# A Charlier-Parseval approach to Poisson approximation and its applications

Vytas Zacharovas, Hsien-Kuei Hwang
Institute of Statistical Science
Academia Sinica
Taipei 115
Taiwan

October 27, 2008



**Abstract**

A new approach to Poisson approximation is proposed. The basic idea is very simple and based on properties of the Charlier polynomials and the Parseval identity. Such an approach quickly leads to new effective bounds for several Poisson approximation problems. A selected survey on diverse Poisson approximation results is also given.

*MSC 2000 Subject Classifications*: Primary 62E17; secondary 60C05 60F05.
*Key words*: Poisson approximation, total variation distance, Wasserstein distance, Kolmogorov distance, point metric, Kullback-Leibler distance, $\chi^2$-distance, Charlier polynomials, Parseval identity, Tauberian theorems

## 1 Introduction

Poisson approximation to many discrete distributions (notably the Poisson-binomial distribution) has received extensive attention in the literature and many different approaches have been proposed. The main problem is to study the closeness between the discrete distribution in question and a suitably chosen Poisson distribution. Applications in diverse problems also stimulated much of its recent interest among probabilists and scientists in applied disciplines. We propose in this paper a new, self-contained approach to Poisson approximation, which leads readily to many new effective bounds for several distances studied before, including total variation, Kolmogorov, Wasserstein, Kullback-Leibler, point metric, and $\chi^2$; see below for more information and references. In addition to the application to these distances, we also attempt to survey most of the quantitative results we collected for the Poisson approximation distances discussed in this paper.

### 1.1 A historical account with brief review of results

We start with a brief historical account of Poisson approximation, focusing particular on the evolution of the total variation distance; a more detailed, technical discussion will be given in Section 6. For other surveys, see [38, 9, 4, 22, 72].

**The early history of Poisson approximation.** Poisson distribution appeared naturally as the limit of the sum of a large number of independent trials each with very small probability of success. Such a limit form, being the most primitive version of Poisson approximation, dates back to at least de Moivre's work [32] in the early eighteenth century and Poisson's book [61] in the nineteenth century. Haight [38] writes: "...although Poisson (or de Moivre) discovered the mathematical expression (1.1-1) [which is $e^{-\lambda}\lambda^k/k!$], Bortkiewicz discovered the probability distribution (1.1-1)." And according to Good [37], "perhaps the Poisson distribution should have been named after von Bortkiewicz (1898) because he was the first to write extensively about rare events whereas Poisson added little to what de Moivre had said on the matter and was probably aware of de Moivre's work;" see also Seneta's account in [74] on Abbe's work. In addition to Bortkiewicz's book [17], another important contribution to the early history of Poisson approximation was made by Charlier [21] for his type B expansion, which will play a crucial role in our development of arguments.

The next half a century or so after Bortkiewicz and Charlier then witnessed an increase of interests in the properties and applications of the Poisson distribution and Charlier's expansion. In particular, Jordan [47] proved the orthogonality of the Charlier polynomials with respect to the Poisson measure, and considered a formal expansion pair, expressing the Taylor coefficients of a given function in terms of series of Charlier polynomials and vice versa. A sufficient condition justifying the validity of such an expansion pair was later on provided by Uspensky [83]; he also derived very precise estimates for the coefficients in the case of binomial distribution. His complex-analytic approach was later on extended by Shorgin [80] to the more general Poisson-binomial distribution (each trial with a different probability; see next paragraph). Schmidt [73] then gives a sufficient and necessary condition for justifying the Charlier-Jordan expansion; see also Boas [13] and the references therein. Prohorov [65] was the first to study, using elementary arguments, the total variation distance between binomial and Poisson distributions, thus upgrading the classical limit theorem to an approximation theorem.

**From classical to modern.** However, a large portion of the development of modern theory of Poisson approximation deviates significantly from the classical line, and much of its modern interest can be attributed to the pioneering paper by Le Cam [54], extending the previous study by Prohorov [65] for binomial distribution. Le Cam considered particularly the sum $S_n$ of $n$ independent Bernoulli random variables with parameters $p_1, p_2, \dots, p_n$, respectively, and proved that the total variation distance

$$d_{TV}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) := \frac{1}{2}\sum_{j\geqslant 0}\left|\mathbb{P}(S_n = j) - e^{-\lambda}\frac{\lambda^j}{j!}\right|$$

between the distribution of $S_n$ (often referred to as the Poisson-binomial distribution) and that of a Poisson with mean $\lambda := \sum_{1\leqslant j\leqslant n} p_j$ is bounded above by

$$d_{TV}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) \leqslant 8\theta,$$

whenever $p_* := \max_j p_j \leqslant 1/4$, where $\theta := \lambda_2/\lambda$, $\lambda_2 := \sum_{1\leqslant j\leqslant n} p_j^2$. He also proved in the same paper the following inequality, now often referred to under his name,

$$d_{TV}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) \leqslant \lambda_2. \tag{1.1}$$

These results were later on further improved in the literature and the approach he used became the source of developments of more advanced tools; see Table 1.1 for a selected list of known results of the simplest form $d_{TV} \leqslant c\theta$.

| Author(s) | Year | $d_{TV} \leqslant$ | Assumption | Approach |
|---|---|---|---|---|
| Le Cam | 1960 | $8\theta$ | $p_* \leqslant \frac{1}{4}$ | Operator and Fourier |
| Kerstan | 1964 | $1.05\theta$ | $p_* \leqslant \frac{1}{4}$ | Operator and Fourier |
| Chen | 1974 | $5\theta$ | | Chen-Stein |
| Barbour and Hall | 1984 | $\theta$ | | Chen-Stein |
| Presman | 1985 | $2.08\theta$ | | Fourier |
| Daley and Vere-Jones | 1988 | $0.71\theta$ | $p_* \leqslant \frac{1}{4}$ | Fourier |

Table 1: *Some results of the form $d_{TV} := d_{TV}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) \leqslant c\theta$. Here $\theta := \lambda_2/\lambda$ and $p_* := \max_j p_j$. It is known that $d_{TV}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) \sim \theta/\sqrt{2\pi e}$ when $\theta \to 0$; see Deheuvels and Pfeifer [30] or Hwang [43]. Numerically, $1/\sqrt{2\pi e} \approx 0.242$.*

Form Table 1.1, we should point out that the leading constant in the first-order estimate for $d_{TV}$ is often less important than the generality of the approach used, although the pursuit for optimal leading constant is of independent interest *per se*. One reason is that if an approach is quickly amended for obtaining higher-order estimates, then one can push the calculations further by obtaining more terms in the asymptotic expansions with smaller and smaller errors, so that the implied constants in the error terms matter less (the derivation of which often involves detailed calculus).

On the other hand, estimates for the total variation distance between the distribution of $S_n$ and a suitably chosen Poisson distribution has been the subject of many papers in the last five decades. Other forms in the literature include $d_{TV} \leqslant \varphi(\theta)$, $d_{TV} \leqslant \varphi(\theta, \max_j p_j)$, $d_{TV} \leqslant \varphi(\theta, \lambda), \dots$, for certain functionals $\varphi$ ($\varphi$ not the same for each occurrence). Thus it is often difficult to compare these results; further complications arise because some metrics are related to others by simple inequalities and the results for one can be transferred to the others; also the complexity of the diverse methods of proof is not easily compared. Despite these, we quickly review those that are pertinent to ours, a more detailed, technical comparative discussion for some of these will be given later; the special case of binomial distribution will however not be compared separately; see, for example, Prohorov [65], Vervaat [84], Romanowska [67], Matsunawa [56], Pfeifer [59], Kennedy and Quine [48], Poor [63].

Kerstan [49] refined some results of Le Cam [54] on $d_{TV}$ by a similar approach. He also derived a second-order estimate. Herrmann [39] further extended results in Kerstan [49] in two directions: to *sums of random variables each assuming finitely many integer values* and, in addition to higher-order estimates from the Charlier expansion, to *signed measures* whose generating functions are of the forms $\exp(\sum_{1\leqslant j\leqslant s}(-1)^{j-1}\lambda_j(z-1)^j/j)$. We will comment on Kerstan's and Herrmann's second-order estimates later. As far as we are aware, Herrmann [39] was the first to use such signed measures for Poisson approximation problems, although such approximations are later on referred to as Kornya-Presman or Kornya-type approximations, the two references being Kornya [52] and Presman [64]. Note that the idea of using other signed measures (binomial) were already discussed in Le Cam [54]. Serfling [75] extended Le Cam's inequality (1.1) to dependent cases; see also [76]. Chen [23] proposed a new approach to Poisson approximation, based on Stein's method of normal approximation (see Stein [78]).

From 1980 on, most of the approaches proposed previously for Poisson approximation problems received much more attention and were further developed and refined. Among these, *the Chen-Stein method* (with or without *couplings*) is undoubtedly the most widely used and the most fruitful one. It is readily amended for dealing with dependent situations, but leads usually to less precise bounds for numerical purposes. On the other hand, direct or indirect classical *Fourier analysis*, although involving less probability ingredient and relying on more explicit forms of generating functions, often gives better numerical bounds.

For these and other approaches (including *semigroup* with Fourier analysis, *information-theoretic*), see Deheuvels and Pfeifer [28], Stein [78], Aldous (1989), Barbour et al. [9], Steele [81], Janson [46], Roos [69, 70], Kontoyiannis et al. [51] and the references therein.

## 1.2 Our new approach

The new approach we are developing in this paper starts from the integral representation for a given sequence $\{A_n\}_{n\geqslant 0}$ (satisfying certain conditions specified in the next section)

$$\sum_{n\geqslant 0}\left|\frac{A_n}{e^{-\lambda}\frac{\lambda^n}{n!}}\right|^2 e^{-\lambda}\frac{\lambda^n}{n!} = \int_0^\infty e^{-r} I(\sqrt{r/\lambda})\,dr, \tag{1.2}$$

where $\lambda > 0$ and

$$I(r) := \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|e^{-\lambda re^{it}}\sum_{j\geqslant 0} A_j(1+re^{it})^j\right|^2 dt.$$

Note that $I(r) = \sum_{n\geqslant 0}|a_n|^2 r^{2n}$, where $a_n$ denotes the coefficient of $z^n$ in the Taylor expansion of $e^{-\lambda z}\sum_{j\geqslant 0} A_j(1+z)^j$. This means that (1.2) can be written in the form

$$\sum_{n\geqslant 0}\frac{|A_n|^2}{e^{-\lambda}\frac{\lambda^n}{n!}} = \sum_{n\geqslant 0}|a_n|^2\frac{n!}{\lambda^n},$$

which, as far as we are aware, already appeared in the paper Pollaczek-Geiringer [62], but no further use of it has been discussed; see also Jacob [45], Schmidt [73], Siegmund-Schultze [77] and the references cited there. Also the series on the right-hand side is in almost all cases we are considering less useful than the integral in 1.2.

The seemingly strange and complicated starting point (1.2) turns out to be very useful for developing effective tools for most Poisson approximation problems. Other ingredients required are surprisingly simple, with very little use of complex analysis. A typical result is of the form

$$d_{TV}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) \leqslant \frac{(\sqrt{e}-1)\theta}{\sqrt{2}(1-\theta)^{3/2}},$$

where $(\sqrt{e}-1)/\sqrt{2} \approx 0.46$; see Theorem 3.4. The relation (1.2), which will be proved below, is based on the orthogonality of Charlier polynomials and Parseval identity; thus we call it *the Charlier-Parseval identity*.

Other features of our approach are: first, it reduces the estimate of the probability distances to that of certain integral representations with a similar form to the right-hand side of (1.2), and thus being of certain Tauberian character; second, it can be readily extended to derive asymptotic expansions; third, the use of the correspondence between Charlier polynomials and Poisson distribution can be quickly amended for other families of orthogonal polynomials and their corresponding probability distributions; fourth, the same idea used applies equally well to the de-Poissonization procedure, and leads to some interesting new results, details being discussed elsewhere.

**Organization of the paper.** This paper is organized as follows. We begin with the development of our approach in the next section. Then except for Section 6, which is focusing on reviewing and comparing with known results, the next three sections consist of applications of our Charlier-Parseval approach: Section 3 to several distances of Poisson approximation to $S_n$ for large $\lambda$, Section 4 to second order estimates, Section 5 to approximations by signed measures.

# 2 The new Charlier-Parseval approach

Crucial to the development our approach is the use of Charlier polynomials, so we first derive a few properties of Charlier polynomials we will need.

## 2.1 Definition and basic properties of Charlier polynomials

The Charlier polynomials $C_k(\lambda, n)$ are defined by

$$\sum_{n\geqslant 0} C_k(\lambda, n)\frac{\lambda^n}{n!}\, z^n = (z-1)^k e^{\lambda z} \qquad (k = 0, 1, \dots). \tag{2.1}$$

Multiplying both sides by $z-1$, we see that

$$\frac{\lambda^{n-1}}{(n-1)!}C_k(\lambda, n-1) - \frac{\lambda^n}{n!}C_k(\lambda, n) = \frac{\lambda^n}{n!}C_{k+1}(\lambda, n), \tag{2.2}$$

which implies that the Charlier polynomials $\varphi_k(n) := C_k(\lambda, n)$ are solutions to the system of difference equations $x\varphi_k(x-1) - \lambda\varphi_k(x) = \lambda\varphi_{k+1}(x)$, with the initial condition $\varphi_0(x) \equiv 1$. In particular,

$$C_1(\lambda, n) = \frac{n-\lambda}{\lambda} \quad \text{and} \quad C_2(\lambda, n) = \frac{n^2 - (2\lambda+1)n + \lambda^2}{\lambda^2}. \tag{2.3}$$

An alternative expression for $C_k(\lambda, n)$ is given by

$$\frac{\lambda^n}{n!}C_k(\lambda, n) = e^{\lambda}\frac{d^k}{d\lambda^k}\, e^{-\lambda}\frac{\lambda^n}{n!},$$

which follows from substituting the relation $(z-1)^k e^{\lambda z} = e^{\lambda}(d^k/d\lambda^k)\, e^{\lambda(z-1)}$ into (2.1).

