Bernoulli **16**(2), 2010, 301–330
DOI: 10.3150/09-BEJ201

A new method for obtaining sharp compound Poisson approximation error estimates for sums of locally dependent random variables

MICHAEL V. BOUTSIKAS[1] and EUTICHIA VAGGELATOU[2]

[1]*Department of Statistics and Insurance Science, University of Piraeus, Karaoli and Dimitriou Str. 80, Piraeus 18534, Greece. E-mail: mbouts@unipi.gr*
[2]*Section of Statistics and Operations Research, Department of Mathematics, University of Athens, Panepistemiopolis, Athens 15784, Greece. E-mail: evagel@math.uoa.gr*

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent or locally dependent random variables taking values in $\mathbb{Z}_+$. In this paper, we derive sharp bounds, via a new probabilistic method, for the total variation distance between the distribution of the sum $\sum_{i=1}^n X_i$ and an appropriate Poisson or compound Poisson distribution. These bounds include a factor which depends on the smoothness of the approximating Poisson or compound Poisson distribution. This "smoothness factor" is of order $\mathrm{O}(\sigma^{-2})$, according to a heuristic argument, where σ^2 denotes the variance of the approximating distribution. In this way, we offer sharp error estimates for a large range of values of the parameters. Finally, specific examples concerning appearances of rare runs in sequences of Bernoulli trials are presented by way of illustration.

Keywords: compound Poisson approximation; coupling inequality; law of small numbers; locally dependent random variables; Poisson approximation; rate of convergence; total variation distance; Zolotarev's ideal metric of order 2

1. Introduction and overview

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent or locally dependent random variables which take values in $\mathbb{Z}_+$. If $X_1, X_2, \ldots, X_n$ rarely differ from zero (that is, $P(X_i \neq 0) \approx 0$), then it is well known that the distribution of their sum can be efficiently approximated by an appropriate Poisson or compound Poisson distribution. This situation appears in a great number of applications involving locally dependent and rare events, such as risk theory, extreme value theory, reliability theory, run and scan statistics, graph theory and biomolecular sequence analysis.

The main method used so far for establishing effective Poisson or compound Poisson approximation results in the case of independent or dependent random variables is the much acclaimed Stein–Chen method (see, for example, Barbour, Holst and Janson (1992), Barbour and Chryssaphinou (2001), Barbour and Chen (2005) and the references therein). Another method for independent random variables is Kerstan's method (see Roos (2003) and the references therein).

In the recent years, an alternative methodology has been developed in a series of papers concerning compound Poisson approximation for sums or processes of dependent random variables, employing probabilistic techniques, that is, properties of certain probability metrics, stochastic orders and coupling techniques (see Boutsikas and Koutras (2000, 2001), Boutsikas and Vaggelatou (2002), Boutsikas (2006)). In this series of papers, the error estimates are, under analogous assumptions, of almost the same nature and the same order as the error estimates developed by the Stein–Chen method. The main shortcoming of these bounds, though, is that they do not incorporate any so-called "magic factor" (however, in the process approximation case treated in Boutsikas (2006), such a factor cannot be present). This factor, also known as a *Stein factor*, appears in approximation error estimates obtained through the Stein–Chen method and decreases as the parameter of the Poisson distribution increases.

The purpose of this work is to derive sharp error bounds for the total variation distance between the distribution of the sum of integer-valued random variables and an appropriate Poisson or compound Poisson distribution. Specifically, by assuming that the random variables $X_1, X_2, \dots, X_n$ are locally dependent (in the strict sense of k-dependence), we derive bounds similar in nature to those obtained by the Stein–Chen method that include a factor analogous to a Stein factor. This factor is better/smaller than the associated Stein factors, thereby offering (for a large range of the values of the parameters) sharper bounds than relative ones derived via the Stein–Chen method. This factor is just the L_1-norm, $\|\Delta^2 f\|_1$, of the second difference of the probability distribution function f of the approximating Poisson or compound Poisson distribution. It decreases as f becomes smoother, which, in our case, usually happens when the variance of the distribution corresponding to f increases. Hence, we shall often refer to this factor as the *smoothness factor*. The methodology we employ is based on a modification of Lindeberg's method, along with the coupling inequality of Lemma 4 and the smoothing inequality (which produces the aforementioned smoothness factor) of Lemma 1.

It is worth pointing out an undesired effect of our treatment, which is an additional term in the proposed bounds that does not appear in Stein–Chen bounds. This term becomes large for a certain range of values of the parameters, but, as we explain in Remark 3 of Section 3, it can be substantially reduced if we possess a simple and effective upper bound for $\|\Delta^2 f\|_1$. Nevertheless, this term is generally negligible, especially for small or moderate values of λ, where λ is the parameter of the approximating Poisson distribution.

It is worth stressing that the error estimates presented in this work have the same optimal order as other bounds obtained through the Stein–Chen method. In fact, bounds derived using the latter method contain an additional $\log\lambda$ term or, worse, an e^{λ} term for certain ranges of the parameters (see Barbour, Chen and Loh (1992), Barbour and Utev (1999), Barbour and Xia (2000) or Barbour and Chryssaphinou (2001) and the references therein). On the other hand, our bounds do not include such terms and they incorporate a better and more natural factor which we conjecture to be optimal.

The paper is organized as follows. In Section 2, we present some already known, as well as new, auxiliary lemmas which concern probability metrics and coupling techniques. These lemmas will be used for the derivation of our main results. In Sections 3 and 4, we present our main results, that is, bounds concerning Poisson and compound Poisson approximation for sums of *independent* and *k-dependent* random variables, respectively. Finally, in Section 5, in order to illustrate the applicability and effectiveness of our main results, we present a simple example of an application which concerns the distribution of appearances of rare runs in sequences of independent and identically distributed (i.i.d.) trials.

2. Preliminary results

Throughout this paper, the abbreviations c.d.f. and p.d.f. will stand for the cumulative distribution function and probability density function, respectively. In addition, $\mathcal{L}X$ or $\mathcal{L}(X)$ will denote the distribution of a random variable X and the notation $X \sim G$ will imply that X follows the distribution G. Moreover, we shall write $Po(\lambda)$ to denote the Poissson distribution with mean λ and $CP(\lambda, F)$ to denote the compound Poisson distribution with Poisson parameter λ and compounding distribution F. In other words, $CP(\lambda, F)$ is the distribution of the random sum $\sum_{i=1}^{N} Z_i$, where $N \sim Po(\lambda)$ and Z_i are i.i.d. random variables with c.d.f. F which are also independent of N. For two functions f and g, the following standard notation will be used:

$$f(t) \sim g(t) \text{ as } t \to t_0 \text{ if } \lim_{t\to t_0} \frac{f(t)}{g(t)} = 1; \qquad f(t) = \mathrm{O}(g(t)) \text{ if } \frac{f(t)}{g(t)} \text{ is bounded.}$$

Moreover, whenever dependence or independence of some random variables is mentioned, it will be immediately assumed that they are defined on the same probability space. Finally, $\lfloor x \rfloor$ denotes the integer part of x and we will assume that $\sum_{i=a}^{b} x_i = 0$ when $a > b$.

2.1. Probability metrics and smoothness factors

In order to quantify the quality of a distribution approximation, the *total variation distance* and *Zolotarev's ideal metric* of order 2 will be used. Since the results of this paper concern discrete distributions, it suffices to consider only the discrete versions of the aforementioned probability metrics.

The *total variation distance* between the distributions $\mathcal{L}X$ and $\mathcal{L}Y$ of two random variables X and Y is defined by

$$d_{\mathrm{TV}}(\mathcal{L}X, \mathcal{L}Y) := \sup_{A\subseteq\mathbb{Z}} |P(X \in A) - P(Y \in A)| = \frac{1}{2} \sum_{k=-\infty}^{\infty} |P(X = k) - P(Y = k)|,$$

whereas the *total variation distance of order* 2 or *Zolotarev's ideal metric of order* 2 (Zolotarev (1983)) is defined by

$$\zeta_2(\mathcal{L}X, \mathcal{L}Y) = \int_{-\infty}^{\infty} |E(X - t)_+ - E(Y - t)_+| \,\mathrm{d}t = \sum_{k=-\infty}^{\infty} \left| \sum_{u=k}^{\infty} \bigl(F_X(u) - F_Y(u)\bigr) \right|,$$

where, as usual, F_X denotes the c.d.f. of the random variable X. Throughout, whenever a $\zeta_2(\mathcal{L}X, \mathcal{L}Y)$ distance appears, it will be implicitly assumed that X, Y *possess finite first and second moments* and that $E(X) = E(Y)$. For a comprehensive exposition on probability metrics and their properties, the interested reader may consult Rachev (1991) and the references therein.

Next, we denote by $\Delta^k f$ the kth order (backward) difference operator over a function $f : \mathbb{Z} \to \mathbb{R}$, that is, $\Delta f(i) = f(i) - f(i - 1)$ and $\Delta^k = \Delta(\Delta^{k-1} f)$, $k = 1, 2, \ldots$ ($\Delta^0 f = f$). The smoothness factor mentioned in the Introduction emerges from the following lemma. Analogous results concerning random variables with a Lebesgue density have been used in the past in order to obtain Berry–Esseen-type results (see Senatov (1980), Rachev (1991) and the references therein).

Lemma 1. *If X, Y, Z are integer-valued random variables (with finite first and second moments) such that $E(X) = E(Y)$ and Z is independent of X, Y, then*

$$d_{\mathrm{TV}}\big(\mathcal{L}(X+Z), \mathcal{L}(Y+Z)\big) \le \tfrac{1}{2}\|\Delta^2 f_Z\|_1 \zeta_2(\mathcal{L}X, \mathcal{L}Y),$$

where f_Z is the p.d.f. of Z and $\|\Delta^2 f_Z\|_1 := \sum_{z\in\mathbb{Z}} |\Delta^2 f_Z(z)|$.

Proof. For any functions $a, b : \mathbb{Z} \to \mathbb{R}$ and $c, d \in \mathbb{Z}$, we have (second-order Abel summation formula)

$$\sum_{z=c}^{d} b_{z-2}\Delta^2 a_z = \sum_{z=c}^{d} a_z \Delta^2 b_z + b_{d-1}\Delta a_d - a_d \Delta b_d + a_{c-1}\Delta b_{c-1} - b_{c-2}\Delta a_{c-1}. \tag{1}$$

Denote by f_W the p.d.f. of any discrete random variable W. If, for fixed k, we now choose

$$a_z = f_Z(z), \qquad b_z = \sum_{i=-\infty}^{z+1} \big(R_X(k-i) - R_Y(k-i)\big),$$

where $R_X(k-z) = \sum_{i=-\infty}^{z-1} f_X(k-i)$, and then take $c \to -\infty$, $d \to \infty$, identity (1) leads to

$$\sum_{z=-\infty}^{\infty} \sum_{i=-\infty}^{z-1} \big(R_X(k-i) - R_Y(k-i)\big)\Delta^2 f_Z(z) = \sum_{z=-\infty}^{\infty} \big(f_X(k-z) - f_Y(k-z)\big) f_Z(z) \tag{2}$$

since all quantities $b_z, a_z, \Delta a_z, \Delta b_z$ vanish as $z \to \infty$ or $z \to -\infty$. Using (2), we get

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathcal{L}(X+Z), \mathcal{L}(Y+Z)\big) &= \frac{1}{2}\sum_{k=-\infty}^{\infty}\left|\sum_{z=-\infty}^{\infty}\big(f_X(k-z) - f_Y(k-z)\big) f_Z(z)\right| \\ &= \frac{1}{2}\sum_{k=-\infty}^{\infty}\left|\sum_{z=-\infty}^{\infty}\Delta^2 f_Z(z)\sum_{i=-\infty}^{z-1}\big(R_X(k-i) - R_Y(k-i)\big)\right| \\ &\le \frac{1}{2}\sum_{z=-\infty}^{\infty}|\Delta^2 f_Z(z)|\sum_{k=-\infty}^{+\infty}\left|\sum_{i=-\infty}^{z-1}\big(R_X(k-i) - R_Y(k-i)\big)\right|. \end{aligned}$$

Finally, setting $s := k - z + 1$ and $u := k - i$ in the second and third summation above yields

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathcal{L}(X+Z), \mathcal{L}(Y+Z)\big) &\le \frac{1}{2}\sum_{z=-\infty}^{\infty}|\Delta^2 f_Z(z)|\sum_{s=-\infty}^{\infty}\left|\sum_{u=s}^{\infty}\big(R_X(u) - R_Y(u)\big)\right| \\ &= \frac{1}{2}\|\Delta^2 f_Z\|_1 \zeta_2(X, Y). \end{aligned}$$

□

If $Z=0$ and $E(X)=E(Y)$, then a simple consequence of the above result is the inequality

$$d_{\mathrm{TV}}(\mathcal{L}X,\mathcal{L}Y)\le \tfrac{1}{2}\|\Delta^2 f_0\|_1\zeta_2(\mathcal{L}X,\mathcal{L}Y)=2\zeta_2(\mathcal{L}X,\mathcal{L}Y), \tag{3}$$

where $f_0:=f_Z$ when $Z=0$. If Z follows a Poisson distribution with parameter λ, then we can find the explicit value of $\|\Delta^2 f_Z\|_1$ and its asymptotic behavior. In the sequel, we shall write $f_{Po(\lambda)}$ instead of f_Z when $Z\sim Po(\lambda)$. As we will see below, it is convenient to first find the L_∞-norm, $\|\Delta f_{Po(\lambda)}\|_\infty$, and then to investigate its relation with the norm $\|\Delta^2 f_{Po(\lambda)}\|_1$.

