

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 for the starting point of the so called *Chen-Stein method for the Poisson approximation* is the following (see Chen (1975)). Let $\{X_i\}_{i=1}^n$ be independent Bernoulli random variables with $\mathbb{E}(X_i) = p_i$. Let $W \triangleq \sum_{i=1}^n X_i$ and $V_i \triangleq \sum_{j \neq i} X_j$ for every $i \in \{1, \dots, n\}$, and $Z \sim \text{Po}(\lambda)$ with mean $\lambda \triangleq \sum_{i=1}^n p_i$. It is easy to show that

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

holds for an arbitrary bounded function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ where $\mathbb{N}_0 \triangleq \{0, 1, \dots\}$. Furthermore (see, e.g., Chapter 2 in Ross and Peköz (2007))

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

which then serves to provide 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 method*, serves as a powerful tool for the derivation of rigorous bounds for various distributional approximations. Nice expositions of this method are provided by, e.g., Arratia et al. (1990), Ross and Peköz (2007) and Ross (2011). Furthermore, some interesting links between the Chen-Stein method and information-theoretic functionals in the context of Poisson and compound Poisson approximations are provided by Barbour et al. (2010).

Throughout this letter, the term ‘distribution’ refers to a discrete probability mass function of an integer-valued random variable. In the following, we introduce 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

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

where the supremum 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

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

so the total variation distance is equal to 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 (see Le Cam (1960)) 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 $d_{\text{TV}}(P_W, \text{Po}(\lambda)) \leq \sum_{i=1}^n p_i^2$ so if, e.g., $X_i \sim \text{Bern}(\frac{\lambda}{n})$ for every $i \in \{1, \dots, n\}$ (referring to the case where W is binomially distributed) then this upper bound is equal to $\frac{\lambda^2}{n}$, decaying to zero as $n \rightarrow \infty$. The following theorem combines Theorems 1 and 2 of Barbour and Hall (1984), 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 $i \in \{1, \dots, n\}$, and $\mathbb{E}(W) = \lambda$. Then, the total variation distance between the probability distribution of W and the Poisson distribution with mean λ satisfies

$$\frac{1}{32} \left(1 \wedge \frac{1}{\lambda}\right) \sum_{i=1}^n p_i^2 \leq d_{\text{TV}}(P_W, \text{Po}(\lambda)) \leq \left(\frac{1 - e^{-\lambda}}{\lambda}\right) \sum_{i=1}^n p_i^2 \quad (5)$$

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 $\frac{1}{14}$ with no explicit proof (see Remark 3.2.2 in Barbour et al. (1992)). This shows that, for independent Bernoulli random variables, these bounds are essentially tight. Furthermore, note that the upper bound in (5) improves Le Cam's inequality; for large values of λ , this improvement is by approximately a factor of $\frac{1}{\lambda}$.

This letter 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 derivation of these new bounds generalizes and improves the analysis by Barbour and Hall (1984), based on the Chen-Stein method for the Poisson approximation. This letter concludes by outlining a use of the new lower bounds for the analysis in Sason (2012), followed by a comparison of the new bounds to previously reported bounds.

This work forms a continuation of the line of work in Barbour and Chen (2005)–Kontoyiannis et al. (2005) where the Chen-Stein method was studied in the context of the Poisson and compound Poisson approximations, and it was linked to an information-theoretic context by Barbour et al. (2010), Kontoyiannis et al. (2005), and Sason (2012).

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 λ satisfies the inequality

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

where

$$K_1(\lambda) \triangleq \sup_{\substack{\alpha_1, \alpha_2 \in \mathbb{R}, \\ \alpha_2 \leq \lambda + \frac{3}{2}, \\ \theta > 0}} \left(\frac{1 - h_\lambda(\alpha_1, \alpha_2, \theta)}{2g_\lambda(\alpha_1, \alpha_2, \theta)} \right) \quad (7)$$

and

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

$$x_+ \triangleq \max\{x, 0\}, \quad x_+^2 \triangleq (x_+)^2, \quad \forall x \in \mathbb{R} \quad (9)$$

$$g_\lambda(\alpha_1, \alpha_2, \theta) \triangleq \max \left\{ \left| \left(1 + \sqrt{\frac{2}{\theta\lambda e}} \cdot |\alpha_1 - \alpha_2| \right) \lambda + \max_{u_i} \{x(u_i)\} \right|, \left| \left(2e^{-\frac{3}{2}} + \sqrt{\frac{2}{\theta\lambda e}} \cdot |\alpha_1 - \alpha_2| \right) \lambda - \min_{u_i} \{x(u_i)\} \right| \right\} \quad (10)$$

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

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

$$c_0 \triangleq (\alpha_2 - \alpha_1)(\lambda - \alpha_2) \quad (13)$$

$$c_1 \triangleq \sqrt{\theta \lambda} (\lambda + \alpha_1 - 2\alpha_2) \quad (14)$$

$$c_2 \triangleq -\theta \lambda. \quad (15)$$

Proof: See Section IV-A. The derivation relies on the Chen-Stein method for the Poisson approximation, and it improves (significantly) the constant in the lower bound of Theorem 2 of Barbour and Hall (1984). ■

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, but they offer a significant improvement in their tightness (see Section V).

Remark 2: The cardinality of the set $\{u_i\}$ in (12) is equal to 3 (see Section IV-A).

Remark 3: The optimization that is required for the computation of K_1 in (7) w.r.t. the three parameters $\alpha_1, \alpha_2 \in \mathbb{R}$ and $\theta \in \mathbb{R}^+$ is performed numerically.

