**ESAIM: Probability and Statistics** October 1999, Vol. 3, p. 131–150
URL: http://www.emath.fr/ps/

# POISSON PERTURBATIONS *

Andrew D. Barbour[1] and Aihua Xia[2]

**Abstract**. Stein's method is used to prove approximations in total variation to the distributions of integer valued random variables by (possibly signed) compound Poisson measures. For sums of independent random variables, the results obtained are very explicit, and improve upon earlier work of Kruopis (1983) and Čekanavičius (1997); coupling methods are used to derive concrete expressions for the error bounds. An example is given to illustrate the potential for application to sums of dependent random variables.

**Résumé**. On utilise la méthode de Stein pour approximer, par rapport à la variation totale, la distribution d'une variable aléatoire aux valeurs entières par une mesure (éventuellement signée) de Poisson composée. Pour les sommes de variables aléatoires indépendantes, on obtient des résultats très explicites ; les estimations de la précision de l'approximation, construites à l'aide de la méthode de "coupling", sont plus exactes que celles de Kruopis (1983) et de Čekanivičius (1997). Un exemple sert à illustrer le potentiel de la méthode envers les sommes de variables aléatoires dépendantes.

**AMS Subject Classification.** 62E17, 60G50, 60F05.

Received January 28, 1999. Revised July 26, 1999.

## 1. Introduction

In a series of papers, beginning with the pioneering work of Presman (1983) and Kruopis (1986), and continuing with those of Čekanavičius *et al.* (see for example, the references in Čekanavičius 1997), it has been shown that signed compound Poisson measures can be used to make very close approximations in total variation to the distributions of sums of independent integer valued random variables. Signed compound Poisson measures $\boldsymbol{\mu}$ on the integers are measures of the form $\boldsymbol{\mu} = \exp\{\lambda(F - E)\}$ for some $\lambda \in \mathbb{R}$, where $F$ is any probability distribution on the integers and $E$ is the unit mass on 0: multiplication is interpreted as convolution, and the exponential is defined through its power series. Such measures may be signed measures if $\lambda < 0$, but it is always the case that $\boldsymbol{\mu}\{\mathbb{Z}\} = 1$.

In this paper, we confine our attention to the very small subset of such measures of the form $\boldsymbol{\pi}_{\mu,a}$, for $\mu > 0$ and $a \in \mathbb{R}$, having generating function

$$\hat{\boldsymbol{\pi}}_{\mu,a}(z) := \sum_{r\geq 0} z^r \boldsymbol{\pi}_{\mu,a}\{r\} = \exp\left\{\mu(z-1) + \tfrac{1}{2}a(z^2-1)\right\}, \tag{1.1}$$

which are concentrated on $\mathbb{Z}_+$, together with their translates

$$\boldsymbol{\pi}^{(m)}_{\mu,a},\ m \in \mathbb{Z}; \qquad \boldsymbol{\pi}^{(m)}_{\mu,a}\{s+m\} := \boldsymbol{\pi}_{\mu,a}\{s\}, \quad s \in \mathbb{Z}_+. \tag{1.2}$$

*Keywords and phrases:* Stein's method, signed compound Poisson measure, total variation, coupling.

* *Andrew Barbour was supported in part by Schweizerischer NF Projekt Nr 20-50686.97.*

[1] Abteilung für Angewandte Mathematik, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; e-mail: adb@amath.unizh.ch

[2] Department of Statistics, School of Mathematics, The University of New South Wales, Sydney 2052, Australia.

The generating function (1.1) is that of the convolution of the Poisson distribution $\mathrm{Po}(\mu)$ and, for $a > 0$, the distribution of $2Z$, where $Z \sim \mathrm{Po}(a/2)$; for $a < 0$, the measure $\boldsymbol{\pi}_{\mu,a}$ must be a signed measure, in view of Raikov's (1938) theorem. For many purposes, this class is already wide enough to give good approximations; indeed, we often assume that $\mu$ is large but $a$ is bounded, in which case $\boldsymbol{\pi}_{\mu,a}$ is a small perturbation of $\mathrm{Po}(\mu)$, a Poisson distribution with large parameter. In these circumstances, $\boldsymbol{\pi}_{\mu,a}\{r\}$ is positive even when $a < 0$, except for very large values of $r$, and the total negative mass is exponentially small with $\mu$: hence the signed measure $\boldsymbol{\pi}_{\mu,a}$ could actually be replaced as an approximation by a probability distribution, with a very small change in total variation.

In order to use measures from the family $\boldsymbol{\pi}_{\mu,a}$ as approximations, we need to have a way of showing how close they are to the distribution of any given random variable. Previous work has largely concentrated on Fourier methods, limiting the scope of applicability. Here, we show how Stein's method can be used. For $a > 0$, the measures $\boldsymbol{\pi}_{\mu,a}$ are compound Poisson distributions, and the general theory in Barbour and Utev (1998, 1999) could be invoked. However, the particular form of the $\boldsymbol{\pi}_{\mu,a}$ allows us to prove better bounds on the solutions of the Stein equation by a much easier argument, which is also valid for $a < 0$, provided that $\mu$ is large enough by comparison to $|a|$. Indeed, the same method can be used to establish bounds on the solutions of the Stein equation (Th. 2.5) which are of optimal order in $\mu$, for a whole class of compound Poisson distributions. Bounds of this order were previously known only for the Poisson distribution.

We first use Stein's method to demonstrate that the measures $\boldsymbol{\pi}_{\mu,a}$ are often close to other, better known distributions, so that approximation with respect to $\boldsymbol{\pi}_{\mu,a}$ can then be more easily understood. We then show how to use the $\boldsymbol{\pi}_{\mu,a}$ to prove approximation theorems with respect to *total variation* distance for sums of independent integer valued random variables, under much the same circumstances as are required for the usual central limit theorem, and we give very explicit error bounds. For sums of independent indicator random variables, our bounds have much better constants than those of Kruopis; for more general summands, explicit bounds seem to be new. In particular, we show that the family (1.2), restricted to have $0 \le a < 1$, can frequently be used to give good approximations. This family consists entirely of probability distributions, and comes as near as possible on $\mathbb{Z}$ to a family of translates of the Poisson distribution $\mathrm{Po}(\mu)$ by any *real* displacement; furthermore, for large $\mu$, these distributions are extremely close to the negative binomial distribution with the same mean and variance. Finally, we use a very simple example to illustrate that approximations of this kind can in principle also be obtained for sums of dependent random variables.

## 2. A Stein equation

It follows directly from the definition in (1.1) that

$$\mu\boldsymbol{\pi}_{\mu,a}\{0\} = \boldsymbol{\pi}_{\mu,a}\{1\}; \quad a\boldsymbol{\pi}_{\mu,a}\{r-2\} + \mu\boldsymbol{\pi}_{\mu,a}\{r-1\} = r\boldsymbol{\pi}_{\mu,a}\{r\},\ r \ge 2, \tag{2.1}$$

from which it follows that $\boldsymbol{\pi}_{\mu,a}\{f\} = 0$ for all $f$ of the form

$$f(j) = ag(j+2) + \mu g(j+1) - jg(j) \tag{2.2}$$

for bounded $g$, where, for a function $f$ and a measure $\boldsymbol{\mu}$, $\boldsymbol{\mu}\{f\} := \sum_r f(r)\boldsymbol{\mu}\{r\}$. This suggests a Stein equation for the measure $\boldsymbol{\pi}_{\mu,a}$. The following lemma gives the necessary properties of the solutions.

**Lemma 2.1.** *For any $\mu > 0$, $a \in \mathbb{R}$ such that $\mu + a > 0$ and $\theta = |a|(\mu+a)^{-1} < 1/2$, and for any bounded $f : \mathbb{Z}_+ \to \mathbb{R}$, there is a solution $g = g_f : \mathbb{Z}_+ \to \mathbb{R}$ to the Stein equation*

$$ag(j+2) + \mu g(j+1) - jg(j) = f(j) - \boldsymbol{\pi}_{\mu,a}\{f\}; \quad j \ge 0, \tag{2.3}$$

*which satisfies*

$$\|g_f\| \le \frac{2}{1-2\theta}\,(\mu+a)^{-1/2}\|f\|; \qquad \|\Delta g_f\| \le \frac{2}{1-2\theta}\,(\mu+a)^{-1}\|f\|, \tag{2.4}$$

*where $\Delta g(j) := g(j+1) - g(j)$ and $\|\cdot\|$ applied to functions denotes the supremum norm.*

*Proof.* We construct $g_f$ as a perturbation of the solution $g_0$ to the Stein equation

$$(\mu+a)g_0(j+1) - jg_0(j) = f(j) - \mathrm{Po}(\mu+a)\{f\}, \quad j \geq 0, \tag{2.5}$$

for the Poisson distribution $\mathrm{Po}(\mu+a)$ with mean $\mu+a$; $g_0$ satisfies (Barbour *et al.*, Lem. 1.1.1)

$$\|g_0\| \leq 2(\mu+a)^{-1/2}\|f\|; \quad \|\Delta g_0\| \leq 2(\mu+a)^{-1}\|f\|. \tag{2.6}$$

To do this, define $Tg$ for any bounded function $g : \mathbb{Z}_+ \to \mathbb{R}$ to be the solution $\tilde{g}$ of the equation

$$(\mu+a)\tilde{g}(j+1) - j\tilde{g}(j) = f(j) - \mathrm{Po}(\mu+a)\{f\} - a\Delta g(j+1) + a\mathrm{Po}(\mu+a)\{\Delta g(\cdot+1)\},\ j \geq 0. \tag{2.7}$$

Consider the sequence $(g_n,\ n \geq 0)$ defined by $g_n = Tg_{n-1}$, $n \geq 1$, with $g_0$ as in (2.5). Then, writing $h_n(j) = g_n(j) - g_{n-1}(j)$, we have

$$(\mu+a)h_n(j+1) - jh_n(j) = -a\Delta h_{n-1}(j+1) + a\mathrm{Po}(\mu+a)\{\Delta h_{n-1}(\cdot+1)\}, \tag{2.8}$$

so that, from (2.6),

$$\|h_n\| \leq 2|a|(\mu+a)^{-1/2}\|\Delta h_{n-1}\|; \tag{2.9}$$

$$\|\Delta h_n\| \leq 2|a|(\mu+a)^{-1}\|\Delta h_{n-1}\| = 2\theta\|\Delta h_{n-1}\|. \tag{2.10}$$

From (2.10) and (2.6), it then follows that

$$\|\Delta h_n\| \leq (2\theta)^n\|\Delta g_0\| \leq 2(2\theta)^n(\mu+a)^{-1}\|f\|;$$

in conjunction with (2.9), this gives

$$\|h_n\| \leq 2(2\theta)^n(\mu+a)^{-1/2}\|f\|.$$

Hence $g_f = \lim_{n\to\infty} g_n$ exists uniformly and satisfies $Tg_f = g_f$, and

$$\|g_f\| \leq \sum_{n\geq 1} 2(2\theta)^n(\mu+a)^{-1/2}\|f\| + \|g_0\| \leq \frac{2}{1-2\theta}(\mu+a)^{-1/2}\|f\|.$$

Furthermore, since $Tg_f = g_f$, it follows from (2.6) and (2.7) that

$$\|\Delta g_f\| \leq 2(\mu+a)^{-1}\{\|f\| + |a|\|\Delta g_f\|\},$$

so that $\|\Delta g_f\| \leq \frac{2}{1-2\theta}(\mu+a)^{-1}\|f\|$. Finally, again since $Tg_f = g_f$, it follows from (2.7) that

$$ag_f(j+2) + \mu g_f(j+1) - jg_f(j) = f(j) - \mathrm{Po}(\mu+a)\{f\} + a\mathrm{Po}(\mu+a)\{\Delta g_f(\cdot+1)\}, \tag{2.11}$$

and applying $\boldsymbol{\pi}_{\mu,a}$ to both sides gives

$$0 = \boldsymbol{\pi}_{\mu,a}\{f\} - \mathrm{Po}(\mu+a)\{f\} + a\mathrm{Po}(\mu+a)\{\Delta g_f(\cdot+1)\},$$

so that, from (2.11), the function $g_f$ indeed satisfies (2.3). □

The following corollary is immediate. The notation $\|\cdot\|$ is used with measures to denote the total variation norm: if $\boldsymbol{\nu}$ is a (signed) measure on $\mathbb{Z}$, with positive and negative parts $\boldsymbol{\nu}^+$ and $\boldsymbol{\nu}^-$, then $\|\boldsymbol{\nu}\| = \boldsymbol{\nu}^+(\mathbb{Z}) - \boldsymbol{\nu}^-(\mathbb{Z})$.

