Non-Uniform Bounds in the Poisson Approximation With Applications to Informational Distances I

S. G. Bobkov, G. P. Chistyakov, and F. Götze

Abstract— **We explore asymptotically optimal bounds for deviations of Bernoulli convolutions from the Poisson limit in terms of the Shannon relative entropy and the Pearson χ^2-distance. The results are based on proper non-uniform estimates for densities. This part deals with the so-called non-degenerate case.**

Index Terms— **Poisson approximation, Rényi, χ^2-divergence.**

I. Introduction

Let $X_1, \dots, X_n$ be independent Bernoulli random variables taking the two values, 1 (interpreted as a success) and 0 (as a failure) with respective probabilities p_j and $q_j = 1 - p_j$. The total number of successes

$$W = X_1 + \cdots + X_n$$

takes values $k = 0, 1, \dots, n$ with probabilities

$$\mathbb{P}\{W = k\} = \sum p_1^{\varepsilon_1} q_1^{1-\varepsilon_1} \dots p_n^{\varepsilon_n} q_n^{1-\varepsilon_n}, \tag{I.1}$$

where the summation runs over all 0-1 sequences $\varepsilon_1, \dots, \varepsilon_n$ such that $\varepsilon_1 + \cdots + \varepsilon_n = k$. Although this expression is difficult to determine in case of arbitrary p_j and large n, it can be well approximated by the Poisson probabilities under quite general assumptions. Putting

$$\lambda = p_1 + \cdots + p_n,$$

let Z be a Poisson random variable with parameter $\lambda > 0$ (for short, $Z \sim P_\lambda$), i.e.,

$$v_k = \mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, \dots$$

It is well-known for a long time that, if $\max_j p_j$ is small, the distribution P_λ approximates the distribution P_W of W, which may be quantified by means of the total variation distance

$$d(W, Z) = \|P_W - P_\lambda\|_{\rm TV}$$
$$= 2 \sup_{A \subset \mathbb{Z}} |\mathbb{P}\{W \in A\} - \mathbb{P}\{Z \in A\}| = \sum_{k=0}^{\infty} |w_k - v_k|,$$

where $w_k = \mathbb{P}\{W = k\}$. In particular, based on Stein-Chen's method, there is the following two-sided bound due to Barbour and Hall involving the functional

$$\lambda_2 = p_1^2 + \cdots + p_n^2.$$

Theorem I.1 [1]. *One has*

$$\frac{1}{32} \min(1, 1/\lambda)\, \lambda_2 \le \frac{1}{2} d(W, Z) \le \frac{1 - e^{-\lambda}}{\lambda} \lambda_2. \tag{I.2}$$

Here, the parameter λ_2, or more precisely – the ratio λ_2/λ (for λ bounded away from zero), plays a similar role as the Lyapunov ratio L_3 in the central limit theorem.

In the i.i.d. case with $p_j = \lambda/n$ and fixed $\lambda > 0$, both sides of (I.2) are of the same order $1/n$. In the case $\lambda \le 1$, the upper bound in (I.2) is sharp also in the sense that the second inequality becomes an equality when $p_1 = \lambda$, $p_j = 0$ $(2 \le j \le n)$.

Theorem I.1 refined many previous results in this direction, starting from bounds for the i.i.d. case by Prokhorov [17] and bounds for the general case by Le Cam [14]. In particular, Le Cam obtained the upper bound

$$d(W, Z) \le 2\lambda_2. \tag{I.3}$$

For large λ Kerstan *et al.* [12], respectively Chen [4] improved these bounds to

$$d(W, Z) \le \frac{2.1}{\lambda} \lambda_2 \quad \text{if} \quad \max_{j \le n} p_j \le \frac{1}{4},$$

respectively

$$d(W, Z) \le \frac{10}{\lambda} \lambda_2.$$

See also [10], [23], [21], [18], [19], [2] and the references therein. A certain refinement of the lower bound in (I.2) was obtained by Sason [20].

While (I.2) provides a sharp estimate for the total variation distance, one may wonder whether or not similar approximation bounds hold for the stronger informational distances. As a first interesting example, one may consider the relative entropy

$$D(W||Z) = \sum_{k=0}^{\infty} w_k \log \frac{w_k}{v_k},$$

often called the Kullback-Leibler distance, or an informational divergence of P_W from P_λ. It dominates the total variation

Manuscript received August 21, 2018; revised April 3, 2019; accepted April 8, 2019. Date of publication April 25, 2019; date of current version August 16, 2019. This work was supported in part by NSF under Grant DMS-1855575, in part by the Russian Academic Excellence Project 5-100 through the HSE University Basic Research Program, and in part by Grant SFB 1283.

S. G. Bobkov is with the School of Mathematics, University of Minnesota, Minneapolis, MN 55455 USA, and also with HSE University, Moscow, Russia (e-mail: bobkov@math.umn.edu).

G. P. Chistyakov and F. Götze are with the Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany (e-mail: chistyak@math.uni-bielefeld.de; goetze@math.uni-bielefeld.de).

Communicated by I. Kontoyiannis, Associate Editor At Large.

Digital Object Identifier 10.1109/TIT.2019.2913313

distance in view of the Pinsker inequality $2\ D(W||Z) \geq d(W,Z)^2$. In this context, lower and upper bounds for the relative entropy were studied by Harremoës [6], [7], and Harremoës and Ruzankin [9]. In particular, in the i.i.d. case $p_j = p$, it was shown in [9] that

$$\frac{-\log(1-p)-p}{2} - \frac{14\ p^2}{n\,(1-p)^3} \leq D(W||Z) \leq \frac{-\log(1-p)-p}{2} - \frac{(1+p)\,p^2}{4n\,(1-p)^3}.$$

If $p = \lambda/n$ with a fixed (or just bounded) value of λ, these estimates provide the following rate for the Poisson approximation

$$D(W||Z) = \frac{\lambda^2}{4n^2} + O(1/n^3) \quad \text{as} \quad n \to \infty. \tag{I.4}$$

The general non-i.i.d. scenario (with not necessarily equal probabilities p_j) has been partially studied as well. A simple upper estimate $D(W||Z) \leq \lambda_2$, analogous to Le Cam's bound (I.3), may be found in [6], cf. also Johnson [11]. It is however not so sharp as (I.4). A tighter upper bound

$$D(W||Z) \leq \frac{1}{\lambda} \sum_{j=1}^{n} \frac{p_j^3}{1-p_j} \tag{I.5}$$

was later derived by Kontoyiannis, Harremoës and Johnson [13]. If $p_j = \lambda/n$ with $\lambda \leq n/2$, it yields $D(W||Z) \leq 2\lambda^2/n^2$ reflecting a correct decay with respect to n up to a constant, according to (I.4). Nevertheless, in the general case, Pinsker's inequality and the bounds (I.2)-(I.3) suggest that a further sharpening such as

$$D(W||Z) \leq A_\lambda \lambda_2^2 \tag{I.6}$$

might be possible by involving λ_2 rather than the functional $\lambda_3 = p_1^3 + \cdots + p_n^3$. To compare the two quantities, note that $\lambda_2^2 \leq \lambda\lambda_3$ (by Cauchy's inequality). Hence, the inequality (I.6) would be sharper compared to (I.5) modulo a λ-dependent factor. An upper bound such as (I.6) may also be inspired by the lower bound

$$D(W||Z) \geq \frac{1}{4}\left(\frac{\lambda_2}{\lambda}\right)^2 \tag{I.7}$$

recently derived by Harremoës, Johnson and Kontoyiannis [8]. It is consistent with (I.4) and also shows that the constant $\frac{1}{4}$ is best possible.

As it turns out, (I.6) does hold in the so-called non-degenerate situation, and in essence, (I.7) may be reversed (we say that the range of (λ, λ_2) is non-degenerate, if $\lambda_2 \leq \kappa\lambda$ with $\kappa \in (0,1)$, or if $\lambda \leq \lambda_0$, implicitly meaning that the resulting inequalities may contain κ or λ_0 as fixed parameters).

In fact, one can further sharpen (I.6) by replacing the relative entropy with the Pearson χ^2-distance, as well as with other Rényi/Tsallis distances. To avoid technical complications, let us restrict ourselves to the χ^2-divergence which is given by

$$\chi^2(W,Z) = \sum_{k=0}^{\infty} \frac{(w_k - v_k)^2}{v_k}.$$

It is a divergence type quantity which dominates the relative entropy: $\chi^2(W,Z) \geq D(W||Z)$. For a general theory of informational distances, we refer interested readers to the recent review by van Erven and Harremoës [5]; an additional material may be found in the books [15], [16], [22], [11].

Here, we reverse the bound (I.7) and prove:

Theorem I.2. *If $\lambda_2 \leq \lambda/2$, then with some absolute constant c*

$$D(W||Z) \leq \chi^2(W,Z) \leq c\left(\frac{\lambda_2}{\lambda}\right)^2. \tag{I.8}$$

The condition $\lambda_2 \leq \lambda/2$ is readily fulfilled as long as $\max_j p_j \leq \frac{1}{2}$ (note that, if $\lambda \leq \frac{1}{2}$, then necessarily $p_j \leq \frac{1}{2}$, so, $\lambda_2 \leq \lambda/2$). Similar bounds as in (I.8) remain to hold under the weaker assumption $\lambda_2 \leq \kappa\lambda$ with a constant $c = c_\kappa$ depending on $\kappa \in (0,1)$, cf. Proposition VI.2 below. This assumption may actually be replaced with the requirement that λ is bounded. More precisely, in the second part of the paper it will be shown that without any restriction, up to some universal factors, we have

$$D(W||Z) \sim \left(\frac{\lambda_2}{\lambda}\right)^2 (1 + \log F),$$
$$\chi^2(W,Z) \sim \left(\frac{\lambda_2}{\lambda}\right)^2 \sqrt{F},$$

where

$$F = \frac{\max(1,\lambda)}{\max(1, \lambda - \lambda_2)}.$$

This shows that in general the bound (I.7) cannot be reversed.