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 Theorem 2 of Barbour and Hall (1984). The following lower bound follows from Theorem 2 via 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

$$\tilde{K}_1(\lambda) \sum_{i=1}^n p_i^2 \leq d_{\text{TV}}(P_W, \text{Po}(\lambda)) \leq \left(\frac{1 - e^{-\lambda}}{\lambda} \right) \sum_{i=1}^n p_i^2 \quad (16)$$

where

$$\tilde{K}_1(\lambda) \triangleq \frac{e}{2\lambda} \frac{1 - \frac{1}{\theta} \left(3 + \frac{7}{\lambda} \right)}{\theta + 2e^{-1/2}} \quad (17)$$

$$\theta \triangleq 3 + \frac{7}{\lambda} + \frac{1}{\lambda} \cdot \sqrt{(3\lambda + 7)[(3 + 2e^{-1/2})\lambda + 7]}. \quad (18)$$

Proof: See Section IV-B. ■

III. OUTLOOK

We conclude our discussion in this letter by outlining a use of the new lower bounds in this work: 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 by Sason (2012). This 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 binary adder multiple-access channel (see Sason (2012)).

IV. PROOFS OF THE NEW BOUNDS

A. Proof of Theorem 2

The proof of Theorem 2 starts similarly to the proof of Theorem 2 of Barbour and Hall (1984). However, it significantly deviates from the original analysis in order to derive an improved lower bound on the total variation distance.

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 $i \in \{1, \dots, n\}$, and $Z \sim \text{Po}(\lambda)$ with mean $\lambda \triangleq \sum_{i=1}^n p_i$. From the basic equation of the Chen-Stein method, equation (1) holds for an arbitrary bounded function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$. Furthermore, it follows from the proof of Theorem 2 of Barbour and Hall (1984) that

$$d_{\text{TV}}(P_W, \text{Po}(\lambda)) \geq \frac{\sum_{j=1}^n \left\{ p_j^2 \mathbb{E}[f(V_j + 2) - f(V_j + 1)] \right\}}{2 \sup_{k \in \mathbb{N}_0} |\lambda f(k + 1) - k f(k)|} \quad (19)$$

which holds, in general, for an arbitrary bounded function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ (see the appendix for a proof, which does not appear in the published version).

At this point, we deviate from the proof of Theorem 2 of Barbour and Hall (1984) by generalizing and refining (in a non-trivial way) the original analysis. The general problem with the current lower bound in (19) is that it is not calculable in closed form for a given f , so one needs to choose a proper function f and derive a closed-form expression for a lower bound on the right-hand side of (19). To this end, let

$$f(k) \triangleq (k - \alpha_1) \exp\left(-\frac{(k - \alpha_2)^2}{\theta\lambda}\right), \quad \forall k \in \mathbb{N}_0 \quad (20)$$

where $\alpha_1, \alpha_2 \in \mathbb{R}$ and $\theta \in \mathbb{R}^+$ are fixed constants (note that θ in (20) needs to be positive for f to be a bounded function). In order to derive a lower bound on the total variation distance, we calculate a lower bound on the numerator and an upper bound on the denominator of the right-hand side of (19) for the function f in (20). Referring to the numerator of the right-hand side of (19) with f in (20), for every $j \in \{1, \dots, n\}$,

$$\begin{aligned} & f(V_j + 2) - f(V_j + 1) \\ &= \int_{V_j + 1 - \alpha_2}^{V_j + 2 - \alpha_2} \frac{d}{du} \left((u + \alpha_2 - \alpha_1) \exp\left(-\frac{u^2}{\theta\lambda}\right) \right) du \\ &= \int_{V_j + 1 - \alpha_2}^{V_j + 2 - \alpha_2} \left(1 - \frac{2u(u + \alpha_2 - \alpha_1)}{\theta\lambda} \right) \exp\left(-\frac{u^2}{\theta\lambda}\right) du \\ &= \int_{V_j + 1 - \alpha_2}^{V_j + 2 - \alpha_2} \left(1 - \frac{2u^2}{\theta\lambda} \right) \exp\left(-\frac{u^2}{\theta\lambda}\right) du - \frac{2(\alpha_2 - \alpha_1)}{\theta\lambda} \int_{V_j + 1 - \alpha_2}^{V_j + 2 - \alpha_2} u \exp\left(-\frac{u^2}{\theta\lambda}\right) du \\ &= \int_{V_j + 1 - \alpha_2}^{V_j + 2 - \alpha_2} \left(1 - \frac{2u^2}{\theta\lambda} \right) \exp\left(-\frac{u^2}{\theta\lambda}\right) du \\ &\quad - (\alpha_2 - \alpha_1) \left[\exp\left(-\frac{(V_j + 2 - \alpha_2)^2}{\theta\lambda}\right) - \exp\left(-\frac{(V_j + 1 - \alpha_2)^2}{\theta\lambda}\right) \right]. \end{aligned} \quad (21)$$

We rely in the following on the inequality

$$(1 - 2x) e^{-x} \geq 1 - 3x, \quad \forall x \geq 0.$$

Applying it to the integral on the right-hand side of (21) gives that

$$\begin{aligned}
& f(V_j + 2) - f(V_j + 1) \\
& \geq \int_{V_j+1-\alpha_2}^{V_j+2-\alpha_2} \left(1 - \frac{3u^2}{\theta\lambda}\right) du - (\alpha_2 - \alpha_1) \left[\exp\left(-\frac{(V_j + 2 - \alpha_2)^2}{\theta\lambda}\right) - \exp\left(-\frac{(V_j + 1 - \alpha_2)^2}{\theta\lambda}\right) \right] \\
& \geq 1 - \frac{(V_j + 2 - \alpha_2)^3 - (V_j + 1 - \alpha_2)^3}{\theta\lambda} \\
& \quad - |\alpha_2 - \alpha_1| \cdot \left| \exp\left(-\frac{(V_j + 2 - \alpha_2)^2}{\theta\lambda}\right) - \exp\left(-\frac{(V_j + 1 - \alpha_2)^2}{\theta\lambda}\right) \right|. \tag{22}
\end{aligned}$$