Lemma 2. *If $f_{Po(\lambda)}$ denotes the probability distribution function of the Poisson distribution with parameter λ, then*

$$\|\Delta f_{Po(\lambda)}\|_\infty=\sup_{k\in\mathbb{Z}_+}|f_{Po(\lambda)}(k)-f_{Po(\lambda)}(k-1)|=\mathrm{e}^{-\lambda}\frac{\lambda^{k_\lambda}}{k_\lambda!}\left(1-\frac{k_\lambda}{\lambda}\right),$$

where $k_\lambda:=\lfloor\lambda-\sqrt{\lambda+1/4}+1/2\rfloor$ for all $\lambda>0$. In particular, $\|\Delta f_{Po(\lambda)}\|_\infty=\mathrm{e}^{-\lambda}$ for $\lambda\le 2$. Furthermore,

$$\|\Delta f_{Po(\lambda)}\|_\infty\sim\frac{1}{\lambda\sqrt{2\pi\mathrm{e}}}\qquad \textit{as } \lambda\to\infty.$$

Proof. It can be easily verified that $\Delta f_{Po(\lambda)}(k)=\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}(1-\frac{k}{\lambda})$, $k\in\{0,1,2,\ldots\}$, while $\Delta f_{Po(\lambda)}(k)=0$ for $k<0$, and also that

$$\Delta^2 f_{Po(\lambda)}(k)=\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\left(1+\frac{k(k-1)}{\lambda^2}-2\frac{k}{\lambda}\right),\qquad k\in\{0,1,2,\ldots\},$$

while $\Delta^2 f_{Po(\lambda)}(k)=0$ for $k<0$. Define $h:\mathbb{R}_+\to\mathbb{R}$ such that $h(x)=\Delta^2 f_{Po(\lambda)}(x)$ (that is, the extension of $\Delta^2 f_{Po(\lambda)}$ over $\mathbb{R}_+$), where $x!$ now denotes the Gamma function $\Gamma(1+x)$. It is easy to verify that h is positive when $0\le x\le\rho_1$, negative when $\rho_1\le x\le\rho_2$ and positive again when $x\ge\rho_2$, where $\rho_1=\rho_1(\lambda)=\lambda-\sqrt{\lambda+1/4}+1/2$ and $\rho_2=\rho_2(\lambda)=\lambda+\sqrt{\lambda+1/4}+1/2$ are the two roots of the equation $h(x)=0$ ($0<\rho_1<\rho_2$). Since h is an extension of $\Delta^2 f_{Po(\lambda)}$, we deduce that $\Delta^2 f_{Po(\lambda)}(k)\ge 0$ when $0\le k\le\rho_1$, $\Delta^2 f_{Po(\lambda)}(k)\le 0$ when $\rho_1\le k\le\rho_2$ and $\Delta^2 f_{Po(\lambda)}(k)\ge 0$ when $k\ge\rho_2$. This implies that $0=\Delta f_{Po(\lambda)}(-1)\le\Delta f_{Po(\lambda)}(0)\le\cdots\le\Delta f_{Po(\lambda)}(\lfloor\rho_1\rfloor)$, while $\Delta f_{Po(\lambda)}(\lfloor\rho_1\rfloor)\ge\Delta f_{Po(\lambda)}(\lfloor\rho_1\rfloor+1)\ge\cdots\ge\Delta f_{Po(\lambda)}(\lfloor\rho_2\rfloor)$ and $\Delta f_{Po(\lambda)}(\lfloor\rho_2\rfloor)\le\Delta f_{Po(\lambda)}(\lfloor\rho_2\rfloor+1)\le\cdots$. Hence, $|\Delta f_{Po(\lambda)}(k)|$ must be maximized at $\lfloor\rho_1\rfloor$ or $\lfloor\rho_2\rfloor$ (since $\Delta f_{Po(\lambda)}(k)\to 0$ as $k\to\infty$). In order to verify that it is maximized at $k_\lambda=\lfloor\rho_1(\lambda)\rfloor$, we shall prove that $g_1(\lambda)>g_2(\lambda)$ for all $\lambda>0$ where $g_1(\lambda)=\lambda|\Delta f_{Po(\lambda)}(\lfloor\rho_1(\lambda)\rfloor)|$ and $g_2(\lambda)=\lambda|\Delta f_{Po(\lambda)}(\lfloor\rho_2(\lambda)\rfloor)|$, that is,

$$\begin{aligned}g_1(\lambda)&=\lambda\mathrm{e}^{-\lambda}\frac{\lambda^{\lfloor\rho_1(\lambda)\rfloor}}{\lfloor\rho_1(\lambda)\rfloor!}\left(1-\frac{\lfloor\rho_1(\lambda)\rfloor}{\lambda}\right)>g_2(\lambda)\\&=-\lambda\mathrm{e}^{-\lambda}\frac{\lambda^{\lfloor\rho_2(\lambda)\rfloor}}{\lfloor\rho_2(\lambda)\rfloor!}\left(1-\frac{\lfloor\rho_2(\lambda)\rfloor}{\lambda}\right),\qquad \lambda>0.\end{aligned}$$

For every $k \in \{0, 1, \dots\}$, $\varepsilon \in [0, 1)$, we have $\lfloor \rho_1(k + \varepsilon + \sqrt{k+\varepsilon}) \rfloor = \lfloor k + \varepsilon \rfloor = k$. Therefore, $\lfloor \rho_1(\lambda) \rfloor = k$ for every $\lambda \in [k + \sqrt{k}, k + 1 + \sqrt{k+1})$. Hence, in this interval, the function $g_1(\lambda)$ is equal to $\lambda e^{-\lambda} \frac{\lambda^k}{k!}(1 - \frac{k}{\lambda})$, differentiable (except at $k + \sqrt{k}$) and concave, and $g_1'(\lambda) = 0$ at $\lambda = a(k) = k + 1/2 + \sqrt{k + 1/4}$. Moreover, $g_1(\lambda) \to g_1(k + 1 + \sqrt{k+1})$ as $\lambda \to k + 1 + \sqrt{k+1}$ and thus $g_1(\lambda)$ is continuous for every $\lambda > 0$. Therefore, $g_1(\lambda) \geq g_1(k + \sqrt{k})$ for every $\lambda \in [a(k-1), a(k)]$, $k \in \{1, 2, \dots\}$. Using the upper bound of Stirling's approximation ($k! \leq k^k e^{-k} \sqrt{2\pi k} e^{1/(12k)}$) and the elementary inequality $\log(1+x) > x - x^2/2 + x^3/3 - x^4/4$, $x > 0$, we get

$$g_1\big(k + \sqrt{k}\big) = e^{-(k+\sqrt{k})} \frac{(k + \sqrt{k})^k}{k!} \sqrt{k} \geq \frac{e^{-\sqrt{k} - 1/(12k)}}{\sqrt{2\pi}} e^{k \log(1 + 1/\sqrt{k})} > \frac{e^{1/(3\sqrt{k}) - 1/(3k)}}{\sqrt{2\pi e}} \geq \frac{1}{\sqrt{2\pi e}}$$

for every $k \geq 1$. Therefore, $g_1(\lambda) > \frac{1}{\sqrt{2\pi e}}$ for every $\lambda \in \bigcup_{k \geq 1} [a(k-1), a(k)] = [1, \infty)$.

Similarly, for every $k \in \{1, 2, \dots\}$, $\varepsilon \in [0, 1)$, we have $\lfloor \rho_2(k + \varepsilon - \sqrt{k+\varepsilon}) \rfloor = \lfloor k + \varepsilon \rfloor = k$. Therefore, $\lfloor \rho_2(\lambda) \rfloor = k$ for every $\lambda \in [k - \sqrt{k}, k + 1 - \sqrt{k+1})$. Moreover, in this interval, the function $g_2(\lambda)$ is equal to $\lambda e^{-\lambda} \frac{\lambda^k}{k!}(\frac{k}{\lambda} - 1)$, differentiable (except at $k - \sqrt{k}$) and concave, and $g_2'(\lambda) = 0$ at $\lambda = k + 1/2 - \sqrt{k + 1/4}$ ($g_2(\lambda)$ is also continuous for every $\lambda > 0$). Therefore, $g_2(\lambda) \leq g_2(k + 1/2 - \sqrt{k + 1/4})$ for every $\lambda \in [k - \sqrt{k}, k + 1 - \sqrt{k+1}]$, $k \in \{1, 2, \dots\}$. Using the lower bound of Stirling's approximation ($k! \geq k^k e^{-k} \sqrt{2\pi k}$) and the elementary inequality $\log(1+x) < x - x^2/2$, $x \in (-1, 0)$, we get, for $k \geq 1$,

$$\begin{aligned} g_2\big(k + 1/2 - \sqrt{k + 1/4}\big) &\leq \frac{\sqrt{k + 1/4} - 1/2}{\sqrt{2\pi k}} e^{-1/2 + \sqrt{k+1/4} + k \log(1 + (1/2 - \sqrt{k+1/4})/k)} \\ &< \frac{\sqrt{k + 1/4} - 1/2}{\sqrt{k}} \frac{e^{(\sqrt{k+1/4} - 1/2)/2k}}{\sqrt{2\pi e}} < \frac{1}{\sqrt{2\pi e}}. \end{aligned}$$

Therefore, $g_2(\lambda) < \frac{1}{\sqrt{2\pi e}}$ for every $\lambda \geq 0$. Hence, $g_2(\lambda) < \frac{1}{\sqrt{2\pi e}} < g_1(\lambda)$ for every $\lambda \geq 1$. It now remains to show that $g_2(\lambda) < g_1(\lambda)$ for every $0 < \lambda < 1$. This is easily verified since $g_1(\lambda) = \lambda e^{-\lambda}$ for $\lambda < 2$, while $g_2(\lambda) = e^{-\lambda} \lambda (1 - \lambda)$ for $\lambda \in [0, 2 - \sqrt{2})$ and $g_2(\lambda) = e^{-\lambda} \lambda^2 (1 - \frac{\lambda}{2})$ for $\lambda \in [2 - \sqrt{2}, 3 - \sqrt{3})$.

Finally, from $(1 + y)^{k_\lambda} = e^{k_\lambda \log(1+y)} = e^{k_\lambda (y - y^2/2 + o(y^2))}$ with $y = (\lambda - k_\lambda)/k_\lambda$, we get

$$e^{k_\lambda - \lambda} (\lambda / k_\lambda)^{k_\lambda} \to e^{-1/2} \qquad \text{as } \lambda \to \infty.$$

From this fact and Stirling's formula, we get, as $\lambda \to \infty$, that

$$\lambda \Delta f_{Po(\lambda)}(k_\lambda) = \lambda e^{-\lambda} \frac{\lambda^{k_\lambda}}{k_\lambda !} \left(1 - \frac{k_\lambda}{\lambda}\right) \sim e^{k_\lambda - \lambda} \left(\frac{\lambda}{k_\lambda}\right)^{k_\lambda} \frac{\lambda - k_\lambda}{\sqrt{2\pi k_\lambda}} \to \frac{1}{\sqrt{2\pi e}}. \qquad \square$$

In the next lemma, we find the explicit value of $\|\Delta^2 f_{Po(\lambda)}\|_1$ and a convenient upper bound in terms of $\|\Delta f_{Po(\lambda)}\|_\infty$.

Lemma 3. *If $f_{Po(\lambda)}$ denotes the p.d.f. of the Poisson distribution with parameter λ, then*

$$\left\|\Delta^2 f_{Po(\lambda)}\right\|_1 = \sum_{z=0}^{+\infty}\left|\Delta^2 f_{Po(\lambda)}(z)\right| = 2\mathrm{e}^{-\lambda}\left(\frac{\lambda^{k_\lambda-1}(\lambda-k_\lambda)}{k_\lambda!} - \frac{\lambda^{u_\lambda-1}(\lambda-u_\lambda)}{u_\lambda!}\right),$$

where $k_\lambda := \lfloor\lambda - \sqrt{\lambda+1/4}+1/2\rfloor$ and $u_\lambda := \lfloor\lambda + \sqrt{\lambda+1/4}+1/2\rfloor$. Moreover,

$$\left\|\Delta^2 f_{Po(\lambda)}\right\|_1 \le 4\left\|\Delta f_{Po(\lambda)}\right\|_\infty \quad \textit{and} \quad \left\|\Delta^2 f_{Po(\lambda)}\right\|_1 \sim \frac{4}{\lambda\sqrt{2\pi\mathrm{e}}} \qquad \textit{as } \lambda\to\infty.$$

Proof. For convenience, we set $k_\lambda := \lfloor\rho_1\rfloor$ and $u_\lambda := \lfloor\rho_2\rfloor$, where $\rho_1 := \lambda - \sqrt{\lambda+1/4}+1/2$ and $\rho_2 := \lambda + \sqrt{\lambda+1/4}+1/2$, and $g(z) := \Delta f_{Po(\lambda)}(z)$. In the proof of Lemma 2, we have seen that $0 = g(-1) \le g(0) \le \cdots \le g(k_\lambda)$, while $g(k_\lambda) \ge g(k_\lambda+1) \ge \cdots \ge g(u_\lambda)$ and $g(u_\lambda) \le g(u_\lambda+1) \le \cdots$. We then have

$$\begin{aligned}\left\|\Delta^2 f_{Po(\lambda)}\right\|_1 &= \sum_{z=0}^{+\infty}|\Delta g(z)| = \sum_{z=0}^{k_\lambda}\Delta g(z) - \sum_{z=k_\lambda+1}^{u_\lambda}\Delta g(z) + \sum_{z=u_\lambda+1}^{+\infty}\Delta g(z)\\ &= \big(g(k_\lambda)-g(-1)\big) - \big(g(u_\lambda)-g(k_\lambda)\big) + \big(0-g(u_\lambda)\big)\\ &= 2\big(g(k_\lambda)-g(u_\lambda)\big).\end{aligned}$$

From the proof of Lemma 2, we also get that

$$g(k_\lambda) = \Delta f_{Po(\lambda)}(k_\lambda) = \max_{z\in\mathbb{Z}_+}\Delta f_{Po(\lambda)}(z) = \left\|\Delta f_{Po(\lambda)}\right\|_\infty; \qquad g(u_\lambda) = \min_{z\in\mathbb{Z}_+}\Delta f_{Po(\lambda)}(z) < 0$$

and $g(k_\lambda) > -g(u_\lambda)$. Therefore, we obtain that $\|\Delta^2 f_{Po(\lambda)}\|_1 \le 4g(k_\lambda) = 4\|\Delta f_{Po(\lambda)}\|_\infty$. The last asymptotic result follows from the fact that $\Delta f_{Po(\lambda)}(u_\lambda) \sim -\lambda^{-1}(2\pi\mathrm{e})^{-1/2}$, which can be proven in exactly the same way as $\Delta f_{Po(\lambda)}(k_\lambda) \sim \lambda^{-1}(2\pi\mathrm{e})^{-1/2}$ was proven in Lemma 2. □

A crude but simple upper bound is $\|\Delta^2 f_{Po(\lambda)}\|_1 \le 4\frac{1-\mathrm{e}^{-3\lambda}}{3\lambda} \le 4(1\wedge\frac{1}{3\lambda})$ for all $\lambda > 0$, whereas $\|\Delta f_{Po(\lambda)}\|_\infty \le 1/(3\lambda)$ for $\lambda \ge 2$.

Remark 1. For distributions other than Poisson, it is not always easy to derive an analytic expression for $\|\Delta f\|_\infty$ or $\|\Delta^2 f\|_1$. Nevertheless, it is always feasible to compute the numeric value of these norms employing numerical or symbolic mathematics software packages (for example, Mathematica, Maple or MATLAB).

