KYUNGPOOK Math. J. 57(2017), 493-502
https://doi.org/10.5666/KMJ.2017.57.3.493
pISSN 1225-6951 eISSN 0454-8124

Krawtchouk Polynomial Approximation for Binomial Convolutions

Hyung-Tae Ha
Department of Applied Statistics, Gachon University, Sungnam-ci, Kyunggi-do, Korea, 13120
e-mail : hthа@gachon.ac.kr

Abstract. We propose an accurate approximation method via discrete Krawtchouk orthogonal polynomials to the distribution of a sum of independent but non-identically distributed binomial random variables. This approximation is a weighted binomial distribution with no need for continuity correction unlike commonly used density approximation methods such as saddlepoint, Gram-Charlier A type(GC), and Gaussian approximation methods. The accuracy obtained from the proposed approximation is compared with saddlepoint approximations applied by Eisinga *et al.* [4], which are the most accurate method among higher order asymptotic approximation methods. The numerical results show that the proposed approximation in general provide more accurate estimates over the entire range for the target probability mass function including the right-tail probabilities. In addition, the method is mathematically tractable and computationally easy to program.

1. Introduction

Statistical methods involving the convolutions of sequences of random numbers are useful in many scientific fields. Especially the convolution of the independent binomial variable is utilized in such as reliability theory of engineering systems, acceptance survey random sampling in connection with ascertained statistical data, number of claims in insurance mathematics and default models for financial credit risks. For instance, Ong [7] utilized the convolution of two binomial variables in several physical and stochastic models, Benneyan and TaŞeli [1] used the convolution for measuring qualities of bundle compliances in health care organizations, and Smalley *et al.* [9] used it to analyze DNA sequence in the context of a genome search.

Received July 7, 2016; revised July 14, 2017; accepted August 9, 2017.
2010 Mathematics Subject Classification: 41A58.
Key words and phrases: Krawtchouk polynomials, distribution approximation, saddlepoint approximation, sum of Independent nonidentical Binomial Distributions.

The exact distribution of the convolution is possible to calculate via combinatorial approach relying on modern computing facilities. In order to illustrate the computational complexity of the combinatorial approach, the exact distribution of a convolution of two discrete random variables Y and Z can be obtained from calculating all the combinations based on the following formula;

$$\mathbf{Pr}(Y+Z=j)=\sum_{i=0}^{j}\mathbf{Pr}(Y=i)\cdot\mathbf{Pr}(Z=j-i),$$

where j is the possible outcomes of the convolution. We easily see that the computation for the convolution can be extremely heavy when the number of components and the size parameters of the components are increasing, which can often occur, for instance, in the reliability for k-out-of n system and number of defaults under credit risk rating system. The exact expressions for the probability mass functions of the general case of the convolutions are not available, but the exact probabilities are feasible to computationally obtain via recursive or combinatorial methods. But their computations are complicated and inefficient, especially when the number of outcomes with non-zero probability is large.

In the first step, the exact probability mass function of a convolution of two binomial variables can be expressed in explicit form using Gaussian hypergeometric function ${}_2F_1$, as shown in Ong [7]. The probability mass function of $S_2 = X_1 + X_2$ where $X_i \sim \mathbf{B}in[m_i,\ p_i]$ is

$$\mathbf{Pr}(S_2=k)=\frac{(m_1+m_2)!}{k!(m_1+m_2-k)!}(1-p_1)^{m_1}p_2^k(1-p_2)^{m_2-k}\times$$
$$ {}_2F_1\Big(-m_1,-k;-m_1-m-2;1-\frac{(1-p_1)p_2}{(1-p_2)p_1}\Big).$$

