P. Vellaisamy, N. S. Upadhye, V. Cekanavicius, On negative binomial approximation, *Теория вероятн. и ее примен.*, 2012, том 57, выпуск 1, 141–154

DOI: 10.4213/tvp4435

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
Том 57 И ЕЕ ПРИМЕНЕНИЯ Выпуск 1
2012

 VELLAISAMY P.*, UPADHYE N. S.**, ČEKANAVIČIUS V.**

ON NEGATIVE BINOMIAL APPROXIMATION

Рассматривается отрицательная биномиальная аппроксимация сумм независимых $\mathbf{Z}_+$-значных случайных величин. С помощью метода Стейна устанавливаются границы ошибок. Свертка отрицательного биномиального и пуассоновского распределений используется в качестве трехпараметрической аппроксимации.

Ключевые слова и фразы: отрицательное биномиальное распределение, отрицательное биномиальное возмущение, метод Стейна, расстояние по вариации.

1. Introduction. It is well known that negative binomial (NB) distribution and its generalizations arise naturally in many fields such as modelling of crash-data, telecommunication networks, population genetics, epidemics and various other related fields. Moreover, since it has a quite simple structure and depends on two parameters only, the NB distribution can be used as approximation, see [6], [11], [12] and the references therein.

In this paper, we investigate NB approximation to the sum of random variables via the Stein's method. Our results deal with a sum of arbitrary independent random variables taking values in $\mathbf{Z}_+ = \{0, 1, \ldots\}$ and having three or four finite moments. We discuss the accuracy that can be achieved by one-parameter and two-parameter NB approximations. Our results resemble the binomial approximation results to the Poisson binomial distribution (see [3, p.189] or [9]), where the approximation is exact when indicators are identically distributed. Since we deal with NB approximation, the role of indicators is played by geometric variables and our approximations are also exact, as expected, when the geometric variables are also identically distributed. Also, the convolution of NB and Poisson distributions is considered, as an example of three-parametric approximation. This approximation is treated as perturbation to the NB law and an appropriate Stein's perturbation technique is used.

*Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India; e-mail: pv@math.iitb.ac.in
**Department of Mathematics and Informatics, Vilnius University, Naugarduko 24, Vilnius 03225, Lithuania; e-mail: vydas.cekanavicius@mif.vu.lt

We now introduce necessary notation. Throughout, assume that $\sum_{j=k}^{m} = 0$, if $m < k$, and

$$d_{\mathrm{TV}}(\mathrm{M}_1, \mathrm{M}_2) = \frac{1}{2}\sum_{k=0}^{\infty} |\mathrm{M}_1\{k\} - \mathrm{M}_2\{k\}| = \sup_A |\mathrm{M}_1\{A\} - \mathrm{M}_2\{A\}|$$

represents the total variation distance between two distributions M_1 and M_2 on $\mathbf{Z}_+$. Here the supremum is taken over all Borel sets. For any bounded function g defined on $\mathbf{Z}_+$, we denote by $\Delta g(j) = g(j+1) - g(j)$ its first forward difference, $\Delta^k g = \Delta(\Delta^{k-1} g)$, and $\|g\| = \sup_{j\geqslant 0} |g(j)|$. Also, $\mathscr{L}(X)$ denotes the distribution of X. We write $X \sim \mathrm{Be}(p)$ for Bernoulli variable with $\mathbf{P}(X=1) = p = 1 - \mathbf{P}(X=0)$.

For real $r > 0$ and $0 < p < 1$, let $Y \sim \mathrm{NB}(r,p)$ denote the NB distribution with

$$\mathbf{P}(Y = k) = \binom{r+k-1}{k} p^r q^k, \qquad k = 0, 1, \ldots,$$

where $q = 1 - p$. Note that r is not necessarily an integer. The notation $X \sim \mathrm{Ge}(p)$ is equivalent to $X \sim \mathrm{NB}(1,p)$.

Throughout the paper, we assume that $X_1, X_2, \ldots, X_n$ are independent random variables, concentrated on $\mathbf{Z}_+$ and having finite second moment. For $1 \le i \le n$, let $\mu_i = \mathbf{E}\,X_i$, $\sigma_i^2 = \mathbf{D}X_i$, $p_{ik} = \mathbf{P}(X_i = k)$, $W = \sum_{i=1}^{n} X_i$ and $W_i = W - X_i$.

2. The Stein operator and its perturbation. First we recall the main facts related to the NB distribution. Let $Y \sim \mathrm{NB}(r,p)$. For any bounded function g on $\mathbf{Z}_+$, define the following Stein operator:

$$(\mathscr{A} g)(j) := q(r+j)g(j+1) - jg(j). \tag{1}$$

It is easy to check that $\mathbf{E}\,(\mathscr{A} g)(Y) = 0$.

For any $A \subset \mathbf{Z}_+$, let $g(\cdot) = g_A(\cdot)$ be a solution of the following equation:

$$(\mathscr{A} g)(j) = \mathbb{I}(j \in A) - \mathrm{NB}(r,p)(A), \qquad j \in \mathbf{Z}_+. \tag{2}$$

Then

$$\|g\| \leqslant \frac{1}{p}\left(1 \wedge \frac{1.75}{\sqrt{\alpha}}\right), \qquad \|\Delta g\| \leqslant \frac{1 - \mathrm{e}^{-\alpha}}{p\alpha} \leqslant \frac{1}{rq}, \tag{3}$$

where $\alpha = rq/p$ (see [6]).