**Corollary 2.2.** *Suppose that $\mu, a \in \mathbb{R}$ satisfy $\mu+a > 0$ and $\theta = |a|(\mu+a)^{-1} < 1/2$. If $W$ is any random variable on $\mathbb{Z}_+$ such that*

$$|\mathbb{E}\{ag(W+2) + \mu g(W+1) - Wg(W)\}| \leq \varepsilon \tag{2.12}$$

*for all $g = g_f$ as in Lemma 2.1 with $\|f\| \le 1$, then $\|\mathcal{L}(W) - \boldsymbol{\pi}_{\mu,a}\| \le \varepsilon$. In particular, if*

$$|\mathbb{E}\{ag(W+2) + \mu g(W+1) - Wg(W)\}| \le \varepsilon_0\|g\| + \varepsilon_1\|\Delta g\| \tag{2.13}$$

*for all bounded $g : \mathbb{Z}_+ \to \mathbb{R}$, it follows that*

$$\|\mathcal{L}(W) - \boldsymbol{\pi}_{\mu,a}\| \le \frac{2}{1-2\theta}\left\{(\mu+a)^{-1/2}\varepsilon_0 + (\mu+a)^{-1}\varepsilon_1\right\}.$$

If $W$ can take values in the whole of $\mathbb{Z}$, the corresponding result is a little more complicated.

**Corollary 2.3.** *Suppose that $\mu, a \in \mathbb{R}$ satisfy $\mu + a > 0$ and $\theta = |a|(\mu+a)^{-1} < 1/2$. Let $W$ be a random variable on $\mathbb{Z}$ such that*

$$|\mathbb{E}\{ag(W+2) + \mu g(W+1) - Wg(W)\}| \le \varepsilon_0\|g\| + \varepsilon_1\|\Delta g\| \tag{2.14}$$

*for all bounded $g : \mathbb{Z} \to \mathbb{R}$. Then it follows that*

$$\|\mathcal{L}(W) - \boldsymbol{\pi}_{\mu,a}\| \le \frac{2}{1-2\theta}\left\{(\mu+a)^{-1/2}\varepsilon_0 + (\mu+a)^{-1}\varepsilon_1 + (1-\theta)\mathbb{P}[W \le -1]\right\}. \tag{2.15}$$

*In particular, if also $\mathbb{E}W = \mu + a$ and $\mathrm{Var}\,W = \mu + 2a$, then*

$$\|\mathcal{L}(W) - \boldsymbol{\pi}_{\mu,a}\| \le \frac{2}{1-2\theta}\left\{(\mu+a)^{-1/2}\varepsilon_0 + (\mu+a)^{-1}[\varepsilon_1 + 1]\right\}. \tag{2.16}$$

*Proof.* For bounded $f : \mathbb{Z}_+ \to \mathbb{R}$, take $g_f$ as given by Lemma 2.1, and extend to $g_f : \mathbb{Z} \to \mathbb{R}$ by setting $g_f(j) = 0$, $j \le 0$. Note that $g_f(0)$ is not actually defined by (2.3), but that, from (2.3) with $j = 0$ and from (2.4),

$$(\mu+a)|g_f(1)| = |f(0) - \boldsymbol{\pi}_{\mu,a}\{f\} - a(g_f(2) - g_f(1))| \le (1 + \|\boldsymbol{\pi}_{\mu,a}\|)\|f\| + |a|\,\|\Delta g_f\| \le \frac{2}{1-2\theta}\|f\|,$$

so that this extension continues to satisfy (2.4), provided that we have $\|\boldsymbol{\pi}_{\mu,a}\| \le (1-2\theta)^{-1}$.

For $a \ge 0$, it is immediate that $\|\boldsymbol{\pi}_{\mu,a}\| = 1 \le (1-2\theta)^{-1}$. For $a < 0$, take any $f$ with $\|f\| \le 1$, and use (2.5) to give

$$\boldsymbol{\pi}_{\mu,a}\{f\} - \mathrm{Po}(\mu+a)\{f\} = \sum_{j\ge 0}((\mu+a)g_0(j+1) - jg_0(j))\boldsymbol{\pi}_{\mu,a}\{j\} = -a\boldsymbol{\pi}_{\mu,a}\{\Delta g_0(\cdot+1)\},$$

the last line coming from (2.2), giving

$$\|\boldsymbol{\pi}_{\mu,a}\| \le 1 + 2|a|(\mu+a)^{-1}\|\boldsymbol{\pi}_{\mu,a}\|$$

from (2.6). Hence, if $a < 0$ and $\theta = |a|(\mu+a)^{-1} < 1/2$, then

$$\|\boldsymbol{\pi}_{\mu,a}\| \le (1-2\theta)^{-1}. \tag{2.17}$$

In terms now of functions on $\mathbb{Z}$, (2.3) is equivalent to

$$(f(j) - \boldsymbol{\pi}_{\mu,a}\{f\})1_{\{j\ge 0\}} = \{ag_f(j+2) + \mu g_f(j+1) - jg_f(j)\}1_{\{j\ge 0\}}, \tag{2.18}$$

defining the value of $h_f(j) = ag_f(j+2) + \mu g_f(j+1) - jg_f(j)$ for $j \ge 0$, the definition of $g_f$ then giving $h_f(-1) = ag_f(1)$ for $j = -1$, and $h_f(j) = 0$ for $j \le -2$. Hence, from (2.14), it follows that, for any bounded $f : \mathbb{Z} \to \mathbb{R}$,

$$|ag_f(1)\mathbb{P}[W = -1] + \mathbb{E}\{(f(W) - \boldsymbol{\pi}_{\mu,a}\{f\})1_{\{W\ge 0\}}\}| \le \varepsilon_0\|g_f\| + \varepsilon_1\|\Delta g_f\|,$$

and hence that

$$|\mathbb{E}f(W) - \boldsymbol{\pi}_{\mu,a}\{f\}| \le 2\|f\|\mathbb{P}[W \le -1] + \frac{2}{1-2\theta}\,|a|(\mu+a)^{-1}\|f\|\mathbb{P}[W=-1]$$
$$+ \frac{2}{1-2\theta}\,\{(\mu+a)^{-1/2}\varepsilon_0 + (\mu+a)^{-1}\varepsilon_1\},$$

completing the proof of (2.15). The particular case then follows by applying Chebyshev's inequality. □

**Remark.** The Chebyshev bound for $\mathbb{P}[W \le -1]$ can of course be improved, if more information about $W$ is available.

The following corollary can also prove useful.

**Corollary 2.4.** *Suppose that $\mu_i, a_i$ satisfy $\mu_i + a_i > 0$ and $\theta_i = |a_i|(\mu_i + a_i)^{-1} < 1/2$, $i = 1, 2$. Then*

$$\|\boldsymbol{\pi}_{\mu_1,a_1} - \boldsymbol{\pi}_{\mu_2,a_2}\| \le k(\mu_1 + a_1)^{-1/2}\{|\mu_1 - \mu_2| + |a_1 - a_2|\},$$

*with $k = 2/(1-2\theta_1)$ if $a_2 \ge 0$, $k = 2/\{(1-2\theta_1)(1-2\theta_2)\}$ if $a_2 < 0$.*

*Proof.* Apply $\boldsymbol{\pi}_{\mu_2,a_2}$ to (2.3) with $a = a_1$ and $\mu = \mu_1$, using the fact that $\boldsymbol{\pi}_{\mu_2,a_2}\{f\} = 0$ for all $f$ as given in (2.2) with $a = a_2$ and $\mu = \mu_2$. This, together with (2.4), gives

$$|\boldsymbol{\pi}_{\mu_2,a_2}\{f\} - \boldsymbol{\pi}_{\mu_1,a_1}\{f\}| \le \{|\mu_1 - \mu_2| + |a_1 - a_2|\}\frac{2}{1-2\theta_1}(\mu_1 + a_1)^{-1/2}\|f\|\,\|\boldsymbol{\pi}_{\mu_2,a_2}\|.$$

The corollary follows by taking arbitrary $f$ such that $\|f\| \le 1$, and using (2.17) to bound $\|\boldsymbol{\pi}_{\mu,a}\|$. □

The perturbation argument of Lemma 2.1 can easily be modified to cover a more general class of compound Poisson distributions. This class is particularly useful in problems where a crude Poisson approximation is to be refined.

**Theorem 2.5.** *Let $\lambda_i \in \mathbb{R}_+$, $i \ge 1$, satisfy*

$$\theta := \left(\sum_{i\ge1} i\lambda_i\right)^{-1} \sum_{i\ge2} i(i-1)\lambda_i < 1/2. \tag{2.19}$$

*Then, for any $A \subset \mathbb{Z}_+$, the Stein equation*

$$\sum_{i\ge1} i\lambda_i g(j+i) - jg(j) = 1_A(j) - \mathrm{CP}\,(\boldsymbol{\lambda})\{A\}, \qquad j \ge 0, \tag{2.20}$$

*for the compound Poisson distribution* $\mathrm{CP}\,(\boldsymbol{\lambda})$ *has a solution $g = g_A$ satisfying*

$$\|g_A\| \le \frac{1}{1-2\theta}\left(\sum_{i\ge1} i\lambda_i\right)^{-1/2}; \qquad \|\Delta g_A\| \le \frac{1}{1-2\theta}\left(\sum_{i\ge1} i\lambda_i\right)^{-1}. \tag{2.21}$$

These simple bounds are in stark contrast to the much more complicated behaviour of the solutions to the Stein equation (2.20) which can occur when $\theta > 1/2$. Note also that, by defining

$$\theta := \left(\sum_{i\ge1} i\lambda_i\right)^{-1} \sum_{i\ge2} i(i-1)|\lambda_i|,$$

signed compound Poisson measures such that $\sum_{i\ge1} i\lambda_i > 0$ and $\theta < 1/2$ could also be covered. If only $\lambda_1$ and $\lambda_2$ are non–zero, the definition of $\theta$ reduces to that of Lemma 2.1.

## 3. Approximating $\boldsymbol{\pi}_{\mu,a}$

The measures $\boldsymbol{\pi}_{\mu,a}$ are not immediately familiar, especially since, for $a < 0$, they are not probability measures. Our first use of the results of the previous section is therefore to show that the $\boldsymbol{\pi}_{\mu,a}$ are often close to more widely used distributions. To start with, we show that, at least when $|a| \ll \mu$, the measures $\boldsymbol{\pi}_{\mu,a}$ can be seen as small perturbations of the Poisson distribution $\mathrm{Po}(\mu)$. More precisely, for $\mu \geq 1$ and $c_1, c_2 \in \mathbb{R}$, let $\boldsymbol{\nu}(\mu, c_1, c_2)$ denote the (still possibly signed) measure on $\mathbb{Z}_+$ defined by

$$\boldsymbol{\nu}(\mu, c_1, c_2)\{s\} = \mathrm{Po}(\mu)\{s\} \left(1 + c_1\mu^{-1}(s-\mu) + \tfrac{1}{2}c_2\mu^{-2}\{(s-\mu)^2 - \mu\}\right), \tag{3.1}$$

$s \in \mathbb{Z}_+$, which satisfies $\boldsymbol{\nu}(\mu, c_1, c_2)\{\mathbb{Z}_+\} = 1$. As for $\boldsymbol{\pi}_{\mu,a}$, if $\mu$ is large and $c_1$, $c_2$ remain bounded, there are probability measures differing in total variation from the measure $\boldsymbol{\nu}(\mu, c_1, c_2)$ by an amount which is exponentially small in $\mu$. Under these circumstances, we show in the next two theorems that measures close to $\boldsymbol{\nu}(\mu, c_1, c_2)$ are also close to the measure $\boldsymbol{\pi}_{\mu',a'}$ with

$$\mu' = \mu + c_1 - (c_2 - c_1^2) \quad \text{and} \quad a' = c_2 - c_1^2. \tag{3.2}$$

We begin with a lemma, showing that $\boldsymbol{\nu}(\mu, c_1, c_2)$ almost satisfies a Stein equation of the form (2.3).

