Soft Computing (2020) 24:17517–17524
https://doi.org/10.1007/s00500-020-05190-2

FOCUS

Discrete uniform and binomial distributions with infinite support

Andrey Pepelyshev[1] · **Anatoly Zhigljavsky**[1]

Published online: 22 July 2020

Abstract

We study properties of two probability distributions defined on the infinite set $\{0, 1, 2, \ldots\}$ and generalizing the ordinary discrete uniform and binomial distributions. Both extensions use the grossone-model of infinity. The first of the two distributions we study is uniform and assigns masses $1/①$ to all points in the set $\{0, 1, \ldots, ① - 1\}$, where ① denotes the grossone. For this distribution, we study the problem of decomposing a random variable ξ with this distribution as a sum $\xi \stackrel{\mathrm{d}}{=} \xi_1 + \cdots + \xi_m$, where $\xi_1, \ldots, \xi_m$ are independent non-degenerate random variables. Then, we develop an approximation for the probability mass function of the binomial distribution $\mathrm{Bin}(①, p)$ with $p = c/①^{\alpha}$ with $1/2 < \alpha \leq 1$. The accuracy of this approximation is assessed using a numerical study.

Keywords Binomial distribution · Poisson approximation · Charlier polynomials

1 Introduction

In this paper, we are interested in properties of two probability distributions defined on the infinite set $\{0, 1, 2, \ldots\}$ and generalizing the ordinary discrete uniform and binomial distributions. Both of these extensions have been recently discussed in Calude and Dumitrescu (2020) and mentioned in Zhigljavsky (2012); both extensions use the notion of grossone. The grossone, introduced in Sergeyev (2013) and denoted by ①, is a model of infinity which, as shown in Sergeyev (2009), Sergeyev (2017) and many other publications can be very useful in solving diverse problems of computational mathematics and optimization; in such applications, ① is used as numerical infinity. Grossone can also be useful as a theoretical model of infinity, see, e.g., (Zhigljavsky 2012; Sergeyev 2017). Some historical, philosophical and logical aspects of grossone have been considered in Lolli (2012), Lolli (2015), Hansson (2020). In Sect. 1, we consider and briefly discuss postulates of ①.

For a positive integer n, the discrete uniform distribution on the set $\{0, 1, \ldots, n - 1\}$ assigns equal mass $1/n$ to all integers $j \in \{0, \ldots, n - 1\}$. We use the notation $\mathrm{DU}(n)$ from Balakrishnan and Nevzorov (2004) for this distribution. An extension of this distribution to the infinite set $\{0, 1, 2 \ldots\}$ is denoted by $\mathrm{DU}(\infty)$ and is used in Bayesian statistics as vague prior (often called 'Jeffrey's prior') for large integer-valued parameters, in particular for the parameter N in the binomial distribution Bin (N, p), see, e.g., (Raftery 1988). An extension of $\mathrm{DU}(n)$ to $\mathrm{DU}(①)$ is straightforward: for a random variable (r.v.) $\xi \sim \mathrm{DU}(①)$, we have $\Pr\{\xi = k\} = 1/(①)$ for all $k \in \{0, 1, \ldots, ① - 1\}$. This distribution has been considered, in particular, in Calude and Dumitrescu (2020). The distribution $\mathrm{DU}(①)$ is easier than $\mathrm{DU}(\infty)$: indeed, $\mathrm{DU}(\infty)$ is a vague (improper) distribution but, if one agrees to perform calculations with ①, $\mathrm{DU}(①)$ is a well-defined distribution, very similar to $\mathrm{DU}(n)$.

In Sect. 2.2, we consider the problem of decomposing a random variable $\xi \sim \mathrm{DU}(①)$ into sums $\xi \stackrel{\mathrm{d}}{=} \xi_1 + \ldots + \xi_m$, where $\xi_1, \ldots, \xi_m$ are independent non-degenerate random variables and the equality $\stackrel{\mathrm{d}}{=}$ means that the distributions of the random variables in the lhs and rhs of the equation are equal. In particular, we shall establish that $\mathrm{DU}(①)$ is not an infinitely divisible distribution which might have been expected in view of results of Warde and Katti (1971).

The probability mass function (pmf) for $\mathrm{Bin}(N, p)$, the binomial distribution with parameters N and p, is

$$b_x = \frac{N!}{(N-x)!\,x!} p^x (1-p)^{N-x}, \quad x = 0, 1, \ldots, N, \qquad (1)$$

Communicated by Yaroslav D. Sergeyev.

✉ Anatoly Zhigljavsky
ZhigljavskyAA@cardiff.ac.uk

Andrey Pepelyshev
pepelyshevan@cardiff.ac.uk

[1] Cardiff University, Cardiff, UK

where N is usually interpreted as the number of Bernoulli trials and p as the probability of success in these trials. We are interested in approximating the binomial probabilities (3) in the case when N is (very) large but p is rather small like $p = c/N^\alpha$ with finite $c > 0$ and $1/2 < \alpha \le 1$. This case is important for understanding the distribution $\mathrm{Bin}(①, p)$, the grossone extension of $\mathrm{Bin}(N, p)$.