In order to proceed, note that if $x_1, x_2 \geq 0$ then (on the basis of the mean-value theorem of calculus)

$$\begin{aligned}
& |e^{-x_2} - e^{-x_1}| \\
& = |e^{-c}(x_1 - x_2)| \quad \text{for some } c \in [x_1, x_2] \\
& \leq e^{-\min\{x_1, x_2\}} |x_1 - x_2|
\end{aligned}$$

which, by applying it to the second term on the right-hand side of (22), gives that for every $j \in \{1, \dots, n\}$

$$\begin{aligned}
& \left| \exp\left(-\frac{(V_j + 2 - \alpha_2)^2}{\theta\lambda}\right) - \exp\left(-\frac{(V_j + 1 - \alpha_2)^2}{\theta\lambda}\right) \right| \\
& \leq \exp\left(-\frac{\min\{(V_j + 2 - \alpha_2)^2, (V_j + 1 - \alpha_2)^2\}}{\theta\lambda}\right) \cdot \left(\frac{(V_j + 2 - \alpha_2)^2 - (V_j + 1 - \alpha_2)^2}{\theta\lambda}\right). \tag{23}
\end{aligned}$$

Since $V_j = \sum_{i \neq j} X_i \geq 0$ then

$$\begin{aligned}
& \min\{(V_j + 2 - \alpha_2)^2, (V_j + 1 - \alpha_2)^2\} \\
& \geq \begin{cases} 0 & \text{if } \alpha_2 \geq 1 \\ (1 - \alpha_2)^2 & \text{if } \alpha_2 < 1 \end{cases} \\
& = (1 - \alpha_2)_+^2 \tag{24}
\end{aligned}$$

where

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

Hence, the combination of the two inequalities in (23)–(24) gives that

$$\begin{aligned}
& \left| \exp\left(-\frac{(V_j + 2 - \alpha_2)^2}{\theta\lambda}\right) - \exp\left(-\frac{(V_j + 1 - \alpha_2)^2}{\theta\lambda}\right) \right| \\
& \leq \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \cdot \left(\frac{|(V_j + 2 - \alpha_2)^2 - (V_j + 1 - \alpha_2)^2|}{\theta\lambda}\right) \\
& = \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \cdot \frac{|2V_j + 3 - 2\alpha_2|}{\theta\lambda} \\
& \leq \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \cdot \frac{2V_j + |3 - 2\alpha_2|}{\theta\lambda} \tag{25}
\end{aligned}$$

and therefore, a combination of the inequalities in (22) and (25) gives that

$$\begin{aligned}
& f(V_j + 2) - f(V_j + 1) \\
& \geq 1 - \frac{(V_j + 2 - \alpha_2)^3 - (V_j + 1 - \alpha_2)^3}{\theta\lambda} \\
& \quad - |\alpha_2 - \alpha_1| \cdot \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \cdot \frac{2V_j + |3 - 2\alpha_2|}{\theta\lambda}. \tag{26}
\end{aligned}$$

Let $U_j \triangleq V_j - \lambda$; then

$$\begin{aligned}
& f(V_j + 2) - f(V_j + 1) \\
& \geq 1 - \frac{(U_j + \lambda + 2 - \alpha_2)^3 - (U_j + \lambda + 1 - \alpha_2)^3}{\theta\lambda} \\
& \quad - |\alpha_2 - \alpha_1| \cdot \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \cdot \frac{2U_j + 2\lambda + |3 - 2\alpha_2|}{\theta\lambda} \\
& = 1 - \frac{3U_j^2 + 3(3 - 2\alpha_2 + 2\lambda)U_j + (2 - \alpha_2 + \lambda)^3 - (1 - \alpha_2 + \lambda)^3}{\theta\lambda} \\
& \quad - |\alpha_2 - \alpha_1| \cdot \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \cdot \frac{2U_j + 2\lambda + |3 - 2\alpha_2|}{\theta\lambda}. \tag{27}
\end{aligned}$$

In order to derive a lower bound on the numerator of the right-hand side of (19), for the function f in (20), we need to calculate the expected value of the right-hand side of (27). To this end, the first and second moments of U_j are calculated as follows:

$$\mathbb{E}(U_j) = \mathbb{E}(V_j) - \lambda = \sum_{i \neq j} p_i - \sum_{i=1}^n p_i = -p_j \tag{28}$$

and

$$\begin{aligned}
& \mathbb{E}(U_j^2) \\
& = \text{Var}(U_j) + (\mathbb{E}(U_j))^2 \\
& = \text{Var}(V_j) + p_j^2 \\
& \stackrel{(a)}{=} \sum_{i \neq j} p_i(1 - p_i) + p_j^2 \\
& = \sum_{i \neq j} p_i - \sum_{i \neq j} p_i^2 + p_j^2 \\
& = \lambda - p_j - \sum_{i \neq j} p_i^2 + p_j^2. \tag{29}
\end{aligned}$$

where equality (a) holds since the binary random variables $\{X_i\}_{i=1}^n$ are independent and $\text{Var}(X_i) = p_i(1 - p_i)$. By taking expectations on both sides of (27), one obtains from (28) and (29) that