But the expression for the exact probability mass function of binomial convolutions with more than three components becomes very complicated and challenging to analytically obtain. On denoting a binomial convolution by $S_n=\sum_{i=1}^n X_i$ where the independent Binomial random sequence of X_i with non-identical parameters m_i and p_i, that is, $X_i \sim \mathrm{Bin}(m_i,\, p_i)$, Shah [8] provided the following two recurrent relations,

$$\mathbf{Pr}(S_n=m)=\sum_{j=1}^{m}\frac{\mathbf{Pr}(S_n=m-j)}{m\,j!}\sum_{i=1}^{n}\frac{\partial^j C_{X_i}(\ln y)}{\partial y^i}\Big|_{y\to 0},$$

where ∂^j denotes the j^{th} order partial differentiation and $C_X(z)=n(pe^z+1-p)$, and

$$\text{(1.1)}\quad \mathbf{Pr}(S_n=x)=\begin{cases}\displaystyle\prod_{i=1}^{n}(1-p_i)^{m_i} & \text{for } x=0\\ \frac{1}{x}\displaystyle\sum_{j=1}^{x}(-1)^{j-1}(\mathbf{Pr}(S_n=x-j)\sum_{i=1}^{n}m_i[\frac{1-p_i}{p_i}]^j) & \text{for } x>0.\end{cases}$$

Efforts for accurate and efficient approximations have been made to overcome this problem of computational complexity. Especially, approximation methods with closed form expressions are recently explored in computational and statistical literatures. See, for instance, Benneyan and TaŞelt [1], Hong [6] and Eisinga *et al.* [4]. Recently, Eisinga *et al.* [4] explained that saddlepoint mass approximations obtain superiority in accuracy to other existing approximation methods including asymptotic distributions such as Binomial approximation, Poisson approximation and Gaussian approximation, and higher order asymptotic approximations such as GC approximation.

In this paper, we aim to provide a new accurate approximation method with relatively simple closed form expression to the probability mass function of the convolution. The new method provides in general more accurate estimates than the saddlepoint approximations of Eisinga *et al.* [4]. And the new approximation is very simple in functional expression and easy to implement in symbolic computational packages. In addition, this approximation doesn't require continuity correction unlike commonly used density approximation methods such as saddlepoint, GC and Gaussian approximations.

The rest of this paper is organized as follows: In Section 2, some benchmark saddlepoint approximation techniques are briefly explained. Section 3 proposes a new approximation methods of the discrete Krawtchouk polynomial approximants. Section 4 illustrated the details of the proposed technique via numerical examples for comparison purposes. And the concluding remarks follow in Section 5.

2. Benchmark Saddlepoint Approximations

Some versions of saddlepoint approximations were proposed by Eisinga *et al.* [4] for obtaining approximate distributions of binomial convolutions. For illustrating the computational comparisons, we review the saddlepoint approximations. The cumulant generating function of the binomial convolution denoted by $\kappa_{S_n}(t)$ is obtained by taking the logarithm of the moment generating function,

$$\kappa_{S_n}(t) = \sum_{i=1}^{n} m_i \ln\left(1 - p_i + p_i e^t\right), \quad t \in (-\infty, \infty). \tag{2.1}$$

Saddlepoint solution ($t = \hat{t}$) is obtained from equating $\kappa^{(1)}_{S_n}(t) = \sum_{i=1}^{n} m_i\, q_i = x$, where $\kappa^{(i)}_{S_n}(\cdot)$ denotes the i^{th} derivative of the cumulant generating function and $q_i = p_i \exp(t)/\{1 - p_i + p_i \exp(t)\}$. The saddlepoint solution can be solved analytically or numerically.

First-order saddlepoint approximation

The saddlepoint approximation to the probability density function is

$$\hat{\mathbf{Pr}}_1(S_n = x) = \left(2\pi\, \kappa^{(2)}_{S_n}(\hat{t})\right)^{-1/2} \exp(\kappa_{S_n}(\hat{t}) - x\, \hat{t}) \ , \tag{2.2}$$

where $\kappa_{S_n}^{(2)}(t) = \sum_{i=1}^{n} m_i \, q_i \, (1 - q_i)$. This saddlepoint approximant is not a *bona fide* distribution that sums to unity. The normalized saddlepoint approximant can be modified by making use of a normalizing constant τ, that is,

$$\bar{\mathbf{Pr}}_1(S_n = x) = \tau \, \hat{\mathbf{Pr}}_1(S_n = x), \tag{2.3}$$

where the normalization constant τ can be calculated by making use of numerical integration methods such as trapezoid rule because the saddlepoint approximation can not be analytically integrable. And the approximate tail probabilities of S_n can also be determined by numerical integration. This tail approximant is denoted by $\bar{\mathbf{Pr}}_2(S_n > x)$ in numerical examples. And it should be mentioned that the saddlepoint solution equation cannot be solved at the endpoints 0 and $\max(x) = \sum_{i=1}^{n} m_i$ of the support of S_n. For a sum of n binomial random variables, the exact boundary probabilities are given by $\mathbf{Pr}(S_n = 0) = \prod_{i=1}^{n} (1 - p_i)^{m_i}$ and $\mathbf{Pr}(S_n = \max(x)) = \prod_{i=1}^{n} p_i^{m_i}$.