According to the central limit theorem, for any p and large N, the distribution $[\mathrm{Bin}(N, p) - Np]/\sqrt{Np(1-p)}$ is approximately the standard normal distribution $\mathcal{N}(0, 1)$ and, therefore, the binomial distribution $\mathrm{Bin}(N, p)$ can be approximated by the normal distribution $\mathcal{N}(Np, Np(1-p))$. However, if p is small then, even for very large N, this normal approximation is very poor, especially in the tails, see for example (Berry 1941). Also, the support of the random variable with distribution $\mathcal{N}(Np, Np(1-p))$ barely resembles the support of $\mathrm{Bin}(N, p)$ and this could be a serious problem in practice. There are many improvements to the normal approximation, see, e.g., (Brown et al. 2001). However, even corrected normal approximations are rather poor in approximating tails; in particular, the approximations based on the Edgeworth expansion do not guarantee that the approximations to individual binomial probabilities are non-negative, see for example (Petrov 1995) for an excellent account of different approximations in the CLT. Even the shape of the normal approximation $\mathcal{N}(Np, Np(1-p))$ may be misleading. Consider, for instance, the skewness which is the widely accepted characteristic of a non-symmetry of distributions. The skewness of $\mathcal{N}(0, 1)$ is zero, whereas the skewness of $[\mathrm{Bin}(①,p)-p\ ①\ ]/\sqrt{p(1-p)①}$ is $\gamma_1 = (2p-1)/\sqrt{p(1-p)①}$. As an example, for $p = \lambda/①$ we have $\gamma_1 = 1/\sqrt{\lambda} + O(①^{-1})$ which shows that even if N is very large, the binomial distribution $\mathrm{Bin}(N, p)$ can still be very asymmetric for small p, even after the renormalization.

Bearing in mind that the normal approximation to $\mathrm{Bin}(N, p)$ cannot be suitably corrected if p is small, in Sect. 3 we will concentrate on correcting the Poisson approximation to $\mathrm{Bin}(N, p)$ assuming that N is very large but p is of order $p = c/N^\alpha$ with finite $c > 0$ and $1/2 < \alpha \le 1$.

One of the central concepts used below is the concept of grossone which has been introduced in Sergeyev (2013), developed in a series of papers by Ya. Sergeyev and coauthors and recently comprehensively reviewed in Sergeyev (2017). Grossone can be defined axiomatically, see (Sergeyev 2017). The two main axioms are given below.

Axiom 1 *(Grossone is 'the largest natural number'). The set of natural numbers is* $\mathbb{N} = \{1, 2, \ldots, ①\}$*, where* ① *is the grossone.*

Axiom 2 *(Divisibility) For any finite positive integer* n*,* ① *is divisible by* n*.*

The grossone models infinity. Similarly, the quantities like $1/①$ and $1/①^2$ model infinitesimals. These models, as comprehensively discussed in Sergeyev (2017), are very useful as theoretical models and as models of numerical infinity and infinitesimals. A very attractive feature of these numerical infinity and infinitesimals is a possibility to operate with them in numerical fashion, exactly as with numbers (rather than with symbols like in MAPLE); this feature is the key concept of the 'infinity computer' discussed in many publications of Ya. Sergeyev and coauthors, see (Sergeyev 2009, 2017).

In mathematics, a more common approach to model infinitesimal quantities is to use the framework of the non-standard analysis. The non-standard analysis approach for modeling infinitesimal probabilities has been recently discussed in Benci et al. (2018). Modeling infinitesimal probabilities with $1/①$ and similar quantities involving ① has also attracted serious attention, see (Calude and Dumitrescu 2020; Sergeyev 2017; Rizza 2018). One of an attractive features of the grossone-based approach is that one may simultaneously work with infinitesimal probabilities of different order like $1/①$ and $1/①^2$. It should be stressed that the grossone-based methodology is different from the approach based on the non-standard analysis, see (Sergeyev 2019).

2 Deconvolutions of discrete distributions

2.1 Deconvolution and infinite divisibility of a discrete r.v.

The concepts of deconvolution of a r.v. and its infinite divisibility are closely related.

A r.v. ξ can be deconvoluted if it can be represented as $\xi \overset{\mathrm{d}}{=} \xi_1 + \xi_2$, where ξ_1 and ξ_2 are independent but not necessarily identically distributed r.v. Let $\phi(t) = E\exp(it\xi)$ be the characteristic function (c.f.) of ξ. Clearly, ξ can be deconvoluted if and only if $\phi(t)$ can be written as a product of two or more characteristic functions of non-degenerate r.v. In Sect. 2.2, we consider the case where ξ is discrete uniform r.v.

Let now ξ be a discrete r.v. taking values $0, 1, \ldots$ with probabilities $p_i = \Pr(\xi = i)$, $i = 0, 1, \ldots$ (here we use the traditional language). This r.v. ξ is infinitely divisible if for any positive integer n there exist i.i.d.r.v. $\xi_{j,n}$ ($j = 1, \ldots, n$) such that $\xi \overset{\mathrm{d}}{=} \xi_{1,n} + \ldots + \xi_{n,n}$. Equivalently, a r.v. ξ is infinitely divisible if and only if $\phi(t) = E\exp(it\xi)$, the characteristic function (c.f.) of ξ satisfies the relation $\phi(t) = [\phi_n(t)]^n$ for any $n = 1, 2, \ldots$, where $\phi_n(t)$ are some characteristic functions.

It yields, in particular, that if a discrete r.v. ξ has a finite support then it cannot be infinitely divisible. In particular, discrete uniform $\mathrm{DU}(n)$ and binomial $\mathrm{Bin}(N, p)$ r.v. are not infinitely divisible. Note that c.f. $\phi_{\mathrm{DU}(n)}(t)$ of $\xi \sim \mathrm{DU}(n)$ and

$\phi_{\mathrm{Bin}(N,p)}(t)$ of $\xi \sim \mathrm{Bin}(N, p)$ are

$$\phi_{\mathrm{DU}(n)}(t)=\frac{1-e^{nit}}{n(1-e^{i}t)}, \quad \phi_{\mathrm{Bin}(N,p)}(t)=\left(1-p+pe^{it}\right)^{N}, \tag{2}$$

see formulas (2.22) and (5.5) in Balakrishnan and Nevzorov (2004).