Since by (2.1)

$$C_k(\lambda, n)\frac{\lambda^n}{n!} = [z^n](z-1)^k e^{\lambda z}, \tag{2.4}$$

where $[z^n]\phi(z)$ denotes the coefficient of $z^n$ in the Taylor expansion of $\phi(z)$, we have, for each fixed $n$,

$$\begin{aligned}\frac{\lambda^n}{n!}\sum_{k\geqslant 0}\frac{\lambda^k}{k!}C_k(\lambda, n)w^k &= [z^n]\sum_{k\geqslant 0}\frac{\lambda^k}{k!}w^k(z-1)^k e^{\lambda z} \\ &= [z^n]e^{-\lambda w + z\lambda(w+1)} \\ &= \frac{\lambda^n}{n!}(1+w)^m e^{-\lambda w}.\end{aligned}$$

It follows that

$$\sum_{n\geqslant 0} C_n(\lambda, k)\frac{\lambda^n}{n!}\, w^n = (1+w)^k e^{-\lambda w}.$$

Comparing this relation with (2.1), we obtain the property $C_k(\lambda, n) = (-1)^{n+k}C_n(\lambda, k)$, for all $k, n \geqslant 0$.

Another important property we will need is the following orthogonality relation (see [79, p. 35]).

**Lemma 2.1.** *The Charlier polynomials are orthogonal with respect to the Poisson measure $e^{-\lambda}\lambda^n/n!$, namely,*

$$\sum_{n\geqslant 0} C_k(\lambda, n)C_\ell(\lambda, n)e^{-\lambda}\frac{\lambda^n}{n!} = \delta_{k,\ell}\frac{k!}{\lambda^k}, \tag{2.5}$$

*where $\delta_{a,b}$ denotes the Kronecker symbol.*

For self-containedness and in view of the importance of this orthogonality relation to our analysis below, we give here a proof similar to the original one by Jordan [47].

*Proof.* We start from the expansion

$$C_k(\lambda, n) = \sum_{0\leqslant j\leqslant k}\binom{k}{j}(-1)^{k-j}\frac{n(n-1)\cdots(n-j+1)}{\lambda^j}, \tag{2.6}$$

which follows directly from (2.4). Differentiating both sides of (2.1) $j$ times with respect to $z$ and substituting $z = 1$, we get

$$\sum_{n\geqslant 0} e^{-\lambda}\frac{\lambda^n}{n!}C_k(\lambda, n)n(n-1)\cdots(n-j+1) = \begin{cases} j! & \text{if } j = k; \\ 0 & \text{if } j < k, \end{cases}$$

which means that the Charlier polynomials $C_k(\lambda, x)$ are orthogonal to any falling factorials of the form $x(x-1)\cdots(x-j+1)$ with $j < k$ with respect to the Poisson measure. Now without loss of generality, we may assume that $\ell \leqslant k$. Then applying (2.6), we get

$$\begin{aligned}\sum_{n\geqslant 0} e^{-\lambda}\frac{\lambda^n}{n!}C_k(\lambda, n)C_\ell(\lambda, n) &= \sum_{0\leqslant j\leqslant \ell}\binom{\ell}{j}(-1)^{\ell-j}\lambda^{-j}\sum_{n\geqslant 0}e^{-\lambda}\frac{\lambda^n}{n!}C_k(\lambda, n)n(n-1)\cdots(n-j+1) \\ &= \sum_{0\leqslant j\leqslant \ell}\binom{\ell}{j}(-1)^{\ell-j}\lambda^{-j}\delta_{k,j}k! \\ &= \delta_{k,\ell}\frac{k!}{\lambda^k}.\end{aligned}$$

This completes the proof. □

## 2.2 The Charlier-Parseval identity

Assume that we have a generating function

$$F(z) = \sum_{n\geqslant 0} A_n z^n,$$

which can be written in the form

$$F(z) = e^{\lambda(z-1)}f(z). \tag{2.7}$$

Let

$$f(z) = \sum_{j\geqslant 0} a_j(z-1)^j.$$

Then, by (2.4), we have *formally* the Charlier-Jordan expansion

$$A_n = e^{-\lambda}\frac{\lambda^n}{n!}\sum_{j\geqslant 0} a_j C_j(\lambda, n), \tag{2.8}$$

and we expect that $A_n$ will be close to $e^{-\lambda}\lambda^n/n!$ if $f(z)$ is close to $1$, or, alternatively, if $a_0$ is close to $1$ and all other $a_j$'s are close to $0$. The following identity provides our first step in quantifying such a heuristic.

**Proposition 2.2** (Charlier-Parseval identity)**.** *Assume that $f(z)$ is analytic in the whole complex plane and satisfies*

$$|f(z)| = O\left(e^{H|z-1|^2}\right), \tag{2.9}$$

*as $|z| \to \infty$. Then for any $\lambda > 2H$*

$$\sum_{n\geqslant 0}\left|\frac{A_n}{e^{-\lambda}\frac{\lambda^n}{n!}}\right|^2 e^{-\lambda}\frac{\lambda^n}{n!} = \int_0^\infty I(\sqrt{r/\lambda})e^{-r}\,dr, \tag{2.10}$$

*where*

$$I(r) := \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(1+re^{it})|^2\,dt. \tag{2.11}$$

*Proof.* Since by definition $I(r) = \sum_{j\geqslant 0}|a_j|^2 r^{2j}$ and the condition (2.9) implies the convergence of the series $\sum_{j\geqslant 0}|a_j|^2 j!/\lambda^j$, it follows that

$$\int_0^\infty I\left(\sqrt{r/\lambda}\right)e^{-r}\,dr = \sum_{j\geqslant 0}|a_j|^2\frac{j!}{\lambda^j}. \tag{2.12}$$

Both the series and the integral are convergent because, by (2.9), $I(r) = O(e^{2Hr^2})$.

Again by definition

$$\sum_{n\geqslant 0} A_n z^n = e^{\lambda(z-1)}\sum_{j\geqslant 0} a_j(z-1)^j.$$

Taking coefficient of $z^n$ on both sides, we obtain (2.8), which can be written as

$$\frac{A_n}{e^{-\lambda}\frac{\lambda^n}{n!}} = \sum_{j\geqslant 0} a_j C_j(\lambda, n),$$

where the convergence of the above series is pointwise. But the convergence of the series in (2.12) implies that the series on the right side also converges in $L_2$-norm with respect to the Poisson measure $e^{-\lambda}\lambda^n/n!$. Thus the Proposition follows from (2.5). □

In the special cases when $F(z) = (z-1)^k e^{\lambda(z-1)}$, or $A_n = C_k(\lambda, n)e^{-\lambda}\lambda^n/n!$, we have the identity

$$\sum_{n\geqslant 0} e^{-\lambda}\frac{\lambda^n}{n!}|C_k(\lambda, n)|^2 = k!\lambda^{-k} \qquad (k = 0, 1, \dots),$$

which is nothing but (2.5) with $k = \ell$. This implies that

$$\sum_{n\geqslant 0} e^{-\lambda}\frac{\lambda^n}{n!}|C_k(\lambda, n)| \leqslant \sqrt{k!}\lambda^{-k/2} \qquad (k = 0, 1, \dots). \tag{2.13}$$

## 2.3 A probabilistic interpretation of the Charlier-Parseval identity

Assume that $F(z)$ is a probability generating function of some non-negative integer valued random variable $X$ having the form

$$F(z) := \sum_{m\geqslant 0} \mathbb{P}(X=m)z^m = e^{\lambda(z-1)} \sum_{j\geqslant 0} a_j (z-1)^j.$$

Applying the Charlier-Parseval identity (2.10) and (2.12) to $F$ gives

$$\sum_{m\geqslant 0} \left| \frac{\mathbb{P}(X=m)}{e^{-\lambda}\frac{\lambda^m}{m!}} - 1 \right|^2 e^{-\lambda}\frac{\lambda^m}{m!} = \sum_{j\geqslant 1} \frac{j!}{\lambda^j}|a_j|^2,$$

provided that both series converge. In view of the orthogonality relations (2.5), the coefficients $a_j$ can be expressed as

$$a_j = \frac{\lambda^j}{j!}\sum_{m\geqslant 0}\mathbb{P}(X=m)C_j(\lambda,m) = \frac{\lambda^j}{j!}\mathbb{E}C_j(\lambda,X).$$

Thus

$$\sum_{m\geqslant 0} \left| \frac{\mathbb{P}(X=m)}{e^{-\lambda}\frac{\lambda^m}{m!}} - 1 \right|^2 e^{-\lambda}\frac{\lambda^m}{m!} = \sum_{j\geqslant 1} \frac{\lambda^j}{j!}\big|\mathbb{E}C_j(\lambda,X)\big|^2.$$

This identity relates the closeness of $X$ to Poisson measure by means of the moments of $X$ since the quantity $\mathbb{E}C_j(\lambda,X)$ is a linear combination of the moments of $X$.

On the other hand, it is also clear, by Cauchy-Schwarz inequality, that the series on the right-hand side satisfies

$$\sum_{j\geqslant 1} \frac{\lambda^j}{j!}\big|\mathbb{E}C_j(\lambda,X)\big|^2 = \sup \frac{\Big(\mathbb{E}\sum_{j\geqslant 1}a_jC_j(\lambda,X)\Big)^2}{\sum_{j\geqslant 1}a_j^2 j!/\lambda^j},$$

where the supremum is taken over all real sequences $\{a_j\}_{j\geqslant 1}$ such that $\sum_{j\geqslant 1}a_j^2j!/\lambda^j<\infty$. Let

$$g(x) := \sum_{j\geqslant 1} a_jC_j(\lambda,x).$$

Then

$$\sup \frac{\Big(\mathbb{E}\sum_{j\geqslant 1}a_jC_j(\lambda,X)\Big)^2}{\sum_{j\geqslant 1}a_j^2 j!/\lambda^j} = \sup_{\mathbb{E}g(\zeta)=0}\frac{\big(\mathbb{E}g(X)\big)^2}{\mathbb{E}g(\zeta)^2},$$

where $\zeta$ is a Poisson random variable with mean $\lambda$.

Applying the difference equation (2.2) for Charlier polynomials and taking into account that $a_0 = \mathbb{E}g(X) = 0$. we then have

$$g(X) = \frac{1}{\lambda}\sum_{j\geqslant 1}a_j\mathbb{E}\big(XC_{j-1}(\lambda,X-1)-\lambda C_{j-1}(\lambda,X)\big) = \frac{1}{\lambda}\big(Xh(X-1)-\lambda h(X)\big),$$

where $h(x)=\sum_{j\geqslant 1}a_kC_{j-1}(\lambda,x)$. Thus we can write

$$\left(\sum_{m\geqslant 0} \left| \frac{\mathbb{P}(X=m)}{e^{-\lambda}\frac{\lambda^m}{m!}} - 1 \right|^2 e^{-\lambda}\frac{\lambda^m}{m!}\right)^{1/2} = \sup \mathbb{E}\big(Xh(X-1)-\lambda h(X)\big),$$

the supremum being taken over all functions $h$ such that $\mathbb{E}\big(\zeta h(\zeta-1)-\lambda h(\zeta)\big)^2=1$. The right-hand side of the last expression is reminiscent of the Chen-Stein equation; see the book [9]; see also Goldstein and Reinert [36] and the references therein for the connection between orthogonal polynomials and Stein's method.

## 2.4 Asymptotic forms of the Charlier-Parseval identity

The identity (2.10) can be readily extended to the following effective (or asymptotic) versions for large $\lambda$.