$$\begin{aligned}
& \mathbb{E}[f(V_j + 2) - f(V_j + 1)] \\
& \geq 1 - \frac{3\left(\lambda - p_j - \sum_{i \neq j} p_i^2 + p_j^2\right) + 3(3 - 2\alpha_2 + 2\lambda)(-p_j) + (2 - \alpha_2 + \lambda)^3 - (1 - \alpha_2 + \lambda)^3}{\theta\lambda} \\
& \quad - |\alpha_2 - \alpha_1| \cdot \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \cdot \left(\frac{-2p_j + 2\lambda + |3 - 2\alpha_2|}{\theta\lambda}\right) \\
& = 1 - \frac{3\lambda + (2 - \alpha_2 + \lambda)^3 - (1 - \alpha_2 + \lambda)^3 - \left[3p_j(1 - p_j) + 3\sum_{i \neq j} p_i^2 + 3(3 - 2\alpha_2 + 2\lambda)p_j\right]}{\theta\lambda} \\
& \quad - \left(\frac{|\alpha_2 - \alpha_1|(2\lambda - 2p_j + |3 - 2\alpha_2|)}{\theta\lambda}\right) \cdot \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right) \\
& \geq 1 - \frac{3\lambda + (2 - \alpha_2 + \lambda)^3 - (1 - \alpha_2 + \lambda)^3 - (9 - 6\alpha_2 + 6\lambda)p_j}{\theta\lambda} \\
& \quad - \left(\frac{|\alpha_2 - \alpha_1|(2\lambda + |3 - 2\alpha_2|)}{\theta\lambda}\right) \cdot \exp\left(-\frac{(1 - \alpha_2)_+^2}{\theta\lambda}\right). \tag{30}
\end{aligned}$$

Therefore, from (30), the following lower bound on the right-hand side of (19) holds

$$\begin{aligned} \sum_{j=1}^n \left\{ p_j^2 \mathbb{E}[f(V_j + 2) - f(V_j + 1)] \right\} &\geq \left(\frac{3(3 - 2\alpha_2 + 2\lambda)}{\theta\lambda} \right) \sum_{j=1}^n p_j^3 \\ &+ \left(1 - \frac{3\lambda + (2 - \alpha_2 + \lambda)^3 - (1 - \alpha_2 + \lambda)^3 + |\alpha_1 - \alpha_2|(2\lambda + |3 - 2\alpha_2|) \exp\left(-\frac{(1-\alpha_2)_+^2}{\theta\lambda}\right)}{\theta\lambda} \right) \sum_{j=1}^n p_j^2. \end{aligned} \quad (31)$$

Note that if $\alpha_2 \leq \lambda + \frac{3}{2}$, which is a condition that is involved in the maximization of (7), then the first term on the right-hand side of (31) can be removed, and the resulting lower bound on the numerator of the right-hand side of (19) takes the form

$$\sum_{j=1}^n \left\{ p_j^2 \mathbb{E}[f(V_j + 2) - f(V_j + 1)] \right\} \geq \left(1 - h_\lambda(\alpha_1, \alpha_2, \theta) \right) \sum_{j=1}^n p_j^2 \quad (32)$$

where the function h_λ is introduced in (8).

We turn now to deriving an upper bound on the denominator of the right-hand side of (19). Therefore, we need to derive a closed-form upper bound on $\sup_{k \in \mathbb{N}_0} |\lambda f(k+1) - k f(k)|$ with the function f in (20). For every $k \in \mathbb{N}_0$,

$$\lambda f(k+1) - k f(k) = \lambda [f(k+1) - f(k)] + (\lambda - k) f(k). \quad (33)$$

In the following, we derive bounds on each of the two terms on the right-hand side of (33), and we start with the first term. Let

$$t(u) \triangleq (u + \alpha_2 - \alpha_1) \exp\left(-\frac{u^2}{\theta\lambda}\right), \quad \forall u \in \mathbb{R}$$

then $f(k) = t(k - \alpha_2)$ for every $k \in \mathbb{N}_0$, and by the mean-value theorem of calculus,

$$\begin{aligned} f(k+1) - f(k) &= t(k+1 - \alpha_2) - t(k - \alpha_2) \\ &= t'(c_k) \quad \text{for some } c_k \in [k - \alpha_2, k + 1 - \alpha_2] \\ &= \left(1 - \frac{2c_k^2}{\theta\lambda} \right) \exp\left(-\frac{c_k^2}{\theta\lambda}\right) + \left(\frac{2(\alpha_1 - \alpha_2)c_k}{\theta\lambda} \right) \exp\left(-\frac{c_k^2}{\theta\lambda}\right). \end{aligned} \quad (34)$$

Referring to the first term on the right-hand side of (34), let

$$p(u) \triangleq (1 - 2u)e^{-u}, \quad \forall u \geq 0$$

then the global maximum and minimum of p over the non-negative real line are obtained at $u = 0$ and $u = \frac{3}{2}$, respectively, and therefore

$$-2e^{-\frac{3}{2}} \leq p(u) \leq 1, \quad \forall u \geq 0.$$

Let $u = \frac{c_k^2}{\theta\lambda}$; then it follows that the first term on the right-hand side of (34) satisfies the inequality

$$-2e^{-\frac{3}{2}} \leq \left(1 - \frac{2c_k^2}{\theta\lambda} \right) \exp\left(-\frac{c_k^2}{\theta\lambda}\right) \leq 1. \quad (35)$$

Furthermore, referring to the second term on the right-hand side of (34), let

$$q(u) \triangleq ue^{-u^2}, \quad \forall u \in \mathbb{R}$$

then the global maximum and minimum of q over the real line are obtained at $u = +\frac{\sqrt{2}}{2}$ and $u = -\frac{\sqrt{2}}{2}$, respectively, and therefore

$$-\frac{1}{2}\sqrt{\frac{2}{e}} \leq q(u) \leq +\frac{1}{2}\sqrt{\frac{2}{e}}, \quad \forall u \in \mathbb{R}.$$