For the study of the asymptotic behavior of D and χ^2 in terms of λ and λ_2, we derive new bounds for the difference between densities of W and Z, that is, for

$$\Delta_k = w_k - v_k = \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}.$$

To this aim, one has to consider different zones of λ's, distinguishing between "small" and "large" values. The case $\lambda \leq \frac{1}{2}$ can be handled directly leading to the non-uniform density bound

$$|\Delta_k| \leq 2\lambda_2\, \mathbb{P}\{k-2 \leq Z \leq k\}.$$

It easily yields sharp upper bounds for all distances as in Theorems I.1-I.2 in the case of small λ, at least up to numerical factors (cf. Propositions III.3 and III.4).

To treat larger values of λ, a more sophisticated analysis in the complex plane is involved – using closeness of generating functions associated with the sequences w_k and v_k. In particular, the following statement may be of independent interest.

Theorem I.3. *For all integers $k \geq 0$,*

$$|\Delta_0| \leq 3\lambda_2\, e^{-\lambda}, \qquad |\Delta_k| \leq 3\lambda_2 \quad (k \geq 1). \tag{I.9}$$

Moreover, putting $\rho = (\lambda - \lambda_2)\, \min\{\frac{k}{\lambda}, \frac{\lambda}{k}\}$, $k \geq 1$, we have

$$|\Delta_k| \leq 7\sqrt{k}\left(\frac{k-\lambda}{\lambda}\right)^2 \lambda_2 \min\left\{1, \rho^{-\frac{1}{2}}\right\} \mathbb{P}\{Z = k\} + 21\, k^{\frac{3}{2}}\, \frac{\lambda_2}{\lambda} \min\left\{1, \rho^{-\frac{3}{2}}\right\} \mathbb{P}\{Z = k\}. \tag{I.10}$$

Let us clarify the meaning of the last bound, assuming that $\lambda_2 \leq \kappa\lambda$ with some constant $\kappa \in (0,1)$. If $k \leq 2\lambda$ and $\lambda \geq \frac{1}{2}$, then with some $c = c_\kappa > 0$, it gives

$$|\Delta_k| \,\leq\, c\,\Big(\frac{(k-\lambda)^2}{\lambda} + 1\Big)\,\frac{\lambda_2}{\lambda}\,\mathbb{P}\{Z = k\},$$

while for $k \geq \lambda \geq \frac{1}{2}$, we also have

$$|\Delta_k| \,\leq\, c\,\Big(\frac{k}{\lambda}\Big)^3\,\lambda_2\,\mathbb{P}\{Z = k\}.$$

Since $|k - \lambda|$ is of order at most $\sqrt{\lambda}$ on a sufficiently large part of $\mathbb{Z}$ measured by P_λ, these non-uniform bounds explain the possibility of upper bounds in Theorem I.2.

The paper is organized as follows. First we describe several general bounds on the probability function of the Poisson law (Section II). In Section III, we consider the deviations Δ_k and prove Theorem I.2 in case $\lambda \leq \frac{1}{2}$. Sections IV-V are devoted to non-uniform bounds and the proof of Theorem I.3, which is used to complete the proof of Theorem I.2 for $\lambda \geq \frac{1}{2}$. Uniform bounds for large λ are discussed in Section VII. There we shall demonstrate that in a typical situation, when the ratio λ_2/λ is small, the Poisson approximation considerably improves the rate of normal approximation described by the Berry-Esseen bound in the central limit theorem.

II. Gaussian Type Bounds on Poisson Probabilities

When bounding the Poisson probabilities

$$v_k = f(k) = \mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!}\,e^{-\lambda}, \qquad k = 0, 1, \dots,$$

with a fixed parameter $\lambda > 0$, it is convenient to use the well-known Stirling-type two-sided bound:

$$\sqrt{2\pi}\,k^{k+\frac{1}{2}}\,e^{-k} \leq k! \leq e\,k^{k+\frac{1}{2}}\,e^{-k} \qquad (k \geq 1). \tag{II.1}$$

In particular, it implies the following Gaussian type estimates.

Lemma II.1. *For all $k \geq 1$,*

$$f(k) \,\leq\, \frac{1}{\sqrt{2\pi k}}. \tag{II.2}$$

Moreover, if $1 \leq k \leq 2\lambda$, then

$$\frac{1}{e\sqrt{k}}\,e^{-\frac{(k-\lambda)^2}{\lambda}} \,\leq\, f(k) \,\leq\, \frac{1}{\sqrt{2\pi k}}\,e^{-\frac{(k-\lambda)^2}{3\lambda}}. \tag{II.3}$$

Here, the lower bound may be improved in the region $k \geq \lambda$ as

$$f(k) \,\geq\, \frac{1}{e\sqrt{k}}\,e^{-\frac{(k-\lambda)^2}{2\lambda}}. \tag{II.4}$$

Proof. Applying the lower estimate in (II.1), we get

$$\begin{aligned} f(k) &\leq \frac{1}{\sqrt{2\pi k}}\,e^{k-\lambda}\,\Big(\frac{\lambda}{k}\Big)^k \\ &= \frac{1}{\sqrt{2\pi k}}\,e^{k-\lambda}\;e^{-k\log(1+\frac{k-\lambda}{\lambda})} \\ &= \frac{1}{\sqrt{2\pi k}}\,e^{\lambda h(\theta)}, \qquad \theta = \frac{k-\lambda}{\lambda}, \end{aligned} \tag{II.5}$$

where

$$h(\theta) = \theta - (1+\theta)\log(1+\theta).$$

The function $h(\theta)$ is concave on the half-axis $\theta \geq -1$, with $h(0) = h'(0) = 0$. Hence, $h(\theta) \leq 0$ for all θ, thus proving the first assertion (II.2).

Assuming that $1 \leq k \leq 2\lambda$ (with $\lambda \geq \frac{1}{2}$), we necessarily have $|\theta| \leq 1$. In this interval, consider the function $T_c(\theta) = h(\theta) + c\theta^2$ with parameter $\frac{1}{4} < c \leq 1$. The second derivative

$$T_c''(\theta) = -\frac{1}{1+\theta} + 2c \qquad (-1 < \theta \leq 1)$$

is vanishing at the point $\theta_0 = \frac{1}{2c} - 1$, while $T_c''(-1) = -\infty$. This means that T_c is concave on $[-1,\theta_0]$ and convex on $[\theta_0, 1]$. Since also $T_c(0) = T_c'(0) = 0$, we have $T_c(\theta) \leq 0$ for all $\theta \in [-1,1]$, if and only if this inequality is fulfilled at $\theta = 1$. But $T_c(1) = 1 - 2\log 2 + c$, so the optimal value is $c = 2\log 2 - 1 = 0.387... > \frac{1}{3}$. Hence, $h(\theta) \leq -\frac{1}{3}\,\theta^2$, and we arrive at the upper bound in (II.3).

Similarly, applying the upper estimate in (II.1), we get

$$f(k) \,\geq\, \frac{1}{e\sqrt{k}}\,e^{k-\lambda}\,\Big(\frac{\lambda}{k}\Big)^k \,=\, \frac{1}{e\sqrt{k}}\,e^{\lambda h(\theta)},$$

where $\theta = \frac{k-\lambda}{\lambda}$. Choosing $c = 1$, consider the function $T(\theta) = h(\theta) + \theta^2$ in the interval $|\theta| \leq 1$. Since $T''(-\frac{1}{2}) = 0$, it is concave on $[-1, -\frac{1}{2}]$ and is convex on $[-\frac{1}{2}, 1]$. Since $T(0) = T'(0) = 0$ and $T(-1) = 0$, this means that $\theta = 0$ is the point of minimum of T. Therefore, $T(\theta) \geq 0$, that is, $h(\theta) \geq -\theta^2$ for all $\theta \in [-1,1]$, giving the lower bound in (II.3).

Finally, to get the refinement (II.4) in the region $k \geq \lambda$, consider the function $T(\theta) = h(\theta) + \frac{1}{2}\,\theta^2$ for $\theta \geq 0$. Since $T(0) = 0$ and $T'(\theta) = \theta - \log(1+\theta) \geq 0$, this function is increasing. Therefore, $T(\theta) \geq 0$, that is, $h(\theta) \geq -\frac{1}{2}\,\theta^2$ for all $\theta \geq 0$. $\square$

III. Elementary Upper Bounds

We keep the same notations as before; in particular,

$$\mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!}\,e^{-\lambda}, \qquad k = 0, 1, \dots,$$

while

$$\mathbb{P}\{W = k\} \,=\, \sum p_1^{\varepsilon_1}\,q_1^{1-\varepsilon_1} \dots p_n^{\varepsilon_n}\,q_n^{1-\varepsilon_n}$$

with summation over all 0-1 sequences $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ such that $\varepsilon_1 + \dots + \varepsilon_n = k$. Clearly, $\mathbb{P}\{W = k\} = 0$ for $k > n$. To eliminate this condition, one may always assume that n is arbitrary, by extending the sequence $(X_1, \dots, X_n)$ to $(X_1, \dots, X_k)$ in case $n < k$ with $p_{n+1} = \dots = p_k = 0$. This does not change the value of W.

First, let us consider the probability that W equals $k = 0$.