A very general sufficient condition for the infinite divisibility of a discrete r.v. has been established in Warde and Katti (1971):

Theorem 1 *see Theorem 2.1 in Warde and Katti (1971). If $p_0 \neq 0$, $p_1 \neq 0$ and the ratios p_{j+1}/p_j form a monotonously non-decreasing sequence ($j = 0, 1, \ldots$), then the r.v. ξ is infinitely divisible.*

One of our aims in this paper is to generalize the concept of infinite divisibility to the case when the vague ∞ is replaced with rigid ① and check if Theorem 1 can still be applied. In Sect. 2.2 below, we will consider DU(①) where we show that formal extension of Theorem 1 to the random variables defined on $\{0, 1, \ldots, ①-1\}$ or $\{0, 1, \ldots, ①\}$ fails. On the other hand, as shown below in this section, the formal extension of Theorem 1 to Bin(①, p) cannot be applied, but the distribution Bin(①, p) is infinitely divisible.

Consider a discrete r.v. ξ taking values either in $\{0, 1, \ldots, ① - 1\}$ or in $\{0, 1, \ldots, ①\}$. In both cases, $\phi(t) = E\exp(it\xi)$, the c.f. of ξ, is well-defined and we can still classify a r.v. ξ as infinitely divisible if $\phi(t) = [\phi_n(t)]^n$ for any finite integer $n = 1, 2, \ldots$, where $\phi_n(t)$ are some characteristic functions.

Consider $\xi \sim \mathrm{Bin}(①, p)$ defined on $\{0, 1, \ldots, ①\}$ with

$$p_j = \frac{①!}{(①-j)!\,j!}\,p^j\,(1-p)^{①-j}, \quad j = 0, 1, \ldots, ①. \tag{3}$$

For the ratios p_{j+1}/p_j, we have

$$\frac{p_{j+1}}{p_j} = \frac{①-j}{j+1}\,\frac{p}{1-p} = \frac{①\,p}{(1-p)(j+1)} - \frac{jp}{(1-p)(j+1)}. \tag{4}$$

As ① is much larger than j for finite j, for such j we can neglect the second term in (4) and we can clearly see that, at least for finite j, the ratios p_{j+1}/p_j are decreasing with j. The formal extension of Theorem 1 is thus not applicable but $\xi \sim \mathrm{Bin}(①, p)$ is clearly infinitely divisible if we assume Axiom 2. This is a direct consequence of (2) with $N = ①$.

2.2 Deconvolution of a discrete uniform distribution DU(①)

Decomposition of the discrete uniform random variable $\xi \sim \mathrm{DU}(n)$ into a sum $\xi \overset{d}{=} \xi_1 + \ldots + \xi_m$, where $\xi_1, \ldots, \xi_m$ are independent non-degenerate random variables, is considered in Zhigljavsky et al. (2016). It is shown in Zhigljavsky et al. (2016) that such decompositions exist if and only if n is a composite number and that the number of different decompositions $\xi \overset{d}{=} \xi_1 + \ldots + \xi_m$ is equal to the number of all ordered factorizations of n. The results of Zhigljavsky et al. (2016) have been recently extended in Gillard and Zhigljavsky (2016), Golyandina and Zhigljavsky (2020) to cover the case of deconvolution of positive definite matrices.

Theorem 1 of Zhigljavsky et al. (2016) states that if $\xi \sim \mathrm{DU}(n)$ can be represented as $\xi \overset{d}{=} \xi_1 + \xi_2$, where ξ_1 and ξ_2 are independent non-degenerate random variables, then both ξ_1 and ξ_2 are uniformly distributed on some subsets of $\{0, 1, \ldots, n-1\}$. Let us show that such random variables ξ_1 and ξ_2 cannot be identically distributed. Indeed, the moment generating function (mgf) of ξ is $\mathrm{F}(s) = (1 + s + \ldots + s^{n-1})/n$. The assumption $\xi = \xi_1 + \xi_2$ with ξ_1, ξ_2 identically distributed implies $\mathrm{F}(s) = \mathrm{F}_0^2(s)$, where $\mathrm{F}_0(s)$ is the mgf of ξ_1 and ξ_2. From the above-mentioned Theorem 1 of Zhigljavsky et al. (2016), $\mathrm{F}_0(s) = (a_0 + a_1 s + \ldots + a_{n-1}s^{n-1})/k$ for $k = \sqrt{n}$ and coefficients $a_j \in \{0, 1\}$. The squared mgf $\mathrm{F}_0^2(s)$ cannot have the form $(1+s+\ldots+s^{n-1})/n$ as in the expansion of $n\mathrm{F}_0^2(s)$ coefficients with some powers will necessarily be larger than 1. Similarly, for any $n, m > 1$ the random variable $\xi \sim \mathrm{DU}(n)$ cannot be represented as a sum $\xi \overset{d}{=} \xi_1 + \ldots + \xi_m$, where $\xi_1, \ldots, \xi_m$ are i.i.d.r.v.

All results of Zhigljavsky et al. (2016) can be extended to the case $n = ①$. Likewise, we arrive at the conclusion of impossibility of representation of $\xi \sim \mathrm{DU}(①)$ in the form of a sum $\xi \overset{d}{=} \xi_1 + \ldots + \xi_m$, where $\xi_1, \ldots, \xi_m$ are independent non-degenerate random variables ($1 < m \leq ①$). This conclusion seems to contradict to Theorem 1 of Sect. 2.1. However, the proof of Theorem 1 of Sect. 2.1 (which is Theorem 2.1 of Warde and Katti (1971)) cannot be extended to the case when $\Pr(\xi = j+1)/\Pr(\xi = j) = 1$ for all j as for the validity of formula (3) of Warde and Katti (1971), at least one of the inequalities $p_{j+1}/p_j \geq p_j/p_{j-1}$ should be strict.