**Lemma 3.1.** *Define $\mu'$ and $a'$ as in (3.2), and suppose that $\mu$, $c_1$ and $c_2$ are such that $\mu' + a' \geq \mu/2$. Let $g : \mathbb{Z}_+ \to \mathbb{R}$ be any function satisfying $\|g\| \leq 4(\mu' + a')^{-1/2}$ and $\|\Delta g\| \leq 4(\mu' + a')^{-1}$, and define $h$ by*

$$h(j) = a'g(j+2) + \mu'g(j+1) - jg(j).$$

*Then $|\boldsymbol{\nu}(\mu, c_1, c_2)\{h\}| \leq k_{\{3.1\}}(\mu, c_1, c_2)\mu^{-3/2}$, where*

$$k_{\{3.1\}}(\mu, c_1, c_2) := 4\left\{(|c_1|\,|2a' + c_2| + |c_2|) + \mu^{-1/2}(2|a'| + |c_1|)|c_2| + \mu^{-1}|a'c_2|\right\}.$$

*Proof.* By definition, we have

$$\boldsymbol{\nu}(\mu, c_1, c_2)\{h\} = \mathbb{E}\Big\{\left(1 + c_1\mu^{-1}(Z-\mu) + \tfrac{1}{2}c_2\mu^{-2}\{(Z-\mu)^2 - \mu\}\right) \times (a'g(Z+2) + \mu'g(Z+1) - Zg(Z))\Big\}, \tag{3.3}$$

where $Z \sim \mathrm{Po}(\mu)$. Now $\mathbb{E}(Zg(Z)) = \mu\mathbb{E}g(Z+1)$ for any $g$ for which the expectations exist; using this identity to eliminate powers of $Z$ in (3.3) yields, after some computation,

$$2\boldsymbol{\nu}(\mu, c_1, c_2)\{h\} = \mathbb{E}\left\{k_1\Delta^2 g(Z+1) + a'c_2\Delta^3 g(Z+1) + c_2\mu^{-1}[a'\Delta^2 g(Z+1) + c_1\Delta g(Z+1) - g(Z+1)]\right\},$$

where $k_1 = c_1(2a' + c_2)$, and $a' = c_2 - c_1^2$, as in (3.2). The last two terms have expectations bounded by $8|c_1c_2|\mu^{-2}$ and $8|c_2|\mu^{-3/2}$ respectively, because of the bounds on $\|g\|$ and $\|\Delta g\|$ and because $\mu' + a' \geq \mu/2$. Then the bound on $\|\Delta g\|$ also shows that

$$|\mathbb{E}\Delta^2 g(Z+1)| \leq 4(\mu' + a')^{-1}\sum_{j\geq 1}|\mathbb{P}[Z=j] - \mathbb{P}[Z=j-1]| \leq 8\sqrt{2e^{-1}}\mu^{-3/2},$$

and that

$$|\mathbb{E}\Delta^3 g(Z+1)| \leq 4(\mu' + a')^{-1}\sum_{j\geq 1}|\mathbb{P}[Z=j] - 2\mathbb{P}[Z=j-1] + \mathbb{P}[Z=j-2]| \leq 16\mu^{-2},$$

by Barbour *et al.*, pp. 222 and 224. The lemma now follows by collecting terms. □

As a consequence of Lemma 3.1, we have the following approximation.

**Theorem 3.2.** *For* $\mu \geq \max\{1, 8|c_2 - c_1^2|, 2|c_1|\}$, *we have*

$$\|\boldsymbol{\nu}(\mu, c_1, c_2) - \boldsymbol{\pi}_{\mu',a'}\| \leq k_{\{3.1\}}(\mu, c_1, c_2)\mu^{-3/2},$$

*where* $\mu'$ *and* $a'$ *are as in (3.2).*

*Proof.* Take any $f : \mathbb{Z}_+ \to \mathbb{R}$ with $\|f\| \leq 1$, and construct $g_f$ using Lemma 2.1 with $\mu'$ for $\mu$ and $a'$ for $a$. Note that, by the assumption on $\mu$, $\mu' + a' \geq \mu/2$ and $4|a'| \leq (\mu' + a')$; hence, from Lemma 2.1, $\|g_f\| \leq 4(\mu' + a')^{-1/2}$ and $\|\Delta g_f\| \leq 4(\mu' + a')^{-1}$. Now apply Lemma 3.1 in conjunction with (2.3) to give

$$|\boldsymbol{\nu}(\mu, c_1, c_2)\{f\} - \boldsymbol{\pi}_{\mu',a'}\{f\}| \leq k_{\{3.1\}}(\mu, c_1, c_2)\mu^{-3/2}, \tag{3.4}$$

from which the theorem follows. □

Note that $k_{\{3.1\}}(\mu, c_1, 0) = 8|c_1|^3$ and that $k_{\{3.1\}}(\mu, 0, c_2) = 4|c_2|(1 + (2\mu^{-1/2} + \mu^{-1})|c_2|)$. Thus $\boldsymbol{\nu}(\mu, c_1, 0)$ is close to $\boldsymbol{\pi}_{\mu',a'}$ for all $|c_1| \ll \mu^{1/2}$, and $\boldsymbol{\nu}(\mu, 0, c_2)$ is close to $\boldsymbol{\pi}_{\mu',a'}$ for all $|c_2| \ll \mu$. Hence, in particular, $\boldsymbol{\pi}_{\mu,a}$ can be reasonably approximated by the member $\boldsymbol{\nu}(\mu + a, 0, a)$ of the family $\boldsymbol{\nu}(\mu, c_1, c_2)$, as long as $|a| \ll \mu$.

The theorems given so far are useful when $\boldsymbol{\pi}_{\mu,a}$ is a genuinely small perturbation of $\mathrm{Po}(\mu)$, which is the case if $|a| \ll \mu$. However, it is at times useful to allow $|a|$ to be of the same order of magnitude as $\mu$, and still have simple probability approximations to $\boldsymbol{\pi}_{\mu,a}$. When $a > 0$, $\boldsymbol{\pi}_{\mu,a}$ is already a probability measure, and so no further approximation is required. However, this family of compound Poisson distributions is less well known than the negative binomial family, using which good approximations can often also be obtained, as in Corollary 4.8. If $a < 0$, the obvious family to use is the binomial $\mathrm{Bi}(n, p)$. Here, the fact that $n$ has to be integral requires some small adjustment. Matching mean and variance requires that $np = \mu + a$ and that $np(1 - p) = \mu + 2a$, or, equivalently, that $n = |a|^{-1}(\mu + a)^2$ and that $p = |a|(\mu + a)^{-1}$. If $n$ so defined is not integral, choose $\varepsilon$ to satisfy $|a|^{-1}(\mu - \varepsilon + a)^2 = n = \lfloor |a|^{-1}(\mu + a)^2 \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer $m \leq x$, and set $p = |a|(\mu - \varepsilon + a)^{-1}$. We can then use $\mathrm{Bi}(n, p)$ to approximate $\boldsymbol{\pi}_{\mu-\varepsilon,a}$, and $\mathrm{Bi}(n, p) * \mathrm{Po}(\varepsilon)$ to approximate $\boldsymbol{\pi}_{\mu,a}$, since the convolution $\boldsymbol{\pi}_{\mu-\varepsilon,a} * \mathrm{Po}(\varepsilon)$, for any $0 \leq \varepsilon \leq \mu$, is just $\boldsymbol{\pi}_{\mu,a}$. Our choice of $\varepsilon$ is typically of order $\mu^{-1}|a|$, and thus the convolution $\mathrm{Po}(\varepsilon) * \mathrm{Bi}(n, p)$ is a rather simple distribution with which to approximate $\boldsymbol{\pi}_{\mu,a}$. Binomial approximation is treated later in Theorem 4.1 as a special case of sums of independent indicators.

The binomial and negative binomial approximations are not the only possible choices. Another possibility is to approximate $\boldsymbol{\pi}_{\mu,a}$ by a translate of a compound Poisson probability distribution, $\boldsymbol{\pi}_{\chi,b}^{(m)}$ as defined in (1.2), with $b \geq 0$. To achieve this, we first need to show that translating $\boldsymbol{\pi}_{\mu,a}$ by 1 changes the measure by at most $O(\mu^{-1/2})$ in total variation.

**Lemma 3.3.** *For any* $\mu > 0$ *and* $a \in \mathbb{Z}$ *such that* $\theta = (\mu + a)^{-1}|a| < 1/2$,

$$\left\| \boldsymbol{\pi}_{\mu,a} - \boldsymbol{\pi}_{\mu,a}^{(1)} \right\| \leq k(\mu + a)^{-1/2}, \tag{3.5}$$

*where* $k$ *may be taken to be* $2/(1 - 2\theta)$ *if* $a > 0$, *and* $2/(1 - 2\theta)^2$ *otherwise.*

*Proof.* Use Lemma 2.1 and (2.2) to give

$$\begin{aligned} \boldsymbol{\pi}_{\mu,a}\{f(\cdot + 1)\} - \boldsymbol{\pi}_{\mu,a}\{f\} &= \sum_{j \geq 0} (a g_f(j + 3) + \mu g_f(j + 2) - (j + 1) g_f(j + 1)) \boldsymbol{\pi}_{\mu,a}\{j\} \\ &= -\boldsymbol{\pi}_{\mu,a}\{g_f(\cdot + 1)\}, \end{aligned}$$

with $\|g_f(\cdot + 1)\| \leq \frac{2}{1-2\theta}(\mu + a)^{-1/2}\|f\|$. Hence

$$|\boldsymbol{\pi}_{\mu,a}\{f(\cdot + 1)\} - \boldsymbol{\pi}_{\mu,a}\{f\}| \leq \frac{2}{1 - 2\theta}(\mu + a)^{-1/2}\|\boldsymbol{\pi}_{\mu,a}\|\,\|f\|.$$

If $a \geq 0$, $\|\boldsymbol{\pi}_{\mu,a}\| = 1$, and if $a < 0$ we have $\|\boldsymbol{\pi}_{\mu,a}\| \leq 1/(1 - 2\theta)$ from (2.17), proving the lemma. □

We are now in a position to prove our approximation by a translate of a compound Poisson distribution. We use $\lceil x \rceil$ to denote the smallest integer $m \geq x$.

**Theorem 3.4.** *Suppose that $a < 0$ and that $\mu \geq 5(|a| + 1)$. Define $m = \lceil |a| \rceil$, $b = a + m = \lceil |a| \rceil - |a|$ and $\chi = \mu - 2m$. Then*

$$\left\| \boldsymbol{\pi}_{\mu,a} - \boldsymbol{\pi}^{(m)}_{\chi,b} \right\| \leq 21(|a| + 1)(\mu + a)^{-3/2}.$$

*If $a \geq 1$ and $\mu \geq 3a$, take $m = -\lfloor a \rfloor$, $b = a - \lfloor a \rfloor$ and $\chi = \mu - 2m$, giving*

$$\left\| \boldsymbol{\pi}_{\mu,a} - \boldsymbol{\pi}^{(m)}_{\chi,b} \right\| \leq 16a(\mu + a)^{-3/2}.$$

**Remark.** If $a \in \mathbb{Z}$, then $b = 0$ and $\boldsymbol{\pi}^{(m)}_{\chi,b}$ is a shifted Poisson distribution. In all cases, $0 \leq b < 1$, so that $\boldsymbol{\pi}_{\chi,b}$ is close to being Poisson if $\chi$ is large. Corollary 4.8 shows that $\boldsymbol{\pi}_{\chi,b}$ is even closer to the negative binomial distribution with the same mean and variance.

