Adv. Appl. Prob. **40,** 1033–1047 (2008)
Printed in Northern Ireland

EXACT VALUES AND SHARP ESTIMATES FOR THE TOTAL VARIATION DISTANCE BETWEEN BINOMIAL AND POISSON DISTRIBUTIONS

JOSÉ A. ADELL,* **
JOSÉ M. ANOZ * *** AND
ALBERTO LEKUONA,* **** *Universidad de Zaragoza*

Abstract

We present a method to obtain both exact values and sharp estimates for the total variation distance between binomial and Poisson distributions with the same mean λ. We give a simple efficient algorithm, whose complexity order is $\sqrt{\lambda}$, to compute exact values. Such an algorithm can be further simplified for moderate sample sizes n, provided that λ is neither close to $l + \sqrt{l}$, $l = 1, 2, \ldots$, from the left nor close to $m - \sqrt{m}$, $m = 2, 3, \ldots$, from the right. Sharp estimates, better than other known estimates in the literature, are also provided. The 0s of the second Krawtchouk and Charlier polynomials play a fundamental role.

Keywords: Poisson approximation; binomial distribution; total variation distance; Krawtchouk polynomial; Charlier polynomial

2000 Mathematics Subject Classification: Primary 62E17
Secondary 60E15

1. Introduction

Since the pioneering papers by Prohorov [17] and Le Cam [16], a lot of work has been done on Poisson approximation for sums of independent random indicators, where the accuracy of the approximation is usually measured in terms of the total variation distance. Many authors have obtained sharp estimates for such a distance by using different approaches, such as the Stein–Chen method (cf. [6], [8], and [10]), semigroup techniques (cf. [12] and [13]), and Charlier expansions (cf. [18] and [19]), among others. By means of analytical techniques, Kennedy and Quine [14] computed the exact total variation distance between binomial and Poisson distributions when their common mean is less than or equal to $2 - \sqrt{2}$.

This paper is concerned with the most paradigmatic example of Poisson approximation, namely, the evaluation of the total variation distance between binomial and Poisson distributions having the same mean λ. Our purpose is twofold. On the one hand, our purpose is to delve into the results of Kennedy and Quine [14]. These authors have obtained exact values for the total variation distance for small values of λ (see Theorem 1 and the comments following Lemma 3.1 of [14]). In Theorems 2.1 and 2.2, below, we give an algorithm to compute exact values for any arbitrary mean λ. On the other hand, our purpose is to present a unified method to provide

Received 4 April 2008; revision received 7 October 2008.
* Postal address: Departamento de Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain.
** Email address: adell@unizar.es
*** Email address: anoz@unizar.es
**** Email address: lekuona@unizar.es

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

both exact values and sharp estimates in the problem under consideration. The basic ideas are the following. Suppose that we want to evaluate $\mathrm{E}\,\phi(Y) - \mathrm{E}\,\phi(X)$ for any given two random variables X and Y, and suitable functions ϕ. Consider a stochastic process $\mathcal{Z} = (Z(t),\ t \geq 0)$ 'connecting' X and Y, in the sense that $Z(0)$ and X have the same law and that $Z(t)$ converges in law to Y as $t \to \infty$. Under such circumstances, we can apply the differential calculus for the linear operator represented by $\mathcal{Z}$, as developed in [1], [3], and [5], in order to obtain a closed-form expression for $\mathrm{E}\,\phi(Z(t)) - \mathrm{E}\,\phi(Z(0)),\ t \geq 0$, taking advantage of the probabilistic structure of $\mathcal{Z}$. Letting $t \to \infty$, we obtain, under appropriate assumptions on ϕ,

$$\mathrm{E}\,\phi(Y) - \mathrm{E}\,\phi(X) = \lim_{t\to\infty} (\mathrm{E}\,\phi(Z(t)) - \mathrm{E}\,\phi(Z(0))). \tag{1.1}$$

Let us sketch the method to compute the right-hand side of (1.1) for the case in which the stochastic process $\mathcal{Z}$ takes values in the set $\mathbb{Z}_+$ of nonnegative integers. In such a case, the problem is to find $\mathbb{Z}_+$-valued stochastic processes $\mathbb{V}_i = (V_i(t),\ t \geq 0)$ and σ-finite measures ν_i defined on $[0, \infty)$, $i = 1, 2$, satisfying the differentiation formula

$$\mathrm{E}\,\phi(Z(t)) - \mathrm{E}\,\phi(Z(s)) = \int_{(s,t]} \mathrm{E}\,\Delta\phi(V_1(u))\,\mathrm{d}\nu_1(u) - \int_{(s,t]} \mathrm{E}\,\Delta\phi(V_2(u))\,\mathrm{d}\nu_2(u) \tag{1.2}$$

for any $0 \leq s < t < \infty$ and any $\phi\colon \mathbb{Z}_+ \to \mathbb{R}$ satisfying appropriate integrability conditions, where Δ is the first-order difference operator defined by $\Delta\phi(n) = \phi(n) - \phi(n-1)$, $n = 1, 2, \dots$. As shown in [1, Example 4.3], to find $\mathbb{V}_i$ and ν_i, $i = 1, 2$, it suffices to check (1.2) for the family of test functions $\phi_\zeta(n) = \mathrm{e}^{\mathrm{i}\zeta n}$, $n \in \mathbb{Z}_+$, $\zeta \in \mathbb{R}$, that is, to check that

$$\mathrm{E}\,\mathrm{e}^{\mathrm{i}\zeta Z(t)} - \mathrm{E}\,\mathrm{e}^{\mathrm{i}\zeta Z(s)} = (1 - \mathrm{e}^{-\mathrm{i}\zeta})\bigg(\int_{(s,t]} \mathrm{E}\,\mathrm{e}^{\mathrm{i}\zeta V_1(u)}\,\mathrm{d}\nu_1(u) - \int_{(s,t]} \mathrm{E}\,\mathrm{e}^{\mathrm{i}\zeta V_2(u)}\,\mathrm{d}\nu_2(u)\bigg)$$

for any $0 \leq s < t < \infty$ and $\zeta \in \mathbb{R}$. Rewriting (1.2) as

$$\begin{aligned}\mathrm{E}\,\phi(Z(t)) - \mathrm{E}\,\phi(Z(s)) &= \mathrm{E}\,\Delta\phi(V_1(s))\nu_1((s,t]) - \mathrm{E}\,\Delta\phi(V_2(s))\nu_2((s,t]) \\ &\quad + \int_{(s,t]} (\mathrm{E}\,\Delta\phi(V_1(u)) - \mathrm{E}\,\Delta\phi(V_1(s)))\,\mathrm{d}\nu_1(u) \\ &\quad - \int_{(s,t]} (\mathrm{E}\,\Delta\phi(V_2(u)) - \mathrm{E}\,\Delta\phi(V_2(s)))\,\mathrm{d}\nu_2(u),\end{aligned}$$

the preceding ideas can be applied again to the processes $\mathbb{V}_i$, $i = 1, 2$, appearing in the integrands of (1.2), arriving at a Taylor's formula of second order for the original process $\mathcal{Z}$, and so on.