An approximate expression for these norms can be easily derived if we assume that the distribution corresponding to the p.d.f. f can be approximated by a normal distribution $N(\mu,\sigma^2)$, for example, due to CLT. In this case, we expect that $\|\Delta f\|_\infty$ and $\|\Delta^2 f\|_1$ would be close to $\|f^{(1)}_{N(\mu,\sigma^2)}\|_\infty$ and $\|f^{(2)}_{N(\mu,\sigma^2)}\|_1$, respectively, where $f^{(k)}_{N(\mu,\sigma^2)}$ denotes the kth order derivative of the p.d.f. of $N(\mu,\sigma^2)$. It is not difficult to verify that, for the normal distribution, we have

$$\|f^{(1)}_{N(\mu,\sigma^2)}\|_\infty = \sup_{x\in\mathbb{R}}\left|f^{(1)}_{N(\mu,\sigma^2)}(x)\right| = \frac{1}{\sigma^2\sqrt{2\pi e}},$$

$$\|f^{(2)}_{N(\mu,\sigma^2)}\|_1 = \int_{-\infty}^{+\infty}\left|f^{(2)}_{N(\mu,\sigma^2)}(x)\right|\mathrm{d}x = 4\|f^{(1)}_{N(\mu,\sigma^2)}\|_\infty.$$

Hence, for distributions similar to the normal with variance σ^2, we expect $\|\Delta^2 f\|_1$ to be nearly equal to $4\sigma^{-2}(2\pi e)^{-1/2}$. This approximation works for the Poisson distribution (as seen in Proposition 3 above) since, for large λ, it is close to a normal distribution with $\sigma^2=\lambda$. According to the above, concerning the compound Poisson distribution, if $CP(\lambda,F)\approx N(\lambda E(W),\lambda E(W^2))$ (with $W\sim F$) then we can expect that, for large λ,

$$\|\Delta^2 f_{CP(\lambda,F)}\|_1 \approx \frac{4}{\lambda E(W^2)\sqrt{2\pi e}}. \tag{4}$$

It is worth stressing that (4) is valid provided the compounding distribution F is such that $CP(\lambda,F)$ is approximately normal. There exist counterexamples showing that (4) is not always valid; see Example 1.3 or 1.4 of Barbour and Utev (1999). Specifically, the $CP(\lambda,F)$ described there cannot be approximated by a normal distribution and, moreover, it can be verified that the corresponding $\|\Delta^2 f_{CP(\lambda,F)}\|_1$ does not decrease as λ increases. Note that Barbour and Utev (1999) use these counterexamples to show that, even for independent X_i's (with $p_i=P(X_i\neq 0)$, $\lambda=\Sigma p_i$), we cannot always prove that $d_{\mathrm{TV}}(\mathcal{L}(\Sigma X_i), CP(\lambda,F)) = \mathrm{O}(\lambda^{-1}\Sigma p_i^2)$ and sometimes (depending on F) the order $\mathrm{O}(\Sigma p_i^2)$ is optimal. Theorem 9 below implies that this d_{TV} is of order $\mathrm{O}(\lambda^{-1}\Sigma p_i^2)$ whenever F is such that $\|\Delta^2 f_{CP(\lambda,F)}\|_1=\mathrm{O}(\lambda^{-1})$ (see also Remark 2 in Section 3).

2.2. Coupling techniques

A coupling of two random vectors $\mathbf{X},\mathbf{Y}\in\mathbb{R}^k$ (to be more exact, of their distributions $\mathcal{L}\mathbf{X}$, $\mathcal{L}\mathbf{Y}$) is considered to be any random vector $(\mathbf{X}',\mathbf{Y}')$ defined over a probability space $(\Omega,\mathfrak{F},P)$ and taking values in a measurable space $(\mathbb{R}^{2k},\mathcal{B}(\mathbb{R}^{2k}))$ with the same marginal distributions as $\mathbf{X},\mathbf{Y}$, that is, $\mathcal{L}\mathbf{X}=\mathcal{L}\mathbf{X}'$and $\mathcal{L}\mathbf{Y}=\mathcal{L}\mathbf{Y}'$. Loosely speaking, a coupling of $\mathbf{X},\mathbf{Y}$ is any "definition" of $\mathbf{X},\mathbf{Y}$ in the same probability space. This definition of coupling can be generalized for n random vectors in an obvious way. A well-known result concerning the d_{TV} is the so-called (*basic*) *coupling inequality*,

$$d_{\mathrm{TV}}(\mathcal{L}\mathbf{X},\mathcal{L}\mathbf{Y}) \le P(\mathbf{X}'\neq\mathbf{Y}'),$$

which is valid for any coupling $(\mathbf{X}',\mathbf{Y}')$ of two random vectors $\mathbf{X},\mathbf{Y}$. It can be proven that we can always construct a coupling $(\mathbf{X}',\mathbf{Y}')$ of $(\mathbf{X},\mathbf{Y})$ such that $d_{\mathrm{TV}}(\mathcal{L}\mathbf{X},\mathcal{L}\mathbf{Y})=P(\mathbf{X}'\neq\mathbf{Y}')$ (for example, see Lindvall (1992), page 18). Such a coupling is called a *maximal coupling* or γ-*coupling* of $\mathbf{X},\mathbf{Y}$. All of the above could be expressed equivalently for probability measures as follows: if P_1, P_2 are two probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, then any probability measure $\hat{P}$ on $(\mathbb{R}^{2k}, \mathcal{B}(\mathbb{R}^{2k}))$ with $\hat{P}(A \times \mathbb{R}^k) = P_1(A)$, $\hat{P}(\mathbb{R}^k \times A) = P_2(A)$ for every $A \in \mathcal{B}(\mathbb{R}^k)$ is called a *coupling* of P_1, P_2. Moreover, it can be proven that there exists a coupling $\hat{P}_\gamma$ of P_1, P_2, called a *maximal coupling* or γ*-coupling*, such that

$$d_{\mathrm{TV}}(P_1, P_2) = 1 - \hat{P}_\gamma\bigl(\{(\mathbf{x}, \mathbf{x}), \mathbf{x} \in \mathbb{R}^k\}\bigr). \tag{5}$$

Obviously, all of the above can be adapted in the obvious way for random vectors taking values in $\mathbb{Z}^k$ and to multivariate distributions over the probability space $(\mathbb{Z}^k, 2^{\mathbb{Z}^k})$.

The following lemmas will play a crucial role for the establishment of our main results. The first inequality of the following lemma is Corollary 4 in Boutsikas (2006). The second inequality of the following lemma is a direct application of Lemma 3 in Boutsikas (2006) with $(\Xi_1', \Xi_2', \Psi_1', \Psi_2') = (\mathbf{Z} + \mathbf{X}, \mathbf{Z} + \mathbf{Y}, \mathbf{W} + \mathbf{X}, \mathbf{W} + \mathbf{Y})$.

Lemma 4. *For any random vectors* $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^k$ *and* $\mathbf{Z}, \mathbf{W} \in \mathbb{R}^r$ *defined on the same probability space, we have that*

(a) $|d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}, \mathbf{X}), \mathcal{L}(\mathbf{Z}, \mathbf{Y})) - d_{\mathrm{TV}}(\mathcal{L}(\mathbf{W}, \mathbf{X}), \mathcal{L}(\mathbf{W}, \mathbf{Y}))| \le 2P(\mathbf{X} \ne \mathbf{Y}, \mathbf{Z} \ne \mathbf{W})$;
(b) $|d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z} + \mathbf{X}), \mathcal{L}(\mathbf{Z} + \mathbf{Y})) - d_{\mathrm{TV}}(\mathcal{L}(\mathbf{W} + \mathbf{X}), \mathcal{L}(\mathbf{W} + \mathbf{Y}))| \le 2P(\mathbf{X} \ne \mathbf{Y}, \mathbf{Z} \ne \mathbf{W})$.

The next inequality follows from the above lemma. It is remarkable that almost the same inequality can be found in Rachev (1991), page 274, and has been applied to derive Berry–Esseen-type results. We present an entirely different proof using maximal couplings.

Lemma 5. *If the random vectors* $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^k$ *are independent of* $\mathbf{Z}, \mathbf{W} \in \mathbb{R}^r$, *then*

$$\begin{aligned}&\bigl|d_{\mathrm{TV}}\bigl(\mathcal{L}(\mathbf{Z} + \mathbf{X}), \mathcal{L}(\mathbf{Z} + \mathbf{Y})\bigr) - d_{\mathrm{TV}}\bigl(\mathcal{L}(\mathbf{W} + \mathbf{X}), \mathcal{L}(\mathbf{W} + \mathbf{Y})\bigr)\bigr|\\ &\quad\le 2d_{\mathrm{TV}}(\mathcal{L}\mathbf{X}, \mathcal{L}\mathbf{Y})d_{\mathrm{TV}}(\mathcal{L}\mathbf{Z}, \mathcal{L}\mathbf{W}).\end{aligned}$$

Proof. Let $(\mathbf{X}^*, \mathbf{Y}^*)$ be a maximal coupling of $\mathcal{L}\mathbf{X}, \mathcal{L}\mathbf{Y}$ and let $(\mathbf{Z}^*, \mathbf{W}^*)$ be a maximal coupling of $\mathcal{L}\mathbf{Z}$, $\mathcal{L}\mathbf{W}$. Next, let $((\mathbf{X}', \mathbf{Y}'), (\mathbf{Z}', \mathbf{W}'))$ be an independent coupling of $\mathcal{L}(\mathbf{X}^*, \mathbf{Y}^*)$, $\mathcal{L}(\mathbf{Z}^*, \mathbf{W}^*)$ (that is, $(\mathbf{X}', \mathbf{Y}')$ is independent of $(\mathbf{Z}', \mathbf{W}')$ and $\mathcal{L}(\mathbf{X}', \mathbf{Y}') = \mathcal{L}(\mathbf{X}^*, \mathbf{Y}^*)$, $\mathcal{L}(\mathbf{Z}', \mathbf{W}') = \mathcal{L}(\mathbf{Z}^*, \mathbf{W}^*)$). Applying Lemma 4, we get

$$\begin{aligned}&\bigl|d_{\mathrm{TV}}\bigl(\mathcal{L}(\mathbf{Z}' + \mathbf{X}'), \mathcal{L}(\mathbf{Z}' + \mathbf{Y}')\bigr) - d_{\mathrm{TV}}\bigl(\mathcal{L}(\mathbf{W}' + \mathbf{X}'), \mathcal{L}(\mathbf{W}' + \mathbf{Y}')\bigr)\bigr|\\ &\quad\le 2P(\mathbf{X}' \ne \mathbf{Y}', \mathbf{Z}' \ne \mathbf{W}') = 2P(\mathbf{X}' \ne \mathbf{Y}')P(\mathbf{Z}' \ne \mathbf{W}')\\ &\quad= 2P(\mathbf{X}^* \ne \mathbf{Y}^*)P(\mathbf{Z}^* \ne \mathbf{W}^*) = 2d_{\mathrm{TV}}(\mathcal{L}\mathbf{X}, \mathcal{L}\mathbf{Y})d_{\mathrm{TV}}(\mathcal{L}\mathbf{Z}, \mathcal{L}\mathbf{W}).\end{aligned}$$

The obvious fact that $\mathcal{L}(\mathbf{Z}' + \mathbf{X}') = \mathcal{L}(\mathbf{Z} + \mathbf{X})$, $\mathcal{L}(\mathbf{Z}' + \mathbf{Y}') = \mathcal{L}(\mathbf{Z} + \mathbf{Y})$, $\mathcal{L}(\mathbf{W}' + \mathbf{X}') = \mathcal{L}(\mathbf{W} + \mathbf{X})$ and $\mathcal{L}(\mathbf{W}' + \mathbf{Y}') = \mathcal{L}(\mathbf{W} + \mathbf{Y})$ completes the proof. □

A direct application of the previous result leads to the following inequality which is valid for any random variables $X, Y \in \mathbb{R}$ independent of another random variable $W \in \mathbb{R}$. Specifically, if we simply set $Z=0$ in Lemma 5 and exploit the fact that $d_{\mathrm{TV}}(\mathcal{L}0,\mathcal{L}W)=P(W\neq 0)$, we derive

$$d_{\mathrm{TV}}(\mathcal{L}X,\mathcal{L}Y)\leq 2d_{\mathrm{TV}}\big(\mathcal{L}X,\mathcal{L}Y\big)P(W\neq 0)+d_{\mathrm{TV}}\big(\mathcal{L}(X+W),\mathcal{L}(Y+W)\big)$$

which, for $P(W\neq 0)<1/2$, implies that

$$d_{\mathrm{TV}}(\mathcal{L}X,\mathcal{L}Y)\leq \big(1-2P(W\neq 0)\big)^{-1}d_{\mathrm{TV}}\big(\mathcal{L}(X+W),\mathcal{L}(Y+W)\big). \tag{6}$$

The next lemma can be considered as a coupling inequality concerning ζ_2, analogous to Lemma 4.