**Proposition 2.3** (Asymptotic forms of the Charlier-Parseval identity)**.** *Let $F(z)$ and $f(z)$ be defined as above. Assume that $f$ is an entire function and satisfies the condition*

$$|f(z)| \leqslant K e^{H|z-1|^2}, \tag{2.14}$$

*for all $z\in\mathbb{C}$, with some positive constants $K$ and $H$. Then uniformly for all $N\geqslant 0$ and $\lambda \geqslant (2+\varepsilon)H$ with $\varepsilon>0$*

$$\sum_{n\geqslant 0}\left|\frac{A_n}{e^{-\lambda}\frac{\lambda^n}{n!}}-\sum_{0\leqslant j\leqslant N} a_j C_j(\lambda,n)\right|^2 e^{-\lambda}\frac{\lambda^n}{n!} \leqslant K^2\frac{2+\varepsilon}{\varepsilon}\left(\frac{(2+\varepsilon)H}{\lambda}\right)^{N+1}, \tag{2.15}$$

$$\sum_{n\geqslant 0}\left|A_n - e^{-\lambda}\frac{\lambda^n}{n!}\sum_{0\leqslant j\leqslant N} a_j C_j(\lambda,n)\right| \leqslant K\sqrt{\frac{2+\varepsilon}{\varepsilon}}\left(\frac{(2+\varepsilon)H}{\lambda}\right)^{(N+1)/2}, \tag{2.16}$$

*and uniformly for all $n\geqslant 0$*

$$\left|A_n - e^{-\lambda}\frac{\lambda^n}{n!}\left(\sum_{0\leqslant j\leqslant N} a_j C_j(\lambda,n)\right)\right| \leqslant K\frac{2+\varepsilon}{\varepsilon}\cdot\frac{1}{\sqrt{\lambda}}\left(\frac{(2+\varepsilon)H}{\lambda}\right)^{(N+1)/2}. \tag{2.17}$$

*Proof.* Applying (2.10) with $\lambda=(2+\varepsilon)H$ and using the upper bound $I(r)\leqslant K^2 e^{2Hr^2}$ (by (2.14)), we get

$$\begin{aligned}\sum_{j\geqslant 0}\frac{|a_j|^2 j!}{\big((2+\varepsilon)H\big)^j} &= \int_0^\infty I\left(\sqrt{\frac{r}{(2+\varepsilon)H}}\right)e^{-r}\,dr\\ &\leqslant K^2\int_0^\infty e^{-r(1-2/(2+\varepsilon))}\,dr\\ &= K^2\frac{2+\varepsilon}{\varepsilon}.\end{aligned}$$

Applying again Proposition 2.2 but to the function $f(z)=g(z)-\sum_{0\leqslant j\leqslant N} a_j(z-1)^j$ and using the above estimate for $\lambda\geqslant(2+\varepsilon)H$, we get

$$\begin{aligned}\sum_{n\geqslant 0}\left|\frac{A_n}{e^{-\lambda}\frac{\lambda^n}{n!}}-\sum_{0\leqslant j\leqslant N} a_j C_j(\lambda,n)\right|^2 e^{-\lambda}\frac{\lambda^n}{n!} &= \sum_{j>N}|a_j|^2\frac{j!}{\lambda^j}\\ &\leqslant \frac{1}{\lambda^{N+1}}\sum_{j>N}\frac{|a_j|^2 j!}{\big((2+\varepsilon)H\big)^{j-(N+1)}}\\ &= \frac{\big((2+\varepsilon)H\big)^{N+1}}{\lambda^{N+1}}\sum_{j>N}\frac{|a_j|^2 j!}{\big((2+\varepsilon)H\big)^j}\\ &\leqslant K^2\frac{2+\varepsilon}{\varepsilon}\left(\frac{(2+\varepsilon)H}{\lambda}\right)^{N+1}.\end{aligned}$$

Thus (2.15) follows and the estimate (2.16) is an immediate consequence of Cauchy-Schwarz inequality.

For (2.17), we apply Proposition 2.2 to the function

$$(1-z)\left(f(z)-\sum_{0\leqslant j\leqslant N} a_j(z-1)^j\right),$$

and obtain

$$\sum_{n\geqslant 0}\left|\frac{A_n-A_{n-1}}{e^{-\lambda}\frac{\lambda^n}{n!}}-\sum_{0\leqslant j\leqslant N} a_j C_{j+1}(\lambda,n)\right|^2 e^{-\lambda}\frac{\lambda^n}{n!}=\sum_{j>N}\frac{|a_j|^2(j+1)!}{\lambda^{j+1}}.$$

By partial summation, (2.2) and Cauchy-Schwarz inequality

$$\begin{aligned}
\left|A_n-e^{-\lambda}\frac{\lambda^n}{n!}\left(\sum_{0\leqslant j\leqslant N} a_j C_j(\lambda,n)\right)\right| &\leqslant \sum_{0\leqslant m\leqslant n}\left|A_m-A_{m-1}-e^{-\lambda}\frac{\lambda^m}{m!}\left(\sum_{0\leqslant j\leqslant N} a_j C_{j+1}(\lambda,m)\right)\right| \\
&\leqslant \left(\sum_{m\geqslant 0}\left|\frac{A_m-A_{m-1}}{e^{-\lambda}\frac{\lambda^m}{m!}}-\sum_{0\leqslant j\leqslant N} a_j C_{j+1}(\lambda,n)\right|^2 e^{-\lambda}\frac{\lambda^m}{m!}\right)^{1/2} \\
&= \left(\sum_{j>N}\frac{|a_j|^2(j+1)!}{\lambda^{j+1}}\right)^{1/2}. \qquad (2.18)
\end{aligned}$$

Now for $\lambda\geqslant(2+\varepsilon)H$

$$\begin{aligned}
\sum_{n\geqslant 0}\frac{|a_n|^2(n+1)!}{\lambda^{N+1}} &= \frac{1}{\lambda}\int_0^\infty I\left(\sqrt{r/\lambda}\right) re^{-r}\,dr \\
&\leqslant \frac{K^2}{\lambda}\int_0^\infty e^{-r(1-2H/\lambda)} r\,dr \\
&= \frac{K^2}{\lambda(1-2H/\lambda)^2}.
\end{aligned}$$

Thus (2.17) follows from substituting this bound into (2.18). □

## 2.5 Some useful estimates of Tauberian type

We now derive a few other effective bounds for certain partial sums or series by applying the Charlier-Parseval bounds we derived above; these bounds are more suitable for use for the diverse Poisson approximation distances we will consider. They are the types of results that have more or less the flavor of typical Tauberian theorems.

Assume that $\zeta_\lambda$ is a Poisson$(\lambda)$ distribution. Denote by

$$Z(n)=\min\left\{\mathbb{P}(\zeta_\lambda\leqslant n),\mathbb{P}(\zeta_\lambda>n)\right\}.$$

It is clear that $Z(n)\leqslant 1/2$.

**Proposition 2.4.** *Let $F, f, A_n, a_n$ and $I$ be defined as in (2.7) and (2.11). Assume that $f(z)$ is an entire function and satisfies the condition (2.9). Then for $\lambda > 2H$ the following inequalities hold. For $n \geqslant 0$,*

$$\sum_{n\geqslant 0} |A_n| \leqslant \left( \int_0^\infty I(\sqrt{r/\lambda}) e^{-r}\, dr \right)^{1/2}, \tag{2.19}$$

$$|A_n| \leqslant \frac{1}{\sqrt{\lambda}} \left( \int_0^\infty I(\sqrt{r/\lambda}) r e^{-r}\, dr \right)^{1/2} \sqrt{Z(n)}. \tag{2.20}$$

*If we additionally assume that $F(1) = 0$, then for $n \geqslant 0$,*

$$\sum_{n\geqslant 0} |A_0 + A_1 + \cdots + A_n| \leqslant \sqrt{\lambda} \left( \int_0^\infty I(\sqrt{r/\lambda}) r^{-1} e^{-r}\, dr \right)^{1/2}, \tag{2.21}$$

$$|A_0 + A_1 + \cdots + A_n| \leqslant \left( \int_0^\infty I(\sqrt{r/\lambda}) e^{-r}\, dr \right)^{1/2} \sqrt{Z(n)}. \tag{2.22}$$

*Proof.* By Cauchy-Schwarz inequality

$$\sum_{n\geqslant 0} |A_n| = \sum_{n\geqslant 0} \frac{|A_n|}{e^{-\lambda}\frac{\lambda^n}{n!}} \left( e^{-\lambda} \frac{\lambda^n}{n!} \right)^{1/2} \left( e^{-\lambda} \frac{\lambda^n}{n!} \right)^{1/2} \leqslant \left( \sum_{n\geqslant 0} \left| \frac{A_n}{e^{-\lambda}\frac{\lambda^n}{n!}} \right|^2 e^{-\lambda} \frac{\lambda^n}{n!} \right)^{1/2}.$$

The upper bound (2.19) then follows from (2.10).

The third inequality (2.21) is proved by applying (2.19) to the function $F_1(z) := F(z)/(1-z)$. Note that the condition $F(1) = 0$ implies that $F_1(z)$ is regular at $z = 1$. With this $F_1$, (2.19) now has the form

$$\sum_{n\geqslant 0} |A_0 + A_1 + \cdots + A_n| \leqslant \left( \int_0^\infty I_1(\sqrt{r/\lambda}) e^{-r}\, dr \right)^{1/2},$$

where

$$I_1(r) = \frac{1}{2\pi r^2} \int_{-\pi}^{\pi} |f(1 + re^{it})|^2\, dt = I(r)/r^2,$$

and (2.21) follows.

For the fourth inequality (2.22), we start from applying the Cauchy-Schwarz inequality, giving

$$|A_0 + A_1 + \cdots + A_n| \leqslant \left( \sum_{j\geqslant 0} \left| \frac{A_j}{e^{-\lambda}\frac{\lambda^j}{j!}} \right|^2 e^{-\lambda} \frac{\lambda^j}{j!} \right)^{1/2} \left( \sum_{0\leqslant j\leqslant n} e^{-\lambda} \frac{\lambda^j}{j!} \right)^{1/2}. \tag{2.23}$$

On the other hand, the condition $F(1) = 0$ implies that $\sum_{j\geqslant 0} A_j = 0$. Consequently,

$$\begin{aligned} |A_0 + A_1 + \cdots + A_n| &= |A_{n+1} + A_{n+2} + \cdots| \\ &\leqslant \left( \sum_{j\geqslant 0} \left| \frac{A_j}{e^{-\lambda}\frac{\lambda^j}{j!}} \right|^2 e^{-\lambda} \frac{\lambda^j}{j!} \right)^{1/2} \left( \sum_{j>n} e^{-\lambda} \frac{\lambda^j}{j!} \right)^{1/2}. \end{aligned} \tag{2.24}$$

Taking the minimum of the two upper bounds (2.23) and (2.24), we obtain (2.22).

Finally, the second inequality (2.20) follows from (2.22) by applying it to the generating function $(1-z)F(z)$ instead of $F(z)$. □

# 3 Applications. I. Distances for Poisson approximation

We apply in this section the diverse tools based on the Charlier-Parseval identity and derive bounds for the closeness between the Poisson-binomial distribution and a Poisson distribution with the same mean. We need a few simple inequalities.

## 3.1 Lemmas

**Lemma 3.1.** *The inequalities*

$$|(1+z)e^{-z}| \leqslant e^{|z|^2/2} \tag{3.1}$$

$$\left|(1+z)e^{-z} + \sum_{0\leqslant j\leqslant m} \frac{j-1}{j!}(-z)^j\right| \leqslant c_m|z|^{m+1}e^{|z|^2/2}, \tag{3.2}$$

*hold for all* $z \in \mathbb{C}$*, where* $m \geqslant 1$ *and*

$$c_m := \frac{1}{m!}\int_0^1 e^{t^2/2}(1-t)^{m-1}(m-1+t)dt. \tag{3.3}$$

*Proof.* Write $z = re^{it}$, where $r > 0$ and $t \in \mathbb{R}$. Then, by $1+x \leqslant e^x$ for $x \in \mathbb{R}$,

$$\begin{aligned}|(1+z)e^{-z}| &= \sqrt{1+2r\cos t + r^2}\, e^{-r\cos t}\\ &\leqslant e^{r\cos t + r^2/2 - r\cos t}\\ &= e^{r^2/2}.\end{aligned}$$

For (3.2), we start with the relation

$$e^z - \sum_{j<m}\frac{z^j}{j!} = \frac{z^m}{(m-1)!}\int_0^1 e^{tz}(1-t)^{m-1}dt,$$

and deduce that

$$(1-z)e^z + \sum_{0\leqslant j\leqslant m}\frac{j-1}{j!}z^j = -\frac{z^{m+1}}{m!}\int_0^1 e^{tz}(1-t)^{m-1}(m-1+t)dt,$$

for $m \geqslant 1$. Thus (3.2) follows from the inequality $|tz| \leqslant |z|^2/2 + t^2/2$. □

*Remark* 3.2. Note that in the proof of (3.2), we have the inequality

$$\frac{1+(x-1)e^x}{x^2e^{x^2/2}} \leqslant c_1 = \sqrt{e}-1 = 0.64872\ldots \qquad (x \in \mathbb{R}),$$

which can easily be sharpened, by elementary calculus, to

$$\frac{1+(x-1)e^x}{x^2e^{x^2/2}} \leqslant 0.63236\ldots.$$

But this improvement over $c_1$ is marginal, so we retain the simpler upper bound $c_1$ in the following use.

The next lemma is crucial in applying our Charlier-Parseval bounds derived above.

**Lemma 3.3.** *The inequality*

$$\left|\prod_{1\leqslant k\leqslant n}(1+v_k)e^{-v_k}-1\right|\leqslant c_1V_2e^{V_2/2}, \tag{3.4}$$

*holds for any complex numbers* $\{v_k\}$, *where*

$$V_m := \sum_{1\leqslant k\leqslant n}|v_k|^m. \tag{3.5}$$

*Proof.* By partial summation

$$\prod_{1\leqslant k\leqslant n}\xi_k-\prod_{1\leqslant k\leqslant n}\eta_k=\sum_{1\leqslant k\leqslant n}(\xi_k-\eta_k)\left(\prod_{1\leqslant j<k}\xi_j\right)\left(\prod_{k<j\leqslant n}\eta_j\right), \tag{3.6}$$

for nonzero $\{\xi_k\}$ and $\{\eta_k\}$. Applying this formula, we get

$$\prod_{1\leqslant k\leqslant n}(1+v_k)e^{-v_k}-1=\sum_{1\leqslant k\leqslant n}\big((1+v_k)e^{-v_k}-1\big)\prod_{1\leqslant j<k}(1+v_j)e^{-v_j}.$$

By the two inequalities (3.1) and (3.2) with $m=1$, we then obtain

$$\left|\prod_{1\leqslant k\leqslant n}(1+v_k)e^{-v_k}-1\right|\leqslant c_1\sum_{1\leqslant k\leqslant n}|v_k|^2\prod_{1\leqslant j<k}e^{|v_j|^2/2},$$

and (3.4) follows. □

## 3.2 New results

We are ready to apply in this section the tools we developed above to derive bounds for several Poisson approximation distances.