If, in addition, $\mathcal{Z}$ is a discrete-time process, i.e.

$$Z(t) = \sum_{m=0}^{\infty} Z(m)\,\mathbf{1}_{(m-1,m]}(t), \qquad t \geq 0,$$

where $\mathbf{1}_A$ stands for the indicator function of the set A, then $\mathbb{V}_i$ is also a discrete-time process and ν_i, $i = 1, 2$, is a discrete measure with support on $\mathbb{Z}_+$ so that (1.2) has the form

$$\begin{aligned}\mathrm{E}\,\phi(Z(k)) - \mathrm{E}\,\phi(Z(n)) &= \sum_{m=n+1}^{k} (\mathrm{E}\,\phi(Z(m)) - \mathrm{E}\,\phi(Z(m-1))) \\ &= \sum_{m=n+1}^{k} (\mathrm{E}\,\Delta\phi(V_1(m))\nu_1(\{m\}) - \mathrm{E}\,\Delta\phi(V_2(m))\nu_2(\{m\}))\end{aligned} \tag{1.3}$$

for any $0 \le n < k$. Since there are many stochastic processes $\mathcal{Z}$ connecting X and Y, it seems that the efficiency of the method depends on the right choice of $\mathcal{Z}$. In this respect, the choice $Z(t) = tX + (1-t)Y$, $0 \le t \le 1$, where X and Y are taken to be independent random variables, is not suitable in general.

The aforementioned differential calculus has been successfully applied in [2] and [4] to give sharp estimates in Poisson and binomial approximations of Poisson and binomial mixtures, respectively, and might find more applications in the future. The preceding ideas, particularly (1.3), are close in spirit to the Lindeberg method considered in [9] to deal with Poisson approximation of sums of independent integer-valued random variables. They are also close to the probabilistic method to evaluate Stein's factors introduced in [22] (see also [7] and [21]). In fact, to estimate the total variation distance between two discrete random variables X and Y, these authors use a birth–death process $\mathcal{Z}$ whose initial and equilibrium distributions coincide with those of X and Y, respectively.

To be more precise, let $\mathbb{Z}_+$ be the set of nonnegative integers and let $\mathbb{N} = \mathbb{Z}_+ \setminus \{0\}$. Let $n \in \mathbb{N}$, and let $0 < t < 1$. Denote by

$$S_n(t) = \sum_{k=1}^{n} \mathbf{1}_{[0,t]}(V_k), \qquad S_0(t) = 0,$$

where $(V_k,\ k \in \mathbb{N})$ is a sequence of independent, identically distributed random variables having uniform distribution on $[0, 1]$. Clearly, $S_n(t)$ has the binomial distribution with parameters n and t. We consider the orthogonal polynomials with respect to $S_n(t)$ defined by

$$Q_m^{(n)}(t; x) = \frac{1}{\binom{n}{m}(t(1-t))^m} \sum_{k=0}^{m} \binom{n-x}{m-k}\binom{x}{k}(-t)^{m-k}(1-t)^k, \tag{1.4}$$

where $m = 0, 1, \ldots, n$. Up to a constant, each one of these polynomials coincides with the Krawtchouk polynomial of the same degree, as defined in [11, p. 161]. As shown there, such polynomials satisfy the orthogonality property

$$\mathrm{E}\, Q_m^{(n)}(t; S_n(t))Q_r^{(n)}(t; S_n(t)) = \frac{\delta_{m,r}}{\binom{n}{m}(t(1-t))^m}, \qquad m, r = 0, 1, \ldots, n. \tag{1.5}$$

On the other hand, let N_λ be a random variable having the Poisson distribution with mean $\lambda > 0$. The orthogonal polynomials with respect to N_λ, named the Charlier polynomials, are explicitly defined by (cf. [11, Chapter VI])

$$C_m(\lambda; x) = \sum_{k=0}^{m} \binom{m}{k}\binom{x}{k} k!\,(-\lambda)^{-k}, \qquad m \in \mathbb{Z}_+, \tag{1.6}$$

or, equivalently, by the three-term recurrence relation

$$-\lambda C_{m+1}(\lambda; x) = (x - m - \lambda)C_m(\lambda; x) + mC_{m-1}(\lambda; x), \qquad m \in \mathbb{N},$$

with initial conditions $C_{-1}(\lambda; x) = 0$ and $C_0(\lambda; x) = 1$. These polynomials fulfill the orthogonality condition

$$\mathrm{E}\, C_m(\lambda; N_\lambda)C_r(\lambda; N_\lambda) = \frac{m!}{\lambda^m}\delta_{m,r}, \qquad m, r \in \mathbb{Z}_+. \tag{1.7}$$

The orthogonal polynomials above, especially $Q_2^{(n)}(t; x)$ and $C_2(\lambda; x)$, will play an important role in dealing with the total variation distance between binomial and Poisson distributions. We refer the reader to the monograph by Schoutens [20] for further properties of general orthogonal polynomials and their connections with stochastic processes and various topics in probability theory.

Finally, for any $m \in \mathbb{Z}_+$, the mth forward differences of a function $\phi\colon \mathbb{Z}_+ \to \mathbb{R}$ are recursively defined by $\Delta^0\phi = \phi$, $\Delta^1\phi(i) = \phi(i+1) - \phi(i)$, $i \in \mathbb{Z}_+$, and $\Delta^{m+1}\phi = \Delta^1(\Delta^m\phi)$ or, equivalently, by

$$\Delta^{m+1}\phi(i) = \sum_{k=0}^{m+1}(-1)^{m+1-k}\binom{m+1}{k}\phi(i+k), \qquad i \in \mathbb{Z}_+. \tag{1.8}$$

From now on, it will be assumed that $n \in \mathbb{N}$, $0 < \lambda < n$, and $p = \lambda/n$. The natural choice of the stochastic process $\mathcal{Z}$ connecting $S_n(p)$ and N_λ is the discrete-time process $\mathcal{Z} = (S_k(\lambda/k))_{k\geq n}$. As shown in (3.2) and (4.1), below, we have

$$\begin{aligned} \mathrm{E}\,\phi(S_n(p)) - \mathrm{E}\,\phi(N_\lambda) &= -\lambda^2 \sum_{k=n}^{\infty} \frac{1}{k(k+1)}\,\mathrm{E}\,U\Delta^2\phi(S_{k-1}(T_k)) \\ &= -\lambda^2 \sum_{k=n}^{\infty} \frac{1}{k(k+1)}\,\mathrm{E}\,U\phi(S_{k+1}(T_k))Q_2^{(k+1)}(T_k; S_{k+1}(T_k)) \end{aligned} \tag{1.9}$$

for any function ϕ for which the preceding expectations exist, where

$$T_k = \frac{\lambda}{k}\left(1 - \frac{UV}{k+1}\right), \qquad k = n, n+1, \dots, \tag{1.10}$$

and U and V are independent, identically distributed random variables having uniform distribution on [0, 1]. Here and hereafter, all of the random variables appearing under the same expectation sign are supposed to be mutually independent. Equation (1.9) is the main tool used to obtain first-order estimates for $\mathrm{E}\,\phi(S_n(p)) - \mathrm{E}\,\phi(N_\lambda)$. By applying the same procedure to each one of the terms on the right-hand side of (1.9), we obtain second-order estimates for $\mathrm{E}\,\phi(S_n(p)) - \mathrm{E}\,\phi(N_\lambda)$, and so on. Replacing ϕ by an indicator function in (1.9), we are able to estimate the total variation distance between $S_n(p)$ and N_λ, as defined by

$$d(S_n(p), N_\lambda) = \frac{1}{2}\sum_{k=0}^{\infty}|\,\mathrm{P}(S_n(p) = k) - \mathrm{P}(N_\lambda = k)| \tag{1.11}$$

or, equivalently, by

$$d(S_n(p), N_\lambda) = \mathrm{P}(S_n(p) \in D_\lambda(n)) - \mathrm{P}(N_\lambda \in D_\lambda(n)), \tag{1.12}$$

where

$$D_\lambda(n) = \{i \in \{0, 1, \dots, n\}\colon\ \mathrm{P}(S_n(p) = i) \geq \mathrm{P}(N_\lambda = i)\}. \tag{1.13}$$

Looking at (1.9) and taking into account the fact that $Q_2^{(n)}(t; x)$ converges to $C_2(\lambda; x)$ as $n \to \infty$ and $nt \to \lambda$, it is quite clear that both exact values and sharp estimates for $d(S_n(p), N_\lambda)$ depend on the second Krawtchouk and Charlier polynomials (more specifically, on the 0s of such polynomials).