The Lugannani and Rice formula with continuity correction is used to determine the saddlepoint approximation to tail probability in discrete setting,

$$\hat{\mathbf{Pr}}_3 = \mathbf{Pr}(S_n \geq v) \approx 1 - \Phi(\hat{w}) + \phi(\hat{w})(\frac{1}{\hat{u}} - \frac{1}{\hat{w}}), \tag{2.4}$$

where $\hat{w} = \sqrt{2(\hat{t}v - \kappa_{S_n}(\hat{t}))} \; \mathtt{sgn}(\hat{t})$, $\hat{u} = (1 - e^{-\hat{t}})\sqrt{\kappa_{S_n}^{(2)}(\hat{t})}$, $\mathtt{sgn}(\hat{t}) = \pm 1, 0$ if $\hat{t}$ is positive, negative, or zero, and $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the corresponding cumulative distribution function. This Lugannani and Rice formula is undefined if $\hat{u} = \hat{w} = 0$, which occurs when $x = E(S_n)$ or $\hat{t} = 0$. The approximation in such case should be

$$\hat{\mathbf{Pr}}_3 = \frac{1}{2} - \{2\pi\}^{-1/2}\{\frac{1}{6}\kappa_{S_n}^{(3)}(0)\{\kappa_{S_n}^{(2)}(0)\}^{-3/2} - \frac{1}{2}\{\kappa_{S_n}^{(2)}(0)\}^{-1/2}\}, \tag{2.5}$$

where $\kappa_{S_n}^{(3)}(0) = \sum_{i=1}^{n} m_i p_i (1 - p_i)(1 - 2p_i)$.

Second-order saddlepoint mass approximation

Since the order of error for the first-order saddlepoint approximation is $O(n^{-1})$, we further minimize the error by making use of the second-order approximation;

$$\hat{\mathbf{Pr}}_2(S_n = x) = \hat{\mathbf{Pr}}_1(S_n = x)\{1 + \frac{1}{8}\frac{\kappa_{S_n}^{(4)}(\hat{t})}{\{\kappa_{S_n}^{(2)}(\hat{t})\}^2} - \frac{5}{24}\frac{\{\kappa_{S_n}^{(2)}(\hat{t})\}^2}{\{\kappa_{S_n}^{(2)}(\hat{t})\}^3} + O(n^{-2})\}, \tag{2.6}$$

where $\kappa^{(3)}_{S_n}(\hat{t}) = \sum_{i=1}^{n} m_i q_i(1-q_i)(1-2q_i)$ and $\kappa^{(4)}_{S_n}(\hat{t}) = \sum_{i=1}^{n} m_i q_i(1-q_i)\{1-6q_i(1-q_i)\}$. And the second order continuity-corrected saddlepoint approximation to the right-tail probability is also given as

$$\text{(2.7)}\quad \hat{\mathbf{Pr}}_4(S_n \geq x) = \hat{\mathbf{Pr}}_3(S_n \geq x) - \phi(\hat{w})\Big\{\frac{1}{\hat{u}_2}(\frac{1}{8}\hat{k}_4 - \frac{5}{24}\hat{k}_3^2) - \frac{1}{\hat{u}_2^3} - \frac{\hat{k}_3}{2\hat{u}_2^2} + \frac{1}{\hat{w}^3}\Big\},$$

where $\hat{u}_2 = \hat{u}\{\kappa^{(2)}_{S_n}(\hat{u})\}^{1/2}, \hat{k}_3 = \kappa^{(3)}_{S_n}(\hat{u})\{\kappa^{(2)}_{S_n}(\hat{u})\}^{-3/2}$ and $\hat{k}_4 = \kappa^{(4)}_{S_n}(\hat{u})\{\kappa^{(2)}_{S_n}(\hat{u})\}^{-2}$.