For $\xi \sim \mathrm{DU}(n)$, the number of different decompositions $\xi \overset{d}{=} \xi_1 + \ldots + \xi_m$ with independent non-degenerate random variables $\xi_1, \ldots, \xi_m$ is equal to the number of all ordered factorizations of ①. In view of Hille (1936) and (Chor et al. 2000), the number of all ordered factorizations of n (and hence the number of different decompositions of $\xi \sim \mathrm{DU}(n)$ as sums of independent non-degenerate random variables) may reach the order n^ρ, where $\rho \simeq 1.72865$ and is defined as the solution of the equation $\zeta(\rho) = 2$, where $\zeta(\cdot)$ is the classical Riemann's zeta-function $\zeta(s)$; this function is unambiguously defined for all s with $\mathrm{Re}(s) > 1$.

Assuming the grossone divisibility axiom, we thus expect for the number of all ordered factorizations of ① to be much larger than ① and reaching ①$^{\rho}$ with $\rho \simeq 1.72865$. Note that this is not a precise statement as the divisibility axiom does not give us enough information about all divisors of ①. Note also that there are difficulties (discussed in Sergeyev (2017), Ya (2011) regarding the use of the zeta-function in the grossone-based universe since many different zeta-functions can be distinguished in the latter.

3 Improving Poisson approximation to the Binomial probabilities

The pmf for Poi(λ), the Poisson distribution with parameter λ, is defined by

$$p_x = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots . \tag{5}$$

There is a lot of literature on the accuracy of Poisson approximation to Bin(N, p), see for example (Hodges and Le Cam 1960; Duembgen et al. 2019). If we are not interested in approximating tails of the Binomial distribution, then Poisson approximation Poi(λ) to Bin(N, p) is rather accurate when $\lambda = pN$ is not too large. However, unless λ is small enough, Poi(λ) does not approximate tails of Bin(N, p) well even if N is very large. There are several approaches for correcting the Poisson approximation including the Stein–Chen method, see (Barbour et al. 1992). Among these asymptotic corrections, the most known is based on the use the expansion with respect to the Charlier polynomials. This expansion has been developed in Uspensky (1931) and can be written as follows:

$$\frac{b_x}{p_x} = \left[1 - \frac{\tilde{c}_2(x;\lambda)}{2N} + \frac{8\tilde{c}_3(x;\lambda)+3\tilde{c}_4(x;\lambda)}{4!\,N^2} - \frac{12\tilde{c}_4(x;\lambda)+8\tilde{c}_5(x;\lambda)+\tilde{c}_6(x;\lambda)}{2\cdot 4!\,N^3} + O\left(\frac{1}{N^4}\right)\right], \tag{6}$$

where $\lambda = p/N$, b_x and p_x are defined by (3) and (5), respectively, $\tilde{c}_j(x;\lambda) = \lambda^j c_j(x;\lambda)$ and $c_j(x;\lambda)$ are the Charlier polynomials

$$c_j(x;\lambda) = \sum_{k=0}^{j} (-1)^k \lambda^{-k} \binom{j}{k}\binom{x}{k} k!\,.$$

We will derive an alternative improvement to the Poisson approximation which, as we will demonstrate, is very accurate at the lower tail of the binomial distribution Bin(N, p) with very large N and very small p; the value of $\lambda = Np$ could be rather large but smaller than $\sqrt{N}$.

To start with, we rearrange the binomial probability b_x as follows

$$b_x = \frac{N!}{(N-x)!\,x!} p^x (1-p)^{N-x} = \frac{\lambda^x}{x!} \underbrace{\frac{N!}{(N-x)!N^x}}_{R} \underbrace{\left(1-\frac{\lambda}{N}\right)^N}_{Q} \underbrace{\left(1-\frac{\lambda}{N}\right)^{-x}}_{T}. \tag{7}$$

Assume N is large enough and $p = \lambda/N$. For all $\lambda x << N$, we have

$$T = \left(1-\frac{\lambda}{N}\right)^{-x} = 1 + \sum_{j=1}^{\infty} T_j(\lambda, x) N^{-j} \tag{8}$$

with

$$T_j(\lambda, x) = \frac{(-1)^j \lambda^j x_{(j)}}{j!} \quad (j = 1, 2, \ldots),$$

where $x_{(j)} = x(x-1)\ldots(x-j+1)$ is the falling factorial. The series (8) converges if $\lambda x/N \to 0$ as $N \to \infty$.

Now, consider Q of (7):

$$Q = \left(1-\frac{\lambda}{N}\right)^N = \exp\left\{N \log\left(1-\frac{\lambda}{N}\right)\right\} = \exp\left\{-N \sum_{k=1}^{\infty} \frac{\lambda^k}{k\,N^k}\right\} = e^{-\lambda} e^{-y},$$

where

$$y = \sum_{k=1}^{\infty} \frac{\lambda^{k+1}}{(k+1)\,N^k} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^k}{(k+1)\,N^k}\,.$$