The paper is organized as follows. In Section 2 we give exact values for $d(S_n(p), N_\lambda)$ by computing the set $D_\lambda(n)$ by means of a simple efficient algorithm whose complexity order is $\sqrt{\lambda}$ (Theorem 2.1). In addition, it is shown in Theorem 2.2 that $D_\lambda(n)$ can be easily described for moderate values of n, provided that λ is neither close to $l+\sqrt{l}$, $l \in \mathbb{N}$, from the left nor close to $m-\sqrt{m}$, $m = 2, 3, \dots$, from the right. Finally, sharp estimates for $d(S_n(p), N_\lambda)$, together with a comparative discussion of known results in the literature, are provided in Theorem 2.3. The proofs of these theorems are postponed to the remaining sections.

2. Main results

Let $k \in \mathbb{N}$ with $k \geq n$, and let $0 < t < 1$. We see from (1.4) that

$$Q_2^{(k+1)}(t; x) = \frac{x^2 - (1+2kt)x + k(k+1)t^2}{k(k+1)(t(1-t))^2}. \tag{2.1}$$

The two 0s of this polynomial are given by

$$x_j^{(k+1)}(t) = \tfrac{1}{2} + kt + (-1)^j \sqrt{kt(1-t) + \tfrac{1}{4}}, \qquad j = 1, 2. \tag{2.2}$$

As follows from (1.9), we are dealing with second Krawtchouk polynomials whose parameter t is randomized by T_k, defined in (1.10). Observe that $\lambda/(k+1) \leq T_k \leq \lambda/k$. For this reason, we denote by

$$r_{1,k}(\lambda) = x_1^{(k+1)}\left(\frac{\lambda}{k}\right) = \frac{1}{2} + \lambda - \sqrt{\lambda\left(1 - \frac{\lambda}{k}\right) + \frac{1}{4}} \tag{2.3}$$

and by

$$r_{2,k}(\lambda) = x_2^{(k+1)}\left(\frac{\lambda}{k+1}\right) = \frac{1}{2} + \lambda\,\frac{k}{k+1} + \sqrt{\lambda\left(1 - \frac{\lambda}{k+1}\right)\frac{k}{k+1} + \frac{1}{4}} \tag{2.4}$$

the smallest 0 of $Q_2^{(k+1)}(\lambda/k; x)$ and the largest 0 of $Q_2^{(k+1)}(\lambda/(k+1); x)$, respectively (see Figure 1). Also, we consider the nonempty set

$$B_\lambda(k) = [\lceil r_{1,k}(\lambda) \rceil, \lfloor r_{2,k}(\lambda) \rfloor] \cap \mathbb{Z}_+, \tag{2.5}$$

where $\lfloor x \rfloor$ is the integer part of x and $\lceil x \rceil$ is the ceiling of x, that is, the smallest integer greater than or equal to x.

In view of (1.10) and (2.1), the random polynomial $Q_2^{(k+1)}(T_k; x)$ converges as $k \to \infty$ to the second Charlier polynomial $C_2(\lambda; x)$ given by

$$C_2(\lambda; x) = \frac{x^2 - (1+2\lambda)x + \lambda^2}{\lambda^2}, \tag{2.6}$$

the 0s of which are

$$r_j(\lambda) = \tfrac{1}{2} + \lambda + (-1)^j \sqrt{\lambda + \tfrac{1}{4}}, \qquad j = 1, 2. \tag{2.7}$$

Also, we define the nonempty set

$$D_\lambda = [\lfloor r_1(\lambda) \rfloor + 1, \lceil r_2(\lambda) \rceil - 1] \cap \mathbb{Z}_+. \tag{2.8}$$

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

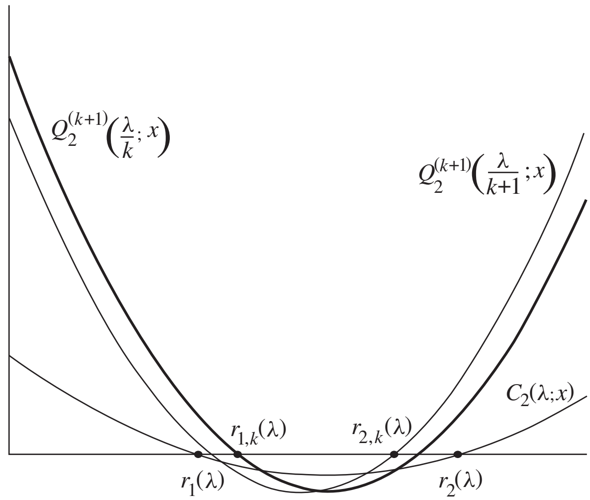

FIGURE 1: Relative position of $Q_2^{(k+1)}(\lambda/k;x)$, $Q_2^{(k+1)}(\lambda/(k+1);x)$, and $C_2(\lambda;x)$ for $\lambda > 2$.

Finally, for $i = 0, 1, \dots, n$, we write

$$P(S_n(p) = i) - P(N_\lambda = i) = P(N_\lambda = i)\left(\frac{c(n,\lambda)}{g_{n,\lambda}(i)} - 1\right), \tag{2.9}$$

where

$$c(n,\lambda) = n!\,e^\lambda\left(1 - \frac{\lambda}{n}\right)^n \quad \text{and} \quad g_{n,\lambda}(i) = (n-i)!\,(n-\lambda)^i. \tag{2.10}$$

In our first main result we give an algorithm to determine the set $D_\lambda(n)$ defined in (1.13), thus allowing us to obtain exact values of $d(S_n(p), N_\lambda)$.