3. Krawtchouk Polynomial Approximation

We propose a new approximation method of Krawtchouk polynomial approximation for the distributions of binomial convolutions. The Krawtchouk orthogonal polynomial of degree k with respect to a binomial variable, on denoting $\mathcal{B}(i\,;\theta, N) = \mathcal{B}(x\,;\theta, N) = \binom{N}{x}\theta^x(1-\theta)^{N-x}$ with a proportion parameter θ and a positive integer valued index parameter N, is given by

$$\text{(3.1)}\qquad \mathcal{H}_k(x;\theta,N) = {}_2F_1(-k, -x; -N; \frac{1}{\theta}),$$

where $k = 0,1,\ldots,N$, $x = 0,1,\ldots,N$, $0 < \theta < 1$ and $N \in \mathbf{N}$. Then, the orthogonality factor, $\mathcal{Q}_k$, of the Krawtchouk polynomial of degree k can be obtained by

$$\sum_{x=0}^{N}\mathcal{B}(x;\theta,N)\mathcal{H}_k(x;\theta,N)\mathcal{H}_j(x;\theta,N) = \begin{cases}\mathcal{Q}_k = \frac{(-1)^k k!}{(-N)_k}\left(\frac{1-\theta}{\theta}\right)^k & \text{for } k = j\\ 0 & \text{for } k \neq j,\end{cases}$$

where $\mathcal{Q}_0 = 1$ and $(a)_b = \prod_{k=0}^{b-1}(a+k)$ is a *Pochhammer* function. Then, the proposed approximant of degree d, denoted by $f_d^{\mathcal{H}}$, can be expressed as

$$\text{(3.2)}\qquad f_d^{\mathcal{H}}(x) = \mathcal{B}(x;\theta,N)\sum_{k=0}^{d}\eta_k\mathcal{H}_k(x;\theta,N),$$

where $\eta_0(=1), \eta_1, \ldots, \eta_d$ form coefficients of the linear combination of the linearly independent Krawtchouk polynomials. For parameter estimations, two parameters of the reference Binomial distribution and the coefficients of the linear combination of the Krawtchouk polynomials can be obtained by method of moments to make use of easily calculable moments of the target convolution. On denoting $\mu_{S_n}(h)$ the h^{th} order moment of the convolution, since the moment generating function of the convolution is available in closed form, we can easily evaluate its higher moments from differentiating the moment generating function, that is,

$$\mu_{S_n}(h) = \frac{d^h\mathcal{M}_{S_n}(z)}{dz^h}\Big|_{z\to 0}.$$

First, the parameters of baseline Binomial distribution are determined as $\hat{N} = \sum_{i=0}^{n} m_i$ and $\hat{\theta} = \frac{\sum_{i=0}^{n} m_i p_i}{\hat{N}}$. Note that N is simply a sum of all Binomial trials because it is just the index parameter. Then, the coefficients η_h of the Krawtchouk polynomial approximant can be accordingly estimated as

$$\eta_h \;=\; \frac{1}{\mathcal{Q}_h} \sum_{j=0}^{k} \delta_{k,j}\ \hat{\mu}_{S_n}(j)\,, \qquad h = 1, \ldots, d, \tag{3.3}$$

where $\delta_{h,k}$ is the coefficient of x^k in the Krawtchouk polynomial of order h, that is, $\mathcal{H}_h(x;\ \hat{\theta},\ \hat{N}) = \sum_{j=0}^{k} \delta_{k,j}\ x^j$. This approximant is very flexible to adapt the higher order exact moments of the target convolution unlike the usual asymptotic approximations that are limited to utilizing only 2 to 3 moments.