Let this time $u = \sqrt{\frac{c_k}{\theta\lambda}}$; then it follows that the second term on the right-hand side of (34) satisfies

$$\left| \left(\frac{2(\alpha_1 - \alpha_2)c_k}{\theta\lambda} \right) \cdot \exp\left(-\frac{c_k^2}{\theta\lambda}\right) \right| \leq \sqrt{\frac{2}{\theta\lambda e}} \cdot |\alpha_1 - \alpha_2|. \quad (36)$$

Hence, on combining the equality in (34) with the two inequalities in (35) and (36), it follows that the first term on the right-hand side of (33) satisfies

$$-\left(2\lambda e^{-\frac{3}{2}} + \sqrt{\frac{2\lambda}{\theta e}} \cdot |\alpha_1 - \alpha_2|\right) \leq \lambda[f(k+1) - f(k)] \leq \lambda + \sqrt{\frac{2\lambda}{\theta e}} \cdot |\alpha_1 - \alpha_2|, \quad \forall k \in \mathbb{N}_0. \quad (37)$$

We continue the analysis by a derivation of bounds on the second term of the right-hand side of (33). For the function f in (20), it is equal to

$$\begin{aligned} & (\lambda - k) f(k) \\ &= (\lambda - k)(k - \alpha_1) \exp\left(-\frac{(k - \alpha_2)^2}{\theta\lambda}\right) \\ &= [(\lambda - \alpha_2) + (\alpha_2 - k)] [(k - \alpha_2) + (\alpha_2 - \alpha_1)] \exp\left(-\frac{(k - \alpha_2)^2}{\theta\lambda}\right) \\ &= \left[(\lambda - \alpha_2)(k - \alpha_2) + (\alpha_2 - \alpha_1)(\lambda - \alpha_2) - (k - \alpha_2)^2 + (\alpha_1 - \alpha_2)(k - \alpha_2)\right] \exp\left(-\frac{(k - \alpha_2)^2}{\theta\lambda}\right) \\ &= \left[\sqrt{\theta\lambda}(\lambda - \alpha_2)v_k - \theta\lambda v_k^2 - \sqrt{\theta\lambda}(\alpha_2 - \alpha_1)v_k + (\alpha_2 - \alpha_1)(\lambda - \alpha_2)\right] e^{-v_k^2}, \quad v_k \triangleq \frac{k - \alpha_2}{\sqrt{\theta\lambda}} \quad \forall k \in \mathbb{N}_0 \\ &= (c_0 + c_1 v_k + c_2 v_k^2) e^{-v_k^2} \end{aligned} \quad (38)$$

where the coefficients c_0, c_1 and c_2 are introduced in Eqs. (13)–(15), respectively. In order to derive bounds on the left-hand side of (38), let us find the global maximum and minimum of the function x in (11):

$$x(u) \triangleq (c_0 + c_1 u + c_2 u^2) e^{-u^2} \quad \forall u \in \mathbb{R}.$$

Note that $\lim_{u \rightarrow \pm\infty} x(u) = 0$ and x is differentiable over the real line, so the global maximum and minimum of x are attained at finite points and their corresponding values are finite. By setting the derivative of x to zero, we have that the candidates for the global maximum and minimum of x over the real line are the real zeros $\{u_i\}$ of the cubic polynomial equation in (12). Note that by their definition in (12), the values of $\{u_i\}$ are *independent* of the value of $k \in \mathbb{N}_0$, and also the size of the set $\{u_i\}$ is equal to 3 (see Remark 2). Hence, it follows from (38) that

$$\min_{i \in \{1,2,3\}} \{x(u_i)\} \leq (\lambda - k) f(k) \leq \max_{i \in \{1,2,3\}} \{x(u_i)\}, \quad \forall k \in \mathbb{N}_0 \quad (39)$$

where these bounds on the second term on the right-hand side of (33) are independent of the value of $k \in \mathbb{N}_0$.

In order to get bounds on the left-hand side of (33), note that from the bounds on the first and second terms on the right-hand side of (33) (see (37) and (39), respectively) then for every $k \in \mathbb{N}_0$

$$\begin{aligned} & \min_{i \in \{1,2,3\}} \{x(u_i)\} - \left(2\lambda e^{-\frac{3}{2}} + \sqrt{\frac{2\lambda}{\theta e}} \cdot |\alpha_1 - \alpha_2|\right) \\ & \leq \lambda f(k+1) - k f(k) \\ & \leq \max_{i \in \{1,2,3\}} \{x(u_i)\} + \lambda + \sqrt{\frac{2\lambda}{\theta e}} \cdot |\alpha_1 - \alpha_2| \end{aligned} \quad (40)$$

which yields that the following inequality is satisfied:

$$\sup_{k \in \mathbb{N}_0} |\lambda f(k+1) - k f(k)| \leq g_\lambda(\alpha_1, \alpha_2, \theta) \quad (41)$$

where the function g_λ is introduced in (10). Finally, by combining the inequalities in Eqs. (19), (32) and (41), the lower bound on the total variation distance in (6) follows. The existing upper bound on the total variation distance in (6) was derived in Theorem 1 of Barbour and Hall (1984) (see Theorem 1 here). This completes the proof of Theorem 2.