*Proof.* Let $Z \sim \boldsymbol{\pi}_{\chi,b}$, so that, from (2.2),

$$\mathbb{E}\{bg(Z + 2) + \chi g(Z + 1) - Zg(Z)\} = 0 \tag{3.6}$$

for all bounded $g$. Now take any bounded $f$, and use Lemma 2.1 to show that

$$\mathbb{E}f(Z + m) - \boldsymbol{\pi}_{\mu,a}\{f\} = \mathbb{E}\{ag_f(Z + m + 2) + \mu g_f(Z + m + 1) - (Z + m)g_f(Z + m)\},$$

where $g_f(\cdot + m)$ is bounded as in (2.4). Applying (3.6), it then follows that

$$\begin{aligned}\mathbb{E}f(Z + m) - \boldsymbol{\pi}_{\mu,a}\{f\} &= \mathbb{E}\{(a - b)g_f(Z + m + 2) + (\mu - \chi)g_f(Z + m + 1) - mg_f(Z + m)\} \\ &= \mathbb{E}\{(a - b + m)g_f(Z + m + 2) + (\mu - \chi - 2m)g_f(Z + m + 1)\} - m\mathbb{E}\Delta^2 g_f(Z + m)\} \\ &= -m\mathbb{E}\Delta^2 g_f(Z + m),\end{aligned}$$

by choice of $\chi$ and $b$. But now

$$\left|\mathbb{E}\Delta^2 g_f(Z + m)\right| \leq \|\Delta g_f\| \left\| \boldsymbol{\pi}_{\chi,b} - \boldsymbol{\pi}^{(1)}_{\chi,b} \right\| \leq 4(\mu + a)^{-1}\|f\|\, 4(\chi + b)^{-1/2} \leq 16\sqrt{5/3}(\mu + a)^{-3/2}\|f\|,$$

from Lemma 2.1 and Lemma 3.3, and because $\chi \geq \mu - 2(|a| + 1) \geq 3\mu/5$ under the given conditions on $\mu$ and $a$; note also that $\chi + b \geq 4|b|$ under these conditions, as is needed to apply Lemma 3.3. Since $0 \leq m < |a| + 1$, the first part of the theorem now follows. The proof of the second part is entirely similar, with the rôles of $\boldsymbol{\pi}_{\mu,a}$ and $\boldsymbol{\pi}_{\chi,b}$ interchanged; now $\chi + b \geq \mu + a \geq 4 \geq 4b$, simplifying the final calculation. □

## 4. Applications: Independent summands

Let $W = \sum_{i=1}^n I_i$ be a sum of independent Be$\,(p_i)$ random variables. By using a signed compound Poisson measure as approximation and matching the first two moments, Kruopis (1986) showed that an error in total variation of order $\lambda\sigma^{-3}\theta_3$ is obtained, where $\lambda = \sum_{i=1}^n p_i$, $\theta_l = \lambda^{-1} \sum_{i=1}^n p_i^l$ and $\sigma^2 = \sum_{i=1}^n p_i(1 - p_i)$. This approximation is extremely accurate. It has an error of at most order $\sigma^{-1}$, as good a rate as for Kolmogorov distance in the usual normal approximation, if the $p_i$ are of order 1, and the error is roughly of order $\lambda^{-1/2}p^2$ if the $p_i$ are small, combining the $\lambda^{-1/2}$ factor with the $O(p^2)$ error of the one term Poisson–Charlier expansion. Here, we use Stein's method, as developed in Section 2, to sharpen his constants.

**Theorem 4.1.** *With notation as above, take $\mu = \lambda(1 + \theta_2)$ and $a = -\lambda\theta_2$, and suppose that $\theta_2 < 1/2$. Then*

$$\|\mathcal{L}(W) - \boldsymbol{\pi}_{\mu,a}\| \leq \frac{2}{1 - 2\theta_2}\theta_3\tau^{-1/2}, \tag{4.1}$$

*where*

$$\tau = \sigma^2 - \max_{1 \leq i \leq n} p_i(1 - p_i).$$

*In particular, if $\mu > 0$ and $a < 0$ are such that $n := (\mu + a)^2/|a| \in \mathbb{Z}_+$ and $p := (\mu + a)^{-1}|a| < 1/2$, then*

$$\|\mathrm{Bi}(n,p) - \boldsymbol{\pi}_{\mu,a}\| \leq \frac{2p^2}{(1-2p)\sqrt{(n-1)p(1-p)}} \,. \tag{4.2}$$

*Proof.* For $W$ as defined above, we have

$$\mathbb{E}Wg(W) = \sum_{i=1}^n p_i \mathbb{E}g(W_i + 1),$$

where $W_i = W - I_i$ is independent of $I_i$. Choosing $\mu$ and $a$ as specified, and for any choices of $\mu_i$ and $a_i$, $1 \leq i \leq n$, such that $\sum_{i=1}^n \mu_i = \mu$ and $\sum_{i=1}^n a_i = a$, we have

$$\begin{aligned}
\mathbb{E}\{Wg(W) - \mu g(W+1) - ag(W+2)\} &= \mathbb{E}\sum_{i=1}^n \Big(p_i g(W_i+1) - \mu_i\{p_i g(W_i+2) + (1-p_i)g(W_i+1)\} \\
&\quad - a_i\{p_i g(W_i+3) + (1-p_i)g(W_i+2)\}\Big) \\
&= \mathbb{E}\left\{-\sum_{i=1}^n a_i p_i \Delta^2 g(W_i+1)\right\} + \sum_{i=1}^n \mathbb{E}\Big\{g(W_i+1)(p_i - \mu_i(1-p_i) \\
&\quad + a_i p_i) + g(W_i+2)(-\mu_i p_i - a_i(1-p_i) - 2a_i p_i)\Big\} \,.
\end{aligned}$$

Taking $\mu_i = p_i + p_i^2$ and $a_i = -p_i^2$, the second expression vanishes, giving

$$|\mathbb{E}\{ag(W+2) + \mu g(W+1) - Wg(W)\}| = \left|\sum_{i=1}^n p_i^3 \mathbb{E}\Delta^2 g(W_i+1)\right|. \tag{4.3}$$

This gives

$$|\mathbb{E}\{ag(W+2) + \mu g(W+1) - Wg(W)\}| \leq \sum_{i=1}^n p_i^3 \|\Delta g\| \, 2d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i+1)),$$

and, since the $W_i$ have unimodal distributions, it follows that

$$d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i+1)) \leq \max_{j\geq 0} \mathbb{P}[W_i = j] \leq e^{-\tau_i} I_0(\tau_i) \leq \frac{1}{2\sqrt{\tau_i}},$$

from Barbour and Jensen (1989), where $\tau_i = \sum_{j\neq i} p_j(1-p_j)$. The theorem now follows from Corollary 2.2. □

The bound (4.1) is rather neater than that of Kruopis (1986), and for small $\theta_2$ it improves on Kruopis's constant by a factor of about 10. If $\theta_2$ is larger, some of this advantage is lost. However, in such cases, approximation by a translated measure can give some improvement. Here, we approximate using the probability distribution nearest to a Poisson that can be obtained by translation from the family $\boldsymbol{\pi}_{\mu,a}$.

**Theorem 4.2.** *Let $m = \lceil \lambda\theta_2 \rceil$, $\mu = \lambda(1+\theta_2) - 2m$ and $0 \leq a = m - \lambda\theta_2 < 1$. Then*

$$\|\mathcal{L}(W) - \boldsymbol{\pi}^{(m)}_{\mu,a}\| \leq 2\theta_2^{-1}\theta_3\tau^{-1/2}\left\{\frac{(1-\theta_2) + \theta_2\lambda^{-1}(1+\sqrt{\tau})/\theta_3}{(1-\theta_2) - 3\lambda^{-1}}\right\}. \tag{4.4}$$

*Proof.* We compare the distribution of $W-m$ on $\mathbb{Z}$ with $\boldsymbol{\pi}_{\mu,a}$, using Corollary 2.3. Arguing as for Theorem 4.1, and writing $g_m(w)$ for $g(w-m)$, we obtain

$$\begin{aligned}
&\mathbb{E}\{(W-m)g(W-m)-\mu g(W-m+1)-ag(W-m+2)\} \\
&= \mathbb{E}\sum_{i=1}^n \Big(\{-a_ip_i\Delta^2 g_m(W_i+1)-m_i(1-p_i)\Delta^2 g_m(W_i)\} \\
&\quad +g_m(W_i+1)\{p_i-2m_i(1-p_i)-m_ip_i+a_ip_i-\mu_i(1-p_i)\} \\
&\quad +g_m(W_i+2)\{-a_i(1-p_i)-2a_ip_i-\mu_ip_i+m_i(1-p_i)\}\Big),
\end{aligned}$$

where $\sum_{i=1}^n m_i = m$, $\sum_{i=1}^n \mu_i=\mu$ and $\sum_{i=1}^n a_i = a$. Taking $m_i=\beta p_i^2$, $a_i=-(1-\beta)p_i^2$ and $\mu_i=p_i+(1-2\beta)p_i^2$ for any $\beta$ makes the last two terms zero, giving

$$\begin{aligned}
&|\mathbb{E}\{(W-m)g(W-m)-\mu g(W-m+1)-ag(W-m+2)\}| \\
&\le \left(|1-\beta|\sum_{i=1}^n p_i^3+|\beta|\sum_{i=1}^n p_i^2(1-p_i)\right)2\|\Delta g\| d_{TV}(\mathcal{L}(W_i),\mathcal{L}(W_i+1)).
\end{aligned}$$

Now take $\beta=(\lambda\theta_2)^{-1}\lceil\lambda\theta_2\rceil$, giving $m$, $\mu$ and $a$ as in the statement of the theorem, and note that $\theta_3\ge\theta_2^2$; this gives

$$|\mathbb{E}\{(W-m)g(W-m)-\mu g(W-m+1)-ag(W-m+2)\}| \le \left\{\frac{a\theta_3}{\theta_2}+\left(\lambda+\frac{a}{\theta_2}\right)(\theta_2-\theta_3)\right\}\tau^{-1/2}\|\Delta g\|,$$

and the theorem follows from Corollary 2.3. □

For large $\lambda$, this bound is preferable to that of Theorem 4.1 when $\theta_2>1/3$. Taking the better of these two bounds always results in an improvement by a factor of at least 3, compared to the bound obtained by Kruopis (1986).

There are analogous results for sums of integer valued random variables which are not restricted to take the values 0 and 1. Suppose that $W=\sum_{i=1}^n Z_i$, where the $Z_i$ are independent and integer valued, and satisfy $\mathbb{E}|Z_i^3|<\infty$. Define

$$\psi_i := \mathbb{E}|Z_i(Z_i-1)(Z_i-2)|+|\mathbb{E}Z_i|\,\mathbb{E}|Z_i(Z_i-1)|+2\mathbb{E}|Z_i|\,|\mathrm{Var}\,Z_i-\mathbb{E}Z_i|; \tag{4.5}$$

$$d_+^{(i)} := d_{TV}(\mathcal{L}(W_i),\mathcal{L}(W_i+1));\quad d_+ := d_{TV}(\mathcal{L}(W),\mathcal{L}(W+1)), \tag{4.6}$$

where $W_i := W-Z_i$.