4. Numerical Examples

We examined numerical comparisons of the proposed methods and several existing methods. The cases from Butler ([3], page 11), Benneyan and TaŞeli [1] and Eisinga *et al.* [4] are revisited for fair comparisons. Each cases are illustrated in the following three examples. The parameter values of n, m_i and p_i are provided in the panels of Table 1.$\sim$ 5. $\mathbf{Pr}(x)$ and *Sim* respectively represent the exact probability obtained from Equation (1.1) and simulation with 10 million replicates. And $\hat{\mathbf{Pr}}_1(x)$, $\bar{\mathbf{Pr}}_1(S)$ and $\bar{\mathbf{Pr}}_2(S)$, which are employed from Butler [3] and Eisinga *et al.* [4], represent saddlepoint approximations from Equations (2.2), (2.3) and (2.5), respectively. And $\bar{\mathbf{Pr}}_6(S)$ represents the GC approximation of order 6 employed by Benneyan and TaŞeli [1]. The single binomial approximation (Bin) with index $\sum m_i$ and probability $\frac{\sum_{i=1}^{n} p_i}{n}$, the Gaussian approximation (ϕ) matching the first and second moments and Poisson approximation (Pois) matching the first moment are also displayed for comparison. Finally, the Krawtchouk polynomial approximant of a degree d ($f_d^{\mathcal{H}}(x)$) are computed. The results of the Krawtchouk polynomial approximants are rounded at the decimals shown in tables in Eisinga *et al.* (2013) for fair comparisons.

Example 1: Revisiting a case of Butler [3]

Butler [3] computed the case of a sum of three binomial variables with $m_i = 6$, 4, 2 and $p_i = 1/15$, 2/15, 4/15 and claimed that unnormalized and normalized saddlepoint approximations, which are comparable, perform better than Poisson and Gaussian approximations. As shown in Table 1, the Krawtchouk polynomial approximants of 3 degree performs moderately well, specially in the mode area, but not extreme well in the right tail. But when the degree of the Krawtchouk polynomial approximant increases, the method provides extremely accurate approximations in both sides of center and right tail. The Krawtchouk polynomial approximant of 6 degree is nearly exact and outperforms any other approximation techniques.

x	$\mathbf{Pr}(\text{x})$	$\hat{\mathbf{Pr}}_1(x)$	$\bar{\mathbf{Pr}}_1(x)$	Pois(x)	$\phi(x)$	$f_3^{\mathcal{H}}(x)$	$f_4^{\mathcal{H}}(x)$	$f_6^{\mathcal{H}}(x)$
1	0.3552	0.3845	0.3643	0.3384	0.3296	0.3555	0.3553	0.3552
3	0.1241	0.1273	0.1206	0.1213	0.1381	0.1240	0.1241	0.1241
5	$0.0^{2}7261$	$0.0^{2}7392$	$0.0^{2}7004$	0.01305	$0.0^{2}2221$	$0.0^{2}7274$	$0.0^{2}7253$	$0.0^{2}7261$
7	$0.0^{3}1032$	$0.0^{3}1052$	$0.0^{4}9963$	$0.0^{3}6682$	$0.0^{5}1369$	$0.0^{4}9876$	$0.0^{3}1038$	$0.0^{3}1032$
9	$0.0^{6}3633$	$0.0^{6}3738$	$0.0^{6}3541$	$0.0^{4}1996$	$0.0^{10}3238$	$0.0^{6}2659$	$0.0^{6}4406$	$0.0^{6}3630$
11	$0.0^{9}2279$	$0.0^{9}2472$	$0.0^{9}2342$	$0.0^{6}3904$	$0.0^{16}2938$	$0.0^{11}6093$	$0.0^{9}6162$	$0.0^{9}2219$

Table 1: The exact and approximated probabilities for the sum of $n = 3$ binomial variables when $m_i = 6, 4, 2$ and $p_i = 1/15, 2/15, 4/15$.

Example 2: Revisiting a case of Benneyan and TaŞeli [1]

We examined numerical comparisons in this case of a sum of 10 binomial variables with the parameter values of m_i = 12, 14, 4, 2, 20, 17, 11, 1, 8, 11 and p_i = 0.074, 0.039, 0.095, 0.039, 0.053, 0.043, 0.067, 0.018, 0.099, 0.045, which was originally used in Benneyan and TaŞeli [1] and revisited in Eisinga *et al.* [4]. Table 2 shows that the Krawtchouk polynomial approximants of 4 and 6 degrees are nearly exact again and don't have even the very little errors that the normalized second-order saddlepoint approximation $\bar{P}_2(s)$, which was claimed to have a superior fit in Eisinga *et al.* [4], showed. More interestingly, very small error for the probability of $x = 19$ from the Krawtchouk polynomial approximants of 4 degree was removed out by increasing the degree of the Krawtchouk polynomial approximant. The Krawtchouk polynomial approximants of 6 degrees provides the same probabilities with the exact ones after rounding.