Lemma 6. *If X,Y,Z,W are real-valued, non-negative random variables defined on the same probability space with finite second moments and $E(X)=E(Y)$, then*

$$\big|\zeta_2\big(\mathcal{L}(X+Z),\mathcal{L}(Y+Z)\big)-\zeta_2\big(\mathcal{L}(X+W),\mathcal{L}(Y+W)\big)\big|\leq E|(X-Y)(Z-W)|. \tag{7}$$

Proof. The distances ζ_2 appearing in (7) are well defined since the random variables $X+Z$, $Y+Z$, $X+W$ and $Y+W$ have finite second moments due to Minkowski's inequality and $E(X+Z)=E(Y+Z)$, $E(X+W)=E(Y+W)$. Set $\mathbf{1}_{[a\leq b]}:=1$ if $a\leq b$ and $\mathbf{1}_{[a\leq b]}:=0$ otherwise. As usual, F_V denotes the c.d.f. of a random variable V. Recall that, for $X,Y\in\mathbb{R}_+$ with $E(X)=E(Y)$, we have

$$\zeta_2(\mathcal{L}X,\mathcal{L}Y)=\int_0^\infty |E(X-s)_+-E(Y-s)_+|\,\mathrm{d}s=\int_0^\infty\left|\int_s^\infty \big(F_X(x)-F_Y(x)\big)\,\mathrm{d}x\right|\mathrm{d}s.$$

Denoting by d the absolute difference in the left-hand side of (7), we have

$$\begin{aligned} d&=\left|\int_0^\infty\left|\int_s^\infty\big(F_{X+Z}(x)-F_{Y+Z}(x)\big)\,\mathrm{d}x\right|\mathrm{d}s-\int_0^\infty\left|\int_s^\infty\big(F_{X+W}(x)-F_{Y+W}(x)\big)\,\mathrm{d}x\right|\mathrm{d}s\right|\\ &\leq\int_0^\infty\left|\left|\int_s^\infty\big(F_{X+Z}(x)-F_{Y+Z}(x)\big)\,\mathrm{d}x\right|-\left|\int_s^\infty\big(F_{X+W}(x)-F_{Y+W}(x)\big)\,\mathrm{d}x\right|\right|\mathrm{d}s. \end{aligned}$$

Using the inequality $||a|-|b||\leq|a-b|$, $a,b\in\mathbb{R}$, we get

$$\begin{aligned} d&\leq\int_0^\infty\left|\int_s^\infty\big(F_{X+Z}(x)-F_{Y+Z}(x)\big)\,\mathrm{d}x-\int_s^\infty\big(F_{X+W}(x)-F_{Y+W}(x)\big)\,\mathrm{d}x\right|\mathrm{d}s\\ &=\int_0^\infty|E(C_s)|\,\mathrm{d}s\leq E\left(\int_0^\infty|C_s|\,\mathrm{d}s\right), \end{aligned}$$

where

$$C_s=\int_s^\infty\big(\mathbf{1}_{[X+Z\leq x]}-\mathbf{1}_{[Y+Z\leq x]}\big)\,\mathrm{d}x-\int_s^\infty\big(\mathbf{1}_{[X+W\leq x]}-\mathbf{1}_{[Y+W\leq x]}\big)\,\mathrm{d}x.$$

Now, if $Z \geq W$, it can be verified that $C_s \geq 0$ for all $s > 0$ and, therefore,

$$\begin{aligned}\int_0^\infty |C_s|\,\mathrm{d}s &= \int_0^\infty x\big(\mathbf{1}_{[X+Z\leq x]} - \mathbf{1}_{[Y+Z\leq x]}\big)\,\mathrm{d}x - \int_0^\infty x\big(\mathbf{1}_{[X+W\leq x]} - \mathbf{1}_{[Y+W\leq x]}\big)\,\mathrm{d}x \\ &= \frac{1}{2}\big(|(X+Z)^2 - (Y+Z)^2| - |(X+W)^2 - (Y+W)^2|\big) \\ &= \frac{1}{2}\big(|(X-Y)(X+Y+2Z)| - |(X-Y)(X+Y+2W)|\big) \\ &= |X-Y|(Z-W).\end{aligned}$$

On the other hand, if $Z \leq W$, then $C_s \leq 0$ for all $s > 0$ and we similarly derive that $\int_0^\infty |C_s|\,\mathrm{d}s = |X-Y|(W-Z)$. Hence, $\int_0^\infty |C_s|\,\mathrm{d}s = |(X-Y)(W-Z)|$ and the proof is completed. □

A direct corollary of Lemma 6 is the following result which will be proven useful when dealing with k-dependent sequences of random variables.

Corollary 7. *If the random variables $X_1, X_2, \ldots, X_i \in \mathbb{R}_+$ are k-dependent with $E(X_i^2) < \infty$ and $l \leq i-k+1$, then*

$$\zeta_2\left(\mathcal{L}\sum_{j=l}^{i} X_j, \mathcal{L}\left(\sum_{j=l}^{i-1} X_j + X_i^{\perp}\right)\right) \leq \sum_{j=i-k+1}^{i-1}\big(E(X_iX_j) + E(X_i)E(X_j)\big),$$

where $X_i^{\perp}$ is a random variable independent of all X_j, $j = 1, 2, \ldots, i$, with $\mathcal{L}X_i = \mathcal{L}X_i^{\perp}$.

Proof. Set $X_{a,b} := \sum_{j=a}^{b} X_j$. Applying Lemma 6 with $X = X_i, Z = X_{l,i-1}, Y = X_i^{\perp}, W = X_{l,i-k}$, we obtain

$$\begin{aligned}&\big|\zeta_2\big(\mathcal{L}X_{l,i}, \mathcal{L}(X_{l,i-1} + X_i^{\perp})\big) - \zeta_2\big(\mathcal{L}(X_{l,i-k} + X_i), \mathcal{L}(X_{l,i-k} + X_i^{\perp})\big)\big| \\ &\quad\leq E|(X_i - X_i^{\perp})(X_{l,i-1} - X_{l,i-k})| = E|(X_i - X_i^{\perp})(X_{i-k+1,i-1})| \\ &\quad\leq \sum_{j=i-k+1}^{i-1}\big(E(X_iX_j) + E(X_i)E(X_j)\big).\end{aligned}$$

Since $X_{l,i-k}$ and X_i are independent, $X_{l,i-k}$ and $X_i^{\perp}$ are independent, and $\mathcal{L}X_i = \mathcal{L}X_i^{\perp}$, we conclude that $\mathcal{L}(X_{l,i-k} + X_i) = \mathcal{L}(X_{l,i-k} + X_i^{\perp})$ and hence we obtain the desired inequality. □

As will be seen in the next section, Lemmas 1 and 5 are sufficient for proving compound Poisson approximation results for sums of *independent* random variables incorporating a smoothness factor. In the case of sums of dependent random variables, though, the following, additional, lemma is needed. The question addressed here is the following: given a random variable X and a random vector $\mathbf{Z}$, can we construct (on the same probability space as $X, \mathbf{Z}$) another random variable Y with a given p.d.f. f such that Y is independent of $\mathbf{Z}$ and $(X, \mathbf{Z})$, $(Y, \mathbf{Z})$ are maximally coupled? In this situation, we could loosely say that we wish to construct a random variable Y (with a given distribution) that resembles X as far as possible, *while remaining independent of* $\mathbf{Z}$. Again, it suffices to restrict our analysis to the discrete case.

Lemma 8. *Let $X \in \mathbb{Z}, \mathbf{Z} \in \mathbb{Z}^k$ be a random variable and a random vector, respectively (defined on the same probability space) and let $f : \mathbb{Z} \to \mathbb{R}_+$ be some given discrete p.d.f. Denote by U a random variable independent of $X, \mathbf{Z}$ that follows the uniform distribution on $(0, 1)$. Then,*

(a) *there exists a function $g : \mathbb{R}^{2+k} \to \mathbb{R}$ such that the random variable $Y = g(U, X, \mathbf{Z})$ has p.d.f. f, Y is independent of $\mathbf{Z}$ and*

$$d_{\mathrm{TV}}(\mathcal{L}(X, \mathbf{Z}), \mathcal{L}(Y, \mathbf{Z})) = P\big((X, \mathbf{Z}) \neq (Y, \mathbf{Z})\big) = P(X \neq Y),$$

in other words, $(X, \mathbf{Z}), (Y, \mathbf{Z})$ are maximally coupled;

(b) *there exists a function $g' : \mathbb{R}^2 \to \mathbb{R}$ such that the random variable $Y' = g'(U, X)$ has p.d.f. f and (X, Y') are maximally coupled, that is, $d_{\mathrm{TV}}(\mathcal{L}X, \mathcal{L}Y') = P(X \neq Y')$.*

Proof. (a) Here, we develop a constructive proof. Denote by $(\Omega, \mathcal{A}, P)$ the probability space on which $X, \mathbf{Z}, U$ are defined and let $f_{X|\mathbf{Z}}(\cdot|\mathbf{z}) = f_{X,\mathbf{Z}}(\cdot, \mathbf{z})/f_{\mathbf{Z}}(\mathbf{z})$ be the conditional p.d.f. of X given $\mathbf{Z} = \mathbf{z}$. Consider the probability measures $P_1^{\mathbf{z}}, P_2$ on the measurable space $(\mathbb{Z}, 2^{\mathbb{Z}})$ generated by $f_{X|\mathbf{Z}}(\cdot|\mathbf{z})$ and f, respectively. According to (5), there exists a maximal coupling of $P_1^{\mathbf{z}}, P_2$. Denote by $h_{\mathbf{z}} : \mathbb{Z}^2 \to \mathbb{R}_+$ the joint p.d.f. corresponding to this maximal coupling. It follows that $\sum_{x \in \mathbb{Z}} h_{\mathbf{z}}(x, y) = f(y)$, $\sum_{y \in \mathbb{Z}} h_{\mathbf{z}}(x, y) = f_{X|\mathbf{Z}}(x|\mathbf{z})$ and

$$d_{\mathrm{TV}}\big(\mathcal{L}(X|\mathbf{Z} = \mathbf{z}), P_2\big) = d_{\mathrm{TV}}(P_1^{\mathbf{z}}, P_2) = 1 - \sum_{x \in \mathbb{Z}} h_{\mathbf{z}}(x, x).$$

We now construct Y as follows. For every $x \in \mathbb{Z}, \mathbf{z} \in \mathbb{Z}^k$, consider the c.d.f.

$$H_{x,\mathbf{z}}(y) := \sum_{i \leq \lfloor y \rfloor} \frac{h_{\mathbf{z}}(x, i)}{f_{X|\mathbf{Z}}(x|\mathbf{z})}, \qquad y \in \mathbb{R},$$

and set $Y(\omega) := H^{-1}_{X(\omega),\mathbf{Z}(\omega)}(U(\omega))$, $\omega \in \Omega$, where $H^{-1}_{x,\mathbf{z}}(y)$ denotes the generalized inverse of $H_{x,\mathbf{z}}(y)$, that is, $H^{-1}_{x,\mathbf{z}}(y) = \inf\{w : H_{x,\mathbf{z}}(w) \geq y\}$. The function $f_{X,Y,\mathbf{Z}}(x, y, \mathbf{z}) := h_{\mathbf{z}}(x, y) f_{\mathbf{Z}}(\mathbf{z})$ is a multivariate discrete p.d.f. and it can be verified that Y and $(X, Y, \mathbf{Z})$ have p.d.f. f and $f_{X,Y,\mathbf{Z}}$, respectively. Indeed,

$$\begin{aligned} P(X = x, Y \leq y, \mathbf{Z} = \mathbf{z}) &= P\big(X = x, H^{-1}_{x,\mathbf{z}}(U) \leq y, \mathbf{Z} = \mathbf{z}\big) \\ &= H_{x,\mathbf{z}}(y) P(X = x, \mathbf{Z} = \mathbf{z}) \end{aligned}$$

and thus, for all x, y, z,

$$\begin{aligned} P(X = x, Y = y, \mathbf{Z} = \mathbf{z}) &= \frac{h_{\mathbf{z}}(x, y)}{f_{X|\mathbf{Z}}(x|\mathbf{z})} P(X = x, \mathbf{Z} = \mathbf{z}) \\ &= h_{\mathbf{z}}(x, y) f_{\mathbf{Z}}(\mathbf{z}) = f_{X,Y,\mathbf{Z}}(x, y, \mathbf{z}). \end{aligned}$$

Also, note that, for all $x, \mathbf{z}$,

$$P(Y=y,\mathbf{Z}=\mathbf{z})=\sum_{x\in\mathbb{Z}} f_{X,Y,\mathbf{Z}}(x,y,\mathbf{z})=\sum_{x\in\mathbb{Z}} h_{\mathbf{z}}(x,y)f_{\mathbf{Z}}(\mathbf{z})=f(y)f_{\mathbf{Z}}(\mathbf{z}),$$

which implies that Y is independent of $\mathbf{Z}$. Furthermore, we derive that, for all $\mathbf{z}$,

$$P(X\neq Y|\mathbf{Z}=\mathbf{z})=1-\sum_{x\in\mathbb{Z}} h_{\mathbf{z}}(x,x)=d_{\mathrm{TV}}\big(\mathcal{L}(X|\mathbf{Z}=\mathbf{z}),\mathcal{L}Y\big)$$

and, therefore,

$$\begin{aligned}
P(X\neq Y)&=\sum_{\mathbf{z}\in\mathbb{Z}^k} P(X\neq Y|\mathbf{Z}=\mathbf{z})f_{\mathbf{Z}}(\mathbf{z})\\
&=\sum_{\mathbf{z}\in\mathbb{Z}^k} d_{\mathrm{TV}}\big(\mathcal{L}(X|\mathbf{Z}=\mathbf{z}),\mathcal{L}Y\big)f_{\mathbf{Z}}(\mathbf{z})\\
&=\sum_{\mathbf{z}\in\mathbb{Z}^k}\frac{1}{2}\sum_{w\in\mathbb{Z}}|P(X=w|\mathbf{Z}=\mathbf{z})-P(Y=w)|f_{\mathbf{Z}}(\mathbf{z})\\
&=\frac{1}{2}\sum_{\mathbf{z}\in\mathbb{Z}^k}\sum_{w\in\mathbb{Z}}|P(X=w,\mathbf{Z}=\mathbf{z})-P(Y=w)f_{\mathbf{Z}}(\mathbf{z})|\\
&=d_{\mathrm{TV}}(\mathcal{L}(X,\mathbf{Z}),\mathcal{L}(Y,\mathbf{Z})).
\end{aligned}$$

(b) This readily follows from part (a) of the lemma by choosing $\mathbf{Z}=\mathbf{0}$. □

3. Compound Poisson approximation for sums of independent random variables

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables which take values in $\mathbb{Z}_+$. We are now ready to exploit the results of the previous section (specifically Lemmas 1 and 5) to derive a simple and, in most cases, sharp upper bound for the total variation distance between the distribution of the sum $\sum_{i=1}^n X_i$ and an appropriate compound Poisson distribution. Before we present this bound, we recall that, (see Boutsikas and Vaggelatou (2002))

$$\zeta_2\left(\mathcal{L}\sum_{i=1}^n X_i, CP\left(\lambda,\frac{1}{\lambda}\sum_{i=1}^n p_i G_i\right)\right)=\frac{1}{2}\sum_{i=1}^n E(X_i)^2, \tag{8}$$

with $p_i := P(X_i\neq 0)$, $\lambda=\sum_{i=1}^n p_i$ and $G_i(x)=P(X_i\le x|X_i\neq 0)$. Naturally, the bound of the following theorem is useful (that is, it tends to 0) when $p_i\approx 0$. Hence, the condition $p_i < \log 2\approx 0.693$ imposed below does not affect the generality of the result. One could easily modify the upper bound (making it a little bit more complicated) so as to eliminate this restriction, but this modification would lead to no practical gain.

Theorem 9. *Let* $X_1, X_2, \dots, X_n$ *be a sequence of independent random variables* (*with finite second moments*) *taking values in* $\mathbb{Z}_+$ *and* $P(X_i \neq 0) =: p_i < \log 2$ (≈ 0.693). *Then*,

$$d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n} X_i, CP(\lambda, F)\right)$$
$$\leq \left(\sum_{i=1}^{n} p_i^2\right)^2 + \frac{1}{4}\left\|\Delta^2 f_{CP(\lambda,F)}\right\|_1 \sum_{i=1}^{n} \frac{E(X_i)^2}{1-2(1-\mathrm{e}^{-p_i})} := UB_{CP},$$

where $\lambda = \sum_{i=1}^{n} p_i$, $F(x) = \sum_{i=1}^{n} \frac{p_i}{\lambda} G_i(x)$ *and* $G_i(x) = P(X_i \leq x | X_i \neq 0)$, $x \in \mathbb{Z}$.