Lemma III.1. *If $\max_j p_j \leq \frac{1}{2}$, then*

$$0 \,\leq\, \mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} \,\leq\, \lambda_2\,e^{-\lambda}.$$

Proof. Expanding the function $p \to -\log(1-p)$ near zero according to the Taylor formula, write

$$\mathbb{P}\{W=0\} = \prod_{j=1}^{n} q_j = e^{-\lambda - S}, \tag{III.1}$$

where

$$S = \sum_{s=2}^{\infty} \frac{1}{s}\,\lambda_s, \quad \lambda_s = p_1^s + \cdots + p_n^s.$$

Since $\lambda_s \leq (\max_j p_j)^{s-2}\,\lambda_2 \leq 2^{-(s-2)}\,\lambda_2$ for $s \geq 2$, we have

$$S \ \leq\ \lambda_2 \sum_{s=2}^{\infty} \frac{2^{-(s-2)}}{s} \ =\ (4\log 2 - 2)\,\lambda_2 \ <\ \lambda_2. \tag{III.2}$$

Hence

$$\mathbb{P}\{Z=0\} - \mathbb{P}\{W=0\} = e^{-\lambda}\,(1-e^{-S}) \leq e^{-\lambda}\,S.$$

□

Note that the condition of Lemma III.1 is fulfilled, if $\lambda \leq \frac{1}{2}$. In that case, the upper bound of the lemma may be reversed up to a numerical factor, for example, in the form

$$\mathbb{P}\{Z=0\} - \mathbb{P}\{W=0\} \ \geq\ 0.47\,\lambda_2\,e^{-\lambda}.$$

Moreover, one can show that

$$\mathbb{P}\{W=1\} - \mathbb{P}\{Z=1\} \ \geq\ 0.42\,\lambda_2\,e^{-\lambda},$$

and if $\lambda \leq 1/8$, then also

$$\mathbb{P}\{Z=2\} - \mathbb{P}\{W=2\} \ \geq\ \frac{17}{49}\,\lambda_2\,e^{-\lambda}.$$

The value $k=2$ turns out to be most essential for obtaining lower bounds, since it immediately yields $d(W,Z) \geq c\lambda_2$ and $D(W||Z) \geq c\,(\frac{\lambda_2}{\lambda})^2$ with some absolute constant $c>0$.

Returning to upper bounds, recall the notation

$$\Delta_k = \mathbb{P}\{W=k\} - \mathbb{P}\{Z=k\}.$$

In order to involve the values $k \geq 1$, we need the following:

Lemma III.2. *If* $\max_j p_j \leq \frac{1}{2}$, *then*

$$|\Delta_1| \ \leq\ \lambda_2\,(\lambda + e - 1)\,e^{-\lambda}. \tag{III.3}$$

Moreover, for any $k \geq 2$,

$$|\Delta_k| \ \leq\ \lambda_2\,\Big(\frac{\lambda^k}{k!} + \frac{e^{\lambda}-1}{\lambda}\,\frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-2}}{(k-2)!}\Big)\,e^{-\lambda}.$$

Proof. Denote by I the collection of all tuples $\varepsilon = (\varepsilon_1,\dots,\varepsilon_n)$ with integer components $\varepsilon_i \geq 0$ such $\varepsilon_1 + \cdots + \varepsilon_n = k$, and let $J = \{\varepsilon \in I : \max_i \varepsilon_i \leq 1\}$. Representing the Poisson random variable $Z \sim P_\lambda$ as $Z = Z_1 + \cdots + Z_n$ with independent summands $Z_i \sim P_{p_i}$, we have that, for any $k = 0,1,\dots$,

$$\mathbb{P}\{Z=k\} \ =\ e^{-\lambda} \sum_{\varepsilon \in I} \frac{p_1^{\varepsilon_1} \dots p_n^{\varepsilon_n}}{\varepsilon_1! \dots \varepsilon_n!}.$$

Hence, we may start with the formula

$$\Delta_k \ =\ e^{-\lambda} \sum_{\varepsilon \in I} \frac{1}{\varepsilon_1! \dots \varepsilon_n!}\ U_\varepsilon \ - \sum_{\varepsilon \in J} U_\varepsilon V_\varepsilon, \tag{III.4}$$

where

$$U_\varepsilon = p_1^{\varepsilon_1} \dots p_n^{\varepsilon_n}, \quad V_\varepsilon = q_1^{1-\varepsilon_1} \dots q_n^{1-\varepsilon_n}.$$

For a 0-1 sequence $\varepsilon = (\varepsilon_1,\dots,\varepsilon_n) \in J$, put

$$L_\varepsilon = \varepsilon_1\ p_1 + \cdots + \varepsilon_n p_n.$$

By the Taylor formula,

$$V_\varepsilon^{-1} \ =\ e^{S_\varepsilon}, \qquad S_\varepsilon = \sum_{s=1}^{\infty} \frac{1}{s} \sum_{j=1}^{n} (1-\varepsilon_j)\,p_j^s.$$

Similarly to (III.1)-(III.2), we have

$$\begin{aligned} S_\varepsilon &= \lambda - L_\varepsilon + \sum_{s=2}^{\infty} \frac{1}{s} \sum_{j=1}^{n} (1-\varepsilon_j)\,p_j^s \\ &= \lambda - L_\varepsilon + \theta\lambda_2 \end{aligned}$$

with some $0 \leq \theta \leq 1$. Therefore,

$$e^{\lambda}\,V_\varepsilon = e^{L_\varepsilon - \theta\lambda_2} \geq 1 + (L_\varepsilon - \theta\lambda_2) \geq 1 + L_\varepsilon - \lambda_2.$$

Moreover, since $L_\varepsilon \leq \min(\lambda,k)$, we have

$$\frac{e^{L_\varepsilon} - 1}{L_\varepsilon} \leq \frac{e^{\min(\lambda,k)} - 1}{\min(\lambda,k)} \equiv c_{k,\lambda},$$

which in turn implies

$$e^{\lambda}\,V_\varepsilon \leq e^{L_\varepsilon} \leq 1 + c_{k,\lambda} L_\varepsilon.$$

The two bounds give $L_\varepsilon - \lambda_2 \leq e^{\lambda}\,V_\varepsilon - 1 \leq c_{k,\lambda} L_\varepsilon$, so that

$$\big|U_\varepsilon - e^{\lambda}\,U_\varepsilon V_\varepsilon\big| \ \leq\ \lambda_2\,U_\varepsilon + c_{k,\lambda}\,U_\varepsilon L_\varepsilon.$$

Next, applying the multinomial formula, we have

$$\sum_{\varepsilon \in J} U_\varepsilon \ \leq\ \sum_{\varepsilon \in I} \frac{p_1^{\varepsilon_1} \dots p_n^{\varepsilon_n}}{\varepsilon_1! \dots \varepsilon_n!} \ =\ \frac{\lambda^k}{k!}$$

and

$$\begin{aligned} \sum_{\varepsilon \in J} U_\varepsilon L_\varepsilon &= \sum_{i=1}^{n} \sum_{\varepsilon \in J} \varepsilon_i\,p_1^{\varepsilon_1} \dots p_{i-1}^{\varepsilon_{i-1}}\,p_i^{\varepsilon_i+1}\,p_{i+1}^{\varepsilon_{i+1}} \dots p_n^{\varepsilon_n} \\ &= \sum_{i=1}^{n} p_i^2 \sum_{\varepsilon \in J,\,\varepsilon_i=1} p_1^{\varepsilon_1} \dots p_{i-1}^{\varepsilon_{i-1}}\,p_{i+1}^{\varepsilon_{i+1}} \dots p_n^{\varepsilon_n} \\ &\leq \sum_{i=1}^{n} p_i^2\,\frac{(\lambda - p_i)^{k-1}}{(k-1)!} \ \leq\ \lambda_2\,\frac{\lambda^{k-1}}{(k-1)!}. \end{aligned}$$

Thus,

$$\sum_{\varepsilon \in J} |U_\varepsilon - e^{\lambda}\,U_\varepsilon V_\varepsilon| \ \leq\ \lambda_2\,\Big(\frac{\lambda^k}{k!} + c_{k,\lambda}\,\frac{\lambda^{k-1}}{(k-1)!}\Big). \tag{III.5}$$

The remaining terms participating in $\mathbb{P}(Z=k)$ correspond to the tuples $\varepsilon \in I$ with $\max_i \varepsilon_i \geq 2$, which is only possible for $k \geq 2$. In that case, restricting for definiteness to the constraint $\varepsilon_n \geq 2$, we have

$$\begin{aligned}\sum_{\varepsilon\in I,\,\varepsilon_n\geq 2} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!} &= \sum_{m=2}^{k} \frac{p_n^m}{m!} \sum_{\varepsilon_1+\dots+\varepsilon_{n-1}=k-m} \frac{p_1^{\varepsilon_1}\dots p_{n-1}^{\varepsilon_{n-1}}}{\varepsilon_1!\dots\varepsilon_{n-1}!} \\ &= \sum_{m=2}^{k} \frac{p_n^m}{m!}\,\frac{(\lambda-p_n)^{k-m}}{(k-m)!} \\ &\leq p_n^2 \sum_{m=2}^{k} \frac{p_n^{m-2}}{(m-2)!}\,\frac{(\lambda-p_n)^{k-m}}{(k-m)!} \\ &= p_n^2\,\frac{\lambda^{k-2}}{(k-2)!}.\end{aligned}$$

Similarly, for any $i = 1,\dots,n$,

$$\sum_{\varepsilon\in I,\,\varepsilon_i\geq 2} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!} \leq p_i^2\,\frac{\lambda^{k-2}}{(k-2)!},$$

and summing over $i \leq n$, we get

$$\sum_{\varepsilon\in I,\,\max\varepsilon_i\geq 2} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!} \leq \lambda_2\,\frac{\lambda^{k-2}}{(k-2)!}.$$

It remains to combine this bound with the bound (III.5) and apply both in (III.4). Then we finally obtain that

$$|\Delta_k| \leq \lambda_2\Big(\frac{\lambda^k}{k!}+c_{k,\lambda}\,\frac{\lambda^{k-1}}{(k-1)!}+1_{\{k\geq 2\}}\frac{\lambda^{k-2}}{(k-2)!}\Big)\,e^{-\lambda}. \quad \text{(III.6)}$$

If $k = 1$, then $c_{1,\lambda} \leq e-1$, and we arrive at the first inequality (III.3). In the case $k \geq 2$, one may use $c_{k,\lambda} \leq \frac{e^\lambda-1}{\lambda}$, and then we arrive at the second inequality of the lemma. □

Note that when $\lambda \leq \frac{1}{2}$, we also have $c_{1,\lambda} \leq 2\,(\sqrt{e}-1)$, and then (III.3) may be replaced with a slightly better bound

$$|\Delta_1| \leq \lambda_2\,(\lambda + 2\,(\sqrt{e}-1))\,e^{-\lambda}. \quad \text{(III.7)}$$

Combining Lemmas III.1–III.2 (cf. (III.6)), we thus obtain the following non-uniform bound on the deviations of Δ_k.