B. Proof of Corollary 1

Corollary 1 follows as a special case of Theorem 2 when the proposed function f in (20) is chosen such that two of its three free parameters (i.e., α_1 and α_2) are determined sub-optimally, and its third parameter (θ) is determined optimally in terms of the sub-optimal selection of the two other parameters. More explicitly, let α_1 and α_2 in (20) be set to be equal to λ (i.e., $\alpha_1 = \alpha_2 = \lambda$). From (13)–(15), this setting implies that $c_0 = c_1 = 0$ and $c_2 = -\theta\lambda < 0$ (since $\theta, \lambda > 0$). The cubic polynomial equation in (12), which corresponds to this (possibly sub-optimal) setting of α_1 and α_2 , is

$$2c_2u^3 - 2c_2u = 0$$

whose zeros are $u = 0, \pm 1$. The function x in (11) therefore takes the form

$$x(u) = c_2u^2e^{-u^2} \quad \forall u \in \mathbb{R}$$

so $x(0) = 0$ and $x(\pm 1) = \frac{c_2}{e} < 0$. This implies that

$$\min_{i \in \{1,2,3\}} x(u_i) = \frac{c_2}{e}, \quad \max_{i \in \{1,2,3\}} x(u_i) = 0,$$

and therefore h_λ and g_λ in (8) and (10), respectively, are simplified to

$$h_\lambda(\lambda, \lambda, \theta) = \frac{3\lambda + 7}{\theta\lambda}, \quad (42)$$

$$g_\lambda(\lambda, \lambda, \theta) = \lambda \max\{1, 2e^{-\frac{3}{2}} + \theta e^{-1}\}. \quad (43)$$

This sub-optimal setting of α_1 and α_2 in (20) implies that the coefficient K_1 in (7) is replaced with a loosened version

$$K'_1(\lambda) \triangleq \sup_{\theta > 0} \left(\frac{1 - h_\lambda(\lambda, \lambda, \theta)}{2g_\lambda(\lambda, \lambda, \theta)} \right). \quad (44)$$

Let $\theta \geq e - \frac{2}{\sqrt{e}}$; then (43) is simplified to $g_\lambda(\lambda, \lambda, \theta) = \lambda(2e^{-\frac{3}{2}} + \theta e^{-1})$. It therefore follows from (6), (7) and (42)–(44) that

$$d_{\text{TV}}(P_W, \text{Po}(\lambda)) \geq \tilde{K}_1(\lambda) \sum_{i=1}^n p_i^2 \quad (45)$$

where

$$\tilde{K}_1(\lambda) = \sup_{\theta \geq e - \frac{2}{\sqrt{e}}} \left(\frac{1 - \frac{3\lambda+7}{\theta\lambda}}{2\lambda(2e^{-\frac{3}{2}} + \theta e^{-1})} \right) \quad (46)$$

and, in general, $K'_1(\lambda) \geq \tilde{K}_1(\lambda)$ due to the above restricted constraint on θ (see (44) versus (46)). Differentiating the function inside the supremum w.r.t. θ and setting its derivative to zero, one gets the following quadratic equation in θ :

$$\lambda\theta^2 - 2(3\lambda + 7)\theta - 2(3\lambda + 7)e^{-1} = 0$$

whose positive solution is the optimized value of θ in (18). Furthermore, it is clear that this value of θ in (18) is larger than, e.g., 3, so it satisfies the constraint in (46). This completes the proof of Corollary 1.

V. A COMPARISON OF THE NEW BOUNDS WITH KNOWN RESULTS

The new lower bounds on the total variation distance in Theorem 2 and Corollary 1 scale like $\sum_{i=1}^n p_i^2$, similarly to the known upper and lower bounds in Theorem 1 that originally appear in Theorems 1 and 2 of Barbour and Hall (1984). However, the new lower bounds offer a significant improvement over the known lower bound in Theorem 1. More explicitly, from Theorems 1 and 2 of Barbour and Hall (1984), the ratio between the upper and lower bounds on the total variation distance (see (5)) is equal to 32 in the two extreme cases where $\lambda \rightarrow 0$ or $\lambda \rightarrow \infty$. In the

following, we calculate the ratio of the same upper bound and the new lower bound in Corollary 1 at these two extreme cases. In the limit where $\lambda \rightarrow \infty$, this ratio tends to

$$\begin{aligned}
& \lim_{\lambda \rightarrow \infty} \frac{\left(\frac{1-e^{-\lambda}}{\lambda}\right) \sum_{i=1}^n p_i^2}{\left(\frac{1-\frac{3\lambda+7}{\lambda\theta}}{2\lambda(2e^{-3/2}+\theta e^{-1})}\right) \sum_{i=1}^n p_i^2} \quad (\theta = \theta(\lambda) \text{ is given in Eq. (18)}) \\
&= \frac{2}{e} \lim_{\lambda \rightarrow \infty} \frac{\theta(2e^{-1/2} + \theta)}{\theta - \left(3 + \frac{7}{\lambda}\right)} \\
&= \frac{6}{e} \left(1 + \sqrt{1 + \frac{2}{3} \cdot e^{-1/2}}\right)^2 \approx 10.539 \tag{47}
\end{aligned}$$

where the last equality follows from (18), since $\lim_{\lambda \rightarrow \infty} \theta = 3 + \sqrt{3(3 + 2e^{-1/2})}$. Furthermore, the limit of this ratio when $\lambda \rightarrow 0$ is equal to

$$\begin{aligned}
& 2 \lim_{\lambda \rightarrow 0} \left(\frac{1-e^{-\lambda}}{\lambda}\right) \lim_{\lambda \rightarrow 0} \left(\frac{\lambda(2e^{-3/2} + \theta e^{-1})}{1 - \frac{3\lambda+7}{\lambda\theta}}\right) \\
&\stackrel{(a)}{=} \frac{28}{e} \lim_{\lambda \rightarrow 0} \left(\frac{2e^{-1/2} + \theta}{\theta - \left(3 + \frac{7}{\lambda}\right)}\right) \\
&\stackrel{(b)}{=} \frac{56}{e} \approx 20.601 \tag{48}
\end{aligned}$$