x	$\mathbf{Pr}(\text{x})$	$\bar{\mathbf{Pr}}_2(x)$	$\bar{\mathbf{Pr}}_6(x)$	Bin(x)	$\phi(x)$	Pois(x)	$f_4^{\mathcal{H}}(x)$	$f_6^{\mathcal{H}}(x)$
1	0.0165	0.0164	0.0172	0.0168	0.0215	0.0187	0.0165	0.0165
3	0.0994	0.0994	0.0986	0.0999	0.0862	0.1021	0.0994	0.0994
5	0.1716	0.1716	0.1719	0.1712	0.1641	0.1673	0.1716	0.1716
7	0.1346	0.1346	0.1346	0.1340	0.1481	0.1305	0.1346	0.1346
9	0.0587	0.0587	0.0590	0.0586	0.0634	0.0594	0.0587	0.0587
11	0.0160	0.0160	0.0156	0.0161	0.0129	0.0177	0.0160	0.0160
13	$0.0^{2}2912$	$0.0^{2}2913$	$0.0^{2}3013$	$0.0^{2}2969$	$0.0^{2}1237$	$0.0^{2}3719$	$0.0^{2}2912$	$0.0^{2}2912$
15	$0.0^{3}3751$	$0.0^{3}3752$	$0.0^{3}4015$	$0.0^{3}3893$	$0.0^{4}5646$	$0.0^{3}5805$	$0.0^{3}3751$	$0.0^{3}3751$
17	$0.0^{4}3543$	$0.0^{4}3544$	$0.0^{4}2621$	$0.0^{4}3762$	$0.0^{5}1222$	$0.0^{4}6995$	$0.0^{4}3543$	$0.0^{4}3543$
19	$0.0^{5}2524$	$0.0^{5}2525$	$0.0^{6}7245$	$0.0^{5}2756$	$0.0^{7}1253$	$0.0^{5}6704$	$0.0^{5}2525$	$0.0^{5}2524$

Table 2: The exact and approximated probabilities for the sum of $n = 10$ binomial variables when m_i = 12, 14, 4, 2, 20, 17, 11, 1, 8, 11 and p_i = 0.074, 0.039, 0.095, 0.039, 0.053, 0.043, 0.067, 0.018, 0.099, 0.045.

In addition, a very recent paper of Butler and Stephens [2] proposed an extremely accurate approximant, namely Kolmogorov approximation. In order to make fair comparisons between the Kolmogorov approximation and our proposed method, we revisit the example of Benneyan and TaŞeli [1], which was also used in Butler and Stephens [2]. Refer Table 7 of Butler and Stephens [2]. We extend more decimals since both methods provide very accurate approximation. As shown in Table 3, they provide identical results except for $\mathbf{Pr}(3)$. Although the proposed method approximates closer ones to the exact $\mathbf{Pr}(3)$ than the Kolmogorov approximation, it is arguable to determine superiority since the approximation errors are extremely small.

x	$\mathbf{Pr}$(x)	K(4)	K(6)	$f_4^{\mathcal{H}}(x)$	$f_6^{\mathcal{H}}(x)$
1	0.01648856	0.0164885	0.0164886	0.0164885	0.0164886
3	0.09937507	0.0993753	0.0993750	0.0993751	0.0993751
5	0.17156980	0.1715700	0.1715700	0.1715700	0.1715700
7	0.13457790	0.1345780	0.1345780	0.1345780	0.1345780

Table 3: Numerical comparison between the Kolmogorov and the proposed approximations.

Example 3: Revisiting cases of Eisinga *et al.* [4]
Eisinga *et al.* [4] examined two interesting cases with extremely small probabilities and long tails for both sides. First, in Table 4, the case with very small probabilities are computed. This case may be interesting to credit risk practitioners because default probabilities of obligors are very small. Poisson distribution is often utilized in this case on the basis of Poisson approximation via probability generating function, see Gordy [5]. Eisinga *et al.* [4] claimed that both the Poisson and the binomial approximations provide superior fits in the center of the distribution, whereas the saddlepoint approximation $\bar{P}_2(s)$ outperforms in the right tail. As shown in Table 4, the proposed Krawtchouk polynomial approximants of 2 and 4 degrees perform extremely well without restriction in any areas. They outperform Poisson, binomial and saddlepoint approximations over the entire range of the distribution including areas of the center and left and right tails. It is also interesting that simulation (Sim) with 10 million replicates is not the best performing in this case. From this result, we can see that accurate approximation methods with closed form expressions are important.