Proposition III.3. *If* $\max_j p_j \leq \frac{1}{2}$, *then, for all* $k \geq 0$,

$$|\Delta_k| \leq \frac{e^\lambda-1}{\lambda}\,\lambda_2\,\mathbb{P}\{k-2\leq Z\leq k\}.$$

The estimates obtained so far are sufficient to establish Theorem I.2 in the case $\lambda \leq \frac{1}{2}$. In fact, one may weaken the latter condition to $\max_j p_j \leq \frac{1}{2}$, as shown in the next statement. To compare the lower and upper bounds, we add the lower bound (I.7) of Harremoës, Johnson and Kontoyiannis.

Proposition III.4. *If* $\max_j p_j \leq \frac{1}{2}$, *then*

$$\frac{1}{4}\Big(\frac{\lambda_2}{\lambda}\Big)^2 \leq D(W||Z) \leq \chi^2(W,Z) \leq C_\lambda\Big(\frac{\lambda_2}{\lambda}\Big)^2,$$

where C_λ *depends on* $\lambda \geq 0$ *as an increasing continuous function with* $C_0 = 2$. *In particular, if* $\lambda \leq \frac{1}{2}$, *then*

$$\chi^2(W,Z) \leq 15\Big(\frac{\lambda_2}{\lambda}\Big)^2.$$

Proof. Applying Lemmas III.1-III.2, we get

$$\lambda_2^{-2}\,e^\lambda\,\chi^2(W,Z) \leq 1 + \frac{1}{\lambda}\,(\lambda+e-1)^2 + \sum_{k=2}^\infty \frac{k!}{\lambda^k}\Big(\frac{\lambda^k}{k!}+c_\lambda\,\frac{\lambda^{k-1}}{(k-1)!}+\frac{\lambda^{k-2}}{(k-2)!}\Big)^2,$$

where $c_\lambda = \frac{e^\lambda-1}{\lambda}$. Expanding the squares of the brackets in this sum results in

$$\begin{aligned}&\sum_{k=2}^\infty \frac{k!}{\lambda^k}\Big(\frac{\lambda^{2k}}{k!^2}+\frac{2c_\lambda\,\lambda^{2k-1}}{k!\,(k-1)!}+\frac{c_\lambda^2\,\lambda^{2k-2}}{(k-1)!^2}+\frac{2\lambda^{2k-2}}{k!\,(k-2)!} \\ &\quad+\frac{2c_\lambda\,\lambda^{2k-3}}{(k-1)!\,(k-2)!}+\frac{\lambda^{2k-4}}{(k-2)!^2}\Big) \\ &= \sum_{k=2}^\infty \frac{\lambda^k}{k!} + 2c_\lambda\sum_{k=2}^\infty \frac{\lambda^{k-1}}{(k-1)!} \\ &\quad+ c_\lambda^2\sum_{k=2}^\infty k\,\frac{\lambda^{k-2}}{(k-1)!} + 2\sum_{k=2}^\infty \frac{\lambda^{k-2}}{(k-2)!} \\ &\quad+ 2c_\lambda\sum_{k=2}^\infty k\,\frac{\lambda^{k-3}}{(k-2)!} + \sum_{k=2}^\infty k(k-1)\,\frac{\lambda^{k-4}}{(k-2)!},\end{aligned}$$

which is the same as

$$\begin{aligned}&3e^\lambda - 1 - \lambda + 2c_\lambda\,(e^\lambda-1) + c_\lambda^2\sum_{k=1}^\infty (k+1)\,\frac{\lambda^{k-1}}{k!} \\ &\quad+ 2c_\lambda\sum_{k=0}^\infty (k+2)\,\frac{\lambda^{k-1}}{k!} + \sum_{k=0}^\infty (k+1)(k+2)\,\frac{\lambda^{k-2}}{k!} \\ &= 3e^\lambda - 1 - \lambda + 2c_\lambda\,(e^\lambda-1) \\ &\quad+ 2c_\lambda e^\lambda\,\frac{2+\lambda}{\lambda} + \frac{2+4\lambda+\lambda^2}{\lambda^2}\,e^\lambda.\end{aligned}$$

Multiplying by λ^2, this gives the desired inequality

$$\lambda^2\lambda_2^{-2}\,\chi^2(W,Z) \leq C_\lambda = \lambda^2 + \lambda\,(\lambda+e-1)^2 + B_\lambda$$

with

$$\begin{aligned}B_\lambda &= \lambda^2\,(3e^\lambda-1-\lambda) + 2\lambda\,(e^\lambda-1)^2 \\ &\quad+ 2\,(2+\lambda)\,e^\lambda\,(e^\lambda-1) + (2+4\lambda+\lambda^2)\,e^\lambda \\ &= \lambda\,(2-\lambda-\lambda^2) - 2\,(1+\lambda-2\lambda^2)\,e^\lambda + 4\,(1+\lambda)\,e^{2\lambda}.\end{aligned}$$

It is easy to check that $\frac{d}{d\lambda}B_\lambda > 0$, so that this function and hence C_λ are increasing in λ, with $C_0 = B_0 = 2$.

For the range $\lambda \leq \frac{1}{2}$, the term $e-1$ appearing in the definition of C_λ may be replaced with $2\,(\sqrt{e}-1)$ (according to the inequality (III.7)), which leads to the constant $C_{1/2} = \frac{1}{2}\,(\frac{1}{2}+2\,(\sqrt{e}-1))^2 + \frac{7}{8} - 2\sqrt{e} + 6\,e < 15$. □

IV. Generating Functions

The probability function $f(k) = \mathbb{P}\{Z = k\}$ of the Poisson random variable $Z \sim P_\lambda$ satisfies the equation $\lambda f(k-1) = kf(k)$ in integers $k \geq 1$, which immediately implies

$$\lambda\,\mathbb{E}\,h(Z+1) = \mathbb{E}\,Zh(Z)$$

for any function h on $\mathbb{Z}$ (as long as the expectations exist). This identity was emphasized by Chen [4] who proposed to consider an approximate equality

$$\lambda\,\mathbb{E}\,h(X+1) \sim \mathbb{E}\,Xh(X)$$

as a characterization of a random variable X being almost Poisson with parameter λ. This idea was inspired by a similar approach of Charles Stein to problems of normal approximation on the basis of the approximate equality $\mathbb{E}\,h'(X) \sim \mathbb{E}\,Xh(X)$.

Another natural approach to Poisson approximation is based on the comparison of characteristic functions. Since the random variables W and Z take non-negative integer values, one may equivalently consider their associated generating functions.

The generating function for the Poisson law P_λ with parameter $\lambda > 0$ is given by

$$\varphi(w) = \mathbb{E}\,w^Z = \sum_{k=0}^{\infty} \mathbb{P}\{Z = k\}\,w^k = e^{\lambda(w-1)} = \prod_{j=1}^{n} e^{p_j(w-1)}, \tag{IV.1}$$

which is an entire function of the complex variable w. Correspondingly, the generating function for the distribution of the random variable $W = X_1 + \dots + X_n$ in (I.1) is

$$g(w) = \mathbb{E}\,w^W = \sum_{k=0}^{\infty} \mathbb{P}\{W = k\}\,w^k = \prod_{j=1}^{n} (q_j + p_j w), \tag{IV.2}$$

which is a polynomial of degree n. Hence, the difference between the involved probabilities may be expressed via the contour integrals by the Cauchy formula

$$\Delta_k = \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\} = \int_{|w|=r} w^{-k}\,(g(w) - \varphi(w))\,d\mu_r(w), \tag{IV.3}$$

where μ_r is the uniform probability measure on the circle $|w| = r$ of an arbitrary radius $r > 0$.

Note that for $w = e^{it}$ with real t, the generating functions φ and g turn into the characteristic functions of Z and W, respectively. Hence, closeness of the distributions of these random variables may be studied as a problem of closeness of their generating functions on the unit circle.