Theorem 2.1. *Let $n \in \mathbb{N}$, and let $0 < \lambda < n$. Then*

(a) $B_\lambda(n) \subseteq D_\lambda(n) \subseteq D_\lambda$;

(b) $D_\lambda(n) = [l_\lambda(n), m_\lambda(n)] \cap \mathbb{Z}_+$, *where*

$$l_\lambda(n) = \min\{i \in [\lfloor r_1(\lambda)\rfloor + 1, \lceil r_{1,n}(\lambda)\rceil] \cap \mathbb{Z}_+ : g_{n,\lambda}(i) \le c(n,\lambda)\}$$

and

$$m_\lambda(n) = \max\{i \in [\lfloor r_{2,n}(\lambda)\rfloor, \lceil r_2(\lambda)\rceil - 1] \cap \mathbb{Z}_+ : g_{n,\lambda}(i) \le c(n,\lambda)\}.$$

Thanks to Theorem 2.1(a) and definition (1.12), we can give the following lower and upper bounds for $d(S_n(p), N_\lambda)$:

$$\begin{aligned}\max\{&P(S_n(p) \in B_\lambda(n)) - P(N_\lambda \in B_\lambda(n)), P(S_n(p) \in D_\lambda) - P(N_\lambda \in D_\lambda)\}\\ &\le d(S_n(p), N_\lambda)\\ &\le P(S_n(p) \in D_\lambda) - P(N_\lambda \in B_\lambda(n)).\end{aligned}$$

On the other hand, an implementation of the backward–forward algorithm in Theorem 2.1(b) should take into account the fact that the function $g_{n,\lambda}(\cdot)$ defined in (2.10) decreases in $[0,\lambda] \cap \mathbb{Z}_+$ and increases in $[\lambda, n] \cap \mathbb{Z}_+$, as well as the recurrence relations

$$g_{n,\lambda}(i+1) = \frac{n-\lambda}{n-i}\,g_{n,\lambda}(i), \qquad i = 0, 1, \dots, n-1.$$

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

In each step, the algorithm allows us to decide whether or not a point i belongs to $D_\lambda(n)$. In the positive case, the contribution of point i to the total variation distance is given in (2.9). We note that Kennedy and Quine [14, p. 398] provided a different algorithm to compute the exact total variation distance between binomial and Poisson distributions when their common mean λ is smaller than $2+\sqrt{2}$, approximately.

The Poisson approximation to the binomial law is usually applied for small or moderate values of λ. In such circumstances, the algorithm in Theorem 2.1(b) gives us the exact value of $d(S_n(p), N_\lambda)$ in a very efficient way. The algorithm is also interesting from a theoretical point of view when we assume that $\lambda = \lambda(n) \to \infty$ and $\lambda(n)/n \to 0$ as $n \to \infty$. To see this, observe that, by Theorem 2.1(a), $\#D_{\lambda(n)}(n) \le \#D_{\lambda(n)}$ and that $\#D_{\lambda(n)}$ has the order of $\sqrt{\lambda(n)}$, as follows from (2.7) and (2.8). This means that if the total variation distance is computed by applying (1.11), n differences of probabilities are needed, whereas if this distance is computed using (1.12) and the algorithm in Theorem 2.1(b), only $\sqrt{\lambda(n)}$ such differences are required. In this sense, the complexity order of the proposed algorithm is $\sqrt{\lambda(n)}$.

As follows from (2.5) and (2.8), the set $B_\lambda(n)$ and, therefore, $D_\lambda(n)$ approaches D_λ as $n \to \infty$. This implies that, under simple sufficient conditions, the set $D_\lambda(n)$ can be computed in an easier way than that in Theorem 2.1(b). In this regard, it will be of interest to describe D_λ in more detail. Noting that the functions $r_j(\lambda)$ in (2.7) are increasing and that

$$r_1(l+\sqrt{l}) = l, \quad l \in \mathbb{Z}_+, \qquad r_2(m-\sqrt{m}) = m, \quad m \in \mathbb{N},$$

we see that

$$\begin{aligned} &D_\lambda = [l+1, m] \cap \mathbb{Z}_+ \\ \text{if} \quad &\lambda \in A_{l,m} = [l+\sqrt{l}, l+1+\sqrt{l+1}) \cap (m-\sqrt{m}, m+1-\sqrt{m+1}] \end{aligned} \tag{2.11}$$

for any $l \in \mathbb{Z}_+$ and $m \in \mathbb{N}$. Observe that the family of sets $(A_{l,m},\ l \in \mathbb{Z}_+,\ m \in \mathbb{N})$ is a partition of $(0,\infty)$ and that the set D_λ is the same for any $\lambda \in A_{l,m}$. Table 1 displays the sets D_λ for $\lambda \in (0, 6]$.

We are in a position to state the following theorem.

TABLE 1: The set D_λ for $\lambda \in (0, 6]$.

λ	D_λ
$(0, 2-\sqrt{2}]$	$\{1\}$
$(2-\sqrt{2}, 3-\sqrt{3}]$	$\{1, 2\}$
$(3-\sqrt{3}, 1+\sqrt{1})$	$\{1, 2, 3\}$
$\{4-\sqrt{4}\}$	$\{2, 3\}$
$(4-\sqrt{4}, 5-\sqrt{5}]$	$\{2, 3, 4\}$
$(5-\sqrt{5}, 2+\sqrt{2})$	$\{2, 3, 4, 5\}$
$[2+\sqrt{2}, 6-\sqrt{6}]$	$\{3, 4, 5\}$
$(6-\sqrt{6}, 7-\sqrt{7}]$	$\{3, 4, 5, 6\}$
$(7-\sqrt{7}, 3+\sqrt{3})$	$\{3, 4, 5, 6, 7\}$
$[3+\sqrt{3}, 8-\sqrt{8}]$	$\{4, 5, 6, 7\}$
$(8-\sqrt{8}, 4+\sqrt{4})$	$\{4, 5, 6, 7, 8\}$
$\{4+\sqrt{4}\}$	$\{5, 6, 7, 8\}$

Theorem 2.2. *Let $n \in \mathbb{N}$, and let $0 < \lambda < n$. Assume that $\lambda \in A_{l,m}$, as defined in (2.11), for some $l \in \mathbb{Z}_+$ and $m \in \mathbb{N}$. Then*

(a) $B_\lambda(n) = D_\lambda(n) = D_\lambda$ *whenever*

$$n \geq \max\left\{\frac{\lambda^2}{l+1-(\lambda-(l+1))^2}, \frac{m(m-1)}{m-(\lambda-m)^2}\right\}; \tag{2.12}$$

(b) $D_\lambda \setminus D_\lambda(n) \subseteq D_\lambda \setminus B_\lambda(n) \subseteq \{l+1, m\}$ *whenever*

$$n \geq \max\left\{\frac{\sqrt{l+1}}{2}(\sqrt{l+1}+1)^2, \frac{(\sqrt{m}+1)(m-2)}{2}\right\}.$$

Concerning Theorem 2.2(a), we mention the following examples (see Table 1). Suppose that $\lambda \in (0, 2-\sqrt{2}]$. Since in this case $l = 0$ and $m = 1$, we have $D_\lambda(n) = D_\lambda = \{1\}$ for $n \geq \lambda/(2-\lambda)$, that is, for $n \geq 1$. This has already been shown in [14]. Assume that $\lambda = 4-\sqrt{4}$. In this case, $l = 1$ and $m = 3$, and, therefore, $D_\lambda(n) = D_\lambda = \{2, 3\}$ for $n \geq 3$. Finally, assume that $\lambda \in [3+\sqrt{3}, 8-\sqrt{8}]$. Since $l = 3$ and $m = 7$, we have $D_\lambda(n) = D_\lambda = \{4, 5, 6, 7\}$ for

$$n \geq \max\left\{\frac{\lambda^2}{4-(\lambda-4)^2}, \frac{42}{7-(\lambda-7)^2}\right\}, \tag{2.13}$$

in particular, for $n \geq 23$, as follows by taking suprema in $\lambda \in [3+\sqrt{3}, 8-\sqrt{8}]$ on the right-hand side of (2.13).