where equalities (a) and (b) hold since, from (18), it follows that $\lim_{\lambda \rightarrow 0}(\lambda\theta) = 14$. This implies that Corollary 1 improves the original lower bound on the total variation distance in Theorem 2 of Barbour and Hall (1984) by a factor of $\frac{32}{10.539} \approx 3.037$ in the limit where $\lambda \rightarrow \infty$, and it also improves it by a factor of $\frac{32}{20.601} \approx 1.553$ if $\lambda \rightarrow 0$ while still having a closed-form expression for the lower bound in Corollary 1. The only reason for this improvement is related to the optimal choice of the free parameter θ in (18), versus its sub-optimal choice in the proof of Theorem 2 of Barbour and Hall (1984). This observation has motivated to further improve the lower bound by introducing the two additional parameters $\alpha_1, \alpha_2 \in \mathbb{R}$ in Theorem 2; these parameters give two additional degrees of freedom in the function f in (20) (according to the proof in Section IV-B, these two parameters are set to be equal to λ for the derivation of the loosened and simplified bound in Corollary 1). The improvement in the lower bound of Theorem 2 (in comparison to Corollary 1) is especially dominant for low values of λ , as is shown in Figure 1. Note, however, that no improvement is obtained for high values of λ (e.g., for $\lambda \geq 20$, as is shown by Figure 1 on noticing that the curves in this plot merge at large values of λ).

The lower bound on the total variation distance in Theorem 2 implies the bound in Corollary 1 (see the proof in Section IV-B). Corollary 1 further implies the lower bound on the total variation distance in Theorem 2 of Barbour and Hall (1984) (see Theorem 1 here). The latter claim follows from the fact that the lower bound in (45) with the coefficient $\tilde{K}_1(\lambda)$ in (46) was loosened in the proof of Theorem 2 of Barbour and Hall (1984) by a sub-optimal selection of the parameter θ , which leads to a lower bound on $\tilde{K}_1(\lambda)$ (the sub-optimal selection of θ in the proof of Theorem 2 of Barbour and Hall (1984) is $\theta = 21 \max\{1, \frac{1}{\lambda}\}$). On the other hand, the optimized value of θ that is used in (18) provides an exact closed-form expression for $\tilde{K}_1(\lambda)$ in (46), and it leads to the derivation of the improved lower bound in Corollary 1.

Theorem 1.2 of Deheuvels and Pfeifer (1986) 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 $\sum_{i=1}^n p_i \rightarrow \infty$ and $\max_{1 \leq i \leq n} p_i \rightarrow 0$ as $n \rightarrow \infty$ then

$$d_{\text{TV}}(P_W, \text{Po}(\lambda)) \sim \frac{1}{\sqrt{2\pi e} \lambda} \sum_{i=1}^n p_i^2. \tag{49}$$

This implies that the ratio of the upper bound on the total variation distance in Theorem 1 of Barbour and Hall (1984) (see Theorems 1 here) and this asymptotic expression is equal to $\sqrt{2\pi e} \approx 4.133$. Therefore, the ratio between

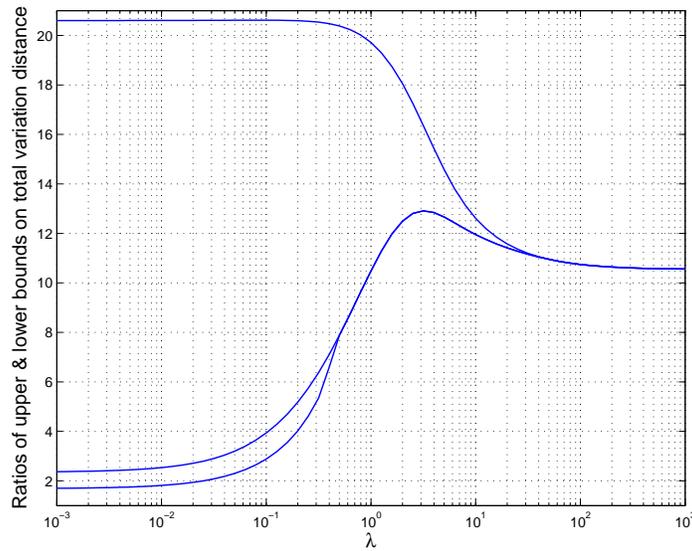


Fig. 1. The figure presents curves that correspond to ratios of upper and lower bounds on the total variation distance between the sum of independent Bernoulli random variables and the Poisson distribution with the same mean λ . The upper bound on the total variation distance for all these three curves is the bound given by Barbour and Hall (see Theorem 1 of Barbour and Hall (1984) or Theorem 1 here). The lower bounds that the three curves refer to are the following. The curve at the bottom (i.e., the one which provides the lowest ratio for a fixed λ) 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 λ corresponds to the looser lower bound obtained when α_1 and α_2 in (7) are set to be equal (i.e., $\alpha_1 = \alpha_2 \triangleq \alpha$ is their common value), and so 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 α is set to be equal to λ ; 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 for 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 Theorems 1 and 2 of Barbour and Hall (1984) (or Theorem 1 here) is roughly equal to 32 for all values of λ .

the exact asymptotic value in (49) and the new lower bound in (6) is equal to $\frac{10.539}{\sqrt{2\pi e}} \approx 2.55$. It therefore follows that, in the limit where $\lambda \rightarrow 0$, 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.