Proof. Let $N_1, N_2, \dots, N_n$ be independent random variables following the compound Poisson distribution with parameters $(p_1, G_1), (p_2, G_2), \dots, (p_n, G_n)$, respectively. We apply the triangle inequality to get the following Lindeberg decomposition of the distance of interest,

$$d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n} X_i, \mathcal{L}\sum_{i=1}^{n} N_i\right) \leq \sum_{m=1}^{n} d_{\mathrm{TV}}\left(\mathcal{L}\left(\sum_{i=1}^{m} X_i + \sum_{i=m+1}^{n} N_i\right), \mathcal{L}\left(\sum_{i=1}^{m-1} X_i + \sum_{i=m}^{n} N_i\right)\right). \quad (9)$$

Furthermore, if we set

$$\mathbf{X}_m := X_m + \sum_{i=m+1}^{n} N_i, \qquad \mathbf{Y}_m := N_m + \sum_{i=m+1}^{n} N_i, \qquad \mathbf{Z}_m := \sum_{i=1}^{m-1} X_i, \qquad \mathbf{W}_m := \sum_{i=1}^{m-1} N_i,$$

then the random variables $\mathbf{X}_m, \mathbf{Y}_m$are independent of $\mathbf{Z}_m, \mathbf{W}_m$ and a direct application of Lemma 5 to $\mathbf{X}_m, \mathbf{Y}_m, \mathbf{Z}_m, \mathbf{W}_m$ reveals that

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(\sum_{i=1}^{m-1} X_i + X_m + \sum_{i=m+1}^{n} N_i\right), \mathcal{L}\left(\sum_{i=1}^{m-1} X_i + N_m + \sum_{i=m+1}^{n} N_i\right)\right) \leq 2a_m b_m + c_m, \quad (10)$$

where

$$a_m := d_{\mathrm{TV}}\left(\mathcal{L}\left(X_m + \sum_{i=m+1}^{n} N_i\right), \mathcal{L}\left(N_m + \sum_{i=m+1}^{n} N_i\right)\right),$$
$$b_m := d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{m-1} X_i, \mathcal{L}\sum_{i=1}^{m-1} N_i\right),$$
$$c_m := d_{\mathrm{TV}}\left(\mathcal{L}\left(\sum_{i=1}^{n} N_i - N_m + X_m\right), \mathcal{L}\sum_{i=1}^{n} N_i\right).$$

Next, let $N_m^{\perp}$ be a random variable independent of all N_i, X_i with $\mathcal{L}N_m^{\perp} = \mathcal{L}N_m$. Applying inequality (6) with $W = N_m^{\perp}$, we derive

$$c_m \le \big(1 - 2(1 - \mathrm{e}^{-p_m})\big)^{-1} d_{\mathrm{TV}}\Bigg(\mathcal{L}\Bigg(\sum_{i=1}^n N_i + X_m\Bigg), \mathcal{L}\Bigg(\sum_{i=1}^n N_i + N_m^{\perp}\Bigg)\Bigg)$$

since $P(N_m^{\perp} \ne 0) = 1 - \mathrm{e}^{-p_m}$. Furthermore, Lemma 1 yields

$$c_m \le \frac{1/2\|\Delta^2 f_{\sum_{i=1}^n N_i}\|_1}{1 - 2(1 - \mathrm{e}^{-p_m})} \zeta_2(\mathcal{L}X_m, \mathcal{L}N_m^{\perp}) = \frac{\|\Delta^2 f_{CP(\lambda,F)}\|_1}{4(1 - 2(1 - \mathrm{e}^{-p_m}))} E(X_m)^2, \tag{11}$$

where we have used (8) to get that $\zeta_2(\mathcal{L}X_m, \mathcal{L}N_m^{\perp}) = \zeta_2(\mathcal{L}X_m, CP(p_m, G_m)) = \frac{1}{2}E(X_m)^2$. On the other hand, we can easily bound the quantities a_m, b_m as follows:

$$a_m \le d_{\mathrm{TV}}(\mathcal{L}X_m, \mathcal{L}N_m) \le p_m^2 \quad \text{and} \quad b_m \le \sum_{i=1}^{m-1} d_{\mathrm{TV}}(\mathcal{L}X_i, \mathcal{L}N_i) = \sum_{i=1}^{m-1} p_i^2. \tag{12}$$

Finally, combining (9)–(12), we get

$$\begin{aligned} d_{\mathrm{TV}}\Bigg(\mathcal{L}\sum_{i=1}^n X_i, \mathcal{L}\sum_{i=1}^n N_i\Bigg) &\le \sum_{m=1}^n (2a_m b_m + c_m) \\ &\le \sum_{m=1}^n \Bigg(2p_m^2 \sum_{i=1}^{m-1} p_i^2 + \frac{\|\Delta^2 f_{CP(\lambda,F)}\|_1}{4(1 - 2(1 - \mathrm{e}^{-p_m}))} E(X_m)^2\Bigg), \end{aligned}$$

which readily leads to the desired inequality since

$$\begin{aligned} \sum_{m=1}^n 2a_m b_m &\le \sum_{m=1}^n 2p_m^2 \sum_{i=1}^{m-1} p_i^2 \\ &= \sum_{m=1}^n p_m^2 \sum_{i=1}^{m-1} p_i^2 + \sum_{m=1}^n p_m^2 \sum_{i=m+1}^n p_i^2 \le \Bigg(\sum_{m=1}^n p_m^2\Bigg)^2. \end{aligned} \qquad \square$$

A straightforward corollary of the above theorem arises when we consider independent Bernoulli random variables. In this case, the distribution of the sum of the binary sequence $X_1, X_2, \dots, X_n$ is also known as a Poisson binomial or generalized binomial distribution and the approximating compound Poisson distribution naturally reduces to an ordinary Poisson distribution.

Corollary 10. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent Bernoulli random variables with $P(X_i = 1) = p_i < \log 2$, $i = 1, 2, \ldots, n$. Then,*

$$d_{\mathrm{TV}}\left(\sum_{i=1}^n X_i, Po(\lambda)\right) \le \left(\sum_{i=1}^n p_i^2\right)^2 + \frac{1}{4}\left\|\Delta^2 f_{Po(\lambda)}\right\|_1 \sum_{i=1}^n \frac{p_i^2}{1 - 2(1 - \mathrm{e}^{-p_i})} := UB_{Po},$$

where $\lambda = \sum_{i=1}^n p_i$ and $\|\Delta^2 f_{Po(\lambda)}\|_1$ is given in Proposition 3.

Remark 2. If we assume that $\sum_{i=1}^n p_i^2 \to 0$ as $n \to \infty$ (implying that $\max_i p_i \to 0$), the first term $(\sum_{i=1}^n p_i^2)^2$ in the upper bound UB_{Po} (Corollary 10) or in UB_{CP} (Theorem 9) tends to 0 at a faster rate than the second term and, therefore, the order of UB_{Po} and UB_{CP} is the same as the order of their second term. That is, for UB_{CP}, we have

$$UB_{CP} \sim \begin{cases} 1/4\left\|\Delta^2 f_{CP(\lambda,F)}\right\|_1 \displaystyle\sum_{i=1}^n E(X_i)^2, & \text{for } \lambda \text{ fixed,} \\ \dfrac{1}{\lambda\mu_2\sqrt{2\pi\mathrm{e}}} \displaystyle\sum_{i=1}^n E(X_i)^2, & \text{when } \lambda \to \infty, \end{cases}$$

where μ_2 denotes the second moment of the compounding distribution F (see Remark 1 above). According to Remark 1, the second asymptotic result for UB_{CP} above (when $\lambda \to \infty$) is valid when $CP(\lambda, F)$ *is close to a normal distribution*. Therefore, we can say that, for independent $X_1, \ldots, X_n \in \mathbb{Z}_+$ with $E(X_i) = \mathrm{O}(p_i)$,

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(\sum_{i=1}^n X_i\right), CP(\lambda, F)\right) = \mathrm{O}\left(\frac{1}{\lambda}\sum_{i=1}^n p_i^2\right),$$

whenever F is such that $CP(\lambda, F) \approx N(\mu, \sigma^2)$ or, more generally, whenever $\|\Delta^2 f_{CP(\lambda,F)}\|_1 = \mathrm{O}(\lambda^{-1})$. Our approach requires the restriction $\sum_{i=1}^n p_i^2 \to 0$ (not only $\max_i p_i \to 0$),but we have reasons to believe (see Remark 3) that this restriction is superfluous and can be weakened. This offers a clue to a question raised by Le Cam (1960) (see also Barbour and Utev (1999) and Roos (2003)) about the form of the compounding distribution F that would permit us to achieve a compound Poisson approximation error order similar to that obtained for Poisson approximation, that is, $\frac{1}{\lambda}\sum_{i=1}^n p_i^2$.

We also point out that the upper bound UB_{Po} of Corollary 10 for the Poisson approximation is similar to the one derived by Deheuvels and Pfeifer (1986), (see also Deheuvels, Pfeifer and Puri (1989)) who employed an entirely different method. The factor $\|\Delta^2 f_{Po(\lambda)}\|_1/4$ appears in the bounds of these articles (in an equivalent form, not recognized as being the L_1-norm of $\Delta^2 f_{Po(\lambda)}/4$) and was proven to be optimal (that is, $d_{\mathrm{TV}} \sim UB_{Po}$; see Deheuvels and Pfeifer (1986)) under the usual asymptotic assumptions. The same argument is possibly true for the more general smoothness factor $\|\Delta^2 f_{CP(\lambda,F)}\|_1/4$.

Remark 3. In the proof of Theorem 9, the quantity $2\sum_{m=1}^n a_m b_m$ (see relation (12)) was bounded rather crudely in order to obtain a closed form upper bound. This resulted in a simple-in-form first term, namely $(\sum_{i=1}^n p_i^2)^2$, in UB_{CP}. If $\sum_{i=1}^n p_i^2 \to 0$, then this term does not have a significant effect on UB_{CP}, but if $\sum_{i=1}^n p_i^2$ is not close to 0, then it may result in a very crude upper bound.

Nevertheless, concerning the Poisson case, if we possessed a simple-in-form upper bound for $\|\Delta^2 f_{Po(\lambda)}\|_1$, we could obtain a better (smaller) bound for the quantity $2\sum_{m=1}^n a_m b_m$. To get an idea of how this can be done, we shall treat the simplest case where $X_1, X_2, \dots, X_n$ are i.i.d. ($p_i = p$) Bernoulli random variables. Recall that, in general, $\frac{1}{4}\|\Delta^2 f_{Po(\lambda)}\|_1 \le (1 \wedge \frac{1}{3\lambda})$ and, therefore,

$$a_m \le 1/2\|\Delta^2 f_{\sum_{i=m+1}^n N_i}\|_1 \zeta_2(X_m, N_m) \le \left(1 \wedge \frac{1}{3(n-m)p}\right)p^2,$$

$$b_m \le \frac{1-\mathrm{e}^{-(m-1)p}}{(m-1)p}\sum_{i=1}^{m-1} p_i^2 \le p.$$

Assuming that $\lambda \ge 1/3 + p$ and taking into account that $\sum_{i=n_1}^{n_2} \frac{1}{i} < \log(\frac{n_2}{n_1-1})$, the sum $2\sum_{m=1}^n a_m b_m$ is bounded above by

$$\begin{aligned} 2\sum_{m=1}^n \left(1 \wedge \frac{1}{3(n-m)p}\right)p^3 &\le \sum_{m=1}^{\lfloor n-1/(3p)\rfloor} \frac{2p^2}{3(n-m)} + \sum_{m=\lfloor n-1/(3p)\rfloor+1}^{n} 2p^3 \\ &\le 2\frac{p^2}{3}\left(\log\frac{3np}{1-3p} + 1\right) + 2p^3, \end{aligned} \tag{13}$$

under the assumption $p < 1/3$. The latter reveals that, when $p_i = p$ and $\lambda > 1/3 + p$, the term $(\sum_{i=1}^n p_i^2)^2 = \lambda^2 p^2$ in Corollary 10 can be substantially reduced to (13), implying that $UB_{Po} \approx \frac{2}{3}p^2(\log 3\lambda + 1) + \frac{1}{\sqrt{2\pi \mathrm{e}}}p$. The above bound could also be reduced (requiring more complicated algebraic manipulations) in the case of non-i.i.d. Bernoulli random variables. For a more general case though, for example, in a compound Poisson approximation, we must first find a suitable general upper bound for $\|\Delta^2 f_{CP(\lambda,F)}\|_1$ which, at the moment, does not seem an easy task and is left for future work.

4. Compound Poisson approximation for sums of k-dependent random variables

In this section, a more general setup is assumed. We are now interested in approximating the distribution of the sum $X_1 + \cdots + X_n$ when the k-dependent X_i's are rarely non-zero. Naturally, we expect that this distribution converges weakly to an appropriate compound Poisson distribution.

Following the same methodological steps as in the proof of Theorem 9, we offer a bound that includes a smoothness factor analogous to a Stein factor. The appearance of such a factor is perhaps the first (for sums of *dependent* random variables) outside the Stein–Chen method. As was mentioned in the Introduction, the smoothness factor we derive is simpler, seems more natural and is better than the corresponding Stein factors. On the other hand, inevitably, an undesired term analogous to $(\sum p_i^2)^2$ of Theorem 9 again appears in the upper bounds.

For convenience, we shall focus our approach on a sequence of independent random variables $Z_1, Z_2, \ldots$ defined over a probability space $(\Omega, \mathcal{A}, P)$ and consider k-dependent random variables of the form $h_i(Z_i, \ldots, Z_{i+k-1})$. This approach is not restrictive since, in almost all applications, local dependency arises in this setup (for example, runs or scan statistics, patterns, reliability theory, graph theory problems, moving sums, etcetera). Specifically, let $Z_1, Z_2, \ldots$ be independent random variables and also let

$$X_i = h_i(Z_i, \ldots, Z_{i+k-1}), \qquad i = 1, 2, \ldots, \tag{14}$$

be a sequence of non-negative, integer-valued random variables, generated by some measurable functions $h_i : \mathbb{R}^k \to \mathbb{Z}$. The above definition implies that X_i is independent of $X_1, \ldots, X_{i-k}$ and $X_{i+k}, \ldots$. Therefore, $X_1, X_2, \ldots$ are "k-dependent" random variables (independent random variables can be considered as 1-dependent). Naturally, the bound offered tends to 0, provided that $P(X_i \neq 0) = p_i \approx 0$. Hence, the condition $\max_i \sum_{j=i-3k+3}^{i} p_j < \log 2 \approx 0.693$ does not affect the generality of the result. We assume that $X_i = 0$ for all $i < 1$.