Let us now describe first steps based on the application of the formula (IV.3). Given complex numbers a_j, b_j ($1 \leq j \leq n$), we have an identity

$$a_1 \dots a_n - b_1 \dots b_n = \sum_{j=1}^{n} (a_j - b_j) \prod_{l<j} b_l \prod_{l>j} a_l \tag{IV.4}$$

with the convention that $\prod_{l<j} b_l = 1$ for $j = 1$ and $\prod_{l>j} a_l = 1$ for $j = n$. It implies

$$\Big|\prod_{j=1}^{n} a_j - \prod_{b=1}^{n} b_j\Big| \leq \sum_{j=1}^{n} |a_j - b_j| \prod_{l<j} |b_l| \prod_{l>j} |a_l|.$$

According to the product representations (IV.1)-(IV.2) to be used in (IV.3), one should choose here $a_j = q_j + p_j w$ and $b_j = e^{p_j(w-1)}$ with $|w| = r$. Then

$$|a_j| \leq q_j + p_j r \leq e^{p_j(r-1)}, \quad |b_j| = e^{p_j(\mathrm{Re}\,w-1)} \leq e^{p_j(r-1)}. \tag{IV.5}$$

Therefore

$$|g(w) - \varphi(w)| \leq \sum_{j=1}^{n} |a_j - b_j| \prod_{l\neq j} e^{p_l(r-1)} = e^{\lambda(r-1)} \sum_{j=1}^{n} |a_j - b_j|\,e^{-p_j(r-1)}. \tag{IV.6}$$

To estimate the terms in this sum, consider the function

$$\xi(u) = 1 + u - e^u = -u^2 \int_0^1 e^{tu}\,(1-t)\,dt \tag{IV.7}$$

of the complex variable u, where the Taylor integral formula is applied in the second representation. If $\mathrm{Re}\,u \leq 0$, then $|u^2\,e^{tu}| = |u|^2\,\exp\{t\,\mathrm{Re}\,u\} \leq |u|^2$, so,

$$|\xi(u)| \leq \frac{1}{2}\,|u|^2, \qquad \mathrm{Re}\,u \leq 0. \tag{IV.8}$$

In particular, for $u = p_j(w-1)$ with $w = \cos\theta + i\sin\theta$, we have

$$|w-1|^2 = (\cos\theta - 1)^2 + \sin^2\theta = 2(1-\cos\theta),$$

hence $|\xi(u)| \leq p_j^2\,(1-\cos\theta)$, and (IV.6) yields

$$|g(w) - \varphi(w)| \leq \sum_{j=1}^{n} |\xi(p_j(w-1))| \leq (1-\cos\theta) \sum_{j=1}^{n} p_j^2 \leq (1-\cos\theta)\,\lambda_2.$$

Integrating over the unit circle in (IV.3), we then arrive at the uniform bound:

Proposition IV.1. *We have*

$$\sup_{k\geq 0}\,|\,\mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}| \leq \lambda_2. \tag{IV.9}$$

This is a weakened variant of Le Cam's bound

$$|\mathbb{P}\{W \in A\} - \mathbb{P}\{Z \in A\}| \leq \lambda_2,$$

specialized to the one-point sets $A = \{k\}$. In order to get a similar bound with arbitrary sets, or develop applications to stronger distances, we need sharper forms of (IV.9), with the right-hand side properly depending on k.

V. Proof of Theorem I.3

Applying (IV.4) with $a_j = q_j + p_j w$ and $b_j = e^{p_j(w-1)}$ in (IV.3), one may write this formula as

$$\Delta_k = \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\} = \sum_{j=1}^n T_j(k), \qquad \text{(V.1)}$$

for all $k = 0, 1, \dots$ with

$$T_j(k) = \int_{|w|=r} w^{-k}\,(a_j - b_j) \prod_{l<j} b_l \prod_{l>j} a_l \, d\mu_r(w), \qquad \text{(V.2)}$$

where the integration is performed over the uniform probability measure μ_r on the circle $|w| = r$. Let us write $w = r(\cos\theta + i\sin\theta)$, $|\theta| < \pi$, and estimate $|T_j(k)|$ by replacing the integrand by its absolute value. Thus, using (IV.5), we get that $|T_j(k)|$ does not exceed

$$r^{-k} \int_{|w|=r} |a_j - b_j| \prod_{l<j} |e^{p_l\,(w-1)}| \prod_{l>j} |q_l + p_l w| \; d\mu_r(w)$$
$$= r^{-k} \int_{|w|=r} |a_j - b_j| \,\exp\Big\{(r\cos\theta - 1)\sum_{l=1}^{j-1} p_l\Big\} \times \prod_{l=j+1}^n |q_l + p_l w| \; d\mu_r(w)$$
$$= r^{-k} \exp\Big\{(r-1)\sum_{l=1}^{j-1} p_l\Big\} \int_{|w|=r} |a_j - b_j| \times \exp\Big\{-2r\sin^2\frac{\theta}{2}\sum_{l=1}^{j-1} p_l\Big\} \prod_{l=j+1}^n |q_l + p_l w| \; d\mu_r(w).$$

Here, in order to estimate $|a_j - b_j|$, let us return to the function $\xi(u)$ introduced in (IV.7), which we need at the values $u_j = p_j(w-1)$ with $|w| = r$.

Case 1: $r \geq 1$. Since $\mathrm{Re}\, u_j \leq p_j(r-1)$, we have, for any $t \in (0,1)$,

$$|u_j^2\, e^{tu_j}| = |u_j|^2\, e^{t\,\mathrm{Re}\, u_j} \leq |u_j|^2\, e^{p_j t(r-1)} \leq |u_j|^2\, e^{p_j(r-1)},$$

so, by (IV.7),

$$|a_j - b_j| = |\xi(u_j)| \leq \frac{1}{2}\, p_j^2\, |w-1|^2\, e^{p_j(r-1)}.$$

Case 2: $0 < r < 1$. Then $\mathrm{Re}\, u_j \leq 0$, so, by (IV.8),

$$|a_j - b_j| = |\xi(u_j)| \leq \frac{1}{2}\, p_j^2\, |w-1|^2.$$

Since $|w-1|^2 = (r-1)^2 + 4r\sin^2(\theta/2)$, we therefore obtain from (V.2) that

$$|T_j(k)| \leq \frac{p_j^2}{2}\, R_j(r)\, r^{-k}\Big((r-1)^2\, I_{j0}(r) + 4r\, I_{j2}(r)\Big), \qquad \text{(V.3)}$$

where

$$R_j(r) = \exp\Big\{(r-1)\sum_{l=1}^{j} p_l\Big\} \prod_{l=j+1}^n (q_l + p_l r)$$

for $r \geq 1$,

$$R_j(r) = \exp\Big\{(r-1)\sum_{l=1}^{j-1} p_l\Big\} \prod_{l=j+1}^n (q_l + p_l r)$$

for $r < 1$, and

$$I_{jm}(r) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Big|\sin\frac{\theta}{2}\Big|^m \exp\Big\{-2\,r\sin^2\frac{\theta}{2}\sum_{l=1}^{j-1} p_l\Big\} \times \prod_{l=j+1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r}\, d\theta.$$

In order to estimate the last integrals for $m = 0$ and $m = 2$, let us first note that

$$|q_l + p_l r e^{i\theta}|^2 = q_l^2 + p_l^2\, r^2 + 2p_l q_l\, r\cos\theta = (q_l + p_l r)^2 - 4\, q_l p_l\, r\sin^2\frac{\theta}{2} = (q_l + p_l r)^2\Big(1 - \frac{4q_l p_l r}{(q_l + p_l r)^2}\,\sin^2\frac{\theta}{2}\Big).$$

Hence, using $1 - x \leq e^{-x}$ $(x \in \mathbb{R})$, we have

$$\prod_{l=j+1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r} = \prod_{l=j+1}^n \Big(1 - \frac{4q_l p_l r}{(q_l + p_l r)^2}\sin^2\frac{\theta}{2}\Big)^{1/2} \leq \exp\Big\{-2\sin^2\frac{\theta}{2}\sum_{l=j+1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2}\Big\}, \qquad \text{(V.4)}$$

so that

$$I_{jm}(r) \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} \Big|\sin\frac{\theta}{2}\Big|^m \exp\Big\{-2\gamma_j(r)\sin^2\frac{\theta}{2}\Big\}\, d\theta \leq \frac{1}{2\pi}\, 2^{-m}\int_{-\pi}^{\pi} |\theta|^m \exp\Big\{-\frac{2}{\pi^2}\,\gamma_j(r)\,\theta^2\Big\}\, d\theta. \qquad \text{(V.5)}$$

Here we applied the inequalities $\frac{2}{\pi}\,t \leq \sin t \leq t$ $(0 \leq t \leq \frac{\pi}{2})$ and used the notation

$$\gamma_j(r) = r\Big(\sum_{l=1}^{j-1} p_l + \sum_{l=j+1}^n \frac{q_l p_l}{(q_l + p_l r)^2}\Big).$$

Thus, we need to bound γ_j from below. If $r \geq 1$, then $q_l + p_l r \leq r$, hence

$$\sum_{l=j+1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2} \geq \frac{1}{r}\sum_{l=j+1}^n q_l p_l.$$

This gives

$$\gamma_j(r) \geq r\sum_{l=1}^{j-1} p_l + \frac{1}{r}\sum_{l=j+1}^n q_l p_l$$
$$= r\sum_{l=1}^{j-1} p_l + \frac{1}{r}\sum_{l=1}^n (p_l - p_l^2) - \frac{1}{r}\sum_{l=1}^j (p_l - p_l^2)$$
$$= \Big(r - \frac{1}{r}\Big)\sum_{l=1}^{j-1} p_l + \frac{1}{r}\sum_{l=1}^{j-1} p_l^2 + \frac{1}{r}\sum_{l=1}^n (p_l - p_l^2) - \frac{1}{r}\, q_j p_j$$
$$\geq \frac{1}{r}\,(\lambda - \lambda_2 - q_j p_j).$$

In case $r \le 1$, we use $q_l + p_l r \le 1$, implying that

$$\sum_{l=j+1}^{n} \frac{q_l p_l}{(q_l + p_l r)^2} \ge \sum_{l=j+1}^{n} q_l p_l.$$

Therefore in this range we have a similar lower bound, namely

$$\begin{aligned} \gamma_j(r) &\ge r \sum_{l=1}^{j-1} p_l + r \sum_{l=j+1}^{n} q_l p_l \\ &= r \sum_{l=1}^{j-1} p_l + r \sum_{l=1}^{n} (p_l - p_l^2) - r \sum_{l=1}^{j} (p_l - p_l^2) \\ &= -r p_j + r \sum_{l=1}^{j} p_l^2 + r \sum_{l=1}^{n} (p_l - p_l^2) \\ &\ge r\,(\lambda - \lambda_2 - q_j p_j). \end{aligned}$$

Since $q_j p_j \le \frac{1}{4}$, both lower bounds yield

$$\gamma_j(r) \ \ge\ \psi(r) - \frac{1}{4}, \quad \psi(r) = \min\{r, 1/r\}\,(\lambda - \lambda_2).$$