Let

$$S_n := X_1+X_2+\cdots+X_n,$$

where the $X_j$'s are independent Bernoulli random variables with

$$\mathbb{P}(X_j=1)=1-\mathbb{P}(X_j=0)=p_j \qquad (1\leqslant j\leqslant n).$$

Then, *here and throughout this section*,

$$F(z) := \sum_{0\leqslant m\leqslant n}\mathbb{P}(S_n=m)z^m=\prod_{1\leqslant j\leqslant n}(q_j+p_jz), \tag{3.7}$$

where $q_j := 1-p_j$. Define $\lambda_m := \sum_{1\leqslant j\leqslant n}p_j^m$, $\lambda=\lambda_1$ and $\theta := \lambda_2/\lambda_1$.

Let $\mathscr{P}(\lambda)$ denote a Poisson distribution with mean $\lambda$.

**Theorem 3.4.** *We have the following estimates:* $(i)$ *for the* $\chi^2$*-distance*

$$d_{\chi^2}(\mathscr{L}(S_n),\mathscr{P}(\lambda)) := \sum_{m\geqslant 0}\left|\frac{\mathbb{P}(S_n=m)}{e^{-\lambda}\frac{\lambda^m}{m!}}-1\right|^2 e^{-\lambda}\frac{\lambda^m}{m!}\leqslant\frac{2c_1^2\theta^2}{(1-\theta)^3};$$

$(ii)$ *for the total variation distance*

$$d_{TV}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) := \frac{1}{2}\sum_{m\geqslant 0}\left|\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!}\right| \leqslant \frac{c_1\theta}{\sqrt{2}(1-\theta)^{3/2}};$$

*and* $(iii)$ *for the Wasserstein (or Fortet-Mourier) distance*

$$d_W(\mathscr{L}(S_n), \mathscr{P}(\lambda)) := \sum_{m\geqslant 0}\left|\mathbb{P}(S_n \leqslant m) - \sum_{j\leqslant m} e^{-\lambda}\frac{\lambda^j}{j!}\right| \leqslant \frac{c_1\lambda_2}{\sqrt{\lambda}(1-\theta)}.$$

*We also have the following non-uniform bounds for* $m \geqslant 0$*:* $(iv)$ *for the Kolmogorov distance*

$$\left|\mathbb{P}(S_n \leqslant m) - \sum_{j\leqslant m} e^{-\lambda}\frac{\lambda^j}{j!}\right| \leqslant \frac{\sqrt{2}c_1\theta}{(1-\theta)^{3/2}}\sqrt{Z(m)};$$

*and* $(v)$ *for the point metric*

$$\left|\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!}\right| \leqslant \frac{\sqrt{6}c_1\theta}{(1-\theta)^2\sqrt{\lambda}}\sqrt{Z(m)}.$$

*Proof.* For $(i)$, we apply (2.10) to the function $F(z) - e^{\lambda(z-1)}$ and use the inequality (3.4) with $v_j = p_j r e^{it}$ to estimate the integral $I$. This yields

$$\begin{aligned} I(r) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\prod_{1\leqslant j\leqslant n}(1+p_j r e^{it})e^{-p_j r e^{it}} - 1\right|^2 dt \\ &\leqslant c_1^2\lambda_2^2 r^4 e^{\lambda_2 r^2} \end{aligned} \tag{3.8}$$

hence

$$\begin{aligned} \int_0^\infty I(\sqrt{r/\lambda})e^{-r}\,dr &\leqslant c_1^2\theta^2\int_0^\infty r^2 e^{-r(1-\theta)}\,dr \\ &= \frac{2c_1^2\theta^2}{(1-\theta)^3}, \end{aligned}$$

and the estimate in $(i)$ for the $\chi^2$-distance follows.

Similarly, the inequalities in $(ii)$ and in $(iv)$ follow from substituting the estimate (3.8) into the two inequalities (2.19) and (2.22) respectively.

As to the non-uniform estimate in $(v)$ for the point metric, we have, again, by (3.8),

$$\begin{aligned} \int_0^\infty I(\sqrt{r/\lambda})re^{-r}\,dr &\leqslant c_1^2\theta^2\int_0^\infty r^3 e^{-r(1-\theta)}\,dr \\ &\leqslant \frac{6c_1^2\theta^2}{(1-\theta)^4}. \end{aligned}$$

Substituting this estimate in (2.20) gives the inequality in $(v)$.

Finally, the upper bound in $(iii)$ for $d_W$ is derived similarly by the inequality (2.21) using again (3.8)

$$\int_0^\infty r^{-1}e^{-r}I(\sqrt{r/\lambda})\,dr \leqslant \frac{c_1^2\theta^2}{(1-\theta)^2}.$$

This completes the proof of the theorem. ∎

The reason of studying the $\chi^2$-distance (also referred to as the quadratic divergence) is at least twofold in addition to its applications in real problems. First, it is structurally simpler than most other distances because it satisfies the following identity.

**Corollary 3.5.** *Let $\{a_j\}$ be given by*

$$F(z) - e^{\lambda(z-1)} = e^{\lambda(z-1)} \sum_{j\geqslant 2} a_j (z-1)^j, \tag{3.9}$$

*where $F$ is given in (3.7). Then the $\chi^2$-distance satisfies the identity*

$$d_{\chi^2}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) = \sum_{j\geqslant 2} \frac{j!}{\lambda^j} |a_j|^2. \tag{3.10}$$

*Proof.* By (3.9), we have

$$\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!} = e^{-\lambda}\frac{\lambda^m}{m!} \sum_{j\geqslant 2} a_j C_j(\lambda, m). \tag{3.11}$$

Then (3.10) follows from (2.12). □

Second, the $\chi^2$-distance is often used to provide bounds for other distances; see [14]. An example is as follows.

**Corollary 3.6.** *The information divergence (or the Kullback-Leibner divergence) satisfies*

$$d_{KL}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) := \sum_{m\geqslant 0} \mathbb{P}(S_n = m) \log\left(\frac{\mathbb{P}(S_n = m)}{e^{-\lambda}\frac{\lambda^m}{m!}}\right) \leqslant \frac{2c_1^2\theta^2}{(1-\theta)^3}. \tag{3.12}$$

*Proof.* Given two sequences of non-negative real numbers $x_j$ and $y_j$ such that

$$x_0 + x_1 + \cdots = 1 \quad \text{and} \quad y_0 + y_1 + \cdots = 1.$$

By the elementary inequality $\log x \leqslant x - 1$, we obtain

$$\sum_{n\geqslant 0} y_n \log\frac{y_n}{x_n} \leqslant \sum_{n\geqslant 0} y_n \left(\frac{y_n}{x_n} - 1\right) = \sum_{n\geqslant 0} \frac{y_n^2}{x_n} - 1 = \sum_{n\geqslant 0} x_n \left(\frac{y_n}{x_n} - 1\right)^2.$$

Thus $d_{KL} \leqslant d_{\chi^2}$. Now (3.12) follows from applying this inequality with $x_m = e^{-\lambda}\lambda^m/m!$ and $y_m = \mathbb{P}(S_n = m)$ and then using the inequality in $(i)$ of Theorem 3.4. □

Since $Z(m) \leqslant 1/2$, from the two non-uniform estimates $(iv)$ and $(v)$ of Theorem 3.4, we easily obtain that the Kolmogorov distance satisfies

$$d_K(\mathscr{L}(S_n), \mathscr{P}(\lambda)) := \sup_m \left|\mathbb{P}(S_n \leqslant m) - e^{-\lambda}\sum_{0\leqslant j\leqslant m}\frac{\lambda^j}{j!}\right| \leqslant \frac{c_1\theta}{(1-\theta)^{3/2}};$$

and the point metric is bounded above by

$$d_P(\mathscr{L}(S_n), \mathscr{P}(\lambda)) := \sup_m \left|\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!}\right| \leqslant \frac{\sqrt{3}c_1\theta}{\sqrt{\lambda}\,(1-\theta)^2}.$$

Note that the estimate so obtained for the Kolmogorov distance is worse than that obtained by the simple relation $d_K \leqslant d_{TV}$ and the estimate $(ii)$ of Theorem 3.4.

The quantity $Z(m)$ can be readily bounded above by the following estimate; see also [9, p. 259] or [44].

**Lemma 3.7.**

$$Z(m) \leqslant e^{-(m-\lambda)^2/(2(m+\lambda))}.$$

*Proof.* Let $r = m/\lambda$. If $m \geqslant \lambda$, then

$$Z(m) \leqslant \mathbb{P}(\zeta_\lambda \geqslant m) \leqslant r^{-m} e^{\lambda(r-1)} = e^{-\lambda\psi(m/\lambda)},$$

where $\psi(x) := 1 - x + x\log x$. We now prove that

$$\psi(x) \geqslant \frac{(1-x)^2}{2(1+x)} \qquad (x > 0), \tag{3.13}$$

or, equivalently,

$$\int_0^x \log(1+t)\mathrm{d}t \geqslant \frac{x^2}{2(2+x)} \qquad (x > -1).$$

To prove (3.13), observe first that $\log(1+t) \geqslant t/(1+t)$ for $t > -1$ since $\int_0^t \log(1+v)\mathrm{d}v \geqslant 0$. Then

$$\int_0^x \log(1+t)\mathrm{d}t \geqslant \int_0^x \frac{t}{1+t}\,\mathrm{d}t,$$

which is bounded below by $x^2/(2(2+x))$ by considering the two cases $x \geqslant 0$ and $x \in (-1, 0]$. Thus, by (3.13),

$$Z(m) \leqslant e^{-(m-\lambda)^2/(2(m+\lambda))}.$$

Similarly, if $m \leqslant \lambda$, then $r < 1$, and

$$Z(m) \leqslant P(\xi_\lambda \leqslant m) \leqslant r^{-m} e^{\lambda(r-1)} = e^{-\lambda\psi(m/\lambda)} \leqslant e^{-(m-\lambda)^2/(2(m+\lambda))}.$$

$\square$

# 4 Applications. II. Second-order estimates

We show in this section that the same approach we developed above can be readily extended for obtaining higher order estimates. For simplicity, we consider only the second-order estimates for which we need only to refine Lemma 3.3. From the formal expansion (3.11), we expect that

$$\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!} \approx a_2 e^{-\lambda}\frac{\lambda^m}{m!} C_2(\lambda, m) + \text{smaller order terms},$$

where $a_2 = -\lambda_2/2$, and the error terms for Poisson approximation would be smaller if we take the term $a_2 e^{-\lambda}\lambda^m C_2(\lambda, m)/m!$ into account.

**Lemma 4.1.** *For any complex numbers $\{v_k\}$, the following inequality holds*

$$\left| \prod_{1\leqslant k\leqslant n} (1+v_k)e^{-v_k} - 1 + \frac{1}{2}\sum_{1\leqslant k\leqslant n} v_k^2 \right| \leqslant \left(\frac{c_1}{4}V_2^2 + c_2 V_3\right) e^{V_2/2}, \tag{4.1}$$

*where $V_m$ is defined in (3.5), $c_1 = \sqrt{e} - 1$ and (see (3.3))*

$$c_2 = \frac{1}{2}\int_0^1 e^{t^2/2}(1-t^2)dt \approx 0.3706.$$

*Proof.* By (3.6),

$$\prod_{1\leqslant k\leqslant n}(1+v_k)e^{-v_k}-1+\frac{1}{2}\sum_{1\leqslant k\leqslant n}v_k^2=\sum_{1\leqslant k\leqslant n}\left((1+v_k)e^{-v_k}-1+\frac{v_k^2}{2}\right)\prod_{1\leqslant j<k}(1+v_j)e^{-v_j}$$
$$-\frac{1}{2}\sum_{1\leqslant k\leqslant n}v_k^2\left(\prod_{1\leqslant j<k}(1+v_k)e^{-v_k}-1\right).$$

By (3.1), (3.2) with $m=2$ and (3.4), we then obtain

$$\left|\prod_{1\leqslant k\leqslant n}(1+v_k)e^{-v_k}-1+\frac{1}{2}\sum_{1\leqslant k\leqslant n}v_k^2\right|\leqslant c_2\sum_{1\leqslant k\leqslant n}|v_k|^3\exp\left(\frac{1}{2}\sum_{1\leqslant j\leqslant k}|v_j|^2\right)$$
$$+\frac{c_1}{2}\sum_{1\leqslant k\leqslant n}|v_k|^2\sum_{j<k}|v_j|^2\exp\left(\frac{1}{2}\sum_{1\leqslant j<k}|v_j|^2\right),$$

and (4.1) follows. □

For simplicity, let

$$P_1(z):=e^{\lambda(z-1)}\left(1-\frac{\lambda_2}{2}(z-1)^2\right).$$

Then

$$[z^m]P_1(z)=e^{-\lambda}\frac{\lambda^m}{m!}\left(1-\frac{\lambda_2}{2}C_2(m,\lambda)\right), \tag{4.2}$$
$$[z^m]\frac{P_1(z)}{1-z}=\sum_{j\leqslant m}e^{-\lambda}\frac{\lambda^j}{j!}+\frac{\lambda_2}{2}C_1(m,\lambda)e^{-\lambda}\frac{\lambda^m}{m!},$$

where $C_1, C_2$ are given in (2.3).

With the inequality (4.1) and Proposition 2.4, we can now refine Theorem 3.4 as follows.