x	Sim	$\mathbf{Pr}$(x)	$\bar{\mathbf{Pr}}_2(x)$	$\bar{\mathbf{Pr}}_6(x)$	Bin(x)	$\phi(x)$	Pois(x)	$f_2^{\mathcal{H}}(x)$	$f_4^{\mathcal{H}}(x)$
1	0.3234	0.3231	0.3227	0.3690	0.3230	0.4496	0.3230	0.3231	0.3231
2	0.0918	0.0924	0.0928	0.0480	0.0923	0.0889	0.0925	0.0924	0.0924
3	0.0176	0.0176	0.0177	0.0284	0.0176	$0.0^2 3059$	0.0176	0.0176	0.0176
4	$0.0^2 2601$	$0.0^2 2514$	$0.0^2 2525$	$0.0^2 2794$	$0.0^2 2508$	$0.0^4 1834$	$0.0^2 2525$	$0.0^2 2514$	$0.0^2 2514$
5	$0.0^3 275$	$0.0^3 2868$	$0.0^3 2881$	$0.0^4 1707$	$0.0^3 2859$	$0.0^7 1915$	$0.0^3 2891$	$0.0^3 2868$	$0.0^3 2868$
6	$0.0^4 29$	$0.0^4 2723$	$0.0^4 2735$	$0.0^7 1167$	$0.0^4 2713$	$0.0^{11} 3482$	$0.0^4 2759$	$0.0^4 2723$	$0.0^4 2723$
7	$0.0^5 5$	$0.0^5 2213$	$0.0^5 2224$	$0.0^{11} 1077$	$0.0^5 2205$	$0.0^{15} 1103$	$0.0^5 2256$	$0.0^5 2213$	$0.0^5 2213$

Table 4: The exact and approximated probabilities for the sum of $n = 10$ binomial variables when $m_i = 120, 140, 40, 20, 200, 170, 110, 10, 80, 110$ and $p_i = 0.0^3 74$, $0.0^3 39$, $0.0^3 95$, $0.0^3 39$, $0.0^3 53$, $0.0^3 43$, $0.0^3 67$, $0.0^3 18$, $0.0^3 99$, $0.0^3 45$.

Table 5 shows a case with extreme long tails for both sides. As expected from the central limit theorem, Gaussian and the GC approximation $\bar{P}_6(S)$ are performing very well. It is surprising that the saddlepoint approximation yields even more accurate results. But note that the Krawtchouk polynomial approximant doesn't work very well in this case. Binomial based approximants seem to have difficulties to capture the long tail behavior of both sides.

x	$\mathbf{Pr}$(x)	$\bar{\mathbf{Pr}}_2(x)$	$\bar{\mathbf{Pr}}_6(x)$	Bin(x)	$\phi(x)$	Pois(x)	$f_6^{\mathcal{H}}(x)$
510	$0.0^5 2363$	$0.0^5 2363$	$0.0^5 2363$	$0.0^4 1058$	$0.0^5 2346$	$0.0^3 5109$	$0.0^5 2898$
520	$0.0^4 3730$	$0.0^4 3730$	$0.0^4 3730$	$0.0^3 1056$	$0.0^4 3706$	$0.0^2 1458$	$0.0^4 3774$
530	$0.0^3 3638$	$0.0^3 3638$	$0.0^3 3638$	$0.0^3 7061$	$0.0^3 3623$	$0.0^2 3436$	$0.0^3 3660$
540	$0.0^2 2195$	$0.0^2 2195$	$0.0^2 2195$	$0.0^2 3162$	$0.0^2 2191$	$0.0^2 6702$	$0.0^2 2244$
550	$0.0^2 8202$	$0.0^2 8202$	$0.0^2 8202$	$0.0^2 9471$	$0.0^2 8201$	0.01087	$0.0^2 8450$
560	0.0190	0.0190	0.0190	0.0190	0.0190	0.0147	0.0196
570	0.0272	0.0272	0.0272	0.0253	0.0272	0.0166	0.0283
580	0.0242	0.0242	0.0242	0.0224	0.0241	0.0157	0.0254
590	0.0133	0.0133	0.0133	0.0132	0.0133	0.0126	0.0142
600	$0.0^2 4501$	$0.0^2 4501$	$0.0^2 4501$	$0.0^2 5141$	$0.0^2 4501$	$0.0^2 8500$	$0.0^2 4907$
610	$0.0^3 9419$	$0.0^3 9419$	$0.0^3 9419$	$0.0^2 1321$	$0.0^3 9460$	$0.0^2 4854$	$0.0^2 1058$
620	$0.0^3 1213$	$0.0^3 1213$	$0.0^3 1213$	$0.0^3 2230$	$0.0^3 1230$	$0.0^3 2353$	$0.0^3 1447$
630	$0.0^5 9581$	$0.0^5 9581$	$0.0^5 9581$	$0.0^4 2463$	$0.0^5 9902$	$0.0^3 9708$	$0.0^4 1352$
640	$0.0^6 4630$	$0.0^6 4630$	$0.0^6 4628$	$0.0^5 1773$	$0.0^6 4931$	$0.0^3 3418$	$0.0^5 1003$