As a result, (V.5) simplifies to

$$\begin{aligned} I_{jm}(r) &\le \frac{1}{2\pi}\, 2^{-m} \sqrt{e} \int_{-\pi}^{\pi} |\theta|^m \exp\Big\{ -\frac{2}{\pi^2}\, \psi(r)\, \theta^2 \Big\}\, d\theta \\ &= \sqrt{e}\, \frac{\pi^m}{4^{m+1}}\, \psi(r)^{-\frac{m+1}{2}} \int_{-2\sqrt{\psi(r)}}^{2\sqrt{\psi(r)}} |x|^m\, e^{-\frac{1}{2}x^2}\, dx. \end{aligned}$$

The last integral may be extended to the whole real line, which makes sense for large values of $\psi(r)$, or one may bound the exponential term in the integrand by 1, which makes sense for small values of $\psi(r)$. These two variants of bounds lead to

$$\begin{aligned} I_{jm}(r) &\le \sqrt{e}\, \frac{\pi^m}{4^{m+1}}\, \psi(r)^{-\frac{m+1}{2}} \\ &\quad \times\ \min\Big\{ \sqrt{2\pi}\, \mathbb{E}\,|\xi|^m,\ \frac{2^{m+2}}{m+1}\, \psi(r)^{\frac{m+1}{2}} \Big\} \\ &\le \sqrt{e}\, \frac{\pi^m}{4^{m+1}}\ \max\Big\{ \sqrt{2\pi}\, \mathbb{E}\,|\xi|^m,\ \frac{2^{m+2}}{m+1} \Big\} \\ &\quad \times\ \min\Big\{ 1,\ \psi(r)^{-\frac{m+1}{2}} \Big\}, \end{aligned}$$

where ξ is a standard normal random variable. In particular, we get the upper bounds

$$\begin{aligned} I_{j0}(r) &\le \sqrt{e}\ \min\big\{1,\ \psi(r)^{-1/2}\big\}, \\ I_{j2}(r) &\le \frac{\sqrt{e}\,\pi^2}{12}\ \min\big\{1,\ \psi(r)^{-3/2}\big\}. \end{aligned}$$

In view of $q_l + p_l r \le e^{(r-1)p_l}$, from the definition of $R_j(r)$ we also have the bound

$$R_j(r) \ \le\ \exp\Big\{ (r-1) \sum_{l=1}^{n} p_l \Big\} \ =\ e^{\lambda(r-1)}$$

in case $r \ge 1$, while for $r \le 1$

$$\begin{aligned} R_j(r) &\le \exp\Big\{ (r-1) \sum_{l \ne j} p_l \Big\} \\ &= e^{\lambda(r-1)}\, e^{-p_j(r-1)} \ \le\ e^{\lambda(r-1)+1}. \end{aligned}$$

Applying these bounds in (V.3), we therefore obtain that

$$\begin{aligned} |T_j(k)| \le \frac{\delta_r}{2}\, p_j^2\, e^{\lambda(r-1)+\frac{1}{2}}\, r^{-k} \Big((r-1)^2 \min\big\{1,\ \psi(r)^{-\frac{1}{2}}\big\} \\ + \frac{\pi^2}{3}\, r \min\big\{1,\ \psi(r)^{-\frac{3}{2}}\big\} \Big), \end{aligned}$$

where $\delta_r = 1$ in case $r \ge 1$ and $\delta_r = e$ for $r < 1$. Summing over $j \le n$ and recalling (V.1), one can estimate $|\Delta_k|$ from above by

$$\begin{aligned} |\Delta_k| \le \lambda_2\, \delta_r\, e^{\lambda(r-1)}\, r^{-k} \Big(\frac{\sqrt{e}}{2}\, (r-1)^2\, \min\big\{1,\ \psi(r)^{-\frac{1}{2}}\big\} \\ + \frac{\sqrt{e}\,\pi^2}{6}\ r\ \min\big\{1,\ \psi(r)^{-\frac{3}{2}}\big\} \Big). \end{aligned} \tag{V.6}$$

Now, letting $r \to 0$ in the case $k = 0$, (V.6) leads to

$$|\Delta_0| \ \le\ \frac{e\sqrt{e}}{2}\, \lambda_2\, e^{-\lambda} \ <\ 3\lambda_2\, e^{-\lambda},$$

and we obtain the first inequality in (I.9). Letting $r \downarrow 1$ in the case $k \ge 1$, (V.6) gives

$$|\Delta_k| \ \le\ \frac{\sqrt{e}\,\pi^2}{6}\, \lambda_2\, e^{-\lambda} \ <\ 3\lambda_2\, e^{-\lambda},$$

which is the second inequality in (I.9).

But, if $k \ge 1$, one may also use (V.6) with $r = \frac{k}{\lambda}$ and apply the bound $k! \le e\, k^{k+\frac{1}{2}}\, e^{-k}$, cf. (II.1), giving

$$\begin{aligned} e^{\lambda(r-1)}\, r^{-k} &= \Big(\frac{e\lambda}{k}\Big)^k e^{-\lambda} \\ &\le e\sqrt{k}\, f(k), \quad f(k) = \frac{\lambda^k}{k!}\, e^{-\lambda}. \end{aligned}$$

To simplify the numerical constants, note that $\frac{1}{2} e^{\frac{5}{2}} < 6.1$ and $\frac{1}{6} e^{\frac{5}{2}} \pi^2 < 20.1$. Recalling that $\psi(r) = \rho$ for $r = k/\lambda$, we finally get the inequality (I.10),

$$\begin{aligned} |\Delta_k| \le \lambda_2 \sqrt{k}\, f(k) \Big(7 \Big(\frac{k-\lambda}{\lambda}\Big)^2 \min\big\{1,\ \rho^{-\frac{1}{2}}\big\} \\ + 21\, \frac{k}{\lambda}\ \min\big\{1,\ \rho^{-\frac{3}{2}}\big\} \Big). \end{aligned} \tag{V.7}$$

□

VI. Consequences of Theorem I.3

Under the natural requirement that λ_2 is bounded away from λ, the bound (V.7) on $\Delta_k = \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}$ may be simplified. As before, we use the notations

$$f(k) = \mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!}\, e^{-\lambda},$$

$$\lambda = p_1 + \cdots + p_n, \quad \lambda_2 = p_1^2 + \cdots + p_n^2.$$

Note that $\lambda_2 \le \lambda$ and recall that $\rho = (\lambda - \lambda_2)\ \min\{\frac{k}{\lambda}, \frac{\lambda}{k}\}$.

Corollary VI.1. *If $\lambda_2 \le \kappa\lambda$, $0 < \kappa < 1$, then for any integer $k \ge 0$,*

$$|\Delta_k| \le \frac{7\ f(k)}{(1-\kappa)^{\frac{3}{2}}} \Big(\frac{(k-\lambda)^2}{\lambda} + 3 \Big) \frac{\lambda_2}{\lambda}\, \max\Big\{ \Big(\frac{k}{\lambda}\Big)^3, 1 \Big\}. \tag{VI.1}$$

In particular, if $k \le 2\lambda$, *then*

$$|\Delta_k| \le \frac{56\, f(k)}{(1-\kappa)^{\frac{3}{2}}} \Big(\frac{(k-\lambda)^2}{\lambda} + 3\Big) \frac{\lambda_2}{\lambda}. \qquad \text{(VI.2)}$$

If $k \ge \lambda \ge 1/2$, *we also have*

$$|\Delta_k| \le \frac{49\, f(k)}{(1-\kappa)^{\frac{3}{2}}} \Big(\frac{k}{\lambda}\Big)^3 \lambda_2. \qquad \text{(VI.3)}$$

Proof. The assumption $\lambda_2 \le \kappa\lambda$ ensures that

$$\rho \ge (1-\kappa)\lambda \min\Big\{\frac{k}{\lambda}, \frac{\lambda}{k}\Big\}.$$

If $1 \le k \le K\lambda$ ($K \ge 1$), then $\frac{k}{\lambda} \le K^2 \frac{\lambda}{k}$ and $\rho \ge \frac{1-\kappa}{K^2}\, k$, so, the right-hand side of (V.7) is bounded from above by

$$\lambda_2 \sqrt{k}\, f(k) \Big(7 \Big(\frac{k-\lambda}{\lambda}\Big)^2 \frac{K}{\sqrt{(1-\kappa)\, k}} + 21\, \frac{k}{\lambda}\, \frac{K^3}{(1-\kappa)^{\frac{3}{2}}\, k^{\frac{3}{2}}}\Big).$$

Choosing $K = \max\{\frac{k}{\lambda}, 1\}$, this expression does not exceed the right-hand side of (VI.1). Thus, the inequality (I.10) yields (VI.1), which in turn immediately implies (VI.2).

In case $k = 0$, we apply (I.9). Since $\frac{(k-\lambda)^2}{\lambda} + 3 \ge \lambda$ for $k = 0$, the right-hand side of (I.10) is dominated by the right-hand side of (VI.1). Thus, we obtain (VI.1) without any constraints on k, and therefore (VI.2) holds true for all $k \le 2\lambda$.

In case $k \ge \lambda$, necessarily $\rho \ge (1-\kappa)\,\lambda^2/k$. Hence, the right-hand side of (V.7) may be bounded from above by

$$\lambda_2 \sqrt{k}\, f(k) \Big(7 \Big(\frac{k-\lambda}{\lambda}\Big)^2 \frac{\sqrt{k}}{\lambda\sqrt{1-\kappa}} + 21\, \frac{k}{\lambda} \cdot \frac{k^{\frac{3}{2}}}{\lambda^3\, (1-\kappa)^{\frac{3}{2}}}\Big).$$

Using $(\frac{k-\lambda}{\lambda})^2 \le \frac{k^2}{\lambda^2}$ to bound the first term in the brackets and $\frac{k}{\lambda} \le 2k$ to bound the second term (using $\lambda \ge \frac{1}{2}$), we obtain the bound (VI.3). □

We are now prepared to extend Proposition III.4 to larger values of λ under the assumption that λ_2/λ is bounded away from 1. The next assertion, when combined with Proposition III.4, yields Theorem I.2 with $c = 15$ in case $\lambda \le \frac{1}{2}$ and $c = 56 \cdot 10^6$ in case $\lambda > \frac{1}{2}$.