Theorem 11. *Let $X_1, X_2, \ldots, X_n \in \mathbb{Z}_+$ be k-dependent random variables* (*defined as in* (14)) *with finite second moments. Let $N_1, \ldots, N_n$ be independent random variables* (*also independent of Z_i*) *with N_i following the $CP(p_i, G_i)$ distribution*, *where $G_i(x) = P(X_i \leq x | X_i \neq 0)$, $x \in \mathbb{R}$ and $p_i = P(X_i \neq 0)$. Then*, *for $m := \max_i \sum_{j=i-3k+3}^{i} p_j < \log 2$,*

$$\begin{aligned} &d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n} X_i, CP(\lambda_n, F_n)\right) \\ &\quad \leq C_n + \frac{\|\Delta^2 f_{CP(\lambda_n, F_n)}\|_1}{2(1 - 2(1 - \mathrm{e}^{-m}))} \sum_{i=1}^{n} \zeta_2\left(\mathcal{L}\sum_{j=i-2k+2}^{i} X_j, \mathcal{L}\left(\sum_{j=i-2k+2}^{i-1} X_j + N_i\right)\right) \\ &\quad := UB'_{CP}, \end{aligned}$$

where

$$\begin{aligned} C_n := 2\sum_{i=1}^{n}&\left(d_{\mathrm{TV}}\left(\mathcal{L}\sum_{j=1}^{i-3k+2} X_j, \mathcal{L}\sum_{j=1}^{i-3k+2} N_j\right) + \sum_{j=i-3k+3}^{i-2k+1} p_j\right) \\ &\times \left(2P\big((X_{i-k+1}, \ldots, X_{i-1}) \neq \mathbf{0}, X_i \neq 0\big) + 2p_i \sum_{j=i-k+1}^{i-1} p_j + p_i^2\right) \end{aligned}$$

and $\lambda_n = \sum_{i=1}^{n} p_i$, $F_n = \sum_{i=1}^{n} \frac{p_i}{\lambda_n} G_i$.

Proof. In order to simplify notation, set $X_{a,b} := \sum_{i=a}^{b} X_i$, $\mathbf{X}_{a,b} := (X_a, X_{a+1}, \ldots, X_b)$, $N_{a,b} := \sum_{i=a}^{b} N_i$ and $\mathbf{Z}_{a,b} := (Z_a, Z_{a+1}, \ldots, Z_b)$. Also, let U_i, U_i^*, $i = 1, 2, \ldots, n$, be independent random variables, also independent of Z_i, N_i following the uniform distribution on $(0, 1)$.

Fix $i \in \{1, 2, \ldots, n\}$. In order to avoid a special treatment for small values of i due to edge effects and to preserve a unified analysis for all i that takes into account edge effects, we simply assume that $X_j = N_j = Z_j = 0$ for $j \le 0$. According to Lemma 8(b) (with f being the p.d.f. of $N_{1,i-3k+2}$), there exists a random variable $N^*_{1,i-3k+2} = g_1(U_{i-3k+2}, X_{1,i-3k+2})$ such that $\mathcal{L}N^*_{1,i-3k+2} = \mathcal{L}N_{1,i-3k+2}$ and $(X_{1,i-3k+2}, N^*_{1,i-3k+2})$ are maximally coupled, that is,

$$d_{\mathrm{TV}}(\mathcal{L}X_{1,i-3k+2}, \mathcal{L}N^*_{1,i-3k+2}) = P(X_{1,i-3k+2} \ne N^*_{1,i-3k+2}).$$

Moreover, according to Lemma 8(a) (with f now being the p.d.f. of N_i), there exists a random variable

$$N_i^* = g_2(U_i^*, X_i, \mathbf{X}_{i-k+1,i-1}, \mathbf{Z}_{i-k+1,i-1})$$

such that $\mathcal{L}N_i^* = \mathcal{L}N_i$, N_i^* is independent of the vector $(\mathbf{X}_{i-k+1,i-1}, \mathbf{Z}_{i-k+1,i-1})$ and

$$d_{\mathrm{TV}}(\mathcal{L}(X_i, \mathbf{X}_{i-k+1,i-1}, \mathbf{Z}_{i-k+1,i-1}), \mathcal{L}(N_i^*, \mathbf{X}_{i-k+1,i-1}, \mathbf{Z}_{i-k+1,i-1})) = P(X_i \ne N_i^*). \quad (15)$$

It is easy to check that, as defined, N_i^* is also independent of $\mathbf{X}_{1,i-1}$. Indeed, if we set $\mathbf{Y} := (\mathbf{Z}_{i-k+1,i-1}, \mathbf{X}_{i-k+1,i-1})$, for all $x, \mathbf{x}$, we have that

$$P(N_i^* = x, \mathbf{X}_{1,i-1} = \mathbf{x}) = \sum_{\mathbf{y}} P(N_i^* = x, \mathbf{Y} = \mathbf{y}, \mathbf{X}_{1,i-1} = \mathbf{x}).$$

We may write $\mathbf{X}_{1,i-1} = \mathbf{g}(\mathbf{Z}_{1,i-k}, \mathbf{Y})$ for some appropriate function $\mathbf{g}$ taking values in $\mathbb{Z}^{i-1}$. Hence, the above sum is equal to

$$\begin{aligned}
&\sum_{\mathbf{y}} P\big(g_2(U_i^*, X_i, \mathbf{Y}) = x, \mathbf{Y} = \mathbf{y}, \mathbf{g}(\mathbf{Z}_{1,i-k}, \mathbf{y}) = \mathbf{x}\big) \\
&\quad = \sum_{\mathbf{y}} P\big(g_2(U_i^*, X_i, \mathbf{Y}) = x, \mathbf{Y} = \mathbf{y}\big) P\big(\mathbf{g}(\mathbf{Z}_{1,i-k}, \mathbf{y}) = \mathbf{x}\big) \\
&\quad = \sum_{\mathbf{y}} P(N_i^* = x) P(\mathbf{Y} = \mathbf{y}) P\big(\mathbf{g}(\mathbf{Z}_{1,i-k}, \mathbf{y}) = \mathbf{x}\big) \\
&\quad = \sum_{\mathbf{y}} P(N_i^* = x) P\big(\mathbf{Y} = \mathbf{y}, \mathbf{g}(\mathbf{Z}_{1,i-k}, \mathbf{y}) = \mathbf{x}\big) \\
&\quad = P(N_i^* = x) \sum_{\mathbf{y}} P(\mathbf{Y} = \mathbf{y}, \mathbf{X}_{1,i-1} = \mathbf{x}) = P(N_i^* = x) P(\mathbf{X}_{1,i-1} = \mathbf{y}),
\end{aligned}$$

which is valid for all $x, \mathbf{y}$ and thus N_i^* is independent of $\mathbf{X}_{1,i-1}$.

Now, applying the inequality (see Lemma 4)

$$d_{\mathrm{TV}}\big(\mathcal{L}(Z + X), \mathcal{L}(Z + Y)\big) \le 2P(Z \ne W, X \ne Y) + d_{\mathrm{TV}}\big(\mathcal{L}(W + X), \mathcal{L}(W + Y)\big)$$

with $Z = X_{1,i-1}$, $X = X_i + N_{i+1,n}$, $Y = N_i^* + N_{i+1,n}$, $W = N^*_{1,i-3k+2} + X_{i-2k+2,i-1}$, we obtain

$$\begin{aligned}
&d_{\mathrm{TV}}\big(\mathcal{L}(X_{1,i-1} + X_i + N_{i+1,n}), \mathcal{L}(X_{1,i-1} + N_i^* + N_{i+1,n})\big)\\
&\quad \le 2P(X_{1,i-1} \neq N^*_{1,i-3k+2} + X_{i-2k+2,i-1}, X_i + N_{i+1,n} \neq N_i^* + N_{i+1,n})\\
&\qquad + d_{\mathrm{TV}}\big(\mathcal{L}(N^*_{1,i-3k+2} + X_{i-2k+2,i} + N_{i+1,n}),\\
&\qquad\qquad \mathcal{L}(N^*_{1,i-3k+2} + X_{i-2k+2,i-1} + N_i^* + N_{i+1,n})\big).
\end{aligned} \tag{16}$$

Note that $\mathcal{L}N_i^* = \mathcal{L}N_i$ and N_i^* is independent of $\mathbf{X}_{1,i-1}$, also that $\mathcal{L}N^*_{1,i-3k+2} = \mathcal{L}N_{1,i-3k+2}$ and $N^*_{1,i-3k+2} = g_1(U_{i-3k+2}, X_{1,i-3k+2})$ is independent of $\mathbf{X}_{i-2k+2,i}$ and N_i^*. Therefore, we have that

$$\begin{aligned}
\mathcal{L}(X_{1,i-1} + N_i^* + N_{i+1,n}) &= \mathcal{L}(X_{1,i-1} + N_i + N_{i+1,n}),\\
\mathcal{L}(N^*_{1,i-3k+2} + X_{i-2k+2,i} + N_{i+1,n}) &= \mathcal{L}(N_{1,i-3k+2} + X_{i-2k+2,i} + N_{i+1,n}),\\
\mathcal{L}(N^*_{1,i-3k+2} + X_{i-2k+2,i-1} + N_i^* + N_{i+1,n}) &= \mathcal{L}(N_{1,i-3k+2} + X_{i-2k+2,i-1} + N_i + N_{i+1,n}).
\end{aligned}$$

Using the above relations, inequality (16) is equivalent to

$$d_{\mathrm{TV}}\big(\mathcal{L}(X_{1,i-1} + X_i + N_{i+1,n}), \mathcal{L}(X_{1,i-1} + N_i + N_{i+1,n})\big) \le 2a_i + b_i, \tag{17}$$

where

$$\begin{aligned}
a_i &= P(X_{1,i-2k+1} \neq N^*_{1,i-3k+2}, X_i \neq N_i^*),\\
b_i &= d_{\mathrm{TV}}\big(\mathcal{L}(N_{1,i-3k+2} + X_{i-2k+2,i-1} + X_i + N_{i+1,n}),\\
&\qquad \mathcal{L}(N_{1,i-3k+2} + X_{i-2k+2,i-1} + N_i + N_{i+1,n})\big).
\end{aligned}$$

The random variables X_i, N_i^* are independent of $X_1, \ldots, X_{i-2k+1}, N^*_{1,i-3k+2}$ and, hence, it is easy to see that

$$\begin{aligned}
a_i &= P(X_{1,i-3k+2} + X_{i-3k+3,i-2k+1} \neq N^*_{1,i-3k+2})P(X_i \neq N_i^*)\\
&\le \big(P(X_{1,i-3k+2} \neq N^*_{1,i-3k+2}) + P(X_{i-3k+3,i-2k+1} \neq 0)\big)P(X_i \neq N_i^*)\\
&\le \left(d_{\mathrm{TV}}(\mathcal{L}X_{1,i-3k+2}, \mathcal{L}N_{1,i-3k+2}) + \sum_{j=i-3k+3}^{i-2k+1} p_j\right)P(X_i \neq N_i^*).
\end{aligned} \tag{18}$$

Using relation (15) above along with $\mathcal{L}(\mathbf{Z}_{i-k+1,i-1}, \mathbf{X}_{i-k+1,i-1}, N_i^*) = \mathcal{L}(\mathbf{Z}_{i-k+1,i-1}, \mathbf{X}_{i-k+1,i-1}, N_i)$, we observe that

$$P(X_i \neq N_i^*) = d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}_{i-k+1,i-1}, \mathbf{X}_{i-k+1,i-1}, X_i), \mathcal{L}(\mathbf{Z}_{i-k+1,i-1}, \mathbf{X}_{i-k+1,i-1}, N_i))$$

and applying Lemma 4 with $\mathbf{X} = X_i, \mathbf{Y} = N_i, \mathbf{Z} = (\mathbf{Z}_{i-k+1,i-1}, \mathbf{X}_{i-k+1,i-1}), \mathbf{W} =$

$(\mathbf{Z}_{i-k+1,i-1}, \mathbf{0})$, we deduce

$$
\begin{aligned}
&P(X_i \neq N_i^*) \\
&\quad \leq 2P(X_i \neq N_i, \mathbf{X}_{i-k+1,i-1} \neq \mathbf{0}) + d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Z}_{i-k+1,i-1}, \mathbf{0}, X_i), \mathcal{L}(\mathbf{Z}_{i-k+1,i-1}, \mathbf{0}, N_i)) \\
&\quad \leq 2P(X_i \neq 0, \mathbf{X}_{i-k+1,i-1} \neq \mathbf{0}) \\
&\qquad + 2P(N_i \neq 0)P(\mathbf{X}_{i-k+1,i-1} \neq \mathbf{0}) + d_{\mathrm{TV}}(\mathcal{L}X_i, \mathcal{L}N_i) \\
&\quad \leq 2P(X_i \neq 0, \mathbf{X}_{i-k+1,i-1} \neq \mathbf{0}) + 2p_i \sum_{j=i-k+1}^{i-1} p_j + p_i^2.
\end{aligned}
\tag{19}
$$

Next, we consider a random variable $N^{\perp}$ with $\mathcal{L}N_i^{\perp} = \mathcal{L}N_i$, independent of all other random variables involved in our analysis. Applying the inequality (6) with $W = N_{i-3k+3,i}$ and assuming that $P(N_{i-3k+3,i} \neq 0) < 1/2$ (which is valid since we have assumed that $m < \log 2$), we get

$$
\begin{aligned}
b_i &= d_{\mathrm{TV}}\big(\mathcal{L}(N_{1,i-3k+2} + X_{i-2k+2,i-1} + X_i + N_{i+1,n}), \\
&\qquad\quad \mathcal{L}(N_{1,i-3k+2} + X_{i-2k+2,i-1} + N_i^{\perp} + N_{i+1,n})\big) \\
&\leq \big(1 - 2P(N_{i-3k+3,i} \neq 0)\big)^{-1} d_{\mathrm{TV}}\big(\mathcal{L}(N_{1,n} + X_{i-2k+2,i}), \mathcal{L}(N_{1,n} + X_{i-2k+2,i-1} + N_i^{\perp})\big).
\end{aligned}
$$

Finally, using Lemma 1, we derive

$$
b_i \leq \frac{1/2\|\Delta^2 f_{CP(\lambda,F)}\|_1}{1 - 2(1 - \mathrm{e}^{-\sum_{j=i-3k+3}^{i} p_j})} \zeta_2\big(\mathcal{L}X_{i-2k+2,i}, \mathcal{L}(X_{i-2k+2,i-1} + N_i)\big). \tag{20}
$$

Combining (17)–(19) with (20), we obtain, for all $i = 1, 2, \ldots, n$, the inequality

$$
\begin{aligned}
&d_{\mathrm{TV}}\big(\mathcal{L}(X_{1,i-1} + X_i + N_{i+1,n}), \mathcal{L}(X_{1,i-1} + N_i + N_{i+1,n})\big) \\
&\quad \leq 2\left(d_{\mathrm{TV}}(\mathcal{L}X_{1,i-3k+2}, \mathcal{L}N_{1,i-3k+2}) + \sum_{j=i-3k+3}^{i-2k+1} p_j\right) \\
&\qquad \times \left(2P(X_i \neq 0, \mathbf{X}_{i-k+1,i-1} \neq \mathbf{0}) + 2p_i \sum_{j=i-k+1}^{i-1} p_j + p_i^2\right) \\
&\qquad + \frac{1/2\|\Delta^2 f_{CP(\lambda,F)}\|_1}{1 - 2(1 - \mathrm{e}^{-m})} \zeta_2\big(\mathcal{L}X_{i-2k+2,i}, \mathcal{L}(X_{i-2k+2,i-1} + N_i)\big)
\end{aligned}
$$

and the final result follows immediately by virtue of the Lindeberg decomposition (triangle inequality)

$$
d_{\mathrm{TV}}(\mathcal{L}X_{1,n}, \mathcal{L}N_{1,n}) \leq \sum_{i=1}^{n} d_{\mathrm{TV}}\big(\mathcal{L}(X_{1,i-1} + X_i + N_{i+1,n}), \mathcal{L}(X_{1,i-1} + N_i + N_{i+1,n})\big). \qquad \square
$$

Remark 4. The upper bound UB'_{CP} in Theorem 11 is composed of two terms, the first of which is the quantity C_n, which is analogous to the term $(\Sigma p_i^2)^2$ appearing in Theorem 9 that concerns the independent summands case. As it was for $(\Sigma p_i^2)^2$, the term C_n tends to 0 faster than the second term of UB'_{CP}, under certain asymptotic conditions. Therefore, under these conditions, the order of UB'_{CP} coincides with the order of the second term.