In general, denote by $n_0(\lambda)$ the smallest integer such that $D_\lambda(n) = D_\lambda$ for $n \geq n_0(\lambda)$. Numerical computations show that $n_0(\lambda)$ is not uniformly bounded when λ varies in a compact set. This is reflected in (2.12). In fact, for each $l \in \mathbb{Z}_+$, $n_0(\lambda)$ tends to ∞ when λ approaches $l+1+\sqrt{l+1}$ from the left. Similarly, for each $m = 2, 3, \dots$, $n_0(\lambda) \to \infty$ as λ tends to $m-\sqrt{m}$ from the right. This explains why the set $D_\lambda(n)$ has no simple form in general.

To illustrate Theorem 2.2(b), assume that $\lambda \in (8-\sqrt{8}, 4+\sqrt{4})$. In this case, $l = 3$ and $m = 8$. Hence, we have from Theorem 2.2(a) and Table 1 that $D_\lambda(n) = D_\lambda = \{4, 5, 6, 7, 8\}$ for

$$n \geq \max\left\{\frac{\lambda^2}{4-(\lambda-4)^2}, \frac{56}{8-(\lambda-8)^2}\right\}.$$

Observe that n is large if λ is close to $4+\sqrt{4}$ from the left or close to $8-\sqrt{8}$ from the right. However, Theorem 2.2(b) states that $D_\lambda \setminus D_\lambda(n) \subseteq \{4, 8\}$ for $n \geq 12$. Therefore, $D_\lambda(n)$ can be determined from D_λ by computing (2.9) for $i = 4, 8$, provided that $n \geq 12$. The important role played by the numbers $l+\sqrt{l}$ and $m-\sqrt{m}$, that is, the endpoints of the set $A_{l,m}$, $l \in \mathbb{Z}_+$, $m \in \mathbb{N}$, has already been noted in [14, p. 398].

Sharp estimates for $d(S_n(p), N_\lambda)$ are given in the following theorem, where it is shown that the set D_λ appears in the leading term of the estimate. To this end, denote by $x \wedge y = \min(x, y)$. For any $m \in \mathbb{N}$, $n = 2, 3, \dots$, and $0 < \lambda < n$, we set

$$f_m(n, \lambda) = 2^{m-1} \wedge \frac{1}{2}\left(\frac{n+2}{n-1}\right)^{3/2}\sqrt{\frac{m!}{\lambda^m(1-\lambda/n)^m}}, \tag{2.14}$$

as well as

$$K_\lambda(n) = \frac{n+2}{2(n+1)}\left(\frac{2\lambda}{3}f_3(n, \lambda) + \frac{\lambda^2}{4}f_4(n, \lambda)\right). \tag{2.15}$$

With this notation, we state the following theorem.

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

Theorem 2.3. *Let $n = 2, 3, \ldots$, let $0 < \lambda < n$, and let $p = \lambda/n$. Then*

$$\left| d(S_n(p), N_\lambda) - p\frac{\lambda\,\mathrm{E}\,|C_2(\lambda; N_\lambda)|}{4} \right| \le K_\lambda(n)p^2, \tag{2.16}$$

where $K_\lambda(n)$ is defined in (2.15).

Thanks to (6.4) and (6.5), below, we can write

$$\begin{aligned}\tfrac{1}{2}\,\mathrm{E}\,|C_2(\lambda; N_\lambda)| &= -\,\mathrm{E}\,\mathbf{1}_{D_\lambda}(N_\lambda)C_2(\lambda; N_\lambda)\\ &= \mathrm{E}\,\Delta^1\,\mathbf{1}_{D_\lambda}(N_\lambda)C_1(\lambda; N_\lambda)\\ &= C_1(\lambda; \lfloor r_1(\lambda)\rfloor)\,\mathrm{P}(N_\lambda = \lfloor r_1(\lambda)\rfloor)\\ &\quad - C_1(\lambda; \lceil r_2(\lambda)\rceil - 1)\,\mathrm{P}(N_\lambda = \lceil r_2(\lambda)\rceil - 1),\end{aligned} \tag{2.17}$$

where $C_1(\lambda; x) = (\lambda - x)/\lambda$, as follows from (1.6). Essentially, the preceding equalities were first obtained in [12].

As discussed in the introduction, several authors have shown similar estimates to that in Theorem 2.3 in the context of Poisson approximation for sums of independent random indicators (cf. [6], [13] and [15]). Specializing such estimates to the case at hand, we see the following. The leading term is the same as that on the left-hand side of (2.16), although written in different ways. However, instead of the upper bound on the right-hand side of (2.16), we find the following alternatives:

$$\begin{aligned}&2.6\,p^2 \quad \text{if} \quad p \le \tfrac{1}{4} \quad \text{(see [15])},\\ &2(1-\mathrm{e}^{-\lambda})p^2\left(\left(1 \wedge \frac{1.4}{\sqrt{\lambda}}\right) + 1 - \mathrm{e}^{-\lambda}\right) \quad \text{(see [6])},\\ &\frac{(2p)^{3/2}}{2(1-\sqrt{2p})} \quad \text{if} \quad p < \frac{1}{2} \quad \text{(see [13])}.\end{aligned}$$

From (2.14) and (2.15), it is readily seen that, for $n \ge 10$, we have

$$K_\lambda(n) \le \frac{4}{11}\left(4\left(\lambda \wedge \frac{\sqrt{2}}{3\sqrt{\lambda(1-p)^3}}\right) + \left(3\lambda^2 \wedge \frac{\sqrt{2}}{(1-p)^2}\right)\right). \tag{2.18}$$

This, together with simple numerical computations performed with MAPLE® 9.01, shows that the upper bound on the right-hand side of (2.16) is always better than the preceding ones for $0 < p \le \frac{1}{2}$ and $n \ge 10$. On the other hand, Roos [19] obtained the estimate

$$d(S_n(p), N_\lambda) \le \frac{3}{4\mathrm{e}}p + \frac{7(3-2\sqrt{p})}{6(1-\sqrt{p})^2}p\sqrt{p}, \tag{2.19}$$

where the leading constant $3/(4\mathrm{e})$ is the best possible (see also [18] for other estimates). Applying the triangular inequality to (2.16), we obtain the upper bound

$$d(S_n(p), N_\lambda) \le p\frac{\lambda\,\mathrm{E}\,|C_2(\lambda; N_\lambda)|}{4} + K_\lambda(n)p^2. \tag{2.20}$$

Observe that the leading constant in (2.20) satisfies (cf. [4] or [19])

$$\frac{1}{4}\sup_{0<\lambda<n} \lambda\,\mathrm{E}\,|C_2(\lambda; N_\lambda)| = \frac{1}{4}\mathrm{E}\,|C_2(1; N_1)| = \frac{3}{4\mathrm{e}}.$$

On the other hand, using (2.6), as well as the central limit theorem and the strong law of large numbers for the standard Poisson process, we have

$$\frac{1}{4}\lim_{\lambda\to\infty}\lambda\,\mathrm{E}\,|C_2(\lambda;\,N_\lambda)| = \frac{1}{4}\lim_{\lambda\to\infty}\mathrm{E}\left|\left(\frac{N_\lambda-\lambda}{\sqrt{\lambda}}\right)^2 - \frac{N_\lambda}{\lambda}\right| = \frac{1}{4}\mathrm{E}\,|Z^2-1| = \frac{1}{\sqrt{2\pi\mathrm{e}}},$$

where Z is a standard normal variable. Finally, for $0<\lambda\le 2-\sqrt{2}$, we have, from the first equality in (2.17), Table 1, and (2.6),

$$\frac{\lambda\,\mathrm{E}\,|C_2(\lambda;\,N_\lambda)|}{4} = -\frac{\lambda}{2}\mathrm{E}\,\mathbf{1}_{\{1\}}(N_\lambda)C_2(\lambda;\,N_\lambda) = \frac{2-\lambda}{2}\lambda\mathrm{e}^{-\lambda}.$$

In other words, the leading constant in (2.20) has the order of λ for a small λ. The preceding comments show that the leading constant in (2.20) is better than $3/(4\mathrm{e})$.