Proposition VI.2. *If* $\lambda \ge \frac{1}{2}$ *and* $\lambda_2 \le \kappa\lambda$, $\kappa \in (0,1)$, *then*

$$\frac{1}{4}\Big(\frac{\lambda_2}{\lambda}\Big)^2 \le D(W||Z) \le \chi^2(W,Z) \le c_\kappa \Big(\frac{\lambda_2}{\lambda}\Big)^2 \qquad \text{(VI.4)}$$

where $c_\kappa = c\,(1-\kappa)^{-3}$ *with, for example,* $c = 7 \cdot 10^6$.

Proof. The leftmost lower bound in (VI.4) is added according to (I.7) (using the Pinsker inequality, it also follows with some constant from Barbour-Hall's lower bound in Theorem I.1). Hence, it remains to show the rightmost upper bound in (VI.4).

Write

$$\chi^2(W,Z) = \sum_{k=0}^{\infty} \frac{\Delta_k^2}{f(k)} = S_1 + S_2 = \Big(\sum_{k=0}^{[2\lambda]} + \sum_{k=[2\lambda]+1}^{\infty}\Big) \frac{\Delta_k^2}{f(k)}.$$

In the range $0 \le k \le [2\lambda]$, we apply the inequality (VI.2) which gives

$$\Delta_k^2 \le \frac{56^2}{(1-\kappa)^3} \Big(\frac{(k-\lambda)^4}{\lambda^2} + 6\, \frac{(k-\lambda)^2}{\lambda^2} + 9\Big) \Big(\frac{\lambda_2}{\lambda}\Big)^2 f(k)^2.$$

Hence

$$S_1 \le \frac{56^2}{(1-\kappa)^3} \Big(\frac{\mathbb{E}\,(Z-\lambda)^4}{\lambda^2} + 6\, \frac{\mathbb{E}\,(Z-\lambda)^2}{\lambda} + 9\Big) \Big(\frac{\lambda_2}{\lambda}\Big)^2.$$

Using the moment formula

$$\mathbb{E}\, Z^m = \lambda(\lambda+1)\dots(\lambda+m-1),$$

we have $\mathbb{E}\,(Z-\lambda)^2 = \lambda$ and $\mathbb{E}\,(Z-\lambda)^4 = 3\lambda(\lambda+2)$, so that

$$\begin{aligned} S_1 &\le \frac{56^2}{(1-\kappa)^3} \Big(\frac{3\lambda\,(\lambda+2)}{\lambda^2} + 15\Big) \Big(\frac{\lambda_2}{\lambda}\Big)^2 \\ &= \frac{18\,816}{(1-\kappa)^3}\, (\lambda^{-1} + 3) \Big(\frac{\lambda_2}{\lambda}\Big)^2 \\ &\le \frac{C_1}{(1-\kappa)^3} \Big(\frac{\lambda_2}{\lambda}\Big)^2 \qquad \text{(VI.5)} \end{aligned}$$

with $C_1 = 94\,080$ (where we used the assumption $\lambda \ge \frac{1}{2}$ on the last step).

In order to estimate S_2, we use the following elementary bound

$$\sum_{k=k_0}^{\infty} k^d f(k) \le k_0^d\, f(k_0) \Big(1 - \frac{\lambda}{k_0} \Big(\frac{k_0+1}{k_0}\Big)^{d-1}\Big)^{-1} \qquad \text{(VI.6)}$$

which holds for any $d = 1, 2, \dots$ when $k_0^d/(k_0+1)^{d-1} > \lambda$. For the proof, write

$$\sum_{k=k_0}^{\infty} k^d f(k) = k_0^d f(k_0)\big(1 + \theta_1 + \theta_1\theta_2 + \dots + \theta_1 \dots \theta_m + \dots\big),$$

where

$$\theta_m = \Big(\frac{k_0+m}{k_0+m-1}\Big)^d \frac{\lambda}{k_0+m}, \qquad m = 1, 2, \dots$$

Since the function $(x+1)^{d-1} x^{-d}$ is decreasing in $x > 0$, we have $1 > \theta_1 > \theta_2 > \dots$ This gives

$$\sum_{k=k_0}^{\infty} k^d f(k) \le k_0^d f(k_0) \Big(1 + \sum_{m=1}^{\infty} \theta_1^m\Big),$$

that is, (VI.6). In particular, for $k_0 = [2\lambda] + 1$ and $\lambda \ge 8$ (with $d = 6$),

$$\Big(1 - \frac{\lambda}{k_0} \Big(\frac{k_0+1}{k_0}\Big)^5\Big)^{-1} < \Big(1 - \frac{1}{2} \Big(\frac{2\lambda+1}{2\lambda}\Big)^5\Big)^{-1} < 3.1.$$

So, by (VI.6), and using $[2\lambda] + 1 \le \frac{17}{8}\lambda$ for the chosen range of λ, we have

$$\begin{aligned} \sum_{k=[2\lambda]+1}^{\infty} k^6\ f(k) &\le 3.1\, ([2\lambda]+1)^6\, f([2\lambda]+1) \\ &\le 3.1 \cdot (17\lambda/8)^6\, f([2\lambda]+1). \end{aligned}$$

Hence, by (VI.3),

$$\begin{aligned} S_2 &= \sum_{k=[2\lambda]+1}^{\infty} \frac{|\Delta_k|^2}{f(k)} \\ &\le \frac{49^2}{(1-\kappa)^3} \sum_{[2\lambda]+1}^{\infty} \Big(\frac{k}{\lambda}\Big)^6 \lambda_2^2 f(k) \\ &\le \frac{C_2\,\lambda_2^2}{(1-\kappa)^3}\, f([2\lambda]+1) \qquad \text{(VI.7)} \end{aligned}$$

with $C_2 = 49^2 \cdot 3.1 \cdot (17/8)^6 < 685\,343$. Asymptotically with respect to large λ, this bound is much better than (VI.4). Applying $f(k) \le \frac{1}{\sqrt{2\pi k}}\, e^{k-\lambda}\,(\frac{\lambda}{k})^k$ as in (II.5) with $k = [2\lambda]+1$ and using $2\lambda \le k \le 2\lambda + 1$, we have

$$\begin{aligned} f([2\lambda]+1) &\le \frac{e}{2\sqrt{\lambda\pi}}\,(e/4)^{\lambda} \\ &\le \frac{e}{2\sqrt{\pi}}\; 8^{3/2} \Big(\frac{e}{4}\Big)^8 \frac{1}{\lambda^2} \;<\; \frac{1}{\lambda^2}. \end{aligned}$$

This gives

$$S_2 \le \frac{C_2}{(1-\kappa)^3} \Big(\frac{\lambda_2}{\lambda}\Big)^2.$$

As a result, we arrive at the desired upper bound in (VI.4).

Finally, let us estimate S_2 for the range $\frac{1}{2} \le \lambda \le 8$. Returning to (VI.7), we have

$$\begin{aligned} S_2 &\le \frac{49^2}{(1-\kappa)^3} \sum_{k=1}^{\infty} \Big(\frac{k}{\lambda}\Big)^6 \lambda_2^2 f(k) \\ &\le \frac{49^2}{(1-\kappa)^3}\, \lambda^{-6}\, \lambda_2^2\, \mathbb{E}Z^6 \;\le\; \frac{C_2'}{(1-\kappa)^3} \Big(\frac{\lambda_2}{\lambda}\Big)^2, \end{aligned}$$

where $C_2' = 49^2 \sup_{\frac{1}{2}\le\lambda\le 4} \psi(\lambda)$, $\psi(\lambda) = \lambda^{-4}\,\mathbb{E}Z^6$. Here

$$\psi(\lambda) = \frac{(\lambda+1)\dots(\lambda+5)}{\lambda^3} = \psi_1(\lambda)\psi_2(\lambda)\psi_3(\lambda)$$

with $\psi_1(\lambda) = 5+\lambda+\frac{4}{\lambda}$, $\psi_2(\lambda) = 7+\lambda+\frac{10}{\lambda}$, $\psi_3(\lambda) = 1+\frac{3}{\lambda}$. All these three functions are convex, while ψ_3 is decreasing. In addition, $\psi_i(\frac{1}{2}) \ge \psi_i(8)$ for $i = 1, 2$. Hence $\psi(\lambda) \le \psi(\frac{1}{2}) = \frac{1}{4}\cdot 11!!$ It follows that $C_2' = 49^2 \cdot \frac{1}{4} \cdot 11!! < 6\,239\,560$, and thus $c = C_1 + C_2''$ is the resulting constant in (VI.4). □

Remark VI.3. Up to a numerical constant, the upper bound in (VI.4) immediately implies an upper bound of Theorem I.1 in case $\lambda \ge \frac{1}{2}$, in view of the relation $d(W,Z)^2 \le \frac{1}{2}\, D(W,Z)$. Indeed, (VI.4) gives $d(W,Z) \le c_\kappa \lambda_2/\lambda$, provided that $\lambda_2 \le \kappa\lambda$. But, in the other case $\lambda_2 \ge \kappa\lambda$, there is nothing to prove, since $d(W,Z) \le 2$. Note also that, for $\lambda \le \frac{1}{2}$, the correct upper bound on the total variation distance is of the form $d(W,Z) \le C\lambda_2$. It may be obtained as a consequence of Lemmas III.1-III.2.