**Theorem 4.2.** *For $\theta<1$, we have the following second-order estimates for $\chi^2$-, total variation and Wasserstein distances, respectively,*

$$\sum_{m\geqslant 0}\frac{\left(\mathbb{P}(S_n=m)-[z^m]P_1(z)\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}}\leqslant\left(\frac{\sqrt{3}\,c_1\theta^2}{\sqrt{2}(1-\theta)^{5/2}}+\frac{\sqrt{6}\,c_2\lambda_3}{\lambda^{3/2}(1-\theta)^2}\right)^2,$$
$$\frac{1}{2}\sum_{m\geqslant 0}\left|\mathbb{P}(S_n=m)-[z^m]P_1(z)\right|\leqslant\frac{\sqrt{3}\,c_1\theta^2}{2\sqrt{2}(1-\theta)^{5/2}}+\frac{\sqrt{3}\,c_2\lambda_3}{\sqrt{2}\lambda^{3/2}(1-\theta)^2},$$
$$\sum_{m\geqslant 0}\left|\mathbb{P}(S_n\leqslant m)-[z^m]\frac{P_1(z)}{1-z}\right|\leqslant\sqrt{\lambda}\left(\frac{\sqrt{3}c_1\theta^2}{2\sqrt{2}(1-\theta)^2}+\frac{\sqrt{2}\,c_2\lambda_3}{\lambda^{3/2}(1-\theta)^{3/2}}\right);$$

*and the second-order non-uniform estimates for Kolmogorov distance and point metric, respectively,*

$$\left|\mathbb{P}(S_n\leqslant m)-[z^m]\frac{P_1(z)}{1-z}\right|\leqslant\sqrt{Z(m)}\left(\frac{\sqrt{3}\,c_1\theta^2}{\sqrt{2}(1-\theta)^{5/2}}+\frac{\sqrt{6}\,c_2\lambda_3}{\lambda^{3/2}(1-\theta)^2}\right),$$

$$\left|\mathbb{P}(S_n=m)-[z^m]P_1(z)\right|\leqslant\sqrt{\frac{Z(m)}{\lambda}}\left(\frac{\sqrt{15}\,c_1\theta^2}{\sqrt{2}(1-\theta)^3}+\frac{2\sqrt{6}\,c_2\lambda_3}{\lambda^{3/2}(1-\theta)^{5/2}}\right).$$

*Proof.* Let

$$F(z)=\prod_{1\leqslant j\leqslant n}(1+p_j(z-1))-e^{\lambda(z-1)}\left(1-\frac{\lambda_2}{2}(z-1)^2\right).$$

Take $v_j=p_j(z-1)$ in inequality (4.1). Then

$$\left|\prod_{1\leqslant j\leqslant n}(1+p_j(z-1))e^{-p_j(z-1)}-1-\frac{\lambda_2}{2}(z-1)^2\right|\leqslant\left(\frac{c_1}{4}\lambda_2^2|z-1|^4+c_2\lambda_3|z-1|^3\right)e^{\frac{\lambda_2}{2}|z-1|^2}.$$

It follows that

$$I(r)\leqslant\left(\frac{c_1}{4}\lambda_2^2r^4+c_2\lambda_3r^3\right)^2e^{\lambda_2r^2}. \tag{4.3}$$

Substituting this upper bound into the identity (2.10) and using the relation (4.2), we obtain

$$\begin{aligned}
&\left(\sum_{m\geqslant 0}\frac{(\mathbb{P}(S_n=m)-[z^m]P_1(z))^2}{e^{-\lambda}\frac{\lambda^m}{m!}}\right)^{1/2}\\
&\quad\leqslant\left(\int_0^\infty\left(\frac{c_1}{4}\theta^2r^2+\frac{c_2\lambda_3}{\lambda^{3/2}}r^{3/2}\right)^2e^{-(1-\theta)r}\,dr\right)^{1/2}\\
&\quad\leqslant\frac{c_1}{4}\theta^2\left(\int_0^\infty r^4e^{-(1-\theta)r}\,dr\right)^{1/2}+\frac{c_2\lambda_3}{\lambda^{3/2}}\left(\int_0^\infty r^3e^{-(1-\theta)r}\,dr\right)^{1/2}\\
&\quad=\frac{c_1}{4}\theta^2\cdot\frac{\sqrt{24}}{(1-\theta)^{5/2}}+\frac{c_2\lambda_3}{\lambda^{3/2}}\cdot\frac{\sqrt{6}}{(1-\theta)^2},
\end{aligned}$$

where we used the Minkowsky inequality. This proves the second-order estimate for the $\chi^2$-distance.

Similarly, the corresponding estimates for the total variation distance and the (non-uniform estimate of the) Kolmogorov distance follow from (4.3) and the two inequalities (2.19) and (2.22), respectively.

For the point metric, we have, using again (4.3) and the inequality (2.20),

$$\begin{aligned}
&\sqrt{\frac{\lambda}{Z(m)}}\,|\mathbb{P}(S_n=m)-[z^m]P_1(z)|\\
&\quad\leqslant\left(\int_0^\infty\left(\frac{c_1}{4}\theta^2r^2+\frac{c_2\lambda_3}{\lambda^{3/2}}r^{3/2}\right)^2re^{-(1-\theta)r}\,dr\right)^{1/2}\\
&\quad\leqslant\frac{c_1}{4}\theta^2\left(\int_0^\infty r^5e^{-(1-\theta)r}\,dr\right)^{1/2}+\frac{c_2\lambda_3}{\lambda^{3/2}}\left(\int_0^\infty r^4e^{-(1-\theta)r}\,dr\right)^{1/2}\\
&\quad=\frac{\sqrt{15}\,c_1\theta^2}{\sqrt{2}(1-\theta)^3}+\frac{2\sqrt{6}\,c_2\lambda_3}{\lambda^{3/2}(1-\theta)^{5/2}}.
\end{aligned}$$

Finally, the second-order estimate for the Wasserstein distance follows from (4.3) and the inequality (2.21)

$$\begin{aligned}
&\lambda^{-1/2}\sum_{m\geqslant 0}\left|\mathbb{P}(S_n\leqslant m)-[z^m]\frac{P_1(z)}{1-z}\right|\\
&\quad\leqslant\left(\int_0^\infty\left(\frac{c_1}{4}\theta^2r^2+\frac{c_2\lambda_3}{\lambda^{3/2}}r^{3/2}\right)^2r^{-1}e^{-(1-\theta)r}\,dr\right)^{1/2}\\
&\quad\leqslant\frac{c_1}{4}\theta^2\left(\int_0^\infty r^3e^{-(1-\theta)r}\,dr\right)^{1/2}+\frac{c_2\lambda_3}{\lambda^{3/2}}\left(\int_0^\infty r^2e^{-(1-\theta)r}\,dr\right)^{1/2}.
\end{aligned}$$

□

**Corollary 4.3.** *The total variation distance between the distribution of $S_n$ and a Poisson distribution of mean $\lambda$ satisfies, for $\theta < 1$,*

$$d_{TV}(S_n, \mathcal{P}(\lambda)) \leqslant \frac{\theta}{2^{3/2}} + \frac{\sqrt{3}\, c_1 \theta^2}{2\sqrt{2}(1-\theta)^{5/2}} + \frac{\sqrt{3}\, c_2 \lambda_3}{\sqrt{2}\lambda^{3/2}(1-\theta)^2}. \tag{4.4}$$

*Proof.* By (2.13) with $k = 2$, we have

$$\frac{1}{2}\sum_{m\geqslant 0} e^{-\lambda}\frac{\lambda^m}{m!}|C_2(\lambda, m)| \leqslant \frac{1}{\sqrt{2}\lambda},$$

and (4.4) follows from the second-order estimate for the total variation distance in Theorem 4.2. □

*Remark* 4.4. One can easily derive, by the difference equation (2.2) of Charlier polynomials with $k = 1$, that (see for example [43])

$$\frac{1}{2}\sum_{m\geqslant 0} e^{-\lambda}\frac{\lambda^m}{m!}|C_2(\lambda, m)| = e^{-\lambda}\left(\frac{\lambda^{m_+-1}}{m_+!}(m_+ - \lambda) + \frac{\lambda^{m_--1}}{m_-!}(\lambda - m_-)\right),$$

where $m_\pm := \lfloor \lambda + \frac{1}{2} \pm \sqrt{\lambda + \frac{1}{4}} \rfloor$. Asymptotically, for large $\lambda$,

$$\frac{1}{2}\sum_{m\geqslant 0} e^{-\lambda}\frac{\lambda^m}{m!}|C_2(\lambda, m)| = \frac{\sqrt{2}}{\sqrt{\pi e}\,\lambda}\left(1 + O\left(\lambda^{-1}\right)\right).$$

By a detailed calculus, Roos [70] showed that

$$\frac{1}{2}\sum_{m\geqslant 0} e^{-\lambda}\frac{\lambda^m}{m!}|C_2(\lambda, m)| \leqslant \frac{3}{2e\lambda}, \tag{4.5}$$

where numerically

$$\left\{\frac{1}{\sqrt{2}}, \frac{3}{2e}, \frac{\sqrt{2}}{\sqrt{\pi e}}\right\} \approx \{0.707, 0.552, 0.484\}.$$

Of course, we can apply Roos's inequality (4.5) and replace the constant $1/2^{3/2} \approx 0.354\ldots$ by $3/(4e) \approx 0.276\ldots$ in the first term of our inequality (4.4).

**Corollary 4.5.** *The $\chi^2$-distance satisfies*

$$d_{\chi^2}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) = \frac{\theta^2}{2}\left(1 + O\left(\frac{\theta}{(1-\theta)^5}\right)\right). \tag{4.6}$$

*Proof.* Note that

$$0 \leqslant \sum_{m\geqslant 0} \frac{\left(\mathbb{P}(S_n = m) - [z^m]P_1(z)\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}} = \sum_{m\geqslant 0} \frac{\left(\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!}\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}} - \frac{\theta^2}{2}.$$

This identity together with the first estimate of Theorem 4.2 and an observation that $\lambda_3 \leqslant \lambda_2^{3/2}$ yields (4.6). □

*Remark* 4.6. An alternative way to prove (4.6) is to use the identity (3.10) and apply the estimate for the coefficients $a_j$ derived in Shorgin [80]

$$|a_j| \leqslant \left(\frac{e\lambda_2}{j}\right)^{j/2} \qquad (j \geqslant 2), \tag{4.7}$$

and obtain

$$\sum_{j\geqslant 3} \frac{j!}{\lambda^j}|a_j|^2 \leqslant \sum_{j\geqslant 3} j!(e/j)^j\theta^j = O\left(\sum_{j\geqslant 3} j^{1/2}\theta^j\right),$$

by Stirling's formula $j! = O(j^{1/2}(j/e)^j)$, $j \geqslant 1$. This and $a_2 = -\lambda_2/2$ give

$$d_{\chi^2}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) = \frac{\theta^2}{2}\left(1 + O\left(\frac{\theta}{(1-\theta)^{3/2}}\right)\right). \tag{4.8}$$

For a further refinement of (4.6), see Corollary 5.3. Note that (4.8) implies that

$$d_{KL}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) \leqslant \frac{\theta^2}{2}\left(1 + O\left(\frac{\theta}{(1-\theta)^{3/2}}\right)\right).$$

# 5 Applications. III. Approximations by signed measures

Since the probability generating function of $S_n$ can be represented as

$$\mathbb{E}z^{S_n} = \exp\left(\sum_{j\geqslant 1} \frac{(-1)^{j-1}}{j}\lambda_j(z-1)^j\right),$$

it is well-known since Herrmann [39] that smaller error terms can be achieved if we use finite number of terms in the exponent to approximate $\mathbb{E}z^{S_n}$; namely,

$$\mathbb{E}z^{S_n} \approx \exp\left(\sum_{1\leqslant j\leqslant k} \frac{(-1)^{j-1}}{j}\lambda_j(z-1)^j\right),$$

for $k \geqslant 1$. Anther advantage of such approximations is that the remainder terms tend to zero not only when $\theta \to 0$ but also when $\lambda \to \infty$ (while $\theta$ remaining, say less than $1-\varepsilon$, $\varepsilon > 0$ being a small number). This gives rise to Poisson approximation via signed measures (sometimes also referred to as compound Poisson approximations); see Cekanavicius [18], Roos [71], Barbour et al. [5] for more information.

Although these approximations are not probability generating functions for $k \geqslant 2$, they can numerically and asymptotically be readily computed. Indeed, for $k = 2$

$$[z^m]e^{\lambda(z-1)-\lambda_2(z-1)^2/2} = e^{-\lambda-\lambda_2/2}\frac{\lambda_2^{m/2}}{m!}H_m\left(\frac{\lambda+\lambda_2}{\sqrt{\lambda_2}}\right),$$

where the $H_m(x)$'s are the Hermite polynomials.

## 5.1 Approximation by $e^{\lambda(z-1)-\lambda_2(z-1)^2/2}$

We consider the simplest case of such forms when $k = 2$.

