 2]) on the improved lower bound on the relative entropy (the former improvement is squared via Pinsker’s inequality or its refinement). Furthermore, by comparing the dotted and solid lines of this figure, it shows that the probability-dependent refinement of Pinsker’s inequality, applied to the Poisson distribution, affects the lower bound for $\lambda < \log(2)$.

IV. EXAMPLES FOR THE USE OF THE NEW BOUNDS

We conclude our discussion in this paper by outlining two uses of the new lower bounds in this work:

1. The use of the new lower bound on the total variation distance for the Poisson approximation of a sum of independent Bernoulli random variables is exemplified in [19]. The latter work introduces new entropy bounds for discrete random variables via maximal coupling, providing 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. The new lower bound on the total variation distance for the Poisson approximation from this work was involved in the calculation of some improved bounds on the difference between the entropy of a sum of independent Bernoulli random variables and the entropy of a Poisson random variable of the same mean. A possible application of the latter problem is related to getting bounds on the sum-rate capacity of a noiseless $K$-user multiple-access channel with binary inputs. For more details, the reader is referred to [19, Section 4].

The use of the new lower bound on the relative entropy for the Poisson approximation of a sum of Bernoulli random variables is exemplified in [18, Section 4.3] in the context of binary hypothesis testing. The impact of the improvement of the lower bound on the relative entropy is also exemplified numerically in this context.

REFERENCES

[1] J. A. Adell, A. Lekouna and Y. Yu, “Sharp bounds on the entropy of the Poisson law and related quantities,” IEEE Trans. on Information Theory, vol. 56, no. 5, pp. 2299–2306, May 2010.
[2] R. Arratia, L. Goldstein and L. Gordon, “Poisson approximation and the Chen–Stein method,” Statistical Science, vol. 5, no. 4, pp. 403–424, November 1990.
[3] A. D. Barbour and L. H. Y. Chen, An Introduction to Stein’s Method, Lecture Notes Series, Institute for Mathematical Sciences, Singapore University Press and World Scientific, 2005.
[4] A. D. Barbour and P. Hall, “On the rate of Poisson convergence,” Mathematical Proceedings of the Cambridge Philosophical Society, vol. 95, no. 3, pp. 473–480, 1984.
[5] A. D. Barbour, L. Holst and S. Janson, Poisson Approximation, Oxford University Press, 1992.
[6] A. D. Barbour, O. Johnson, I. Kontoyiannis and M. Madiman, “Compound Poisson approximation via information functionals,” Electronic Journal of Probability, vol. 15, paper no. 42, pp. 1344–1369, August 2010.
[7] L. H. Y. Chen, “Poisson approximation for dependent trials,” Annals of Probability, vol. 3, no. 3, pp. 534–545, June 1975.
[8] T. M. Cover and J. A. Thomas, Elements of Information Theory, John Wiley and Sons, second edition, 2006.
[9] P. Deheuvels and D. Pfeifer, “A semigroup approach to Poisson approximation,” Annals of Probability, vol. 14, no. 2, pp. 663–676, April 1986.
[10] P. Harremoës and C. Vignat, “Lower bounds on information divergence,” IEEE Trans. on Information Theory, February 2011. Online available at http://arxiv.org/pdf/1102.2536.pdf.
[11] P. Harremoës and P. S. Ruzankin, “Rate of convergence to Poisson law in terms of information divergence,” IEEE Trans. on Information Theory, vol. 50, no. 9, pp. 2145–2149, September 2004.
[12] P. Harremoës, personal communications, July 2012.
[13] I. Kontoyiannis, P. Harremoës and O. Johnson, “Entropy and the law of large numbers,” IEEE Trans. on Information Theory, vol. 51, no. 2, pp. 466–472, February 2005.
[14] L. Le Cam, “An approximation theorem for the Poisson binomial distribution,” Pacific Journal of Mathematics, vol. 10, no. 4, pp. 1181–1197, Spring 1960.
[15] E. Ordentlich and M. J. Weinberger, “A distribution dependent refinement of Pinsker’s inequality,” IEEE Trans. on Information Theory, vol. 51, no. 5, pp. 1836–1840, May 2005.
[16] S. M. Ross and E. A. Peköz, A Second Course in Probability, Probability Bookstore, 2007.
[17] N. Ross, “Fundamentals of Stein’s Method,” Probability Surveys, vol. 8, pp. 210–293, 2011.
[18] I. Sason, “An information-theoretic perspective of the Poisson approximation via the Chen-Stein method,” submitted to the IEEE Trans. on Information Theory, June 2012. [Online]. Available: http://arxiv.org/abs/1206.6811
[19] I. Sason, “Entropy bounds for discrete random variables via coupling,” submitted to the IEEE Trans. on Information Theory, September 2012. [Online]. Available: http://arxiv.org/abs/1209.5259