**Theorem 4.3.** *For any $m\in\mathbb{Z}$ such that*

$$(2/3)\mathrm{Var}\,W<\mathbb{E}W+m<2\mathrm{Var}\,W, \tag{4.7}$$

*we have*

$$\|\mathcal{L}(W+m)-\boldsymbol{\pi}_{\mu+2m,a-m}\|\le\frac{2}{(1-2\theta_m)(\mathbb{E}W+m)}\left\{\left(|m|d_++\sum_{i=1}^n\psi_i d_+^{(i)}\right)+e_m(W)\right\}, \tag{4.8}$$

*where*

$$\mu=2\mathbb{E}W-\mathrm{Var}\,W;\ a=\mathrm{Var}\,W-\mathbb{E}W;\ \theta_m=|\mathrm{Var}\,W-(\mathbb{E}W+m)|/(\mathbb{E}W+m), \tag{4.9}$$

*and where $e_m(W)=0$ if $W+m\ge 0$ a.s., and $e_m(W)=1$ otherwise.*

*Proof.* Newton's expansion with remainder gives

$$g(z+l)=g(z+1)+(l-1)\Delta g(z+1)+\eta(g,z,l), \tag{4.10}$$

where

$$\eta(g,z,l) = \begin{cases} \sum_{s=1}^{l-2}(l-1-s)\Delta^2 g(z+s), & l \geq 3; \\ 0, & l = 2,1; \\ \sum_{s=0}^{-l}(-l-s+1)\Delta^2 g(z-s), & l \leq 0. \end{cases} \tag{4.11}$$

Hence, for any bounded $g$, $l \in \mathbb{Z}$ and $1 \leq i \leq n$, it follows that

$$|\mathbb{E}g(W_i+l) - (l-1)\mathbb{E}g(W_i+2) - (2-l)\mathbb{E}g(W_i+1)| \leq \|\Delta g\|\,(l-1)(l-2)d_+^{(i)}, \tag{4.12}$$

where we have used the general inequality

$$|\mathbb{E}\Delta^2 g(U+j)| \leq 2\|\Delta g\| d_{TV}(\mathcal{L}(U), \mathcal{L}(U+1)). \tag{4.13}$$

Fix any $1 \leq i \leq n$. Then applying (4.12) with $l = j$ gives

$$\mathbb{E}\{Z_i g(W)\} = \sum_j j q_{ij}\mathbb{E}g(W_i+j) = \sum_j j(j-1)q_{ij}\mathbb{E}g(W_i+2) + \sum_j j(2-j)q_{ij}\mathbb{E}g(W_i+1) + \phi_{1i}, \tag{4.14}$$

where $q_{ij} = \mathbb{P}[Z_i = j]$ and

$$|\phi_{1i}| \leq \|\Delta g\| d_+^{(i)} \mathbb{E}|Z_i(Z_i-1)(Z_i-2)|. \tag{4.15}$$

Then, taking $l = j+1$ and $W_i$ for $z$ in (4.10), we have

$$\mathbb{E}g(W+1) = \sum_j q_{ij}\mathbb{E}g(W_i+j+1) = \sum_j q_{ij}\{j\mathbb{E}g(W_i+2) + (1-j)\mathbb{E}g(W_i+1)\} + \phi_{2i}, \tag{4.16}$$

where

$$\phi_{2i} = \sum_j q_{ij}\mathbb{E}\eta(g, W_i, j+1); \tag{4.17}$$

whereas, with $l = j+2$, we obtain

$$\mathbb{E}g(W+2) = \sum_j q_{ij}\mathbb{E}g(W_i+j+2) = \sum_j q_{ij}\{(j+1)\mathbb{E}g(W_i+2) - j\mathbb{E}g(W_i+1)\} + \phi_{3i}, \tag{4.18}$$

where

$$\phi_{3i} = \sum_j q_{ij}\mathbb{E}\eta(g, W_i, j+2). \tag{4.19}$$

Careful calculation now shows that, for any choices of $\mu_i$ and $a_i$,

$$|\mu_i\phi_{2i} + a_i\phi_{3i}| \leq \|\Delta g\| d_+^{(i)} \{|\mu_i + a_i|\mathbb{E}|Z_i(Z_i-1)| + 2|a_i|\mathbb{E}|Z_i|\}. \tag{4.20}$$

Thus, from (4.14, 4.16), and (4.18), it follows that

$$\begin{aligned} a_i\mathbb{E}g(W+2) + \mu_i\mathbb{E}g(W+1) - \mathbb{E}\{Z_i g(W)\} &= \mathbb{E}g(W_i+2)\{\mu_i\mathbb{E}Z_i + a_i(1+\mathbb{E}Z_i) - \mathbb{E}\{Z_i(Z_i-1)\}\} \\ &+ \mathbb{E}g(W_i+1)\{\mu_i(1-\mathbb{E}Z_i) - a_i\mathbb{E}Z_i - \mathbb{E}\{Z_i(2-Z_i)\}\} + a_i\phi_{3i} + \mu_i\phi_{2i} - \phi_{1i}, \end{aligned}$$

and taking

$$a_i = \operatorname{Var} Z_i - \mathbb{E}Z_i; \qquad \mu_i = 2\mathbb{E}Z_i - \operatorname{Var} Z_i,$$

the coefficients of $\mathbb{E}g(W_i+2)$ and $\mathbb{E}g(W_i+1)$ vanish. This then implies, using (4.15) and (4.20), that

$$|a_i\mathbb{E}g(W+2)+\mu_i\mathbb{E}g(W+1)-\mathbb{E}\{Z_ig(W)\}|\le\|\Delta g\|d_+^{(i)}\psi_i. \tag{4.21}$$

Finally, again from (4.13), we have

$$|m\mathbb{E}\Delta^2 g(W)|\le\|\Delta g\|\,|m|d_+; \tag{4.22}$$

hence, adding (4.21) over $1\le i\le n$ and then subtracting (4.22), we find that

$$|(a-m)\mathbb{E}g(W+2)+(\mu+2m)\mathbb{E}g(W+1)-(W+m)\mathbb{E}g(W)|\le\|\Delta g\|\left\{|m|d_++\sum_{i=1}^n\psi_i d_+^{(i)}\right\}, \tag{4.23}$$

with $a=\sum_{i=1}^n a_i=\mathrm{Var}\,W-\mathbb{E}W$ and $\mu=\sum_{i=1}^n\mu_i=2\mathbb{E}W-\mathrm{Var}\,W$ as in (4.9). For any $m$ such that (4.7) is satisfied, the quantity $\theta_m$ defined in (4.9) is less than $1/2$, and we can apply Corollary 2.3 with $a-m$ for $a$ and $\mu+2m$ for $\mu$, proving the theorem. If $W+m\ge 0$ a.s., then Corollary 2.2 can be applied instead. □

There is considerable flexibility inherent in Theorem 4.3, both in the choice of $m$ and in the fact that the $Z_i$ need not be centred. The two corollaries that follow are chosen to illustrate standard situations.

**Corollary 4.4.** *If* $(2/3)\mathrm{Var}\,W<\mathbb{E}W<2\mathrm{Var}\,W$*, then*

$$\|\mathcal{L}(W)-\boldsymbol{\pi}_{\mu,a}\|\le\frac{2}{(1-2\theta)\mathbb{E}W}\left\{\left(\sum_{i=1}^n\psi_i d_+^{(i)}\right)+e_0(W)\right\}, \tag{4.24}$$

*where* $\theta=(\mathbb{E}W)^{-1}|\mathrm{Var}\,W-\mathbb{E}W|$ *and* $\mu$ *and* $a$ *are as defined in (4.9).*

**Corollary 4.5.** *If* $\mathrm{Var}\,W\ge 3$*, then*

$$\|\mathcal{L}(W+m)-\boldsymbol{\pi}_{\mu+2m,a-m}\|\le\frac{2}{(\mathrm{Var}\,W-3)}\left\{\left(|m|d_++\sum_{i=1}^n\psi_i d_+^{(i)}\right)+e_m(W)\right\}, \tag{4.25}$$

*where* $m=\lfloor\mathrm{Var}\,W-\mathbb{E}W\rfloor$ *and* $\mu$ *and* $a$ *are as defined in (4.9). Here, we always have* $0\le a-m<1$ *and* $\mathrm{Var}\,W-2\le\mu+2m\le\mathrm{Var}\,W$.

For comparison with the usual central limit theorem, Corollary 4.5 is appropriate. Setting $s_n^2=\sum_{i=1}^n\mathrm{Var}\,Z_i$ and $\Gamma_n=\sum_{i=1}^n\mathbb{E}|Z_i-\mathbb{E}Z_i|^3$ as usual, the numerator in Corollary 4.5 is bounded by $C\Gamma_n+1$ for some universal constant $C$, so that, when $s_n^2\to\infty$, the error bound is of order $\Gamma_n s_n^{-2}\max_{1\le i\le n}d_+^{(i)}$. This differs from the usual Lyapounov ratio $\Gamma_n s_n^{-3}$ in that $\max_{1\le i\le n}d_+^{(i)}$ replaces a factor of $s_n^{-1}$. It is easy to see that a change is required here, when proving approximation bounds with respect to total variation distance, because, if $Z_i\sim 2\mathrm{Be}\,(p)$ for all $i$, then $\Gamma_n s_n^{-3}=O(n^{-1/2})$, whereas $W$ is concentrated on the even integers, and is therefore far from any of the measures $\boldsymbol{\pi}_{\mu,a}$. Nonetheless, Proposition 4.6 below shows that $\max_{1\le i\le n}d_+^{(i)}$ is frequently of order $O(n^{-1/2})$, so that the classical rate of approximation is recovered; in particular, this is so if $Z_i\sim F$ for all $i$, for any strongly aperiodic distribution $F$ on the integers.

Corollary 4.4 exploits the fact that the $\psi_i$ are small if the $Z_i$ mostly take the values 0 and 1. For example, suppose that

$$\begin{aligned}&\mathbb{P}[Z_i=1]=p_i;\quad \mathbb{P}[Z_i=2]=q_i\le cp_i^2\quad\text{for some } c>0;\\ &\mathbb{P}[Z_i=0]=1-(p_i+q_i),\end{aligned}$$

where the $p_i$ and $q_i$ are such that $\sum_{i=1}^n(p_i+q_i)\ge 2$ and that

$$\theta=(\mathbb{E}W)^{-1}|\mathrm{Var}\,W-\mathbb{E}W|\le\frac{\sum_{i=1}^n|2q_i-(p_i+2q_i)^2|}{\sum_{i=1}^n(p_i+2q_i)}\le\frac{1}{4}.$$

Then

$$\psi_i = (p_i + 2q_i)\left\{2q_i + |2q_i - (p_i + 2q_i)^2|\right\} = O(p_i^3),$$

so that the bound given in Corollary 4.4 can be shown to be uniformly of order $\left(\sum_{i=1}^n p_i^3 \Big/ \sum_{i=1}^n p_i\right) \max_{1\le i\le n} d_+^{(i)}$. Proposition 4.6 below now shows that

$$\max_{1\le i\le n} d_+^{(i)} \le \left(\sum_{i=1}^n (p_i + q_i) - 1\right)^{-1/2},$$

provided that $p_i \le 1/\max\{4, c\}$ for all $i$, so that we recover a bound of order $\left(\sum_{i=1}^n p_i^3\right)\left(\sum_{i=1}^n p_i\right)^{3/2}$, as in Theorem 4.1.

Our estimates of the distances $d_+ = d_{TV}(\mathcal{L}(W), \mathcal{L}(W+1))$ and $d_+^{(i)} = d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i+1))$ are obtained by coupling arguments. The following proposition serves as a simple example of what can be obtained.