Finally, the remainder term in (2.20) has the order of p^2 with constant $K_\lambda(n)$ satisfying (2.18). This means that such a term is better than the second term in (2.19).

3. Auxiliary results

For the sake of concreteness, we always denote by ϕ a function $\phi\colon\mathbb{Z}_+\to[0,1]$, although the results in this section essentially hold for arbitrary exponentially bounded functions. It follows from (1.8) that

$$|\Delta^{m+1}\phi(i)|\le 2^m,\qquad m,i\in\mathbb{Z}_+. \tag{3.1}$$

We start with the following lemma.

Lemma 3.1. *For any $m\in\mathbb{N}$, $n\in\mathbb{Z}_+$, and $0<t<1$, we have*

$$|\mathrm{E}\,\Delta^m\phi(S_n(t))|\le 2^{m-1}\wedge\frac{1}{2\sqrt{\binom{n+m}{m}(t(1-t))^m}}.$$

Proof. From [2, Lemma 7] we have

$$\mathrm{E}\,\Delta^m\phi(S_n(t)) = \mathrm{E}\,\phi(S_{n+m}(t))Q_m^{(n+m)}(t;\,S_{n+m}(t)). \tag{3.2}$$

On the other hand, let $A=\{x\in\mathbb{R}\colon Q_m^{(n+m)}(t;x)\ge 0\}$. The orthogonality condition in (1.5) gives us

$$\begin{aligned}\tfrac{1}{2}\mathrm{E}\,|Q_m^{(n+m)}(t;\,S_{n+m}(t))| &= \mathrm{E}\,\mathbf{1}_A(S_{n+m}(t))Q_m^{(n+m)}(t;\,S_{n+m}(t))\\ &= -\mathrm{E}\,\mathbf{1}_{\mathbb{R}\setminus A}(S_{n+m}(t))Q_m^{(n+m)}(t;\,S_{n+m}(t)).\end{aligned}$$

This, together with the fact that $0\le\phi\le 1$, yields

$$\begin{aligned}&|\mathrm{E}\,\phi(S_{n+m}(t))Q_m^{(n+m)}(t;\,S_{n+m}(t))|\\ &\quad\le\max\{\mathrm{E}\,\mathbf{1}_A(S_{n+m}(t))Q_m^{(n+m)}(t;\,S_{n+m}(t)),\,-\mathrm{E}\,\mathbf{1}_{\mathbb{R}\setminus A}(S_{n+m}(t))Q_m^{(n+m)}(t;\,S_{n+m}(t))\}\\ &\quad=\tfrac{1}{2}\mathrm{E}\,|Q_m^{(n+m)}(t;\,S_{n+m}(t))|.\end{aligned} \tag{3.3}$$

Thus, the conclusion follows from (3.1)–(3.3) and Hölder's inequality.

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

Let $n \in \mathbb{N}$, and let $0 < \lambda < n$. Suppose that $k \in \mathbb{N}$ with $k \geq n$. Let U and V be two independent, identically distributed random variables having uniform distribution on [0, 1], and consider a sequence $(p_l)_{l \geq k} \subseteq (0, 1)$. For any $l \geq k$, we define

$$\lambda_l = l p_l, \qquad R_l = p_l + (p_{l+1} - p_l)U, \qquad T_l = p_l + (p_{l+1} - p_l)UV. \tag{3.4}$$

Lemma 3.2. *Let $m \in \mathbb{Z}_+$ and $k \in \mathbb{N}$. If $\lambda_l \to \lambda$ as $l \to \infty$ then*

$$\begin{aligned}\mathrm{E}\,\Delta^m \phi(S_k(p_k)) - \mathrm{E}\,\Delta^m \phi(N_\lambda) &= \sum_{l=k}^{\infty} (\lambda_l - \lambda_{l+1})\,\mathrm{E}\,\Delta^{m+1}\phi(S_l(R_l)) \\ &\quad + \sum_{l=k}^{\infty} \lambda_l (p_{l+1} - p_l)\,\mathrm{E}\,U\Delta^{m+2}\phi(S_{l-1}(T_l)).\end{aligned} \tag{3.5}$$

If, in addition, $(\lambda_l)_{l \geq k}$ is increasing, $(p_l)_{l \geq k}$ is decreasing, and $\lambda/(l+2) \leq p_l \leq \lambda/(l+1)$, $l = k, k+1, \ldots$, then

$$\begin{aligned}|\mathrm{E}\,\Delta^m \phi(S_k(p_k)) - \mathrm{E}\,\Delta^m \phi(N_\lambda)| &\leq (\lambda - \lambda_k) f_{m+1}(k+1, \lambda) \\ &\quad + \frac{1}{2} f_{m+2}(k+1, \lambda) \sum_{l=k}^{\infty} \lambda_l (p_l - p_{l+1}),\end{aligned} \tag{3.6}$$

where $f_m(\cdot, \cdot)$ is defined in (2.14).

Proof. For any $x, y \in (0, 1)$, we have, by calculus (see also [2, Theorem 1]),

$$\begin{aligned}\mathrm{E}\,\phi(S_n(x)) - \mathrm{E}\,\phi(S_{n+1}(y)) &= (nx - (n+1)y)\,\mathrm{E}\,\Delta^1 \phi(S_n(x + (y-x)U)) \\ &\quad + nx(y-x)\,\mathrm{E}\,U\Delta^2 \phi(S_{n-1}(x + (y-x)UV)), \qquad n \in \mathbb{N}.\end{aligned}$$

This, together with (3.4) and the telescoping sum,

$$\mathrm{E}\,\phi(S_k(p_k)) - \mathrm{E}\,\phi(N_\lambda) = \sum_{l=k}^{\infty} (\mathrm{E}\,\phi(S_l(p_l)) - \mathrm{E}\,\phi(S_{l+1}(p_{l+1}))),$$

shows (3.5). To show (3.6), observe that we have, by assumption and (3.4),

$$\frac{\lambda}{l+3} \leq p_{l+1} \leq U_l \leq p_l \leq \frac{\lambda}{l+1}, \qquad l = k, k+1, \ldots,$$

where $U_l = R_l$ or $U_l = T_l$. Hence, estimate (3.6) follows from Lemma 3.1, (2.14), and (3.5). The proof is complete.