**Lemma 5.1.** *The inequality*

$$\left|\prod_{1\leqslant k\leqslant n}(1+v_k)e^{-v_k} - \exp\left(-\frac{1}{2}\sum_{1\leqslant k\leqslant n} v_k^2\right)\right| \leqslant \left(c_2V_3 + \frac{1}{8}V_4\right)e^{V_2/2} \tag{5.1}$$

*holds for any complex numbers $\{v_k\}$, where $V_m$ is given in (3.5) and $c_2$ in (3.3).*

*Proof.* Again by (3.6),

$$\begin{aligned}
&\prod_{1\leqslant k\leqslant n}(1+v_k)e^{-v_k} - \prod_{1\leqslant k\leqslant n} e^{-v_k^2/2}\\
&\quad = \sum_{1\leqslant k\leqslant n}\left((1+v_k)e^{-v_k} - e^{-v_k^2/2}\right)\left(\prod_{1\leqslant j<k}(1+v_j)e^{-v_j}\right)\left(\prod_{k<j\leqslant n} e^{-v_j^2/2}\right).
\end{aligned}$$

Now

$$\begin{aligned}
\left|(1+z)e^{-z} - e^{-z^2/2}\right| &= \left|(1+z)e^{-z} - 1 + \frac{z^2}{2} - \left(e^{-z^2/2} - 1 + \frac{z^2}{2}\right)\right|\\
&= \left|-\frac{z^3}{2}\int_0^1 (1-t^2)e^{-tz}dt - \frac{z^4}{4}\int_0^1(1-t)e^{-tz^2/2}dt\right|\\
&\leqslant c_2|z|^3 e^{|z|^2/2} + \frac{|z|^4}{8}e^{|z|^2/2}.
\end{aligned}$$

This and the inequality (3.1) yield (5.1). $\square$

Let

$$P_2(z) := e^{\lambda(z-1)-\lambda_2(z-1)^2/2}.$$

**Theorem 5.2.** *Assume that $\theta < 1$. Then*

$$\begin{aligned}
\sum_{m\geqslant 0}\frac{(\mathbb{P}(S_n = m) - [z^m]P_2(z))^2}{e^{-\lambda}\frac{\lambda^m}{m!}} &\leqslant \frac{\lambda_3^2}{\lambda^3}\left(\frac{\sqrt{6}\,c_2}{(1-\theta)^2} + \frac{\sqrt{3\theta}}{2\sqrt{2}(1-\theta)^{5/2}}\right)^2,\\
\sum_{m\geqslant 0}|\mathbb{P}(S_n = m) - [z^m]P_2(z)| &\leqslant \frac{\lambda_3}{\lambda^{3/2}}\left(\frac{\sqrt{6}\,c_2}{(1-\theta)^2} + \frac{\sqrt{3\theta}}{2\sqrt{2}(1-\theta)^{5/2}}\right),\\
\sum_{m\geqslant 0}\left|\mathbb{P}(S_n \leqslant m) - [z^m]\frac{P_2(z)}{1-z}\right| &\leqslant \frac{\lambda_3}{\lambda}\left(\frac{\sqrt{2}\,c_2}{(1-\theta)^{3/2}} + \frac{\sqrt{3\theta}}{4\sqrt{2}(1-\theta)^2}\right),\\
\left|\mathbb{P}(S_n \leqslant m) - [z^m]\frac{P_2(z)}{1-z}\right| &\leqslant \frac{\lambda_3}{\lambda^{3/2}}\sqrt{Z(m)}\left(\frac{\sqrt{6}\,c_2}{(1-\theta)^2} + \frac{\sqrt{3\theta}}{2\sqrt{2}(1-\theta)^{5/2}}\right),\\
|\mathbb{P}(S_n = m) - [z^m]P_2(z)| &\leqslant \frac{\lambda_3}{\lambda^2}\sqrt{Z(m)}\left(\frac{2\sqrt{6}\,c_2}{(1-\theta)^{5/2}} + \frac{\sqrt{15\theta}}{2\sqrt{2}(1-\theta)^3}\right).
\end{aligned}$$

*Proof.* All estimates follow similarly as the proof of Theorem 3.4 but with

$$F(z) = \prod_{1\leqslant j\leqslant n} (1 + p_j(z-1)) - e^{\lambda(z-1)-\lambda_2(z-1)^2/2}.$$

For the first two estimates of the theorem, we apply the inequality (5.1), which gives

$$I(r) \leqslant \left(c_2\lambda_3 r^3 + \frac{1}{8}\lambda_4 r^4\right)^2 e^{\lambda_2 r^2}.$$

By the inequality $\lambda_4 \leqslant \lambda_3\sqrt{\lambda_2}$, we obtain

$$\begin{aligned}\left(\sum_{m\geqslant 0} \frac{\left(\mathbb{P}(S_n = m) - [z^m]P_2(z)\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}}\right)^{1/2} &\leqslant \frac{\lambda_3}{\lambda^{3/2}}\left(\int_0^\infty \left(c_2 r^{3/2} + \frac{\sqrt{\theta}}{8} r^2\right)^2 e^{-(1-\theta)r^2}\right)^{1/2} \\ &\leqslant \frac{\lambda_3}{\lambda^{3/2}}\left(\frac{c_2\sqrt{6}}{(1-\theta)^2} + \frac{\sqrt{24\theta}}{8(1-\theta)^{5/2}}\right).\end{aligned}$$

Then we apply Proposition 2.4. The other estimates are similarly proved. ☐

**Lemma 5.3.** *For any $\theta < 1$, we have*

$$\sum_{m\geqslant 0} \frac{\left(e^{-\lambda}\frac{\lambda^m}{m!} - [z^m]P_2(z)\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}} = \frac{1}{\sqrt{1-\theta^2}} - 1. \tag{5.2}$$

*Proof.* Applying (2.10) and (2.12) to the function

$$F(z) = e^{\lambda(z-1)} - P_2(z) = e^{\lambda(z-1)}\left(\sum_{k\geqslant 1}\left(\frac{\lambda_2}{2}\right)^k \frac{(z-1)^{2k}}{k!}\right),$$

we obtain

$$\sum_{m\geqslant 0} \frac{\left(e^{-\lambda}\frac{\lambda^m}{m!} - [z^m]P_2(z)\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}} = \sum_{k\geqslant 1}\left(\frac{\theta}{2}\right)^{2k}\frac{(2k)!}{(k!)^2} = \frac{1}{\sqrt{1-\theta^2}} - 1.$$

☐

**Corollary 5.4.** *For $\theta < 1$,*

$$\left|\left(\sum_{m\geqslant 0} \frac{\left(\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!}\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}}\right)^{1/2} - \left(\frac{1}{\sqrt{1-\theta^2}} - 1\right)^{1/2}\right| \leqslant \frac{\lambda_3}{\lambda^{3/2}}\left(\frac{c_2\sqrt{6}}{(1-\theta)^2} + \frac{\sqrt{24\theta}}{8(1-\theta)^{5/2}}\right). \tag{5.3}$$

*Proof.* By applying the Minkowsky inequality and the first estimate of Theorem 5.2, we obtain

$$\begin{aligned}&\left|\left(\sum_{m\geqslant 0} \frac{\left(\mathbb{P}(S_n = m) - e^{-\lambda}\frac{\lambda^m}{m!}\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}}\right)^{1/2} - \left(\sum_{m\geqslant 0} \frac{\left(e^{-\lambda}\frac{\lambda^m}{m!} - [z^m]P_2(z)\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}}\right)^{1/2}\right| \\ &\leqslant \left(\sum_{m\geqslant 0} \frac{\left(\mathbb{P}(S_n = m) - [z^m]P_2(z)\right)^2}{e^{-\lambda}\frac{\lambda^m}{m!}}\right)^{1/2} \\ &\leqslant \frac{\lambda_3}{\lambda^{3/2}}\left(\frac{c_2\sqrt{6}}{(1-\theta)^2} + \frac{\sqrt{24\theta}}{8(1-\theta)^{5/2}}\right).\end{aligned}$$

Consequently, by (5.2), we obtain (5.3). ☐

Note that (5.3) implies that, for all $\theta < 1$,

$$d_{\chi^2}(\mathscr{L}(S_n), \mathscr{P}(\lambda)) = \left(\frac{1}{\sqrt{1-\theta^2}} - 1\right)\left(1 + O\left(\frac{\lambda_3}{\lambda_2\sqrt{\lambda}(1-\theta)^5}\right)\right).$$

On the other hand, by the inequality $d_{\chi^2} \geqslant 4d_{TV}^2$ (which following from (2.12) and (2.19)), we obtain another upper bound for $d_{TV}$.

**Corollary 5.5.** *For $\theta < 1$,*

$$d_{TV}(S_n, \mathcal{P}(\lambda)) \leqslant \frac{1}{2}\left(\frac{1}{\sqrt{1-\theta^2}} - 1\right)^{1/2} + \frac{\lambda_3}{\lambda^{3/2}}\left(\frac{c_2\sqrt{6}}{2(1-\theta)^2} + \frac{\sqrt{24\theta}}{16(1-\theta)^{5/2}}\right).$$

# 6 Comparative discussions

We review briefly some known results in the literature and compare them in this section. For simplicity, we write $d_*$ for $d_*(\mathscr{L}(S_n), \mathscr{P}(\lambda))$ throughout this section, where $d_*$ represents one of the distances we discuss.

Among the five measures of closeness of Poisson approximation $\{d_{\chi^2}, d_{TV}, d_W, d_K, d_P\}$, the estimation of the three $\{d_{\chi^2}, d_K, d_P\}$ is generally simpler in complexity since they can all be easily bounded above by explicit summation or integral representations: see (3.10) for $d_{\chi^2}$, (6.2) for $d_K$ and (6.3) for $d_P$.

In addition to the Poisson approximations to $\mathscr{L}(S_n)$ we consider in this paper, many other different types of approximations to $\mathscr{L}(S_n)$ were proposed in the literature; these include Poisson with different mean, compound Poisson, translated Poisson, large deviations, other perturbations of Poisson, binomial, compound binomial, etc. They are too numerous to be listed and compared here; see, for example, Barbour et al. [9], Roos [69, 72], Barbour and Chryssaphinou [7], Barbour and Chen [6], Röllin [66] and the references therein.

## 6.1 The $\chi^2$-distance and the Kullback-Leibner divergence

Borisov and Vorozheĭkin [14] showed that $d_{\chi^2} \sim \theta^2/2$ under the assumption that $\theta = o(\lambda^{-1/7})$. They also derived in the same paper the identity (3.10) in the special case when all $p_j$'s are equal. More refined estimates were then given. The estimate (4.6) we obtained is more general and stronger.

The Kullback-Leibner divergence has been widely studied in the information-theoretic literature and many results are known. The connection between $d_{TV}$ and $d_{KL}$ for general distributions also received much attention since they can be used to bridge results in probability theory and in information theory; see the survey paper Fedotov et al. [34] for more information and references. One such tool studied is Pinsker's inequality $d_{TV} \leqslant \sqrt{d_{KL}/2}$ (see [34]). Note that in the case of $S_n$, this inequality implies that $d_{TV} \leqslant \sqrt{d_{\chi^2}/2}$, while we have $d_{TV} \leqslant \sqrt{d_{\chi^2}}/2$ by (2.12) and (2.19).

Kontoyiannis et al. [51] recently proved, by an information-theoretic approach, that

$$d_{KL} \leqslant \frac{1}{\lambda}\sum_{1\leqslant j\leqslant n}\frac{p_j^3}{1-p_j}.$$

The right-hand side in the above inequality is, by Cauchy-Schwarz inequality, always larger than $\theta^2$, provided that at least one of the $p_j$'s is nonzero, and can be considerably larger than our estimate (3.12) for certain cases. Indeed, take for example $p_j = 1/\sqrt{j+1}$. Then

$$d_{KL} \leqslant \frac{1}{\lambda} \sum_{1 \leqslant j \leqslant n} \frac{p_j^3}{1-p_j} \asymp \frac{1}{\sqrt{n}},$$

where the symbol "$a_n \asymp b_n$" means that $a_n$ is asymptotically of the same order as $b_n$. Our result (3.12) yields in this case the estimate

$$d_{KL} \leqslant \frac{2c_1^2\theta^2}{(1-\theta)^3} \asymp \frac{\log^2 n}{n}.$$

## 6.2 The total variation distance

We mentioned in Introduction some results in Le Cam [54] and other refinements in the literature of the form $d_{TV} \leqslant c\theta$. We briefly review and compare here other results for $d_{TV}$.

**First- and second-order estimates.** Kerstan [49], in addition to proving that $d_{TV} \leqslant 0.6\theta$ (which was later on corrected to $1.05$ by Barbour and Hall [8]), he also proved the second-order estimate

$$\sum_{j \geqslant 0} \left| \mathbb{P}(S_n = j) - e^{-\lambda} \frac{\lambda^j}{j!} \left(1 - \frac{\lambda_2}{2} C_2(\lambda, j)\right) \right| \leqslant 1.3 \frac{\lambda_3}{\lambda} + 3.9\theta^2.$$

Similar estimates were derived later in Herrmann [39], Chen [23], Barbour and Hall [8]. The order of the error terms is however not optimal for large $\lambda$; see Theorem 4.2.

Many fine estimates were obtained in the series of papers by Deheuvels, Pfeifer and their co-authors. In particular, Deheuvels and Pfeifer [30] proved $d_{TV} \leqslant \theta/(1-\sqrt{2\theta})$ for $\theta < 1/2$ and the second-order estimate

$$\sum_{j \geqslant 0} \left| \mathbb{P}(S_n = j) - e^{-\lambda} \frac{\lambda^j}{j!} \left(1 - \frac{\lambda_2}{2} C_2(\lambda, j)\right) \right| \leqslant \frac{(2\theta)^{3/2}}{1-\sqrt{2\theta}},$$

for $\theta < 1/2$, the order of the error terms being tight. For many other estimates (including higher-order ones), see [30, 31]. Their approach is based on a semi-group formulation, followed by applying the fine estimates of Shorgin [80], which in turn were obtained by the complex-analytic approach of Uspensky [83]. Following a similar approach, Witte [86] gives an upper bound of the form

$$d_{TV} \leqslant \frac{e^{2p_*}\theta}{\sqrt{2\pi}(1-2e^{2p_*}\theta)},$$

for $\theta < \frac{1}{2}e^{-2p_*}$, as well as other more complicated ones. Another very different form for $d_{TV}$ can be found in Weba [85], which results from combining several known estimates.

By refining further Deheuvels and Pfeifer's approach, Roos [69, 70] deduced several precise estimates for $d_{TV}$ and other distances. In particular, he showed that

$$d_{TV} \leqslant \left(\frac{3}{4e} + \frac{7(3-2\sqrt{\theta})}{6(1-\sqrt{\theta})^2}\sqrt{\theta}\right)\theta,$$

when $\theta < 1$; see [70] and the references therein. The proof of this estimate is based on a second-order approximation; see (4.5).