**Proposition 4.6.** *Suppose that $Z_i$, $1 \le i \le n$, are independent integer valued random variables, and set $u_i = 1 - d_{TV}(\mathcal{L}(Z_i), \mathcal{L}(Z_i+1))$, $U = \sum_{i=1}^n \min\{u_i, 1/2\}$. Then, if $W = \sum_{i=1}^n Z_i$, we have*

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(W+1)) \le U^{-1/2}.$$

*Hence also, if $W_i = W - Z_i$, we have*

$$\max_{1\le i\le n} d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i+1)) \le (U-1)^{-1/2}.$$

*Proof.* First suppose that $u_i \le 1/2$ for all $i$. Then the Mineka coupling (Lindvall 1992, Sect. II.14) shows that

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(W+1)) \le \mathbb{P}[T > n],$$

where $T$ is the time at which a simple symmetric random walk $(S_m,\ m \ge 0)$ with $S_0 = 0$ and

$$\mathbb{P}[S_{m+1} - S_m = 1] = \mathbb{P}[S_{m+1} - S_m = -1] = \tfrac{1}{2}(1 - \mathbb{P}[S_{m+1} = S_m]) = \tfrac{1}{2}u_i \tag{4.26}$$

first hits the level 1. But, by the reflection principle,

$$\mathbb{P}[T \le n] = 2\mathbb{P}[S_n \ge 2] + \mathbb{P}[S_n = 1],$$

and hence, again by symmetry,

$$\mathbb{P}[T > n] = \mathbb{P}[S_n \in \{0, -1\}] \le 2\max_j \mathbb{P}[S_n = j].$$

The proposition then follows from Lemma 4.7, because $u_i \le 1/2$ for all $i$. If, for any $i$, $u_i \ge 1/2$, the Mineka coupling can be modified in such a way that (4.26) holds with $\frac{1}{2}u_i$ replaced by $1/4$ for such $i$, so that Lemma 4.7 can still be applied. □

**Lemma 4.7.** *For the random walk $(S_m,\ m \ge 0)$ defined above, if also $\max_{1\le i\le n} u_i \le 1/2$, then*

$$\max_j \mathbb{P}[S_n = j] \le \tfrac{1}{2}U^{-1/2}.$$

*Proof.* By Fourier inversion, since $u_i(1 - \cos t) \le 1$ under the stated condition on the $u_i$, we have

$$\begin{aligned}\max_j \mathbb{P}[S_n = j] &\le \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\prod_{i=1}^n (1 - u_i(1 - \cos t))\right| dt \ \le\ \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left\{-\sum_{i=1}^n u_i(1 - \cos t)\right\} dt \\ &= e^{-U} I_0(U) \le \tfrac{1}{2}U^{-1/2},\end{aligned}$$

where $I_0$ is a modified Bessel function (Abramowitz and Stegun 1964, Sect. 9.6), and the lemma follows. $\square$

**Remark.** The proposition can of course be extended by the use of blocks, if too many of the $u_i$ are zero. As discussed in Lindvall (1992, Sect. II.12–14), if the $Z_i$ are independent and identically distributed with a strongly aperiodic distribution $F$, then

$$d_{TV}(\mathcal{L}(W), \mathcal{L}(W+1)) \leq cn^{-1/2}$$

for some constant $c = c(F) < \infty$. If $Z_i$ has a unimodal distribution, then $u_i = 1 - \max_j \mathbb{P}[Z_i = j]$.

As a final result in this section, we show that the negative binomial distribution $\mathrm{NB}\,(k,p)$ can be closely approximated by a measure of the form $\boldsymbol{\pi}_{\mu,a}$, if $p < 1/3$. Here, for any $k > 0$ and $0 \leq p < 1$, we define

$$\mathrm{NB}\,(k,p)\{l\} := (1-p)^k \binom{k+l-1}{l} p^l, \quad l \in \mathbb{Z}_+.$$

**Corollary 4.8.** *For any $k > 0$ and $0 < p < 1/3$,*

$$\|\mathrm{NB}\,(k,p) - \boldsymbol{\pi}_{\mu,a}\| \leq \frac{4p^2}{(1-p)(1-3p)} \frac{1}{\sqrt{k \log(1/(1-p))}},$$

*where $\mu = kp(1-2p)(1-p)^{-2}$ and $a = kp^2(1-p)^{-2}$.*

*Proof.* Take an arbitrary $n > 0$, and let $W = \sum_{i=1}^n Z_i$, where the $Z_i \sim \mathrm{NB}\,(kn^{-1}, p)$ are independent. Note that $W \geq 0$ a.s., and that the conditions of Corollary 4.4 are satisfied if $0 < p < 1/3$. Now, for large $n$,

$$\mathbb{E}Z_i(Z_i-1)(Z_i-2) = \frac{2k}{n}\left(\frac{p}{1-p}\right)^3 + O(n^{-2});$$
$$\mathbb{E}Z_i = \frac{k}{n}\left(\frac{p}{1-p}\right) \quad \text{and} \quad |\mathrm{Var}\, Z_i - \mathbb{E}Z_i| = O(n^{-1}),$$

showing that

$$\sum_{i=1}^n \psi_i = 2k\left(\frac{p}{1-p}\right)^3 + O(n^{-1}).$$

But for all $n$ large enough,

$$1 - d_{TV}(\mathcal{L}(Z_i), \mathcal{L}(Z_i+1)) = \mathbb{P}[Z_i \geq 1] = 1 - (1-p)^{k/n} = kn^{-1}\log(1/(1-p)) + O(n^{-2}),$$

so that, from Proposition 4.6,

$$d_+^{(i)} \leq \left\{k \log(1/(1-p)) + O(n^{-1})\right\}^{-1/2}.$$

Noting also that $\theta = p/(1-p)$, the conclusion now follows from Corollary 4.4 and by letting $n \to \infty$. $\square$

**Remark.** If $0 < a < 1$ and $\mu$ is large, we can take $k = (\mu+a)^2/a$ and $p = a/(\mu+2a)$ in Corollary 4.8, showing that

$$\|\boldsymbol{\pi}_{\mu,a} - \mathrm{NB}\,(k,p)\| \leq \frac{4a^2}{(\mu^2-a^2)\sqrt{a^{-1}(\mu+a)^2 \log\{(\mu+2a)/(\mu+a)\}}} \leq 4/\{\mu^{1/2}(\mu^2-1)\}.$$

Thus, for large $\mu$, the translated compound Poisson distribution $\boldsymbol{\pi}_{\mu,a}$ with $0 < a < 1$ is approximated extremely closely, to order $O(\mu^{-5/2})$, by the negative binomial distribution $\mathrm{NB}\,(k,p)$ with large $k$ and small $p$ as given above.

## 5. Applications: 2-runs

The theory developed here can also be used to obtain sharper error estimates of approximations to sums of dependent indicator random variables. Not surprisingly, the dependent case is much more complicated than the independent case. To illustrate how the theory works here, we consider a very simple problem: the approximation to the number of 2-runs of 1's in a sequence of independent indicator random variables $\xi_i$ with $\mathbb{P}[\xi_i = 1] = p_i$, $1 \le i \le n$. This problem has the advantage of having been well studied previously. In particular, when $p_i = p$ for $1 \le i \le n$, compound Poisson approximation in total variation to the distribution of $W$ has been examined in Arratia *et al.* (1990), Roos (1993) and Eichelsbacher and Roos (1998). To avoid edge effects, we treat $i + nj$ as $i$ for $1 \le i \le n$, $j = 0, \pm 1, \pm 2, \ldots$ Define $I_i = \xi_i \xi_{i-1}$ and $W = \sum_{i=1}^n I_i$, so that $W$ is our random variable of interest; note that $\mathbb{E}I_i = p_{i-1}p_i$ and $\mathbb{E}W = \sum_{i=1}^n p_{i-1}p_i$.

Our argument is based on showing that, for suitably chosen $\mu$ and $a$, the expression

$$\mathbb{E}\{ag(W+2) + \mu g(W+1) - Wg(W)\} \tag{5.1}$$

can be bounded in the form given in (2.13), so that Corollary 2.2 can be applied. When computing the expectations in (5.1), the aim is first to use the local dependence structure to reduce all of them as far as possible to linear combinations of $\mathbb{E}g(X+1)$ and $\mathbb{E}\Delta g(X+1)$, for some suitable random variable $X$. We then pick $\mu$ and $a$ to make the coefficient of $\mathbb{E}g(X+1)$ vanish in (5.1), and reorganize the coefficients of $\mathbb{E}\Delta g(X+1)$ to reduce the term into $\mathbb{E}\Delta^2 g(Y+1)$, for another suitable random variable $Y$. All terms then involve expectations of the form $\mathbb{E}\Delta^2 g(X+j)$, bounded by using the inequality

$$\left|\mathbb{E}\Delta^2 g(X+j)\right| \le \|\Delta g\| \, \|\mathcal{L}(X+1) - \mathcal{L}(X)\|, \tag{5.2}$$

which is of the form needed to apply Corollary 2.2. The next lemma shows how the total variation distance in (5.2) can be translated into an explicit function of $p_i$.

**Lemma 5.1.** *Let $(\eta_m,\ m \ge 1)$ be independent indicator random variables with $\mathbb{P}(\eta_m = 1) = \alpha_m$, $m \ge 1$, and set $\eta_0 = 0$, i.e. $\alpha_0 = 0$, and $Y_m = \sum_{i=1}^m \eta_i \eta_{i-1}$. Then, for each $n \ge 2$,*

$$b_n(\alpha_1, \alpha_2, \ldots, \alpha_n) := \|\mathcal{L}(Y_n) - \mathcal{L}(Y_n+1)\| \le \frac{4.6}{\sqrt{\sum_{i=1}^n (1-\alpha_{i-2})^2 \alpha_{i-1}(1-\alpha_{i-1})\alpha_i}}.$$

*Proof.* We construct a suitable coupling. Let $(\zeta_m,\ m \ge 1)$ be an independent copy of $(\eta_m,\ m \ge 1)$. Set $\eta_0' = 0$, and define

$$\eta_i' = \begin{cases} \zeta_i, & \text{if } \eta_{i-1} = \eta_{i-1}' = 0; \\ \eta_i, & \text{otherwise;} \end{cases} \tag{5.3}$$

then define $Y_m' = 1 + \sum_{i=1}^m \eta_i' \eta_{i-1}'$. Let $D_m = Y_m - Y_m'$ and $\delta_m = D_m - D_{m-1}$, then $\delta_m$ takes values $0, \pm 1$, and

$$\{\delta_i \ne 0\} = \{\eta_{i-2} = \eta_{i-2}' = 0, |\eta_{i-1} - \eta_{i-1}'| = 1, \eta_i = 1\} \cdot \tag{5.4}$$

Set $R_i = 1_{\{\delta_i \ne 0\}}$ and $R = \sum_{i=2}^n R_i$. For each $i \ge 2$, we have

$$\mathbb{E}R_i = \mathbb{P}(\delta_i \ne 0) = 2\alpha_{i-1}(1-\alpha_{i-1})\alpha_i \mathbb{P}(\eta_{i-2} = \eta_{i-2}' = 0),$$

so

$$\mathbb{E}R \ge 2\sum_{i=2}^n (1-\alpha_{i-2})^2 \alpha_{i-1}(1-\alpha_{i-1})\alpha_i. \tag{5.5}$$

Direct expansion gives

$$\mathrm{Var}(R) = \sum_{i=2}^n \mathbb{E}R_i(1-\mathbb{E}R_i) + 2\sum_{i=2}^n \sum_{i<j\le n} [\mathbb{E}(R_iR_j) - \mathbb{E}R_i \mathbb{E}R_j]. \tag{5.6}$$

However $R_iR_j = 0$ unless $j \geq i+3$, when, from (5.3), we have

$$\mathbb{E}(R_j \mid \eta_i = \eta_i' = 1) = \mathbb{E}(R_j \mid \eta_i = 0, \eta_i' = 1) = \mathbb{E}(R_j \mid \eta_i = 1, \eta_i' = 0);$$

then it also follows that

$$\mathbb{E}(R_iR_j) = \mathbb{P}(R_j = 1|R_i = 1)\mathbb{P}(R_i = 1) = \mathbb{P}(R_j = 1|\eta_i = \eta_i' = 1)\mathbb{E}R_i,$$

which in turn implies that

$$\begin{aligned} |\mathbb{E}(R_iR_j) - \mathbb{E}R_i\mathbb{E}R_j| &\leq (\mathbb{E}R_i)|\mathbb{E}(R_j|\eta_i = \eta_i' = 1) - \mathbb{E}R_j| \\ &= (\mathbb{E}R_i)|\mathbb{E}(R_j|\eta_i = \eta_i' = 1) - \mathbb{E}(R_j|\eta_i = \eta_i' = 0)|\mathbb{P}(\eta_i = \eta_i' = 0). \end{aligned} \tag{5.7}$$