4. Proof of Theorem 2.1

Choosing $m = 0$, $k = n$, and $p_l = \lambda/l$, $l \geq k$, in (3.5), and recalling (3.2), we have

$$\mathrm{E}\,\phi(S_n(p)) - \mathrm{E}\,\phi(N_\lambda) = -\lambda^2 \sum_{k=n}^{\infty} \frac{1}{k(k+1)}\,\mathrm{E}\,U\phi(S_{k+1}(T_k))Q_2^{(k+1)}(T_k; S_{k+1}(T_k)), \tag{4.1}$$

where

$$T_k = \frac{\lambda}{k}\left(1 - \frac{UV}{k+1}\right), \qquad k \geq n. \tag{4.2}$$

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

On the other hand, it follows from (2.3)–(2.8) that $(B_\lambda(k))_{k\geq n}$ is a nondecreasing sequence of sets such that

$$\bigcup_{k=n}^{\infty} B_\lambda(k) = D_\lambda. \tag{4.3}$$

We will firstly show that $B_\lambda(n) \subseteq D_\lambda(n)$. Let $i \in B_\lambda(n)$. We claim (see Figure 1) that

$$Q_2^{(k+1)}(t;i) \leq 0, \qquad \frac{\lambda}{k+1} \leq t \leq \frac{\lambda}{k},\ k \geq n. \tag{4.4}$$

Indeed, the functions $x_j^{(k+1)}(t)$ defined in (2.2) are increasing in t. Therefore, if $\lambda/(k+1) \leq t \leq \lambda/k$, we have, by virtue of (2.3) and (2.4),

$$x_1^{(k+1)}(t) \leq x_1^{(k+1)}\left(\frac{\lambda}{k}\right) = r_{1,k}(\lambda) \leq r_{2,k}(\lambda) = x_2^{(k+1)}\left(\frac{\lambda}{k+1}\right) \leq x_2^{(k+1)}(t). \tag{4.5}$$

As stated in (4.3), $(B_\lambda(k))_{k\geq n}$ is nondecreasing and, therefore, $i \in B_\lambda(k)$, $k \geq n$. From (2.1), (2.2), and (4.5), this means that $Q_2^{(k+1)}(t;i) \leq 0$, thus showing claim (4.4).

By (4.2), we have $\lambda/(k+1) \leq T_k \leq \lambda/k$, $k \geq n$. Thus, applying (4.1) with $\phi = \mathbf{1}_{\{i\}}$, and taking into account (4.4), we see that $\mathrm{P}(S_n(p) = i) \geq \mathrm{P}(N_\lambda = i)$. This shows that $B_\lambda(n) \subseteq D_\lambda(n)$.

Let $i \in \mathbb{Z}_+ \setminus D_\lambda$. To show that $i \in \mathbb{Z}_+ \setminus D_\lambda(n)$, we distinguish the following two cases.

Case 1: $\lambda > 2$. Let $k \geq n$, and let $\lambda/(k+1) \leq t \leq \lambda/k$. It can be checked from (2.2)–(2.4) and (2.7) that

$$r_1(\lambda) < x_1^{(k+1)}\left(\frac{\lambda}{k+1}\right) \leq x_1^{(k+1)}(t) < x_2^{(k+1)}(t) \leq x_2^{(k+1)}\left(\frac{\lambda}{k}\right) < r_2(\lambda).$$

By virtue of (2.1) and (2.2), this means that

$$Q_2^{(k+1)}(t;i) > 0, \qquad \frac{\lambda}{k+1} \leq t \leq \frac{\lambda}{k},\ k \geq n. \tag{4.6}$$

As above, (4.6) implies that $i \in \mathbb{Z}_+ \setminus D_\lambda(n)$.

Case 2: $\lambda \leq 2$. The previous argument partially fails now, because in the case at hand we have $x_1^{(k+1)}(\lambda/(k+1)) \leq r_1(\lambda)$. As follows from Table 1, either $i = 0$ or $i \geq r_2(\lambda)$. Since $0 \notin D_\lambda(n)$, it can be assumed that $i \geq r_2(\lambda)$. In such a case, (4.6) is again true and, therefore, $i \in \mathbb{Z}_+ \setminus D_\lambda(n)$. We have shown that $D_\lambda(n) \subseteq D_\lambda$.

To show part (b), observe that the function $g_{n,\lambda}(\cdot)$ defined in (2.10) decreases in $[0,\lambda] \cap \mathbb{Z}_+$ and increases in $[\lambda, n] \cap \mathbb{Z}_+$. Also, it follows from (2.3) and (2.4) that $r_{1,n}(\lambda) < \lambda < r_{2,n}(\lambda)$. Therefore, part (b) follows from (2.5), (2.8), (2.9), and Theorem 2.1(a). This completes the proof.

5. Proof of Theorem 2.2

Assume that $\lambda \in A_{l,m}$, as defined in (2.11), for some $l \in \mathbb{Z}_+$ and $m \in \mathbb{N}$. Clearly, $B_\lambda(n) = D_\lambda(n) = D_\lambda$ whenever $B_\lambda(n) = D_\lambda$ or, equivalently, whenever

$$[\lceil r_{1,n}(\lambda) \rceil, \lfloor r_{2,n}(\lambda) \rfloor] \cap \mathbb{Z}_+ = [l+1, m] \cap \mathbb{Z}_+,$$

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

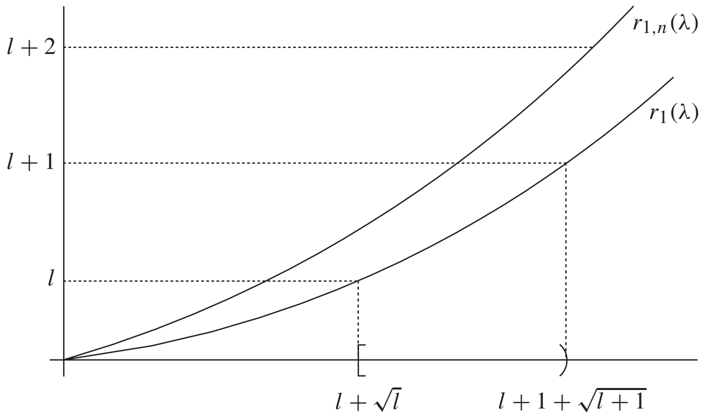

FIGURE 2: Relative position of $r_{1,n}(\lambda)$ and $r_1(\lambda)$.

as follows from (2.5) and (2.11). By (2.3) and (2.7), the functions $r_{1,n}(\lambda)$ and $r_1(\lambda)$ are strictly increasing in λ, and $r_{1,n}(\lambda) \geq r_1(\lambda)$ (see Figure 2). Thus, $\lceil r_{1,n}(\lambda)\rceil = l+1$ if and only if $r_{1,n}(\lambda) \leq l+1$. By (2.3) and some straightforward computations, this is equivalent to

$$n \geq \frac{\lambda^2}{l+1-(\lambda-(l+1))^2}. \tag{5.1}$$

Similarly, the functions $r_{2,n}(\lambda)$ and $r_2(\lambda)$ are strictly increasing in λ, and $r_{2,n}(\lambda) \leq r_2(\lambda)$. Hence, $\lfloor r_{2,n}(\lambda)\rfloor = m$ if and only if $r_{2,n}(\lambda) \geq m$ or, equivalently,

$$n \geq \frac{m(m-1)}{m-(\lambda-m)^2}, \tag{5.2}$$

as follows from (2.4) and some easy computations. Thus, (2.12) follows from (5.1) and (5.2). This shows part (a).