VII. Uniform Bounds. Comparison With Normal Approximation

A different choice of the parameter r in the proof of Theorem I.3 may provide various uniform bounds in the Poisson approximation, like in the next assertion. Using the $L^\infty(\mu)$-norm with respect to the counting measure μ on $\mathbb{Z}$, let us focus on the deviations of the densities of W and Z and the deviations of their distribution functions. These distances are thus given by

$$\begin{aligned} M(W,Z) &= \sup_{k\ge 0}\, |\mathbb{P}\{W=k\} - \mathbb{P}\{Z=k\}|, \\ K(W,Z) &= \sup_{k\ge 0}\, |\mathbb{P}\{W\le k\} - \mathbb{P}\{Z\le k\}|. \end{aligned}$$

Putting $r = 1$ in (V.6), we arrive at the next assertion which sharpens Proposition IV.1.

Theorem VII.1. *We have*

$$M(W,Z) \;\le\; \frac{\sqrt{e}\,\pi^2}{6}\,\lambda_2 \min\big\{1, (\lambda-\lambda_2)^{-\frac{3}{2}}\big\}. \qquad \text{(VII.1)}$$

This uniform bound is not new; with a non-explicit numerical factor, it corresponds to Theorem 3.1 in Čekanavicius [3], p. 53. For $\lambda \le 1$, this relation simplifies to

$$M(W,Z) \;\le\; \frac{\sqrt{e}\,\pi^2}{6}\,\lambda_2,$$

which cannot be improved (modulo a numerical factor) in view of the lower bounds on $|\Delta_k|$ with $k = 0, 1, 2$ mentioned in Section III. We also have a similar bound for the Kolmogorov distance, $K(W,Z) \le C\lambda_2$, which follows from the upper bound for the stronger total variation distance as in Theorem I.1.

When, however, λ is large (and say all $p_j \le \frac{1}{2}$), one would expect to achieve more accurate bounds when replacing the Poisson approximation for P_W by the normal law $N(\lambda,\lambda)$ with mean λ and variance λ. Indeed, suppose, for example, that $p_j = \frac{1}{2}$, so that W has a binomial distribution with parameters $(n, \frac{1}{2})$, while the approximating Poisson distribution has parameter $\lambda = n/2$ with $\lambda_2 = n/4$. Here, (I.2) only yields $d(W,Z) \sim 1$, which means that there is no Poisson approximation with respect to the total variation! Nevertheless, the approximation is still meaningful in a weaker sense in terms of the Kolmogorov distance K, as well as in terms of M. In this case, both P_W and P_λ are almost equal to $N(\lambda,\lambda)$, and the Berry-Esseen theorem provides a correct bound $K(W,Z) \le \frac{c}{\sqrt{n}}$ via the triangle inequality for K. Since $M \le 2K$ (which holds true for all probability distributions on $\mathbb{Z}$), we also have $M(W,Z) \le \frac{c}{\sqrt{n}}$. Note that this inequality also follows from Theorem VII.1. Indeed, when $\lambda_2 \le \frac{1}{2}\,\lambda$, (VII.1) is simplified to

$$M(W,Z) \;\le\; \frac{\sqrt{2e}\,\pi^2}{3}\,\frac{\lambda_2}{\lambda^{\frac{3}{2}}}, \qquad \text{(VII.2)}$$

which yields a correct order for growing n. Thus, the two approaches are equivalent for this particular (i.i.d.) example.

To realize whether or not the normal approximation is better or worse than the Poisson approximation in the general non-i.i.d. situation (that is, with different p_j's), let us evaluate the corresponding Lyapunov ratio in the central limit theorem and apply the Berry-Esseen bound $K(W,N_\lambda) \le cL_3$, where the random variable N_λ is distributed according to $N(\lambda,\lambda)$.

Since $\mathrm{Var}(W) = \sum_{j=1}^n p_j q_j = \lambda - \lambda_2$, the Lyapunov ratio for the sequence $X_1, \dots, X_n$ is given by

$$L_3 = \frac{1}{\mathrm{Var}(W)^{\frac{3}{2}}} \sum_{j=1}^n \mathbb{E}\,|X_j - \mathbb{E}X_j|^3 = \frac{1}{(\lambda - \lambda_2)^{\frac{3}{2}}} \sum_{j=1}^n (p_j^2 + q_j^2)\, p_j q_j \ \le \ \frac{1}{\sqrt{\lambda - \lambda_2}}$$

(note that $\frac{1}{2} \le p_j^2 + q_j^2 \le 1$). Hence $K(W, N_\lambda) \le \frac{c}{\sqrt{\lambda - \lambda_2}}$, up to some absolute constant $c > 0$. A similar bound holds for Z as well when representing W as the sum of n independent Poisson random variables Z_j with parameters p_j. Namely, for the sequence $Z_1, \dots, Z_n$, we have

$$L_3 \ = \ \frac{1}{\mathrm{Var}(Z)^{\frac{3}{2}}} \sum_{j=1}^n \mathbb{E}\,|Z_j - \mathbb{E}Z_j|^3 \ \le \ \frac{c}{\lambda^{\frac{3}{2}}} \sum_{j=1}^n p_j \ = \ \frac{c}{\sqrt{\lambda}}.$$

Therefore, $K(Z, N_\lambda) \le \frac{c}{\sqrt{\lambda}}$ and hence, by the triangle inequality, $K(W, Z) \le \frac{c}{\sqrt{\lambda - \lambda_2}}$. In particular, in a typical situation where $\lambda_2 \le \frac{1}{2}\lambda$, the normal approximation yields

$$M(W, Z) \ \le \ \frac{c}{\sqrt{\lambda}} \qquad \text{(VII.3)}$$

with some absolute constant c. But, this bound is surprisingly worse than (VII.2) as long as $\lambda_2 = o(\lambda)$.

Consider as an example $p_j = 1/(2\sqrt{j})$ for $j = 1, \dots, n$. Then $\lambda \sim \sqrt{n}$, $\lambda_2 \sim \log n$, and we get $M(W, Z) \le cn^{-3/4} \log n$ in (VII.2), while (VII.3) only yields $M(W, Z) \le cn^{-1/4}$. This example is also illustrative when comparing Theorem I.2 with (I.5). The first one provides a correct asymptotic

$$D(W, Z) \sim \frac{\log^2 \ n}{n}$$

(within absolute factors), while (I.5) only gives $D(W, Z) \le c$.

Acknowledgment

The authors would like to thank Igal Sason and two referees for valuable comments and drawing our attention to additional references related to the Poisson approximation in informational distances.

References

[1] A. D. Barbour and P. Hall, "On the rate of Poisson convergence," *Math. Proc. Cambridge Philos. Soc.*, vol. 95, no. 3, pp. 473–480, 1984.

[2] A. D. Barbour, L. Holst, and S. Janson, "Poisson Approximation - Oxford University Press," in *Oxford Science Publications*. New York, NY, USA: Oxford University Press, 1992, p. 277.

[3] V. Čekanavičius, "Approximation methods in probability theory," in *Universitext*. Springer, 2016, p. 274.

[4] L. H. Y. Chen, "Poisson approximation for dependent trials," *Ann. Probab.*, vol. 3, no. 3, pp. 534–545, 1975.

[5] T. van Erven and P. Harremos, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.

[6] P. Harremoes, "Binomial and Poisson distributions as maximum entropy distributions," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 2039–2041, Jul. 2001.

[7] P. Harremo, "Convergence to the poisson distribution in information divergence," Math. Dept., Univ. Copenhagen, København, Denmark, Tech. Rep., Feb. 2003.

[8] P. Harremoës, O. Johnson, and I. Kontoyiannis. (Jan. 2016). "Thinning and information projections." [Online]. Available: https://arxiv.org/abs/1601.04255

[9] P. Harremoes and P. S. Ruzankin, "Rate of convergence to Poisson law in terms of information divergence," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2145–2149, Sep. 2004.

[10] J. L. Hodges, Jr., and L. Le Cam, "The Poisson approximation to the Poisson binomial distribution," *Ann. Math. Statist.*, vol. 31, no. 3, pp. 737–740, 1960.

[11] O. Johnson, *Information Theory and the Central Limit Theorem*. London, U.K.: Imperial College Press, 2004, p. 209.

[12] J. Kerstan, "Verallgemeinerung eines Satzes von Prochorow und Le Cam," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 2, pp. 173–179, Jan. 1964.

[13] I. Kontoyiannis, P. Harremoes, and O. Johnson, "Entropy and the law of small numbers," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 466–472, Feb. 2005.

[14] L. Le Cam, "An approximation theorem for the Poisson binomial distribution," *Pacific J. Math.*, vol. 10, no. 4, pp. 1181–1197, 1960.

[15] M. Le LeCam, "Asymptotic methods in statistical decision theory," in *Springer Series in Statistics*. New York, NY, USA: Springer-Verlag, 1986, p. 742.

[16] F. Liese and I. Vajda, *Convex Statistical Distances. With German, French and Russian Summaries. Teubner-Texte ZUR Mathematik [Teubner Texts in Mathematics]*, vol. 95. Leipzig, Germany: B. G. Teubner Verlag, 1987, p. 224.

[17] Y. V. Prokhorov, "Asymptotic behavior of the binomial distribution," *Uspekhi Mat. Nauk* vol. 8, no. 3, pp. 135–142, 1953.

[18] M. Romanowska, "A note on the upper bound for the distance in total variation between the binomal and the Poisson distribution," *Statistica Neerlandica*, vol. 31, no. 3, pp. 127–130, 1977.

[19] B. Roos, "On the rate of multivariate Poisson convergence," *J. Multivariate Anal.*, vol. 69, no. 1, pp. 120–134, 1999.

[20] I. Sason, "Improved lower bounds on the total variation distance for the Poisson approximation," *Statist. Probab. Lett.*, vol. 83, no. 10, pp. 2422–2431, 2013.

[21] R. J. Serfling, "A general Poisson approximation theorem," *Ann. Probab.*, vol. 3, no. 4, pp. 726–731, 1975.

[22] I. Vajda, *Theory of Statistical Inference and Information*. Norwell, MA, USA: Kluwer Academic Publishers, 1989, p. 432.

[23] W. Vervaat, "Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution," *Statistica Neerlandica*, vol. 23, no. 1, pp. 79–86, 1969.