Now let $U = \min\{k \geq i+1 : \ \eta_k = \zeta_k\}$. Then $U$ is independent of $\eta_i$ and $\eta_i'$, with $\mathbb{P}[U \geq j-1] = \prod_{k=i+1}^{j-2}\{2\alpha_k(1-\alpha_k)\}$, and

$$\mathcal{L}((\eta_k, \eta_k')_{k\geq l} \mid U = l, \eta_i = \eta_i' = 1) = \mathcal{L}((\eta_k, \eta_k')_{k\geq l} \mid U = l, \eta_i = \eta_i' = 0)$$

for all $i+1 \leq l \leq j-2$. Hence, for each such $l$,

$$\mathbb{E}(R_j|U = l, \eta_i = \eta_i' = 1) = \mathbb{E}(R_j|U = l, \eta_i = \eta_i' = 0),$$

and it follows from the fact that $2\alpha_l(1-\alpha_l) \leq 1/2$ and (5.7) that

$$|\mathbb{E}(R_iR_j) - \mathbb{E}R_i\mathbb{E}R_j| \leq (\mathbb{E}R_i)\mathbb{P}(\eta_i = \eta_i' = 0)\mathbb{P}(U \geq j-1) \leq \mathbb{E}(R_i)\mathbb{P}(\eta_i = \eta_i' = 0)2^{-(j-i-2)}, \ j \geq i+3.$$

Noting that $R_i \geq 0$, we get from (5.6)

$$\mathrm{Var}(R) \leq 3\mathbb{E}R. \tag{5.8}$$

Taking $\tau_0 = 0$, define the stopping times

$$\tau_j = \min\{m > \tau_{j-1} : \delta_m \neq 0\}, \qquad j \geq 1.$$

Then $R = r$ is equivalent to $\tau_r \leq n < \tau_{r+1}$, and the conditional distribution $\mathcal{L}((\delta_{\tau_j}, 1 \leq j \leq r)|R = r)$ is a uniform distribution on $\{-1,1\}^r$, so the conditional distribution $\mathcal{L}(D_{\tau_j}, 0 \leq j \leq r|R = r) = \mathcal{L}(Z_j, 0 \leq j \leq r)$, where $(Z_j, \ j \geq 0)$ is a random walk with $Z_0 = -1$ and $\mathbb{P}(Z_i - Z_{i-1} = 1) = \mathbb{P}(Z_i - Z_{i-1} = -1) = 1/2$, for all $i \geq 1$. Define

$$J = \min\{j \geq 1 : \ D_{\tau_j} = 0\},$$

and let

$$Y_m'' = \begin{cases} Y_m', & m < \tau_J; \\ Y_m, & m \geq \tau_J. \end{cases}$$

Then, because of the coupling, it follows that

$$\mathcal{L}(Y_n'') = \mathcal{L}(Y_n') = \mathcal{L}(Y_n + 1),$$

and that

$$\|\mathcal{L}(Y_n) - \mathcal{L}(Y_n + 1)\| = \|\mathcal{L}(Y_n) - \mathcal{L}(Y_n'')\| \leq 2\mathbb{P}[Y_n \neq Y_n'']. \tag{5.9}$$

On the other hand, as in the proof of Proposition 4.6, it follows from the reflection principle for the symmetric Bernoulli random walk that

$$\begin{aligned} \mathbb{P}(Y_n \neq Y_n''|R = r) &= \mathbb{P}(\max_{1\leq j\leq r} D_{\tau_j} \leq -1|R = r) = \mathbb{P}(\max\{Z_i, \ 0 \leq i \leq r\} \leq -1) = \mathbb{P}[Z_r \in \{-2,-1\}] \\ &= \max_j \mathbb{P}(Z_r = j) \leq \sqrt{\frac{2}{\pi r}} \leq \frac{0.8}{\sqrt{r}}, \end{aligned}$$

for all $r \geq 1$, which yields, for any constant $0 < \kappa < 1$,

$$\mathbb{P}(Y_n'' \neq Y_n) = \sum_{r=0}^{\infty} \mathbb{P}(Y_n'' \neq Y_n | R = r)\mathbb{P}(R = r) \leq \mathbb{P}(R \leq \kappa \mathbb{E}R) + \frac{0.8}{\sqrt{\kappa \mathbb{E}R}} \leq \frac{3}{(1-\kappa)^2 \mathbb{E}R} + \frac{0.8}{\sqrt{\kappa \mathbb{E}R}}, \quad (5.10)$$

this last because of Chebyshev's inequality and (5.8).

If

$$s(\alpha) := \sum_{i=1}^{n} (1-\alpha_{i-2})^2 \alpha_{i-1}(1-\alpha_{i-1})\alpha_i \leq (4.6/2)^2 = 5.29,$$

the bound given in the lemma is clearly true, so we assume henceforth that $s(\alpha) > 5.29$, and thus $\mathbb{E}R > 10.58$ from (5.5). Choosing $\kappa = 0.2197412784$, we thus find that

$$\frac{3}{(1-\kappa)^2 \mathbb{E}R} + \frac{0.8}{\sqrt{\kappa \mathbb{E}R}} \leq \frac{1}{\sqrt{\mathbb{E}R}} \left\{ \frac{3}{(1-\kappa)^2 \sqrt{10.58}} + \frac{0.8}{\sqrt{\kappa}} \right\}, \quad (5.11)$$

and the lemma follows from (5.9–5.11).

□

**Theorem 5.2.** *Let $W$ denote the number of 2–runs, as defined above, and let*

$$a = \sum_{i=1}^{n} p_{i-1}p_i[(1-p_{i-1})p_{i-2} + (1-p_i)p_{i+1} - p_{i-1}p_i],$$

$\mu = \sum_{i=1}^{n} p_{i-1}p_i - a$ *and*

$$\gamma = \sum_{i=1}^{n} (1-p_{i+1})^2 p_i(1-p_i)p_{i-1} - 6 \max_{1 \leq j \leq n} (1-p_{j+1})^2 p_j(1-p_j)p_{j-1}.$$

*If $\theta = |a|/(\mu + a) < \frac{1}{2}$, then*

$$\|\mathcal{L}(W) - \boldsymbol{\pi}_{\mu,a}\| \quad \leq \quad \frac{9.2 \sum_{i=1}^{n} [3p_{i-2}p_{i-1}p_i p_{i+1} + p_{i-1}^3 p_i^3 + 4p_{i-1}^2 p_i^2 p_{i+1} + 4p_{i-2}p_{i-1}^2 p_i^2 + 7p_{i-3}p_{i-2}^2 p_{i-1}^2 p_i]}{(1-2\theta)(\mu+a)\sqrt{\gamma}}.$$

*In particular, if $p_i = p < \frac{1}{4}$ for all $1 \leq i \leq n$, and $n > 7$, then*

$$\|\mathcal{L}(W) - \boldsymbol{\pi}_{\mu,a}\| \leq \frac{27.6p^2 + 73.6(p^3 + p^4)}{(1-2\theta)\sqrt{(n-6)(1-p)^3 p^2}}.$$

*Proof.* For $1 \leq i \leq n-1$, let

$$V_i = W - \xi_{i-2}\xi_{i-1} - \xi_{i-1}\xi_i - \xi_i\xi_{i+1}, \quad X_i = V_i - \xi_{i-3}\xi_{i-2} - \xi_{i+1}\xi_{i+2}; \quad (5.12)$$

$$V_i^{1-} = V_i - \xi_{i-3}\xi_{i-2}, \quad V_i^{1+} = V_i - \xi_{i+1}\xi_{i+2}; \quad (5.13)$$

$$V_i^{2-} = V_i^{1-} - \xi_{i-4}\xi_{i-3}, \quad V_i^{2+} = V_i^{1+} - \xi_{i+2}\xi_{i+3}. \quad (5.14)$$

To simplify the typography, we drop $i$ from $V_i$ and pick it up when we need it. Using the fact that $V_i$, $\xi_{i-1}$ and $\xi_i$ are independent, we have

$$\begin{aligned} g(W+2) &= \{g(W+2)[\xi_{i-1}\xi_i + (1-\xi_{i-1})\xi_i + \xi_{i-1}(1-\xi_i) + (1-\xi_{i-1})(1-\xi_i)]\} \\ &= \{g(V+\xi_{i-2}+\xi_{i+1}+3)\xi_{i-1}\xi_i\} + \{g(V+\xi_{i+1}+2)(1-\xi_{i-1})\xi_i\} \\ &\quad + \{g(V+\xi_{i-2}+2)\xi_{i-1}(1-\xi_i)\} + \{g(V+2)(1-\xi_{i-1})(1-\xi_i)\}. \end{aligned}$$

Expressing as much of this as possible in terms of second differences, we obtain

$$\begin{aligned} g(W+2) &= \{\Delta^2 g(V+\xi_{i-2}+\xi_{i+1}+1)\xi_{i-1}\xi_i\} \\ &\quad +2\{[g(V+\xi_{i-2}+\xi_{i+1}+2)-g(V+\xi_{i+1}+2)-g(V+\xi_{i-2}+2)+g(V+2)]\xi_{i-1}\xi_i\} \\ &\quad -\{[g(V+\xi_{i-2}+\xi_{i+1}+1)-g(V+\xi_{i+1}+1)-g(V+\xi_{i-2}+1)+g(V+1)]\xi_{i-1}\xi_i\} \\ &\quad +\{[g(V+\xi_{i+1}+2)-g(V+2)-g(V+\xi_{i+1}+1)+g(V+1)](1+\xi_{i-1})\xi_i\} \\ &\quad +\{[g(V+\xi_{i-2}+2)-g(V+2)-g(V+\xi_{i-2}+1)+g(V+1)]\xi_{i-1}(1+\xi_i)\} \\ &\quad +\{[g(V+\xi_{i-2}+1)-g(V+1)]\xi_{i-1}\}+\{[g(V+\xi_{i+1}+1)-g(V+1)]\xi_i\} \\ &\quad +\{\Delta g(V+1)(1+\xi_{i-1}\xi_i)\}+g(V+1). \end{aligned}$$

Noting that the second and third terms are 0 unless both $\xi_{i-2}$ and $\xi_{i+1}$ are equal to 1, and other terms can be worked out in the same way, we have

$$\begin{aligned} \mathbb{E}g(W+2) &= p_{i-1}p_i\mathbb{E}\Delta^2 g(V+\xi_{i-2}+\xi_{i+1}+1)+2p_{i-1}p_i\mathbb{E}\{\Delta^2 g(V+2)\xi_{i-2}\xi_{i+1}\} \\ &\quad +\mathbb{E}\{\Delta^2 g(V+1)[-p_{i-1}p_i\xi_{i-2}\xi_{i+1}+(1+p_{i-1})p_i\xi_{i+1}+p_{i-1}(1+p_i)\xi_{i-2}]\} \\ &\quad +\mathbb{E}\{\Delta g(V+1)(p_{i-1}\xi_{i-2}+p_i\xi_{i+1}+1+p_{i-1}p_i)\}+\mathbb{E}g(V+1). \end{aligned}$$