Using the expansion $e^{-y} = 1 + \sum_{m=1}^{\infty}(-1)^m y^m/m!$ and writing y^m in the form of the multiple sum

$$y^m = \lambda^m \sum_{k_1=1}^{\infty} \ldots \sum_{k_m=1}^{\infty} \frac{\lambda^{k_1}\ldots\lambda^{k_m}}{N^{k_1}\ldots N^{k_m}(k_1+1)\ldots(k_m+1)}\,,$$

we obtain

$$\left(1-\frac{\lambda}{N}\right)^N = e^{-\lambda}\left[1 + \sum_{m=1}^{\infty} \frac{(-1)^m \lambda^m}{m!} \sum_{k_1=1}^{\infty} \ldots \sum_{k_m=1}^{\infty} \frac{\lambda^{k_1}\ldots\lambda^{k_m}}{N^{k_1}\ldots N^{k_m}(k_1+1)\ldots(k_m+1)}\right].$$

Collecting the terms by powers of $1/N$, we obtain

$$Q=\left(1-\frac{\lambda}{N}\right)^{N}=e^{-\lambda}\left[1+\sum_{j=1}^{\infty}Q_j(\lambda)N^{-j}\right], \tag{9}$$

where the first four polynomials $Q_j(\lambda)$ are

$$\begin{aligned}
Q_1(\lambda)&=-\frac{1}{2}\,\lambda^2,\\
Q_2(\lambda)&=\frac{1}{4!}\,\lambda^3\,(3\,\lambda-8)\,,\\
Q_3(\lambda)&=-\frac{1}{2\cdot 4!}\,\lambda^4\,(\lambda-2)\,(\lambda-6)\,,\\
Q_4(\lambda)&=\frac{1}{8\cdot 6!}\,\lambda^5\left(15\,\lambda^3-240\,\lambda^2+1040\,\lambda-1152\right).
\end{aligned}$$

The series (9) converges if $\sqrt{\lambda}/N\to 0$ as $N\to\infty$.

For the term R of (7), we have:

$$R=\frac{N!}{(N-x)!N^x}=\prod_{i=0}^{x-1}\left(1-\frac{i}{N}\right)=1+\sum_{j=1}^{\infty}R_j(x)N^{-j}, \tag{10}$$

where the first four polynomials $R_j(x)$ are

$$\begin{aligned}
R_1(x)&=-\frac{1}{2}\,x(x-1),\\
R_2(x)&=\frac{1}{4!}\,x(x-1)(x-2)(3x-1),\\
R_3(x)&=-\frac{1}{2\cdot 4!}\,x^2(x-1)^2(x-2)(x-3),\\
R_4(x)&=\frac{1}{8\cdot 6!}x(x-1)(x-2)(x-3)(x-4)\\
&\quad(15x^3-30x^2+5x+2).
\end{aligned}$$

Combining (7)–(10), we obtain the following expansion for the probability b_x:

$$\frac{b_x}{p_x}=\left[1+\sum_{j=1}^{\infty}Q_j(\lambda)N^{-j}\right]\left[1+\sum_{j=1}^{\infty}R_j(x)N^{-j}\right]\left[1+\sum_{j=1}^{\infty}T_j(\lambda,x)N^{-j}\right]. \tag{11}$$

All the series converge if $\lambda x/N\to 0$ as $N\to\infty$. If $\lambda x/N$ does not tend to 0 as $N\to\infty$, but x/N and $\sqrt{\lambda}/N$ do, then we recommend to use the approximation

$$\frac{b_x}{p_x}=\left[1+\sum_{j=1}^{\infty}Q_j(\lambda)N^{-j}\right]\left[1+\sum_{j=1}^{\infty}R_j(x)N^{-j}\right]\left(1-\frac{\lambda}{N}\right)^{-x}, \tag{12}$$

where the term T of (7) is not expanded. Numerical results show that for x in the left tail of the binomial distribution, the approximations (11) and (12) practically coincide if we get enough terms in the expansion for T.

Let us rewrite the result (11) in terms of the binomial probabilities of $\mathrm{Bin}(\text{①},\lambda/\text{①})$ keeping only two main terms in the expansion:

$$\begin{aligned}
\frac{b_x}{p_x}=1&-\frac{(\lambda+x)^2-x}{2\text{①}}\\
&+\frac{3(\lambda+x)^4+9x^2-2(4\lambda^3+9\lambda^2x+6\lambda x^2+5x^3+x)}{24\text{①}^2}\\
&+O\left(\frac{1}{\text{①}^3}\right).
\end{aligned} \tag{13}$$

This formula makes sense in the grossone universe. Indeed, both λ and x could be infinitely large but $(\lambda+x)^2/\text{①}$ should be kept infinitesimal as otherwise the second and third terms in the rhs of (13) become large. In particular, the expansion (13) is valid if $\lambda=c_1\text{①}^{\alpha}$ and $x=c_2\text{①}^{\beta}$ where c_1 and c_2 are finite constants and both α and β are smaller than $1/2$.

4 Numerical study

In Figs. 1, 2 and 3, we demonstrate accuracy of several approximations for b_x in (3) by plotting the ratio $\tilde{b}_x/b_x$ where $\tilde{b}_x$ is an approximation for b_x. We have chosen the following coding for the x axis: "0" is the mean $\lambda=Np$ of the distribution $\mathrm{Bin}(N,p)$ and $j=\pm 1,\pm 2,\ldots$ denote the points $\lambda+js$, where $s=\sqrt{Np(1-p)}$ is the standard deviation of $\mathrm{Bin}(N,p)$.