Note that since $d_{TV} \leqslant 1$, any result of the form $d_{TV} \leqslant \varphi(\theta)\theta$ for $\theta \leqslant \theta_1$, $\theta_1 \in (0,1)$, also leads to an upper bound of the form $d_{TV} \leqslant c\theta$, where

$$c = \sup_{0 \leqslant t \leqslant \theta_0} \varphi(t),$$

$\theta_0 := \min\{\theta_1, \theta_2\}$, $\theta_2 \in (0,1)$ solving the equation $t\varphi(t) = 1$.

Higher-order approximations based on Charlier expansion are studied in Herrmann [39], Barbour [3], Deheuvels and Pfeifer [30], Barbour et al. [9], Roos [69, 71].

**Approximations by signed measures.** Herrmann [39] proved that, when specializing to the case of $S_n$,

$$\sum_{m \geqslant 0} \left| \mathbb{P}(S_n = m) - [z^m] e^{\lambda(z-1) - \lambda_2 (z-1)^2/2} \right| = O\left(\frac{\lambda_3}{\lambda}\right),$$

the rate being $\lambda^{1/2}$ away from optimal; see Theorem 5.2. Presman [64] considered the binomial case and derived an optimal error bound. Kruopis [53] extended further Presman's analysis and derived

$$\begin{aligned} &\sum_{m \geqslant 0} \left| \mathbb{P}(S_n = m) - [z^m] e^{\lambda(z-1) - \lambda_2 (z-1)^2/2} \right| \\ &\qquad \leqslant 10 \varpi \lambda_3 \min\left\{1.2\sigma^{-3} + 4.2\lambda_2\sigma^{-6}, 2 + \sigma^2 + 3.4\lambda_2\right\}, \end{aligned}$$

where $\sigma := \sqrt{\lambda - \lambda_2}$ and

$$\varpi := \max_{1 \leqslant j \leqslant n} \sup_{0 \leqslant t \leqslant 1} e^{2p_j t(1-p_j t)}, \tag{6.1}$$

which was in turn refined by Borovkov [15]. Hipp [41] discussed similar expansions for compound Poisson distributions and attributed the idea to Kornya [52], but his bounds are weaker for large $\lambda$ in the special case of $S_n$; see also Čekanavičius [18]. Barbour and Xia [11] proved, as a special case of their general results, that

$$\sum_{m \geqslant 0} \left| \mathbb{P}(S_n = m) - [z^m] e^{\lambda(z-1) - \lambda_2 (z-1)^2/2} \right| \leqslant \frac{4\lambda_3}{\lambda^{3/2}(1-2\theta)\sqrt{1 - \theta - \max_j p_j(1-p_j)/\lambda}},$$

when $\theta < 1/2$. An extensive study was carried out by Čekanavičius in a series of papers dealing mainly with Kolmogorov's problem of approximating convolutions by infinitely divisible distributions; see Čekanavičius [18, 19] and the references cited there. Approximation results using signed compound measures under more general settings than $S_n$ are derived in Borovkov and Pfeifer [16], Roos [71, 72] and Čekanavičius [19], Barbour et al. [5].

**Other uniform asymptotic approximations.** The estimate $d_{TV} \sim \theta/\sqrt{2\pi e}$ holds whenever $\theta \to 0$. A uniform estimate of the form

$$d_{TV} = \theta J(\theta)\left(1 + O\left(\lambda^{-1}\right)\right),$$

as $\lambda \to \infty$, was recently derived in [44], where

$$J(\theta) := \frac{1}{\theta}\left(\Phi\left(\sqrt{\frac{1}{\theta}\log\frac{1}{1-\theta}}\right) - \Phi\left(\sqrt{\frac{1-\theta}{\theta}\log\frac{1}{1-\theta}}\right)\right),$$

$\Phi$ being the standard normal distribution function. Other more general and more uniform approximations were also derived in [44].

## 6.3 The Wasserstein distance

Deheuvels and Pfeifer [30] proved the asymptotic equivalent $d_W \sim \lambda_2/\sqrt{2\pi\lambda}$, when $\lambda_2/\sqrt{\lambda} \to \infty$, improving earlier results in Deheuvels and Pfeifer [29]. They also obtained many other estimates, including the following second-order one

$$\left| d_W - \lambda_2 e^{-\lambda} \frac{\lambda^{\lceil \lambda \rceil}}{\lceil \lambda \rceil !} \right| \leqslant \frac{2^{5/2}\lambda^{1/2}\theta^{3/2}}{1-\sqrt{2\theta}},$$

for $|\theta| \leqslant 1/2$. Then Witte [86] gave the bound

$$d_W \leqslant -\frac{\sqrt{e\lambda}}{2\sqrt{2\pi}} \log\left(1 - 2e^{2p_*}\theta\right),$$

for $\theta < \frac{1}{2}e^{-2p_*}$. Xia [87] showed that $d_W \leqslant \lambda_2/\sqrt{\lambda(1-\theta)}$; see also Barbour and Xia [12] for the estimate $d_W \leqslant 8\lambda_2/(3\sqrt{2e\lambda})$. The strongest results including more precise higher-order approximations were derived by Roos (1999, 2001), where, in particular,

$$d_W \leqslant \left( \frac{1}{\sqrt{2e}} + \frac{8(2-\theta)}{5(1-\sqrt{\theta})^2}\sqrt{\theta} \right) \frac{\lambda_2}{\sqrt{\lambda}}.$$

For other results in connection with Wasserstein metrics, see Deheuvels et al. [27], Hwang [43], Čekanavičius and Kruopis [20].

## 6.4 The Kolmogorov distance

It is known, by definition and Newton's inequality (see Comtet [24, p. 270] or Pitman [60]), that $d_K \leqslant d_{TV} \leqslant 2d_K$; see Daley and Vere-Jones [26], Ehm [33], Roos [70]. Thus all upper estimates for $d_{TV}$ translate directly to those for $d_K$ and vice versa. Also many approximation results in probability theory for sums of independent random variables apply to $S_n$. Both types of results are not listed and discussed here; see for example Arak and Zaĭtsev [2].

Up to now, we only consider non-uniform bounds for $d_K$. However, effective uniform bounds can be easily derived based on the Fourier inversion formula

$$\begin{aligned} d_K &= \sup_m \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-imt} \frac{\mathbb{E}e^{itS_n} - e^{\lambda(e^{it}-1)}}{1-e^{it}} dt \right| \\ &\leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{\lambda(\cos t - 1)}}{|1-e^{it}|} \left| \prod_{1 \leqslant j \leqslant n} \left(1 + p_j(e^{it}-1)\right) e^{-p_j(e^{it}-1)} - 1 \right| dt. \end{aligned} \tag{6.2}$$

From (6.2) and (3.4), we have

$$d_K \leqslant \frac{c_1}{\pi}\lambda_2 \int_0^{\pi} \left|1 - e^{it}\right| e^{-\sigma^2(1-\cos t)} dt,$$

which, by the simple inequalities $|1 - e^{it}| \leqslant |t|$ and $1 - \cos t \geqslant 2t^2/\pi^2$ for $t \in [-\pi, \pi]$, leads to

$$d_K \leqslant \frac{c_1}{\pi}\lambda_2 \int_0^{\infty} t e^{-2\sigma^2 t^2/\pi^2} dt = \frac{c_1 \pi \theta}{4(1-\theta)},$$

where $c_1\pi/4 \approx 0.51$. Although this bound is worse than some known ones such as $d_K \leqslant 0.36\theta$ in Daley and Vere-Jones [26], its derivation is very simple and self-contained, the order being also tight. Furthermore, the leading constant $c_1\pi/4$ can be lowered, say to $0.363c_1 < 0.24$, by a more careful analysis but we are not pursuing this further here. Note that it is known that $d_K \sim \theta/(2\sqrt{2\pi e})$, as $\theta = o(1)$, see Deheuvels and Pfeifer [30], Hwang [43], where $1/(2\sqrt{2\pi e}) \approx 0.121$.

In a little known paper, Makabe [55] gives a systematic study of $d_K$ using standard Fourier analysis, improving earlier results by Kolmogorov [50], Le Cam [54], Hodges and Le Cam [42]. In particular, he first derived a second-order estimate from which he deduced that $d_K \leqslant 3.7\theta$ and

$$d_K \leqslant \frac{\theta}{2} + O\left(\theta^2 + p_*\theta\right).$$

For $p_* < 1/5$, he also provided a one-page proof of

$$d_K \leqslant \frac{5\theta}{4(1-2p_*-5\theta/2)} \leqslant \frac{25\theta}{12-50\theta}.$$

A Le Cam-type inequality of the form $d_K \leqslant 2\lambda_2/\pi$ was given in Franken [35], which was later refined to $d_K \leqslant \lambda_2/2$ in Serfling [76]; see also Daley [25]. Franken [35] also proves the estimate

$$d_K \leqslant \frac{c}{\pi}\left(1-e^{-\lambda(1-\theta)}\right)\frac{\theta}{1-\theta},$$

for an explicitly given $c$, as well as higher-order terms for $d_K$ based on Charlier expansions. His bound together with $d_K \leqslant 1$ implies $d_K \leqslant 1.9\theta$, improving previous estimates by Le Cam and Makabe.

Shorgin [80] derived an asymptotic expansion for the distribution of $S_n$; in particular, as a simple application of his bounds for $|a_j|$ (see (3.9)) and $|C_k(\lambda, m)|$,

$$d_K \leqslant \left(\frac{1}{2} + \sqrt{\frac{\pi}{8}}\right)\frac{\theta}{1-\sqrt{\theta}},$$

where $1/2 + \sqrt{\pi/8} \approx 1.31$. In Hipp [40], the upper bound

$$d_K \leqslant \frac{\pi}{4\lambda(1-\theta)} \sum_{1\leqslant j\leqslant n} \frac{p_j^2}{1-p_j},$$

was given, so that if $p_* \leqslant 1/4$, then

$$d_K \leqslant \frac{\pi\theta}{3(1-\theta)} \leqslant \frac{1.05\theta}{1-\theta}.$$

A bound of the form

$$d_K \leqslant \frac{2}{\pi}\min\left\{\frac{\sqrt{e}\theta}{2(1-\theta)}, \lambda_2\right\}$$

was given in Kruopis [53], where he also derived

$$\sup_m \left|\mathbb{P}(S_n \leqslant m) - [z^m]\frac{P_2(z)}{1-z}\right| \leqslant \frac{2}{3}\varpi\lambda_3 \min\left\{\frac{1}{\sqrt{\pi}\lambda^{3/2}(1-\theta)^{3/2}}, 1\right\},$$

where $\varpi$ is defined in (6.1). Deheuvels and Pfeifer deduced several estimates for $d_K$; in particular (see [30, 31])

$$\sup_m \left|\mathbb{P}(S_n \leqslant m) - [z^m]\frac{P_1(z)}{1-z}\right| \leqslant \frac{5}{3}\left(\frac{\theta^2}{(1-\sqrt{\theta})} + \frac{\lambda_3}{\lambda^{3/2}}\right);$$

Note that this can also be written as

$$\left|d_K - \frac{\theta}{2}\, e^{-\lambda} \max\left\{\frac{\lambda^{\ell_+}}{\ell_+!}(\ell_+ - \lambda), \frac{\lambda^{\ell_-}}{\ell_-!}(\lambda - \ell_-)\right\}\right| \leqslant \frac{5}{3}\left(\frac{\theta^2}{(1-\sqrt{\theta})} + \frac{\lambda_3}{\lambda^{3/2}}\right),$$

where $\ell_\pm := \lfloor \lambda + 1/2 \pm \sqrt{\lambda + 1/4}\rfloor$.

Witte [86] then derived the estimate

$$d_K \leqslant \frac{\sqrt{e}(1+\sqrt{\pi/2})e^{2p_*}}{2\sqrt{2\pi}(1-e^{2p_*}\theta)}\,\theta,$$

for $\theta < e^{-p_*}$; see also Weba [85]. Roos [69, 70] gives, among several other fine estimates,

$$d_K \leqslant \left(\frac{1}{2e} + \frac{6}{5(1-\sqrt{\theta})}\sqrt{\theta}\right)\theta.$$

Non-uniform estimates are derived in Teerapabolarn and Neammanee [82] for general dependent summands, which is of the form in the case of $S_n$

$$\left|\mathbb{P}(S_n \leqslant m) - e^{-\lambda}\sum_{0\leqslant j\leqslant m}\frac{\lambda^j}{j!}\right| \leqslant \left(1-e^{-\lambda}\right)\theta \min\left\{1, \frac{e^\lambda}{m+1}\right\},$$

generally weaker than our bounds in Theorems 3.4 and 4.2.

## 6.5 The point probabilities

As for $d_K$ above, the point metric can also be readily estimated by using the integral representation

$$d_P \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\lambda(\cos t-1)}\left|\prod_{1\leqslant j\leqslant n}\left(1+p_j(e^{it}-1)\right)e^{-p_j(e^{it}-1)} - 1\right| dt, \tag{6.3}$$

and (3.4), and we obtain for example

$$d_P \leqslant \frac{c_1\pi^{5/2}\theta}{8\sqrt{2\lambda}(1-\theta)^{3/2}}.$$

Classical local limit theorems for probabilities of moderate or large deviations can also be used to give effective bounds for the point metric $d_P := \max_m |\mathbb{P}(S_n = m) - e^{-\lambda}\lambda^m/m!|$; they are not discussed here.

Results for $d_P$ were derived in Franken [35] but are too complicated to be described here. Kruopis [53] gives the estimate

$$d_P \leqslant \min\left\{\frac{\sqrt{e}\theta}{\sqrt{\pi\lambda}(1-\theta)^{3/2}}, \lambda_2\right\},$$

as well as

$$\sup_m |\mathbb{P}(S_n = m) - [z^m]P_2(z)| \leqslant \frac{8\varpi}{3\pi}\lambda_3 \min\left\{\frac{1}{\lambda^2(1-\theta)^2}, \frac{4}{3}\right\}.$$

Barbour and Jensen [10] derived an asymptotic expansion; see also [3].