Remark 5. If $X_1, X_2, \dots, X_n$ are k-dependent Bernoulli random variables then, similarly to Corollary 10, Theorem 11 implies a Poisson approximation result. Specifically, Theorem 11 can now be written with $Po(p_i)$ in place of $CP(p_i, G_i)$ and $Po(\lambda_n)$ in place of $CP(\lambda_n, F_n)$. Consequently, the norm $\|\Delta^2 f_{Po(\lambda_n)}\|_1$ will appear instead of $\|\Delta^2 f_{CP(\lambda_n,F_n)}\|_1$.

The upper bound UB'_{CP} in Theorem 11 may seem difficult to apply in its present form. For this reason, we present the following corollary which provides two slightly worse, but more easily computable, upper bounds. The bound (a) is valid without any assumption on the form of dependence among the X_i's. The bound (b) is smaller than (a), but is valid only when the X_i's exhibit a certain weak form of positive/negative dependence. We recall that two random variables Y_1, Y_2 are called *positively quadrant dependent* (PQD) if

$$P(Y_1 \ge x_1, Y_2 \ge x_2) \ge P(Y_1 \ge x_1)P(Y_2 \ge x_2) \qquad \text{for all } x_1, x_2 \tag{21}$$

and *negatively quadrant dependent* (*NQD*) if (21) holds, but with the inequality sign reversed. Manifestly, if $X_1, X_2, \dots, X_n$ are associated (resp., negatively associated), then the random variables $X_j + X_2 + \cdots + X_{i-1}$ and X_i are PQD (resp., NQD) for every $1 \le j < i \le n$. Therefore, part (b) of the next corollary remains valid under the stronger condition of association or negative association of X_i's.

Corollary 12. (a) *Let $X_1, X_2, \dots, X_n \in \mathbb{Z}_+$ be k-dependent random variables* (*defined as in* (14)) *with finite second moments. Then, for $m := \max_i \sum_{j=i-3k+3}^{i} p_j < \log 2$, $p_i = P(X_i \ne 0)$,*

$$\begin{aligned} &d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^n X_i, CP(\lambda_n, F_n)\right) \\ &\quad \le C_n + \frac{\|\Delta^2 f_{CP(\lambda_n,F_n)}\|_1}{2(1-2(1-\mathrm{e}^{-m}))}\sum_{i=1}^n\left(\sum_{j=i-k+1}^{i-1}\big(E(X_iX_j)+E(X_i)E(X_j)\big)+\frac{1}{2}E(X_i)^2\right) \\ &\quad := UB''_{CP}, \end{aligned}$$

where

$$\begin{aligned} C_n := 2\sum_{i=1}^n&\left(2\sum_{j=1}^{i-3k+2}\left(\sum_{t=j-k+1}^{j-1}\big(P(X_tX_j\ne 0)+p_tp_j\big)+\frac{1}{2}p_j^2\right)+\sum_{j=i-3k+3}^{i-2k+1}p_j\right) \\ &\times\left(2\sum_{j=i-k+1}^{i-1}\big(P(X_j\ne 0, X_i\ne 0)+p_ip_j\big)+p_i^2\right) \end{aligned}$$

and $\lambda_n = \sum_{i=1}^n p_i$, $F_n = \sum_{i=1}^n \frac{p_i}{\lambda_n}G_i$, $G_i(x) = P(X_i \le x | X_i \ne 0)$, $x \in \mathbb{R}$ ($X_i = 0$ for all $i < 1$).

(b) *If, in addition, the random variables* $X_j + \cdots + X_{i-1}$ *and* X_i *are PQD or NQD for every* $1 \le j < i \le n$, *then the bound* UB''_{CP} *in* (a) *is valid with* $|\mathrm{Cov}(X_i, X_j)|$ *in place of* $E(X_i X_j) + E(X_i)E(X_j)$ *and*

$$\sum_{t=j-k+1}^{j-1} |\mathrm{Cov}(X_j, X_t)| + \frac{1}{2}E(X_j)^2 \quad \textit{in place of} \quad \sum_{t=j-k+1}^{j-1} \big(P(X_t X_j \neq 0) + p_t p_j\big) + \frac{1}{2}p_j^2.$$

Proof. (a) This follows readily from Theorem 11 by applying Corollary 7 above, Corollary 7 in Boutsikas (2006) and the fact that

$$\zeta_2\left(\mathcal{L}\left(\sum_{j=l}^{i-1} X_j + X_i^{\perp}\right), \mathcal{L}\left(\sum_{j=l}^{i-1} X_j + N_i\right)\right) \le \zeta_2(\mathcal{L}X_i^{\perp}, \mathcal{L}N_i) = \frac{1}{2}E(X_i)^2$$

($N_i \sim CP(p_i, G_i)$), which is a consequence of the regularity property of ζ_2 combined with equality (8).

(b) This is again a direct consequence of Theorem 11. Set $W = \sum_{j=i-2k+2}^{i-1} X_j$ and let $X_i^{\perp}$ be a random variable independent of all other random variables involved in our analysis with $\mathcal{L}X_i = \mathcal{L}X_i^{\perp}$. Assume that $X_j + \cdots + X_{i-1}$ and X_i are PQD for all $j < i$. Thus, W and X_i are PQD and hence $W + X_i$ is larger than $W + X_i^{\perp}$ with respect to the convex order (see Section 3.3 in Boutsikas and Vaggelatou (2002)). Therefore,

$$\zeta_2\big(\mathcal{L}(W + X_i), \mathcal{L}(W + X_i^{\perp})\big) = \frac{1}{2}\big(\mathrm{Var}(W + X_i) - \mathrm{Var}(W + X_i^{\perp})\big) = \sum_{j=i-k+1}^{i-1} \mathrm{Cov}(X_i, X_j).$$

Since W is independent of $X_i^{\perp}, N_i$, the regularity property of ζ_2 and equality (8) guarantee that

$$\zeta_2\big(\mathcal{L}(W + X_i^{\perp}), \mathcal{L}(W + N_i)\big) \le \zeta_2(\mathcal{L}X_i^{\perp}, \mathcal{L}N_i) = \tfrac{1}{2}E(X_i)^2.$$

Hence, using the triangle inequality and the above two equalities, we deduce that

$$\zeta_2\big(\mathcal{L}(W + X_i), \mathcal{L}(W + N_i)\big) \le \sum_{j=i-k+1}^{i-1} \mathrm{Cov}(X_i, X_j) + \frac{1}{2}E(X_i)^2.$$

Furthermore, from (3) and Theorem 7 in Boutsikas and Vaggelatou (2002), we get that

$$\begin{aligned} d_{\mathrm{TV}}&\left(\mathcal{L}\sum_{j=1}^{i-3k+2} X_j, CP(\lambda_{i-3k+2}, F_{i-3k+2})\right) \\ &\le 2\zeta_2\left(\mathcal{L}\sum_{j=1}^{i-3k+2} X_j, CP(\lambda_{i-3k+2}, F_{i-3k+2})\right) \\ &= 2\sum_{t=2}^{i-3k+2}\sum_{j=t-k+1}^{t-1} \mathrm{Cov}(X_j, X_t) + \sum_{j=1}^{i-3k+2} E(X_j)^2. \end{aligned}$$

Similar reasoning proves the NQD random variables case (in place of all $\mathrm{Cov}(X_j, X_t)$, we now get $-\mathrm{Cov}(X_j, X_t) > 0$). □

5. Illustrating applications

The purpose of this section is to illustrate the applicability and effectiveness of the results presented in the previous sections. These results are applicable to a wide variety of problems involving locally dependent random variables that rarely differ from zero (for example, in risk theory, extreme value theory, reliability theory, run and scan statistics, graph theory and biomolecular sequence analysis). The approximation method described in this paper, as with almost all other methods used for Poisson approximation in the past, requires the computation of only the first- and second-order moments of the variables involved. From this fact, it is understood that the bounds presented can be applied almost directly to many of the problems where other Poisson approximation methods have been elaborated in the past, for example, the Stein–Chen method. The main benefit of the present method is the smoothness factor that substantially improves the approximation error bound in many cases, while the main disadvantage is the additional term C_n. Therefore, the conclusion here is that we usually obtain improved bounds for moderate or small values of λ.

5.1. The number of overlapping runs of length k in i.i.d. trials

Let $\{Z_i\}_{i\in\mathbb{Z}}$ be a sequence of i.i.d. binary trials with outcomes 0 (failure) and 1 (success), and where $P(Z_i = 1) = p = 1 - q$. We are interested in approximating the distribution of the number of (rare) *overlapping* success runs of length k within trials $1, 2, \ldots, n$. This problem has been studied in various ways by many authors in the past; see, for example, Barbour, Holst and Janson (1992), Balakrishnan and Koutras (2002) and the references therein. We shall first derive a Poisson and then a compound Poisson approximation.

(a) *Poisson approximation*. If we assume that $p \to 0$ and $n \to \infty$, then the occurrences of success runs are rare and asymptotically independent, and a Poisson approximation seems suitable. We use the binary random variables

$$X_i = Z_i Z_{i+1} \cdots Z_{i+k-1}, \qquad i = 1, 2, \ldots, n-k+1.$$

Obviously, the random variable $\sum_{i=1}^{n-k+1} X_i$ counts the total number of appearances of overlapping success runs with length k which appear within the first n trials. The random variables $X_1, X_2, \ldots, X_{n-k+1}$ are k-dependent, can be written as in (14) and are associated as coordinatewise non-decreasing functions of independent random variables. Thus, they satisfy the dependence condition required by Corollary 12(b). A direct application of this corollary for

$m = (3k-2)p^k < \log 2$ yields

$$
\begin{aligned}
d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n-k+1} X_i, Po(\lambda)\right) &\le C_{n-k+1} + \frac{1/2\|\Delta^2 f_{Po(\lambda)}\|_1}{1-2(1-\mathrm{e}^{-m})} \\
&\quad \times \left(\sum_{i=2}^{n-k+1}\sum_{j=\max\{1,i-k+1\}}^{i-1}(p^{i-j+k}-p^{2k}) + \frac{n-k+1}{2}p^{2k}\right) \\
&\le C_{n-k+1} + \frac{\lambda p\|\Delta^2 f_{Po(\lambda)}\|_1}{2q(1-2(1-\mathrm{e}^{-m}))}\left(1-\left(k-2+\frac{1}{q}\right)qp^{k-1}\right),
\end{aligned}
$$

where $p_i = P(X_i = 1) = p^k$, $\lambda = (n-k+1)p^k$ and

$$
\begin{aligned}
C_{n-k+1} &= 2\sum_{i=1}^{n-k+1}\left(2\sum_{t=2}^{i-3k+2}\sum_{j=\max\{1,t-k+1\}}^{t-1}(p^{t-j+k}-p^{2k}) + \sum_{j=1}^{i-3k+2}p^{2k} + (k-1)p^k\right) \\
&\quad \times \left(2\sum_{j=\max\{1,i-k+1\}}^{i-1}p^{i-j+k} + (2k-1)p^{2k}\right) \\
&\le 4\frac{\lambda^2p^2}{q}\left(1-p^{k-1}-q\left(k-\frac{3}{2}\right)p^{k-1} + \frac{q(k-1)p^{k-1}}{\lambda}\right) \\
&\quad \times \left(1-p^{k-1}+q\left(k-\frac{1}{2}\right)p^{k-1}\right).
\end{aligned}
$$

Therefore, for $m = (3k-2)p^k < \log 2$,

$$
\begin{aligned}
&d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n-k+1} X_i, Po(\lambda)\right) \\
&\quad \le UB_{n,p} := 4\frac{\lambda^2p^2}{q}\left(1+\frac{qkp^{k-1}}{\lambda}\right)(1+qkp^{k-1}) + \frac{\lambda p\|\Delta^2 f_{Po(\lambda)}\|_1}{2q(1-2(1-\mathrm{e}^{-m}))}.
\end{aligned}
$$

In addition, if $n \to \infty$, $p \to 0$ ($k > 1$ fixed), then

$$
UB_{n,p} \sim \begin{cases} \dfrac{\lambda\|\Delta^2 f_{Po(\lambda)}\|_1}{2q}p, & \text{when } \lambda \text{ is fixed,} \\ \dfrac{2}{q\sqrt{2\pi e}}p, & \text{when } \lambda \to \infty \text{ and } p\lambda^2 \to 0, \end{cases}
$$

where $\|\Delta^2 f_{Po(\lambda)}\|_1$(which is less than $4(1 \wedge \frac{1}{3\lambda})$) is given in Proposition 3. For the same distance, a bound obtained by the Stein–Chen method (see, for example, Barbour, Holst and Janson (1992), page 163) is nearly equal to $2p/q$, which, provided that $p\lambda^2 \approx 0$ and for moderate or large values of λ, is nearly four times larger than $UB_{n,p}$ ($\sqrt{2\pi e} \approx 4.1327$). For $\lambda \approx 1$, it is nearly three times larger.