To show part (b), it follows from (2.5) and (2.11) that $D_\lambda \setminus B_\lambda(n) \subseteq \{l+1, m\}$ whenever $r_{1,n}(\lambda) \leq l+2$ and $r_{2,n}(\lambda) \geq m-1$. In particular (see Figure 2), whenever $r_{1,n}(l+1+\sqrt{l+1}) \leq l+2$ and $r_{2,n}(m-\sqrt{m}) \geq m-1$. Again, by (2.3) and (2.4), this last statement is equivalent to

$$n \geq \max\left\{\frac{(l+1+\sqrt{l+1})^2}{2\sqrt{l+1}}, \frac{(\sqrt{m}+1)(m-2)}{2}\right\},$$

thus showing (b). The proof is complete.

6. Proof of Theorem 2.3

We claim that

$$\left|\mathrm{E}\,\phi(S_n(p)) - \mathrm{E}\,\phi(N_\lambda) + \frac{\lambda^2}{2n}\mathrm{E}\,\Delta^2\phi(N_\lambda)\right| \leq \frac{n+2}{2(n+1)}p^2\left(\frac{2\lambda}{3}f_3(n,\lambda) + \frac{\lambda^2}{4}f_4(n,\lambda)\right). \tag{6.1}$$

https://doi.org/10.1239/aap/1231340163 Published online by Cambridge University Press

In fact, as in the proof of Theorem 2.1, we have, from (3.5),

$$\begin{aligned}\mathrm{E}\,\phi(S_n(p)) - \mathrm{E}\,\phi(N_\lambda) &= -\lambda^2 \sum_{k=n}^{\infty} \frac{1}{k(k+1)}\,\mathrm{E}\,U\,\Delta^2\phi(S_{k-1}(T_k)) \\ &= -\frac{\lambda^2}{2n}\,\mathrm{E}\,\Delta^2\phi(N_\lambda) \\ &\quad - \lambda^2 \sum_{k=n-1}^{\infty} \frac{1}{(k+1)(k+2)}\,\mathrm{E}\,U(\Delta^2\phi(S_k(T_{k+1})) - \Delta^2\phi(N_\lambda)), \end{aligned} \tag{6.2}$$

where T_k is defined in (4.2). Fix $k \geq n-1$, and consider the sequence $(\tilde{p}_l)_{l\geq k}$ given by $\tilde{p}_l = T_{l+1}$, $l \geq k$. By (4.2), this sequence satisfies the requirements of Lemma 3.2 to obtain estimate (3.6) and, in addition,

$$\lambda - kT_{k+1} \leq \frac{\lambda}{k+1}(1+UV), \qquad lT_{l+1}(T_{l+1} - T_{l+2}) \leq \frac{\lambda^2}{(l+1)(l+2)}, \quad l \geq k.$$

Hence, applying (3.6), we have

$$\begin{aligned}&|\mathrm{E}\,U(\Delta^2\phi(S_k(T_{k+1})) - \Delta^2\phi(N_\lambda))| \\ &\quad\leq \frac{\lambda}{k+1}\,\mathrm{E}\,U(1+UV)f_3(k+1,\lambda) + \frac{\lambda^2}{4} f_4(k+1,\lambda) \sum_{l=k}^{\infty} \frac{1}{(l+1)(l+2)} \\ &\quad\leq \left(\frac{2\lambda}{3} f_3(n,\lambda) + \frac{\lambda^2}{4} f_4(n,\lambda)\right)\frac{1}{k+1}, \end{aligned} \tag{6.3}$$

where the second inequality holds because $f_m(\cdot,\lambda)$ is decreasing. Claim (6.1) follows from (6.2) and (6.3).

On the other hand, it is known (cf. [8, Lemma 9.4.4]) that

$$(-1)^m\,\mathrm{E}\,\Delta^m\phi(N_\lambda) = \mathrm{E}\,\phi(N_\lambda)C_m(\lambda; N_\lambda), \qquad m \in \mathbb{Z}_+. \tag{6.4}$$

By (1.7), (2.7), and (2.8), we have

$$\sup_{A\subseteq\mathbb{Z}_+} |\mathrm{E}\,\mathbf{1}_A(N_\lambda)C_2(\lambda; N_\lambda)| = -\mathrm{E}\,\mathbf{1}_{D_\lambda}(N_\lambda)C_2(\lambda; N_\lambda) = \tfrac{1}{2}\,\mathrm{E}\,|C_2(\lambda; N_\lambda)|. \tag{6.5}$$

Choosing $\phi = \mathbf{1}_A$ in (6.1), taking suprema in $A \subseteq \mathbb{Z}_+$, and applying (6.4) and (6.5), we obtain estimate (2.16). This concludes the proof.

Acknowledgements

The authors would like to thank the anonymous referees for their careful reading of the manuscript and for their suggestions, which greatly improved the final outcome. This work has been supported by research grants MTM2005-08376-C02-01 and DGA E-64, and by FEDER funds.

References

[1] Adell, J. A. (2008). A differential calculus for linear operators represented by stochastic processes. Preprint. Available at http://www.unizar.es/galdeano/preprints/lista.html.

[2] Adell, J. A. and Anoz, J. M. (2008). Signed binomial approximation of binomial mixtures via differential calculus for linear operators. *J. Statist. Planning Infer.* **138,** 3687–3698.

[3] Adell, J. A. and Lekuona, A. (2000). Taylor's formula and preservation of generalized convexity for positive linear operators. *J. Appl. Prob.* **37,** 765–777.

[4] Adell, J. A. and Lekuona, A. (2005). Sharp estimates in signed Poisson approximation of Poisson mixtures. *Bernoulli* **11,** 47–65.

[5] Adell, J. A. and Pérez-Palomares, A. (1999). Stochastic orders in preservation properties by Bernstein-type operators. *Adv. Appl. Prob.* **31,** 492–509.

[6] Barbour, A. D. and Hall, P. (1984). On the rate of Poisson convergence. *Math. Proc. Camb. Phil. Soc.* **95,** 473–480.

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

[8] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation* (Oxford Stud. Prob. **2**). Clarendon Press, Oxford.

[9] Borisov, I. S. and Ruzankin, P. S. (2002). Poisson approximation for expectations of unbounded functions of independent random variables. *Ann. Prob.* **30,** 1657–1680.

[10] Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Prob.* **3,** 534–545.

[11] Chihara, T. S. (1978). *An Introduction to Orthogonal Polynomials*. Gordon and Breach, New York.

[12] Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation. *Ann. Prob.* **14,** 663–676.

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

[14] Kennedy, J. E. and Quine, M. P. (1989). The total variation distance between the binomial and Poisson distributions. *Ann. Prob.* **17,** 396–400.

[15] Kerstan, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam. *Z. Wahrscheinlichkeitsth.* **2,** 173–179.

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

[17] Prohorov, Y. V. (1953). Asymptotic behavior of the binomial distribution. *Uspehi Matem. Nauk* **8,** 135–142.

[18] Roos, B. (1999). Asymptotic and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. *Bernoulli* **5,** 1021–1034.

[19] Roos, B. (2001). Sharp constants in the Poisson approximation. *Statist. Prob. Lett.* **52,** 155–168.

[20] Schoutens, W. (2000). *Stochastic Processes and Orthogonal Polynomials* (Lecture Notes Statist. **146**). Springer, New York.

[21] Weinberg, G. V. (2000). Stein factor bounds for random variables. *J. Appl. Prob.* **37,** 1181–1187.

[22] Xia, A. (1999). A probabilistic proof of Stein's factors. *J. Appl. Prob.* **36,** 287–290.