Asymptotically, as $\theta \to 0$,

$$d_P \sim \frac{\theta}{2\sqrt{2\pi\lambda}},$$

see Roos [68], where he also derived a second-order estimate for $d_P$, which was later refined in [69, 70]. In particular,

$$d_P \leqslant \left( \frac{1}{2} \left( \frac{3}{2e} \right)^{3/2} + \frac{6 - 4\sqrt{\theta}}{3(1-\sqrt{\theta})^2} \sqrt{\theta} \right) \frac{\theta}{\sqrt{\lambda}}.$$

A non-uniform bound was given in Neammanee [57, 58] of the form

$$\left| \mathbb{P}(S_n = m) - e^{-\lambda} \frac{\lambda^m}{m!} \right| \leqslant \min \left\{ m^{-1}, \lambda^{-1} \right\} \lambda_2,$$

whenever $\lambda \leqslant 1$.

# References


[1] Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer-Verlag, New York.

[2] Arak, T. V. and Zaĭtsev, A. Yu. (1988). *Uniform Limit Theorems for Sums of Independent Random Variables*. Proc. Steklov Inst. Math. 1988, no. 1 (174).

[3] Barbour, A. D (1987). Asymptotic expansions in the Poisson limit theorem. *Ann. Probab.* **15** 748–766.

[4] Barbour, A. D (2001). Topics in Poisson approximation. Handbook of Statist., 19, Stochastic Processes: Theory and Methods (D. N. Shanbhag and C. R. Rao, eds.), pp. 79–115, North-Holland, Amsterdam.

[5] Barbour, A. D., Čekanavičius, V. and Xia, A. (2007). On Stein's method and perturbations. *ALEA Lat. Am. J. Probab. Math. Stat.* **3** 31–53.

[6] Barbour, A. D and Chen, L. H.-Y. (2005). *Stein's Method and Applications*. Singapore University Press and World Scientific Publishing Co., Singapore.

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

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

[9] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford Science Publications, Clarendon Press, Oxford.

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

[11] Barbour, A. D. and Xia, A. (1999). Poisson perturbations. *ESAIM Probab. Statist.* **3** 131–150.

[12] Barbour, A. D. and Xia, A. (2006). On Stein's factors for Poisson approximation in Wasserstein distance. *Bernoulli* **12** 943–954.

[13] BOAS, R. P., JR. (1949). Representation of probability distributions by Charlier series. *Ann. Math. Statist.* **20** 376–392.

[14] BORISOV, I. S. and VOROZHEĬKIN, I. S. (2008). Accuracy of approximation in the Poisson theorem in terms of the $\chi^2$-distance. *Sib. Math. J.* **49** 5–17.

[15] BOROVKOV, K. A. (1989). On the problem of improving Poisson approximation. *Theory Probab. Appl.* **33** 343–347 (1989).

[16] BOROVKOV, K. AND PFEIFER, D. (1996). On improvements of the order of approximation in the Poisson limit theorem. *J. Appl. Probab.* **33** 146–155.

[17] BORTKIEWICZ, L. VON (1898). *Das Gesetz der kleinen Zahlen*. B. G. Teubner, Leipzig.

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

[19] ČEKANAVIČIUS, V. (2004). On local estimates and the Stein method. *Bernoulli* **10** 665–683.

[20] ČEKANAVIČIUS, V. AND KRUOPIS, J. (2000). Signed Poisson approximation: a possible alternative to normal and Poisson laws. *Bernoulli* **6** 591–606.

[21] CHARLIER, C. V. L. (1905a). Die zweite Form des Fehlergesetzes. *Arkiv for Matematik, Astronomi och Fysik*, 2 (No. 15) 1–8.

[22] CHATTERJEE, S., DIACONIS, P. AND MECKES, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2** 64–106.

[23] CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

[24] COMTET, L. (1974). Advanced combinatorics. Revised and enlarged edition. D. Reidel Publishing Co., Dordrecht.

[25] DALEY, D. J. (1980). A note on bounds for the supremum metric for discrete random variables. *Math. Nachr.* **99** 95–98.

[26] DALEY, D. J. AND VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II. General Theory and Structure*. Second edition. Springer, New York.

[27] DEHEUVELS, P., KARR, A., PFEIFER, D. AND SERFLING, R. (1988). Poisson approximations in selected metrics by coupling and semigroup methods with applications. *J. Statist. Plann. Inference* **20** 1–22.

[28] DEHEUVELS, P. AND PFEIFER, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* **14** 663–676.

[29] DEHEUVELS, P. AND PFEIFER, D. (1986). Operator semigroups and Poisson convergence in selected metrics. *Semigroup Forum* **34** 203–224. (Errata: *Semigroup Forum* **35** 251).

[30] DEHEUVELS, P. AND PFEIFER, D. (1988). On a relationship between Uspensky's theorem and Poisson approximation. *Ann. Inst. Statist. Math.* **40** 671–681.

[31] DEHEUVELS, P., PFEIFER, D. AND PURI, M. L. (1989). A new semigroup technique in Poisson approximation. *Semigroup Forum* **38** 189–201.

[32] DE MOIVRE, A. (1756). *The Doctrine of Chances*. Third Edition, W. Pearson, London; available at `www.ibiblio.org/chance`.

[33] EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16.

[34] FEDOTOV, A. A., HARREMOËS, P. AND TOPSØE, F. (2003). Refinements of Pinsker's inequality. *IEEE Trans. Inform. Theory* **49** 1491–1498.

[35] FRANKEN, P. (1964). Approximation des Verteilungen von Summen unabhängiger nichtnegativer ganzzahler Zufallsgrössen durch Poissonsche verteilungen. *Math. Nachr.* **23** 303–340.

[36] GOLDSTEIN, L. AND REINERT, G. (2005). Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18** 237–260.

[37] GOOD, I. J. (1986). Some statistical applications of Poisson's work. (With comments by Persi Diaconis and Eduardo Engel, Herbert Solomon, C. C. Heyde, and Nozer D. Singpurwalla, and with a reply by the author.) *Statist. Sci.* **1** 157–180.

[38] HAIGHT, F. A. *Handbook of the Poisson Distribution*. John Wiley & Sons, Inc., New York-London-Sydney.

[39] HERRMANN, H. (1965). Variationsabstand zwischen der Verteilung einer Summe unabhängiger nichtnegativer ganzzahliger Zufallsgrössen und Poissonschen Verteilungen. *Math. Nachr.* **29** 265–289.

[40] HIPP, C. (1985). Approximation of aggregate claims distributions by compound Poisson distributions. *Insurance Math. Econom.* **4** 227–232. (correction note: **6** 165, 1987).

[41] HIPP, C. (1986). Improved approximations for the aggregate claims distribution in the individual model. *ASTIN Bull.* **16** 89–100.

[42] HODGES, J. L., JR. AND LE CAM, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist. 31* 737–740.

[43] HWANG, H.-K. (1999). Asymptotics of Poisson approximation to random discrete distributions: an analytic approach. *Adv. in Appl. Probab.* **31** 448–491.

[44] HWANG, H.-K. and ZACHAROVAS, V. (2008). Uniform asymptotics of Poisson approximation to the Poisson-binomial distribution. Manuscript submitted for publication.

[45] JACOB, M. (1933). Sullo sviluppo di una curva di frequenze in serie di Charlier tipo B. *Giornale dell'Istituto Italiano degli Attuari. Istituto Italiano degli Attuari, Roma* **4** 221–234.

[46] JANSON, S. (1994). Coupling and Poisson approximation. *Acta Appl. Math.* **34** 7–15.

[47] JORDAN, C. (1926). Sur la probabilité des épreuves répétées, le théorème de Bernoulli et son inversion. *Bull. Soc. Math. France* **54** 101–137.

[48] KENNEDY, J. E. AND QUINE, M. P. (1989). The total variation distance between the binomial and Poisson distributions. *Ann. Probab.* **17** 396–400.

[49] KERSTAN, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam. *Z. Wahrsch. Verw. Gebiete* **2** 173–179.

[50] KOLMOGOROV, A. N. (1956). Two uniform limit theorems for sums of independent random variables. *Theory Probab. Appl.* **1** 384–394.

[51] KONTOYIANNIS, I., HARREMÖES, P. and JOHNSON, O. (2005). Entropy and the law of small numbers. *IEEE Trans. Inform. Theory* **51** 466–472.

[52] KORNYA, P. S. (1983). Distribution of aggregate claims in the individual risk theory model. *Trans. Soc. Actuaries* **35** 823–858.

[53] KRUOPIS, Y. (1986). The accuracy of approximation of the generalized binomial distribution by convolutions of Poisson measures. *Lith. Math. J.* **26** 37–49.

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

[55] MAKABE, H. (1962). On the approximations to some limiting distributions with some applications. *Kōdai Math. Sem. Rep.* **14** 123–133.

[56] MATSUNAWA, T. (1982). Uniform $\varphi$-equivalence of probability distributions based on information and related measures of discrepancy. *Ann. Inst. Statist. Math.* **34** 1–17.

[57] NEAMMANEE, K. (2003). A nonuniform bound for the approximation of Poisson binomial by Poisson distribution. *Int. J. Math. Math. Sci.* no. 48, 3041–3046.

[58] NEAMMANEE, K. (2003). Pointwise approximation of Poisson distribution. *Stoch. Model. Appl.* **6** 20–26.

[59] PFEIFER, D. (1985). A semigroup setting for distance measures in connexion with Poisson approximation. *Semigroup Forum* **31** 201–205.

[60] PITMAN, J. (1997). Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Combin. Theory Ser. A* **77** 279–303.

[61] POISSON, S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile: précedés des règles générales du calcul des probabilités*, Bachelier, Paris.

[62] POLLACZEK-GEIRINGER, H. (1928). Die Charlier'sche Entwicklung willkürlicher Verteilungen. *Skandinavisk Aktuarietidskrift*, **11** 98–111.

[63] POOR, H. V. (1991). The maximum difference between the binomial and Poisson distributions. *Statist. Probab. Lett.* **11** 103–106.

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

[65] PROHOROV, Y. V. (1953). Asymptotic behavior of the binomial distribution. *Uspekhi Matematicheskikh Nauk* **8** 135–142. Also in *Selected Translations in Mathematical Statistics and Probability*, volume 1, 87–95.

[66] RÖLLIN, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614.

[67] ROMANOWSKA, M. (1977). A note on the upper bound for the distance in total variation between the binomial and the Poisson distribution. *Statist. Neerlandica* **31** 127–130.

[68] ROOS, B. (1995). A semigroup approach to Poisson approximation with respect to the point metric. *Statist. Probab. Lett.* **24** 305–314.

[69] ROOS, B. (1999). Asymptotics and sharp bounds in the Poisson approximation to the Poisson-binomial distribution. *Bernoulli* **5** 1021-1034.

[70] ROOS, B. (2001). Sharp constants in the Poisson approximation *Statist. Probab. Lett.* **52** 155-168.

[71] ROOS, B. (2004). Poisson approximation via the convolution with Kornya-Presman signed measures. *Theory Probab. Appl.* **48** 555–560.

[72] ROOS, B (2007). On variational bounds in the compound Poisson approximation of the individual risk model. *Insurance Math. Econom.* **40** 403–414.

[73] SCHMIDT, E. (1933). Über die Charlier-Jordansche Entwicklung einer willkürlichen Funktion nach der Poissonschen Funktion und ihren Ableitungen. *Z. f. angew. Math.* **13** 139–142.

[74] SENETA, E. Modern probabilistic concepts in the work of E. Abbe and A. De Moivre. *Math. Sci.* **8** 75–80.

[75] SERFLING, R. J. (1974). Probability inequalities for the sum in sampling without replacement. *Ann. Statist.* **2** 39–48.

[76] SERFLING, R. J. (1978). Some elementary results on Poisson approximation in a sequence of Bernoulli trials. *SIAM Rev.* **20** 567–579.

[77] SIEGMUND-SCHULTZE, R. (1993). Hilda Geiringer-von Mises, Charlier series, ideology, and the human side of the emancipation of applied mathematics at the University of Berlin during the 1920s. *Historia Math.* **20** 364–381.

[78] STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.

[79] SZEGÖ, G. (1939). *Orthogonal Polynomials.* American Mathematical Society, New York.

[80] SHORGIN, S. Y. (1977). Approximation of a generalized Binomial distribution. *Theory Probab. Appl.* **22** 846–850.

[81] STEELE, J. M. (1994). Le Cam's inequality and Poisson approximations. *Amer. Math. Monthly* **101** 48–54.

[82] TEERAPABOLARN, K. AND NEAMMANEE, K. (2006). Poisson approximation for sums of dependent Bernoulli random variables. *Acta Math. Acad. Paedagog. Nyhazi. (N.S.)* **22** 87–99.

[83] USPENSKY, J. V. (1931). On Ch. Jordan's series for probability. *Ann. Math.* **32** 306–312.

[84] VERVAAT, W. (1969). Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution. *Statist. Neerlandica* **23** 79–86.

[85] WEBA, M. (1999). Bounds for the total variation distance between the binomial and the poisson distribution in the case of medium-sized success probabilities. *J. Appl. Probab.* **36** 97–104.

[86] WITTE, H.-J. (1990). A unification of some approaches to Poisson approximation. *J. Appl. Probab.* **27** 611–621.

[87] XIA, A. (1997). On the rate of Poisson process approximation to a Bernoulli process. *J. Appl. Probab.* **34** 898–907.