Acknowledgment: The anonymous reviewer is acknowledged for suggestions that led to improvement of the presentation of this paper. This research work was supported by the Israeli Science Foundation (ISF), grant number 12/12.

REFERENCES

- Arratia, R., Goldstein, L., Gordon, L., 1990. Poisson approximation and the Chen-Stein method. *Statistical Science* 5, 403–424.
- Barbour, A.D., Chen, L. H. Y., 2005. *An Introduction to Stein’s Method*. Lecture Notes Series, Institute for Mathematical Sciences, Singapore University Press and World Scientific.
- Barbour, A.D., Hall, P., 1984. On the rate of Poisson convergence. *Mathematical Proceedings of the Cambridge Philosophical Society* 95, 473–480.
- Barbour, A.D., Holst, L., Janson, S., 1992. *Poisson Approximation*. Oxford University Press.
- Barbour, A. D., Johnson, O., Kontoyiannis, I., Madiman, M., 2010. Compound Poisson approximation via information functionals. *Electronic Journal of Probability* 15, 1344–1369.
- Chen, L.H.Y., 1975. Poisson approximation for dependent trials. *Annals of Probability* 3, 534–545.
- Deheuvels, P., Pfeifer, D., 1986. A semigroup approach to Poisson approximation. *Annals of Probability* 14, 663–676.
- Kontoyiannis, I., Harremoës, P., Johnson, O., 2005. Entropy and the law of small numbers. *IEEE Trans. on Information Theory* 51, 466–472.
- Le Cam, L., 1960. An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics* 10, 1181–1197.
- Ross, S.M., Peköz, E.A. 2007. *A Second Course in Probability*. Probability Bookstore.
- Ross, N., 2011. *Fundamentals of Stein’s Method*. *Probability Surveys* 8, 210–293.
- Sason, I., 2012. Entropy bounds for discrete random variables via coupling. [Online]. Available: <http://arxiv.org/abs/1209.5259>.

APPENDIX

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 $i \in \{1, \dots, n\}$, and $Z \sim \text{Po}(\lambda)$ with mean $\lambda \triangleq \sum_{i=1}^n p_i$. From the basic equation of the Chen-Stein method, equation (1) holds for an arbitrary bounded function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$. Furthermore, from the proof of Theorem 2 of Barbour and Hall (1984),

$$\begin{aligned}
& \mathbb{E}[\lambda f(W+1) - Wf(W)] \\
&= \sum_{j=1}^n p_j \mathbb{E}[f(W+1)] - \sum_{j=1}^n \mathbb{E}[X_j f(W)] \\
&= \sum_{j=1}^n p_j \mathbb{E}[f(W+1)] - \sum_{j=1}^n p_j \mathbb{E}[f(W) | X_j = 1] \\
&= \sum_{j=1}^n p_j \mathbb{E}[f(W+1)] - \sum_{j=1}^n p_j \mathbb{E}[f(V_j+1) | X_j = 1] \\
&\stackrel{(a)}{=} \sum_{j=1}^n p_j \mathbb{E}[f(W+1) - f(V_j+1)] \\
&= \sum_{j=1}^n p_j^2 \mathbb{E}[f(W+1) - f(V_j+1) | X_j = 1] \\
&= \sum_{j=1}^n p_j^2 \mathbb{E}[f(V_j+2) - f(V_j+1) | X_j = 1] \\
&\stackrel{(b)}{=} \sum_{j=1}^n p_j^2 \mathbb{E}[f(V_j+2) - f(V_j+1)] \tag{50}
\end{aligned}$$

where equalities (a) and (b) hold since X_j and V_j are independent random variables for every $j \in \{1, \dots, n\}$. By subtracting (1) from (50), it follows that for an arbitrary bounded function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$

$$\mathbb{E}[\lambda f(W+1) - Wf(W)] - \mathbb{E}[\lambda f(Z+1) - Zf(Z)] = \sum_{j=1}^n p_j^2 \mathbb{E}[f(V_j+2) - f(V_j+1)]. \tag{51}$$

In the following, an upper bound on the left-hand side of (51) is derived, based on total variation distance between the two distributions of W and Z .

$$\begin{aligned}
& \mathbb{E}[\lambda f(W+1) - Wf(W)] - \mathbb{E}[\lambda f(Z+1) - Zf(Z)] \\
&= \sum_{k=0}^{\infty} (\lambda f(k+1) - kf(k)) (\mathbb{P}(W=k) - \mathbb{P}(Z=k)) \\
&\leq \sum_{k=0}^{\infty} |\lambda f(k+1) - kf(k)| |\mathbb{P}(W=k) - \mathbb{P}(Z=k)| \tag{52}
\end{aligned}$$

$$\begin{aligned}
& \leq \sup_{k \in \mathbb{N}_0} |\lambda f(k+1) - kf(k)| \sum_{k=0}^{\infty} |\mathbb{P}(W=k) - \mathbb{P}(Z=k)| \\
&= 2d_{\text{TV}}(P_W, \text{Po}(\lambda)) \sup_{k \in \mathbb{N}_0} |\lambda f(k+1) - kf(k)| \tag{53}
\end{aligned}$$

where the last equality follows from (4). Hence, the combination of (51) and (53) gives the following lower bound on the total variation distance:

$$d_{\text{TV}}(P_W, \text{Po}(\lambda)) \geq \frac{\sum_{j=1}^n \left\{ p_j^2 \mathbb{E}[f(V_j+2) - f(V_j+1)] \right\}}{2 \sup_{k \in \mathbb{N}_0} |\lambda f(k+1) - kf(k)|}$$

which holds, in general, for an arbitrary bounded function $f : \mathbb{N}_0 \rightarrow \mathbb{R}$. This proves inequality (19).