In Figs. 1, 2 and 3, we have chosen $p=0.03$ and $N=10^3, 10^4, 10^5$ so that $\lambda=30$, 300 and 3000. The normal approximation (depicted by black solid line) is always very poor. Due to incorrect skewness, the normal approximation considerably underestimates the probabilities b_x for $x<\lambda$ overestimates them for $x>\lambda$. We have tried to use several improved normal approximations, in particular, the ones based on the Edgeworth expansion, but these expansions were not much better and in many cases some of the estimators $\tilde{b}_x$ became negative. Uncorrected Poisson approximation (blue dashed line) is slightly more accurate than normal, but it is still quite poor. The corrected Poisson approximations (6), based on the expansion with respect to the Charlier polynomials, significantly improve the Poisson approximation. The first-order Charlier approximation (red dotted line), where we keep only the term $\tilde{c}_2(x;\lambda)/2N$ in (6), works well for x within the 3σ-interval; the second-order Charlier approximation is very accurate x within the 4σ-interval.

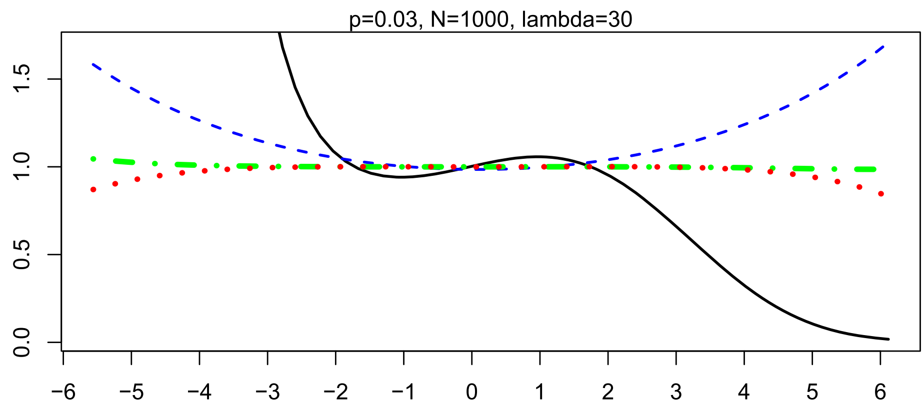

Fig. 1 Ratios $\tilde{b}_x/b_x$, where $\tilde{b}_x$ is computed using the approximations: normal (black solid), Poisson (blue dashed), first-order Charlier (red dotted) and second-order Charlier (green dash-dotted); $N = 1000$, $p = 0.03$ (color figure online)

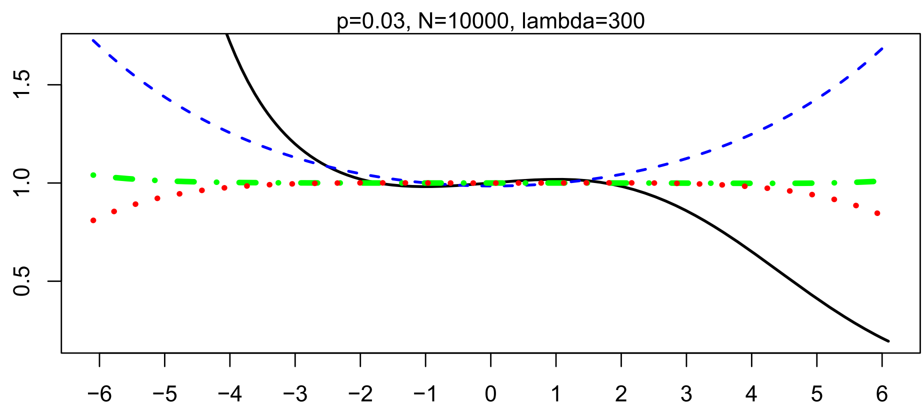

Fig. 2 Ratios $\tilde{b}_x/b_x$, where $\tilde{b}_x$ is computed using the approximations: normal (black solid), Poisson (blue dashed), first-order Charlier (red dotted) and second-order Charlier (green dash-dotted); $N = 10^4$, $p = 0.03$ (color figure online)

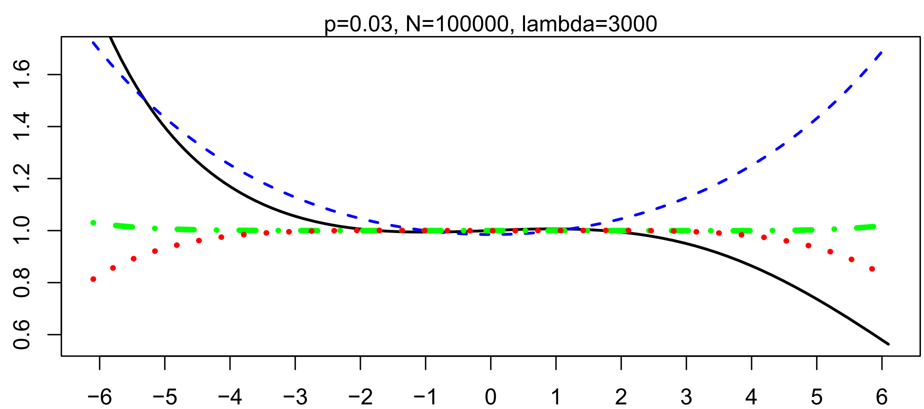

Fig. 3 Ratios $\tilde{b}_x/b_x$, where $\tilde{b}_x$ is computed using the approximations: normal (black solid), Poisson (blue dashed), first-order Charlier (red dotted) and second-order Charlier (green dash-dotted); $N = 10^5$, $p = 0.03$ (color figure online)

The style of Figs. 4, 5 and 6 is similar to Figs. 1, 2 and 3. In these figures, we also demonstrate accuracy of several approximations $\tilde{b}_x$ for binomial probabilities b_x by plotting the ratio $\tilde{b}_x/b_x$ where $\tilde{b}_x$ is an approximation for b_x. Similarly to Figs. 1, 2 and 3, for the x-axis, $j = 0, \pm 1, \pm 2, \dots$ denote the points $\lambda + js$, where $s = \sqrt{Np(1-p)}$. In Figs. 4, 5 and 6, the smallest value at the x-axis corresponds to $x = 0$ in (3) so that we show the values of the ratios $\tilde{b}_x/b_x$ for $x = 0, 1, \dots, \lambda + 6\sqrt{Np(1-p)}$.