Table 5: The exact and approximated probabilities for the sum of $n = 10$ binomial variables when m_i = 120, 140, 40, 20, 200, 170, 110, 10, 8, 110 and p_i =0.74, 0.39, 0.95, 0.39, 0.53, 0.43, 0.67, 0.18, 0.99, 0.45.

5. Concluding Remarks

In this paper, a new approximation technique for the density and distribution functions of sums of independent but nonidentical Binomial random variables was proposed by making use of discrete Krawtchouck orthogonal polynomials. The proposed Krawtchouck polynomial approximants don't require continuity correction unlike commonly utilized methods due to the property of discrete Krawtchouck orthogonal polynomials. Note that the Krawtchouck polynomial approximant utilizes only first several moments whereas the saddlepoint approximation requires cumulant generating function, from which all series of moments can be generated. In the numerical examples, the Krawtchouck polynomial approximants utilized just 4 or 6 moments. Therefore, we conclude that even though the proposed method utilized less information of the Binomial convolution than saddlepoint approximations, the proposed method was shown to provide the most accurate estimates except for a case with extremely long tails in both sides. In addition, it should be mentioned that the proposed method is mathematically simple and computationally easy to program.

Acknowledgements. The author wishes to express thanks to have the presentation opportunity and constructive discussions in the third international conference on Numerical Analysis and Optimization, which was held at Sultan Qaboos University, Sultanate of Oman, from 5 to 9 January 2014. We also thank the editor and an anonymous referee for helpful comments and suggestions concerning our paper. This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A2059794).

References

[1] J. C. Benneyan and A. TaŞeli, *Exact and approximate probability distributions of evidence-based bundle composite compliance measures*, Health Care Manag. Sci., **13**(2010), 193–209.

[2] K. Butler and M. A. Stephens, *The distribution of a sum of independent binomial random variables*, Methodol Comput. Appl. Probab., **in press**, doi:10.1007/s11009-016-9533-4.

[3] R. W. Butler, *Saddlepoint Approximations with Applications*, Cambridge University Press, Cambridge, 2007.

[4] R. Eisinga, M. T. Grotenhuis and B. Pelzer, *Saddlepoint approximations for the sum of independent non-identically distributed binomial random variables*, Stat. Neerl., **67**(2013), 190–201.

[5] M. B. Gordy, *Saddlepoint approximation of CreditRisk$^+$*, J. Bank. Financ., **26**(2002), 1335–1353.

[6] Y. Hong, *On computing the distribution function for the sum of independent and non-identical random indicators*, Technical Report No. 11-2, Department of Statistics, Virginia Tech, Blacksburg, VA, 2011.

[7] S. H. Ong, *Some stochastic models leading to the convolution of two binomial variables*, Stat. Probab. Lett., **22**(1995), 161–166.

[8] B. K. Shah, *On the distribution of the sum of independent integer valued random variables*, Am. Stat., **27**(1973), 123–124.

[9] S. L. Smalley, J. A. Woodward and C. G. S. Palmer, *A general statistical model for detecting complex-trait loci by using affected relative pairs in a genome search*, Am. J. Hum. Genet., **58**(1996), 844–860.