(b) *Compound Poisson approximation*. The bound described in (a) cannot help when we assume that $n \to \infty, k \to \infty$ and p is fixed. Under these conditions, the occurrences of success runs are again rare, but they are no longer asymptotically independent. This happens because if a success run occurs (starts) at trial i (that is, $Z_i = \cdots = Z_{i+k-1} = 1$), then, with probability p, we shall also observe an overlapping success run starting at position $i + 1$, and so forth. Thus, when a success run is observed at some trial, it is likely that a number of success runs will follow at the next trials. This "cluster" of adjacent success runs is usually called a "clump". So, now that $n \to \infty$ and $k \to \infty$, we expect that the occurrences of clumps are *rare* and asymptotically independent, while each clump consists of an asymptotically geometrically distributed number of overlapping success runs. Obviously, this situation readily calls for a compound Poisson approximation result. To achieve this, let $Y_1, Y_2, \ldots, Y_{n-k+1}$ represent the sizes of the clumps started at trials $1, 2, \ldots, n - k + 1$, respectively. If $Y_i = 0$, then we obviously mean that no clump has started at position i. This well-known technique is called "declumping". More formally, set

$$Y_i := (1 - Z_{i-1}) \sum_{r=0}^{n-i-k+1} \prod_{j=i}^{i+k+r-1} Z_j, \qquad i = 2, 3, \ldots, n - k + 1, \quad \text{and}$$

$$Y_1 := \sum_{r=0}^{n-k} \prod_{j=1}^{k+r} Z_j$$

to be the size of a clump starting at position i (that is, the number of adjacent overlapping success runs until trial n). Clearly, $\sum_{i=1}^{n-k+1} Y_i$ is equal to $\sum_{i=1}^{n-k+1} X_i$, the total number of overlapping success runs within trials $1, 2, \ldots, n$. In this case, it is computationally more convenient to use the *stationary*, *locally dependent* random variables

$$Y_i' := (1 - Z_{i-1}) \sum_{r=0}^{k-1} \prod_{j=i}^{i+k+r-1} Z_j, \qquad i = 1, 2, \ldots, n - k + 1,$$

which represent the *truncated* sizes of clumps (their sizes cannot be greater than k) starting at positions $1, 2, \ldots, n - k + 1$. In order to obtain stationarity, we have also allowed the last clumps to extend further than trial n. When k, n increase so that the expected number of runs $(n - k + 1)p^k$ remains bounded, the processes $\mathbf{Y} = (Y_i)$, $\mathbf{Y}' = (Y_i')$ rarely differ. This is expressed by the following inequality (see Boutsikas (2006), page 511):

$$d_{\mathrm{TV}}(\mathcal{L}(\mathbf{Y}), \mathcal{L}(\mathbf{Y}')) \le P(\mathbf{Y} \ne \mathbf{Y}') \le (n - 2k + 1)qp^{2k} + 2p^{k+1}. \tag{22}$$

We can now use Corollary 12(a) to establish an upper bound for $d_{\mathrm{TV}}(\mathcal{L}(\sum Y_i'), CP)$. We verify that the random variables $Y_1', Y_2', \ldots, Y_{n-k+1}' \in \mathbb{Z}_+$ can be written as in (14) and that they are also $2k$-dependent. Obviously, $p_i = P(Y_i' \ne 0) = qp^k$. For $m = (6k - 2)qp^k < \log 2$, Corollary 12(a) yields the inequality

$$d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n-k+1} Y_i', CP(\lambda, F_k)\right) \le C_{n-k+1} + \frac{\|\Delta^2 f_{CP(\lambda,F_k)}\|_1}{2(1-2(1-\mathrm{e}^{-m}))}$$
$$\times \sum_{i=1}^{n-k+1}\left(\sum_{j=i-2k+1}^{i-1}\big(E(Y_i'Y_j') + E(Y_i')E(Y_j')\big) + \frac{1}{2}E(Y_i')^2\right)$$

(we assume that $Y_i' = 0$ for $i < 1$) with

$$C_{n-k+1} \le 2\sum_{i=1}^{n-k+1}\left(2\sum_{j=2}^{i-6k+2}\sum_{t=j-2k+1}^{j-1}\big(P(Y_t'Y_j' \ne 0) + (qp^k)^2\big) + \sum_{j=1}^{i-6k+2}(qp^k)^2 + 2kqp^k\right)$$
$$\times\left(2\sum_{j=i-2k+1}^{i-1} P(Y_j' \ne 0, Y_i' \ne 0) + 4k(qp^k)^2\right)$$

and $\lambda = (n-k+1)qp^k$, $F_k(x) = P(Y_i' \le x | Y_i' \ne 0)$, $x \in \mathbb{R}$. Notice that, for $i \ge 2k$, $P(Y_i' \ne 0, Y_j' \ne 0)$ is now equal to q^2p^{2k} for $j = i-2k+1, \ldots, i-k-1$, while it vanishes when $j = i-k, \ldots, i-1$. Moreover, $E(Y_i') = q\Sigma_{r=0}^{k-1} p^{k+r} = p^k(1-p^k)$, $i = 1, 2, \ldots, n-k+1$, whereas ($i \ge 2k$)

$$E(Y_j'Y_i') = E\left((1-Z_{j-1})\left(\prod_{l=j}^{j+k-1} Z_l + \prod_{l=j}^{j+k} Z_l + \cdots + \prod_{l=j}^{i-2} Z_l\right)Y_i'\right)$$
$$= qp^k\frac{1-p^{i-j-k}}{1-p}E(Y_i') = p^{2k}(1-p^{i-j-k})(1-p^k), \qquad i-2k+1 \le j \le i-k-1$$

and $E(Y_j'Y_i') = 0$ for $i-k \le j \le i-1$. So, for $i \ge 2k$, we get

$$\sum_{j=i-2k+1}^{i-1}\big(E(Y_i'Y_j') + E(Y_i')E(Y_j')\big)$$
$$= p^{2k}(1-p^k)\left(\left(k-1-p\frac{1-p^{k-1}}{1-p}\right) + (2k-1)(1-p^k)\right) \le p^{2k}(3k-2)$$

and, thus, for $m = (6k-2)qp^k < \log 2$,

$$d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n-k+1} Y_i', CP(\lambda, F_k)\right)$$
$$\le UB_{n,k} := \left(1 + \frac{2}{3\lambda}\right)(6\lambda kqp^k)^2 + \frac{1/4\|\Delta^2 f_{CP(\lambda,F_k)}\|_1}{1-2(1-\mathrm{e}^{-m})}\frac{\lambda}{q}(6k-3)p^k, \tag{23}$$

where $\lambda = (n-k+1)qp^k$ and, for $x = 1, 2, \ldots, k-1$,

$$F_k(x) = P(Y_i' \le x | Y_i' \ne 0) = P\left(1 + \prod_{j=i+k}^{i+k} Z_j + \cdots + \prod_{j=i+k}^{i+2k-2} Z_j \le x\right) = 1 - p^x$$

is the geometric distribution truncated at k ($F_k(k) = 1$). It can be verified that for large λ, $CP(\lambda, F_k) \approx N(\lambda E(W), \lambda E(W^2))$ with $W \sim F_k$ and, according to Remark 1, we expect that

$$\left\| \Delta^2 f_{CP(\lambda, F_k)} \right\|_1 \sim \frac{4}{\lambda E(W^2)\sqrt{2\pi e}} \qquad \text{as } \lambda \to \infty.$$

In order to illustrate the above asymptotic relation, we present below a table with the exact value of the norm $\|\Delta^2 f_{CP(\lambda,F_k)}\|_1$ and its approximation $4/(\lambda E(W^2)\sqrt{2\pi e})$ for several values of λ, p (see Table 1). We assume that $k \to \infty$, that is, F_k is the ordinary geometric distribution and thus $E(W) = 1/q$, $V(W) = p/q^2$ and $E(W^2) = (1+p)/q^2$.

As expected, the above approximation is satisfactory for moderate and large values of λ. Moreover, we observe that it becomes better when p decreases. Assuming that $n, k \to \infty$ with $p \in (0, 1)$ fixed, the compound Poisson approximation error bound in (23) is of order

$$UB_{n,k} \sim \begin{cases} \frac{1}{4}\left\| \Delta^2 f_{CP(\lambda,F_k)} \right\|_1 \frac{\lambda}{q} 6kp^k, & \text{when } \lambda = (n-k+1)qp^k \text{ is fixed,} \\ \frac{6q}{(1+p)\sqrt{2\pi e}} kp^k, & \text{when } \lambda \to \infty, \text{ such that } \lambda^2 kp^k \to 0. \end{cases}$$

For almost the same distance as in (23), the Stein–Chen method offers a bound UB_{CS} such that

$$UB_{CS} \sim \frac{\log^+(\lambda q(1-2p))}{q^2(1-2p)} 6kp^k \qquad \text{when } p \le \frac{1}{3} \quad \text{or} \quad UB_{CS} \sim \frac{6q}{1-5p} kp^k \qquad \text{when } p \le \frac{1}{5}$$

(see, for example, Barbour and Chryssaphinou (2001)). Note that for values of $p > 1/3$, the Stein–Chen method yields bounds of order $O(kp^k + e^{-a_k\lambda})$ or $O(\lambda kp^k)$. The $UB_{n,k}$ is smaller provided that $\lambda^2 kp^k \approx 0$ and is of order $O(kp^k)$ for all values of p.

It is worth mentioning that here, instead of Corollary 12(a), we could employ Corollary 12(b) to obtain a bound even better than $UB_{n,k}$. Specifically, it can be proven that for every $1 \le j <$

Table 1.

	$\lambda = 1$		$\lambda = 5$		$\lambda = 10$		$\lambda = 100$	
	norm	approx.	norm	approx.	norm	approx.	norm	approx.
$p = 0.2$	0.97120	0.516204	0.115414	0.103241	0.054341	0.051620	0.005189	0.005162
$p = 0.5$	1.10364	0.161314	0.040737	0.032263	0.017866	0.016131	0.001628	0.001613
$p = 0.8$	1.32437	0.021509	0.019508	0.004302	0.002474	0.002151	0.000218	0.000215

$i \le n$, the random variables $Y'_j + \cdots + Y'_{i-1}$ and Y'_i are NQD. Hence, we can use Corollary 12(b) and, following an exact parallel to the above procedure, we derive the improved bound

$$UB'_{n,k} := 12\left(1 + \frac{1}{kq} + \frac{2q^2}{\lambda}\right)(\lambda k p^k)^2 + \frac{1/2\|\Delta^2 f_{CP(\lambda,F)}\|_1}{1 - 2(1 - \mathrm{e}^{-m})}\left(1 + \frac{1+p}{2kq}\right)\frac{\lambda}{q} k p^k,$$

which, asymptotically, is about three times smaller than $UB_{n,k}$.

Finally, we can approximate $\sum_{i=1}^{n-k+1} Y_i = \sum_{i=1}^{n-k+1} X_i$, the total number of overlapping success runs within trials $1, 2, \ldots, n$, by $CP(\lambda, G)$, where G denotes the ordinary geometric distribution with parameter p. In this case, $CP(\lambda, G)$ is also known as the Pólya–Aeppli distribution with parameters λ, p and will be denoted by $PA(\lambda, p)$. Using the triangle inequality, the distance $d_{\mathrm{TV}}(\mathcal{L}\sum_{i=1}^{n-k+1} X_i, PA(\lambda, p))$ is bounded above by

$$d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n-k+1} Y_i, \mathcal{L}\sum_{i=1}^{n-k+1} Y'_i\right) + d_{\mathrm{TV}}\left(\mathcal{L}\sum_{i=1}^{n-k+1} Y'_i, CP(\lambda, F_k)\right) + d_{\mathrm{TV}}(CP(\lambda, F_k), PA(\lambda, p)).$$

The first d_{TV} is bounded by (22), the second bounded by (23), whereas for the third, we have (W_i, U_i are independent random variables with $W_i \sim F_k$ and $U_i \sim G$)

$$d_{\mathrm{TV}}(CP(\lambda, F_k), PA(\lambda, p)) = d_{\mathrm{TV}}\left(\sum_{i=1}^{N} W_i, \sum_{i=1}^{N} U_i\right) \le \lambda d_{\mathrm{TV}}(W_1, U_1) = \lambda p^k.$$

Acknowledgments

The second author was partially supported by the University of Athens Research Grant 70/4/8810.

References

Balakrishnan, N. and Koutras, M.V. (2002). *Runs, Scans and Applications*. New York: Wiley. MR1882476

Barbour, A.D. and Chen, L.H.Y. (2005). *Stein's Method and Applications*. Hackensack, NJ: Singapore Univ. Press and World Scientific Publishing. MR2201882

Barbour, A.D., Chen, L.H.Y. and Loh, W.-L. (1992). Compound Poisson approximation for non-negative random variables via Stein's method. *Ann. Probab.* **20** 1843–1866. MR1188044

Barbour, A.D. and Chryssaphinou, O. (2001). Compound Poisson approximation: A user's guide. *Ann. Appl. Probab.* **11** 964–1002. MR1865030

Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford: Clarendon Press. MR1163825

Barbour, A.D. and Xia, A. (2000). Estimating Stein's constants for compound Poisson approximation. *Bernoulli* **6** 581–590. MR1777683

Barbour, A.D. and Utev, S. (1999). Compound Poisson approximation in total variation. *Stochastic Process. Appl.* **82** 89–125. MR1695071

Boutsikas, M.V. (2006). Compound Poisson process approximation for locally dependent real-valued random variables via a new coupling inequality. *Bernoulli* **12** 501–514. MR2232729

Boutsikas, M.V. and Koutras, M.V. (2000). A bound for the distribution of the sum of discrete associated or negatively associated random variables. *Ann. Appl. Probab.* **10** 1137–1150. MR1810868

Boutsikas, M.V. and Koutras, M.V. (2001). Compound Poisson approximation for sums of dependent random variables. In *Probability and Statistical Models with Applications*: *A Volume in Honor of Prof. T. Cacoullos* (C.A. Charalambides, M.V. Koutras and N. Balakrishnan, eds.) 63–86. Boca Raton, FL: Chapman and Hall/CRC press.

Boutsikas, M.V. and Vaggelatou, E. (2002). On the distance between convex-ordered random variables. *Adv. in Appl. Probab.* **34** 349–374. MR1909919

Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* **14** 663–676. MR0832029

Deheuvels, P., Pfeifer, D. and Puri, M.L. (1989). A new semigroup technique in Poisson approximation. *Semigroup Forum* **38** 189–201. MR0976202

Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197. MR0142174

Lindvall, T. (1992). *Lectures on the Coupling Method*. New York: Wiley. MR1180522

Rachev, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models*. Chichester: Wiley. MR1105086

Roos, B. (2003). Kerstan's method for compound Poisson approximation. *Ann. Probab.* **31** 1754–1771. MR2016599

Senatov, V.V. (1980). Uniform estimates of the rate of convergence in the multi-dimensional central limit theorem. *Theory Probab. Appl.* **25** 745–759.

Zolotarev, V.M. (1983). Probability metrics. *Theory Probab. Appl.* **28** 278–302. MR0700210

Received March 2008 and revised November 2008