In similar fashion, we have

$$\begin{aligned} \mathbb{E}g(W+1) &= \mathbb{E}\{g(V+\xi_{i-2}+\xi_{i+1}+2)\xi_{i-1}\xi_i\}+\mathbb{E}\{g(V+\xi_{i+1}+1)(1-\xi_{i-1})\xi_i\} \\ &\quad +\mathbb{E}\{g(V+\xi_{i-2}+1)\xi_{i-1}(1-\xi_i)\}+\mathbb{E}\{g(V+1)(1-\xi_{i-1})(1-\xi_i)\} \\ &= \mathbb{E}\{[g(V+\xi_{i-2}+\xi_{i+1}+2)-g(V+\xi_{i+1}+2)-g(V+\xi_{i-2}+2)+g(V+2)]\xi_{i-1}\xi_i\} \\ &\quad +\mathbb{E}\{[g(V+\xi_{i-2}+2)-g(V+\xi_{i-2}+1)-g(V+2)+g(V+1)]\xi_{i-1}\xi_i\} \\ &\quad +\mathbb{E}\{[g(V+\xi_{i+1}+2)-g(V+\xi_{i+1}+1)-g(V+2)+g(V+1)]\xi_{i-1}\xi_i\} \\ &\quad +\mathbb{E}\{[g(V+\xi_{i+1}+1)-g(V+1)]\xi_i\}+\mathbb{E}\{[g(V+\xi_{i-2}+1)-g(V+1)]\xi_{i-1}\} \\ &\quad +\mathbb{E}\{\Delta g(V+1)\xi_{i-1}\xi_i\}+\mathbb{E}g(V+1) \\ &= \mathbb{E}\{\Delta^2 g(V+2)(p_{i-1}p_i\xi_{i-2}\xi_{i+1})\}+\mathbb{E}\{\Delta^2 g(V+1)p_{i-1}p_i(\xi_{i-2}+\xi_{i+1})\} \\ &\quad +\mathbb{E}\{\Delta g(V+1)(p_{i-1}\xi_{i-2}+p_i\xi_{i+1}+p_{i-1}p_i)\}+\mathbb{E}g(V+1), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}I_i g(W) &= p_{i-1}p_i\mathbb{E}\{g(V+\xi_{i-2}+\xi_{i+1}+1)[\xi_{i-2}\xi_{i+1}+(1-\xi_{i-2})\xi_{i+1} \\ &\quad +\xi_{i-2}(1-\xi_{i+1})+(1-\xi_{i-2})(1-\xi_{i+1})]\} \\ &= p_{i-1}p_i\mathbb{E}\{g(V+3)\xi_{i-2}\xi_{i+1}+g(V+2)[\xi_{i-2}(1-\xi_{i+1})+(1-\xi_{i-2})\xi_{i+1}] \\ &\quad +g(V+1)(1-\xi_{i-2})(1-\xi_{i+1})\} \\ &= p_{i-1}p_i[\mathbb{E}\{\Delta^2 g(V+1)\xi_{i-2}\xi_{i+1}\}+\mathbb{E}\{\Delta g(V+1)(\xi_{i-2}+\xi_{i+1})\}+\mathbb{E}g(V+1)]. \end{aligned}$$

Collecting these three expansions, and for any choices of $a_i$ and $\mu_i$, we find that

$$\begin{aligned} \mathbb{E}[a_i g(W+2) &+ \mu_i g(W+1)-I_i g(W)] = a_i p_{i-1}p_i\mathbb{E}\Delta^2 g(V+\xi_{i-2}+\xi_{i+1}+1) \\ &+ (\mu_i+2a_i)p_{i-1}p_i\mathbb{E}\{\Delta^2 g(V+2)\xi_{i-2}\xi_{i+1}\} \\ &+ \mathbb{E}\{\Delta^2 g(V+1)[a_i(1+p_{i-1})p_i\xi_{i+1}+a_i p_{i-1}(1+p_i)\xi_{i-2} \\ &+ \mu_i p_{i-1}p_i(\xi_{i-2}+\xi_{i+1})-(1+a_i)p_{i-1}p_i\xi_{i-2}\xi_{i+1}]\} \\ &+ \mathbb{E}\{\Delta g(V+1)[a_i+(\mu_i+a_i)p_{i-1}p_i+p_{i-1}(a_i+\mu_i-p_i)\xi_{i-2} \\ &+ p_i(a_i+\mu_i-p_{i-1})\xi_{i+1}]\}+\mathbb{E}g(V+1)[(\mu_i+a_i)-p_{i-1}p_i]. \end{aligned} \tag{5.15}$$

We now choose $a_i = p_{i-1}p_i[(1-p_{i-1})p_{i-2}+(1-p_i)p_{i+1}-p_{i-1}p_i]$ and $\mu_i = p_{i-1}p_i - a_i$ so that $\sum_{i=1}^n a_i = a$ and $\sum_{i=1}^n \mu_i = \mu$, then the last term of (5.15) vanishes. Then we apply Lemma 5.1 to bound (5.15). The first term of (5.15) is bounded by

$$|a_i|p_{i-1}p_i\|\Delta g\|b_n(1,p_{i-2},p_{i-3},...,p_1,p_n,...,p_{i+1},1).$$

By (5.12–5.14) the second and third terms of (5.15) can be respectively reduced and bounded as

$$
\begin{aligned}
&(p_{i-1}p_i + |a_i|)\left(\Pi_{j=i-2}^{i+1}p_j\right)|\mathbb{E}\Delta^2 g(X_i + \xi_{i+2} + \xi_{i-3} + 2)| \\
&\qquad \le (p_{i-1}p_i + |a_i|)\left(\Pi_{j=i-2}^{i+1}p_j\right)\|\Delta g\| b_{n-2}(1, p_{i-3}, p_{i-4}, ..., p_1, p_n, ..., p_{i+2}, 1)
\end{aligned}
$$

and

$$
\begin{aligned}
&|(a_i p_i + p_{i-1}^2 p_i^2)p_{i+1}\mathbb{E}\Delta^2 g(V_i^{1+} + \xi_{i+2} + 1) + (a_i p_{i-1} + p_{i-1}^2 p_i^2)p_{i-2}\mathbb{E}\Delta^2 g(V_i^{1-} + \xi_{i-3} + 1) \\
&\qquad - (1 + a_i)\left(\Pi_{j=i-2}^{i+1}p_j\right)\mathbb{E}\Delta^2 g(X_i + \xi_{i+2} + \xi_{i-3} + 1)| \\
&\qquad \le (|a_i|p_i + p_{i-1}^2 p_i^2)p_{i+1}\|\Delta g\| b_{n-2}(p_{i-2}, p_{i-3}, ..., p_1, p_n, ..., p_{i+2}, 1) \\
&\qquad + (|a_i|p_{i-1} + p_{i-1}^2 p_i^2)p_{i-2}\|\Delta g\| b_{n-2}(1, p_{i-3}, ..., p_1, p_n, ..., p_{i+1}) \\
&\qquad + (1 + |a_i|)\left(\Pi_{j=i-2}^{i+1}p_j\right)\|\Delta g\| b_{n-2}(1, p_{i-3}, ..., p_1, p_n, ..., p_{i+2}, 1),
\end{aligned}
$$

while the fourth term of (5.15) becomes

$$
\begin{aligned}
&|p_{i-1}p_i(1 - p_{i-1})\mathbb{E}\{\Delta g(V_i + 1)(p_{i-2} - \xi_{i-2})\} + p_{i-1}p_i(1 - p_i)\mathbb{E}\{\Delta g(V_i + 1)(p_{i+1} - \xi_{i+1})\}| \\
&\qquad \le |p_{i-1}p_i(1 - p_{i-1})p_{i-2}\mathbb{E}\{\Delta g(V_i + 1) - \Delta g(V_i^{1-} + \xi_{i-3} + 1)\}| \\
&\qquad + |p_{i-1}p_i(1 - p_i)p_{i+1}\mathbb{E}\{\Delta g(V_i + 1) - \Delta g(V_i^{1+} + \xi_{i+2} + 1)\}| \\
&\qquad = \left(\Pi_{j=i-3}^{i}p_j\right)(1 - p_{i-1})(1 - p_{i-2})|\mathbb{E}\{\Delta^2 g(V_i^{2-} + \xi_{i-4} + 1)\}| \\
&\qquad + \left(\Pi_{j=i-1}^{i+2}p_j\right)(1 - p_i)(1 - p_{i+1})|\mathbb{E}\{\Delta^2 g(V_i^{2+} + \xi_{i+3} + 1)\}| \\
&\qquad \le \left(\Pi_{j=i-3}^{i}p_j\right)(1 - p_{i-1})(1 - p_{i-2})b_{n-3}(1, p_{i-4}, ..., p_1, p_n, ..., p_{i+1}) \\
&\qquad + \left(\Pi_{j=i-1}^{i+2}p_j\right)(1 - p_i)(1 - p_{i+1})b_{n-3}(p_{i-2}, ..., p_1, p_n, ..., p_{i+3}, 1).
\end{aligned}
$$

Because all the above $b_i's$ can be bounded by $\frac{4.6}{\sqrt{\gamma}}$ and $|a_i| \le p_{i-2}p_{i-1}p_i + p_{i-1}p_ip_{i+1} + p_{i-1}^2 p_i^2$, summarizing the above information, the theorem follows from adding up the bounds for $1 \le i \le n$, after some calculation. □

The approximation obtained improves in a number of ways on those previously known. The previous study has been concentrated on independent and identically distributed Bernoulli random variables only. For this particular case, the simplest good bound is that of Roos (1993, Th. 3.C); it has an explicit constant, but is only of order $O(p^2 \log(np^2))$ when $np^2 \to \infty$. The bound given in Eichelsbacher and Roos (1998) almost always has the better order $O(p^2)$, though the constant is complicated to write down. Here, we have a relatively simple constant, which can be improved if it is assumed for instance that $\sum_{i=2}^{n}(1 - \alpha_{i-2})^2\alpha_{i-1}(1 - \alpha_{i-1})\alpha_i$ is large, together with an order $O(p^2/\sqrt{np^2})$, which is even better than $O(p^2)$ when $np^2 \to \infty$. This is rather impressive precision. Curiously, the variance of the compound Poisson approximations differs from the true variance by an amount of order $np^4$, and this is presumably responsible for the fact that an approximation of better order than $O(p^2)$ is not obtained.

## References


[1] M. Abramowitz and I.A. Stegun, *Handbook of mathematical functions.* Dover, New York (1964).

[2] R. Arratia, A.D. Barbour and S. Tavaré, The number of components in a logarithmic combinatorial structure. *Ann. Appl. Probab.*, to appear.

[3] R. Arratia, L. Goldstein and L. Gordon, Poisson approximation and the Chen–Stein method. *Stat. Science* **5** (1990) 403–434.

[4] A.D. Barbour and J.L. Jensen, Local and tail approximations near the Poisson limit. *Scand. J. Statist.* **16** (1989) 75–87.

[5] A.D. Barbour and S. Utev, Solving the Stein equation in compound Poisson approximation. *Adv. in Appl. Probab.* **30** (1998) 449–475.

[6] A.D. Barbour and S. Utev, Compound Poisson approximation in total variation. *Stochastic Process. Appl.*, to appear.

[7] V. Čekanavičius, Asymptotic expansions in the exponent: A compound Poisson approach. *Adv. in Appl. Probab.* **29** (1997) 374–387.

[8] P. Eichelsbacher and M. Roos, Compound Poisson approximation for dissociated random variables *via* Stein's method (1998) preprint.

[9] J. Kruopis, Precision of approximations of the generalized Binomial distribution by convolutions of Poisson measures. *Lithuanian Math. J.* **26** (1986) 37–49.

[10] T. Lindvall, *Lectures on the coupling method.* Wiley, New York (1992).

[11] E.L. Presman, Approximation of binomial distributions by infinitely divisible ones. *Theory. Probab. Appl.* **28** (1983) 393–403.

[12] D.A. Raikov, On the decomposition of Gauss and Poisson laws. *Izv. Akad. Nauk Armyan. SSR Ser. Mat.* **2** (1938) 91–124.

[13] M. Roos, Stein–Chen method for compound Poisson approximation. Ph.D. Dissertation, University of Zürich (1993).