In Figs. 4, 5 and 6, we do not use uncorrected normal and Poisson approximations as these approximations are poor. We use three approximations based on the expansion (6): the first-order Charlier approximation (red dotted line), second-order Charlier (green dash-dotted) and third-order Charlier (magenta long dashed). We also use the expansion (12) (brown dashed line), where we keep the first four terms in the expansions for Q and R.

The corrected Poisson approximations (6), based on the expansion with respect to the Charlier polynomials, significantly improve the Poisson approximation. The first-order Charlier approximation works well for x within the 3σ-interval, but the second-order and especially third-order Charlier approximations are much more accurate. The new approximation (11) is basically exact at the lower tail of the binomial distribution and outperforms the Charlier approximations.

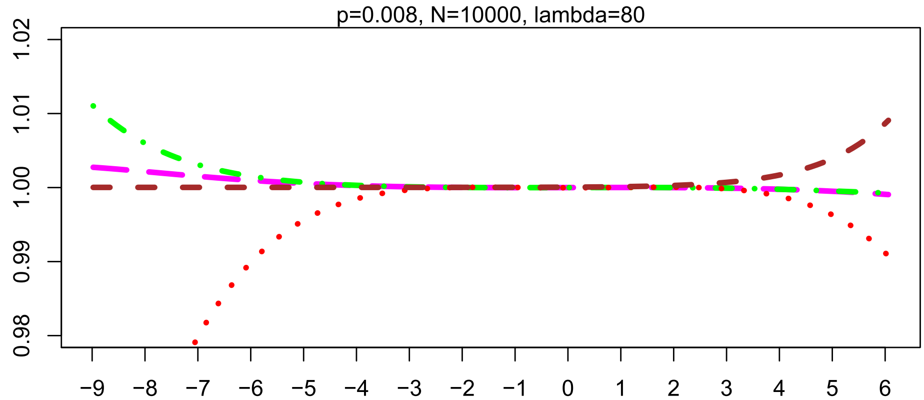

Fig. 4 Ratios $\tilde{b}_x/b_x$, where $\tilde{b}_x$ is computed using the approximations: (12) (brown dashed), first-order Charlier (red dotted), second-order Charlier (green dash-dotted), third-order Charlier (magenta long dashed); $N = 10^4$, $p = 0.008$ (color figure online)

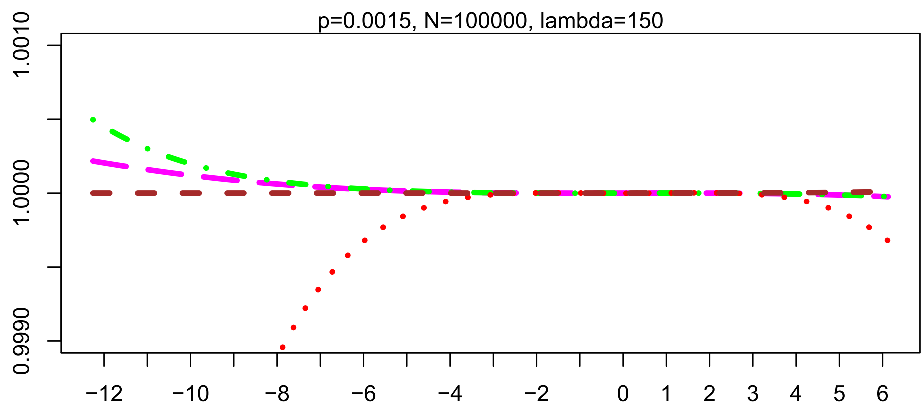

Fig. 5 Ratios $\tilde{b}_x/b_x$, where $\tilde{b}_x$ is computed using the approximations: (12) (brown dashed), first-order Charlier (red dotted), second-order Charlier (green dash-dotted), third-order Charlier (magenta long dashed); $N = 10^5$, $p = 0.0015$ (color figure online)

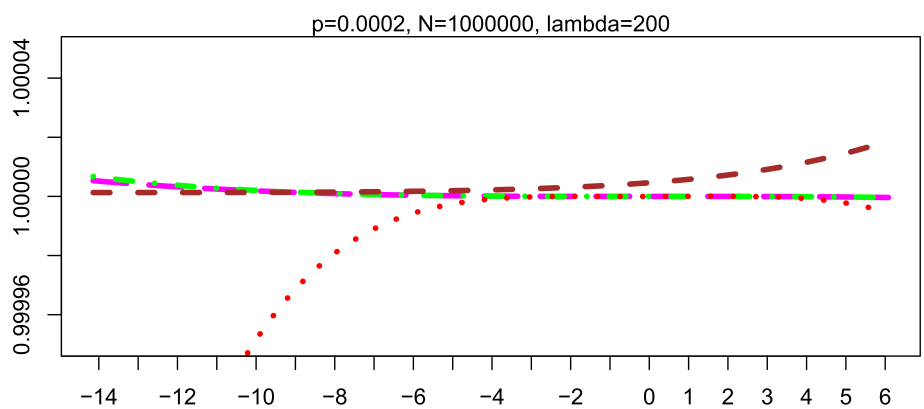

Fig. 6 Ratios $\tilde{b}_x/b_x$, where $\tilde{b}_x$ is computed using the approximations: (12) (brown dashed), first-order Charlier (red dotted), second-order Charlier (green dash-dotted), third-order Charlier (magenta long dashed); $N = 10^6$, $p = 0.0002$ (color figure online)

Conclusion

We study properties of two probability distributions defined on the infinite set $\{0, 1, 2, \ldots\}$ and generalizing the ordinary discrete uniform and binomial distributions. Both extensions use the notion of grossone denoted by ①. The uniform distribution assigns masses $1/①$ to all points in the set $\{0, 1, \ldots, ① - 1\}$. For this distribution, we study the problem of decomposing a r.v. ξ with this distribution as a sum $\xi \stackrel{d}{=} \xi_1 + \ldots + \xi_m$, where $\xi_1, \ldots, \xi_m$ are independent non-degenerate r.v. We establish that, under the validity of the grossone divisibility axiom, such decompositions exist, but all r.v. ξ_j in the decomposition $\xi \stackrel{d}{=} \xi_1 + \ldots + \xi_m$ must have different distributions and, as a corollary, that the discrete uniform distribution on the set $\{0, 1, \ldots, ① - 1\}$ is not infinitely divisible, where the natural extension of the notion of infinite divisibility (introduced in Sect. 2.1) is used.

Then, we study the accuracy of different approximations for the probability mass function of the binomial distribution Bin$(①, p)$ with $p = c/①^{\alpha}$ with $1/2 < \alpha \leq 1$. We demonstrate that the normal and uncorrected Poisson approximations are rather poor and develop a new approximation which is demonstrated to be extremely accurate on the lower tail of Bin$(①, p)$. We compare the accuracy of the developed approximation with the corrected Poisson approximations constructed from the expansion with respect to the Charlier polynomials. The accuracy of approximations is assessed on the base of a numerical study. To derive approximations, we use asymptotic expansions formulated in the standard language, but the final results we translate into the language of grossone.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

References

Balakrishnan N, Nevzorov V (2004) A primer on statistical distributions. Wiley, New Jersey
Barbour AD, Holst L, Janson S (1992) Poisson approximation. Oxford University Press, Oxford
Benci V, Horsten L, Wenmackers S (2018) Infinitesimal probabilities. Br J Philos Sci 69(2):509–552
Berry A (1941) The accuracy of the Gaussian approximation to the sum of independent variates. Trans Am Math Soc 49:122–136
Brown LD, Cai T, DasGupta A (2001) Interval estimation for a binomial proportion. Stat Sci 16:101–117
Calude CS, Dumitrescu M (2020) Infinitesimal probabilities based on grossone. SN Comput Sci 1(1):36
Chor B, Lemke P, Mador Z (2000) On the number of ordered factorizations of natural numbers. Discrete Math 214(1–3):123–133
Duembgen L, Samworth R, Wellner J (2019) Bounding distributional errors via density ratios. arXiv preprint arXiv:1905.03009
Gillard JW, Zhigljavsky Anatoly A (2016) Weighted norms in subspace-based methods for time series analysis. Numer Linear Algebra Appl 23(5):947–967
Golyandina N, Zhigljavsky A (2020) Blind deconvolution of covariance matrix inverses for autoregressive processes. Linear Algebra Appl 593:188–211
Hansson SO (2020) Revising probabilities and full beliefs. J Philos Logic. https://doi.org/10.1007/s10992-020-09545-w
Hille E (1936) A problem in factorisatio numerorum. Acta Arith 2:134–144
Hodges JL, Le Cam L (1960) The poisson approximation to the poisson binomial distribution. Ann Math Stat 31(3):737–740
Lolli G (2012) Infinitesimals and infinites in the history of mathematics: a brief survey. Appl Math Comput 218(16):7979–7988
Lolli G (2015) Metamathematical investigations on the theory of grossone. Appl Math Comput 255:3–14
Petrov VV (1995) Limit theorems of probability theory: sequences of independent random variables. Oxford Science Publications, Oxford
Raftery A (1988) Inference for the binomial N parameter: a hierarchical Bayes approach. Biometrika 75(2):223–228
Rizza D (2018) A study of mathematical determination through Bertrand's Paradox. Philos Math 26(3):375–395
Sergeyev YD (2013) Arithmetic of infinity. Edizioni Orizzonti Meridionali, CS, 2003, 2nd ed
Sergeyev Ya (2009) Numerical point of view on Calculus for functions assuming finite, infinite, and infinitesimal values over finite, infinite, and infinitesimal domains. Nonlinear Anal Ser A Theory Methods Appl 71(12):e1688–e1707
Sergeyev Ya (2017) Numerical infinities and infinitesimals: methodology, applications, and repercussions on two Hilbert problems. EMS Surv Mathe Sci 4:219–320
Sergeyev YaD (2019) Independence of the grossone-based infinity methodology from non-standard analysis and comments upon logical fallacies in some texts asserting the opposite. Found Sci 24(1):153–170
Uspensky JV (1931) On Ch. Jordan's series for probability. Ann Math 32(2):306–312
Warde WD, Katti SK (1971) Infinite divisibility of discrete distributions, ii. Ann Math Stat 42(3):1088–1090
Ya D Sergeyev (2011) On accuracy of mathematical languages used to deal with the Riemann zeta function and the Dirichlet eta function. P-Adic Numbers Ultrametric Anal Appl 3(2):129–148
Zhigljavsky A (2012) Computing sums of conditionally convergent and divergent series using the concept of grossone. Appl Math Comput 218(16):8064–8076
Zhigljavsky A, Golyandina N, Gryaznov S (2016) Deconvolution of a discrete uniform distribution. Stat Prob Lett 118:37–44

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.