The Stein method is based on the fact that the total variation distance between the distribution of any nonnegative integer-valued variable X and $\mathrm{NB}(r,p)$ can be replaced by estimates of $|\mathbf{E}\,(\mathscr{A} g)(X)|$. Indeed, if for some $\varepsilon_1 \geqslant 0$ and $\varepsilon_2 \geqslant 0$,

$$|\mathbf{E}\,(\mathscr{A} g)(X)| \leqslant \varepsilon_1 \|g\| + \varepsilon_2 \|\Delta g\|, \tag{4}$$

then

$$d_{\mathrm{TV}}(\mathscr{L}(X), \mathrm{NB}(r,p)) \leqslant \frac{\varepsilon_1}{p}\left(1 \wedge \frac{1.75}{\sqrt{rq/p}}\right) + \frac{\varepsilon_2}{rq}. \tag{5}$$

Therefore, the problem of obtaining the estimate of the error term in total variation reduces to that of obtaining (4) with small ε_1, ε_2. We exemplify this approach in the next two subsections.

The NB distribution has two parameters which can be utilized for matching of two moments of the approximated distribution. The natural next step is to discuss an approximation with more parameters. However, we encounter a serious problem of obtaining estimates similar to (3). Brown and Xia [7] proved very sharp bounds for solution of (2), when the Stein operators are of the form $(\mathscr{A}g)(j) = a_j g(j+1) - b_j g(j)$. A partial success was achieved for compound Poisson distribution, see [5]. Unfortunately, none of the mentioned results can be applied in our case. Therefore, we use the perturbation technique. Poisson perturbation was introduced by Barbour and Xia in [4] and was later generalized in [1]. The essence of perturbation technique can be summarized in the following way: if approximation has the Stein operator $\mathscr{A}_1$ which is close to some other Stein's operator with known properties, then we can use these properties at the expense of additional restrictive assumptions and larger constants.

The main result of [1] is formulated in very general terms, which we reformulate for the case of NB distribution and total variation metric. In the following, we assume that $\mathscr{A}$ is defined by (1) and g is any bounded function defined on $\mathbf{Z}_+$. Let M be a measure of finite variation defined on $\mathbf{Z}_+$ and let $\mathscr{A}_1$ be its Stein operator defined by

$$\sum_{k=0}^{\infty} (\mathscr{A}_1 g)(k)\mathrm{M}\{k\} = 0.$$

Moreover, let there exist operator U defined on a set of all bounded functions with support $\mathbf{Z}_+$ such that

$$\mathscr{A}_1 = \mathscr{A} + U, \qquad \|Ug\| \leqslant \tilde{\varepsilon}\|\Delta g\|, \quad \tilde{\varepsilon} < rq. \tag{6}$$

Now, for any $A \subset \mathbf{Z}_+$, let $g(\cdot) = g_A(\cdot)$ be a solution of the following equation:

$$(\mathscr{A}_1 g)(j) = \mathbb{I}(j \in A) - \mathrm{M}(A), \qquad j \in \mathbf{Z}_+.$$

If a random variable X on $\mathbf{Z}_+$, for some $\varepsilon \geqslant 0$, satisfies inequality

$$|\mathbf{E}\,(\mathscr{A}_1 g)(X)| \leqslant \varepsilon\|\Delta g\|, \tag{7}$$

then

$$d_{\mathrm{TV}}(\mathscr{L}(X), \mathrm{M}) \leqslant \frac{\varepsilon}{rq - \tilde{\varepsilon}}. \tag{8}$$

It must be mentioned that there are some slight differences in our formulation from the one given in Theorem 2.4 from [1]. We used assumption $\|Ug\| \leqslant \tilde{\varepsilon}\|\Delta g\|$ instead of a weaker one related to the exact operator norm of U.

3. One-parametric approximation. Our first result deals with one-parametric NB approximation of a sum of independent random variables. Let $Y \sim \mathrm{NB}(r,p)$, where r and p are such that

$$\frac{rq}{p} = \sum_{i=1}^{n} \mu_i, \qquad q = 1 - p. \tag{9}$$

We have quite modest assumptions about the moments of X_i. Moreover, the approximation is flexible in a sense that one can choose different r and p satisfying (9). For example, one can take $r = n$. However, Y matches one moment of W only. Consequently, in general, one can expect results comparable to Poisson approximation, but not to the normal one.

Theorem 3.1. *Let* $\mathbf{E}\,X_i^2 < \infty$ $(i = 1, 2, \ldots, n)$ *and* $Y \sim \mathrm{NB}(r,p)$, *where* r *and* p *satisfy* (9). *Then the following estimate holds*:

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant \frac{1}{rq} \sum_{i=1}^{n} \sum_{k=1}^{\infty} k|p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}|. \tag{10}$$

R e m a r k 3.1. (i) If $X_i \sim \mathrm{NB}(r_i, p)$, then by choosing $r = r_1 + r_2 + \cdots + r_n$, we get $d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) = 0$, as expected.

(ii) If $X_i \sim \mathrm{Be}(p_i)$, then

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant \frac{\sum_{i=1}^{n} p_i^2}{\sum_{i=1}^{n} p_i} + \frac{q}{p}.$$

Though we can choose q to be small, it is clear that (see [2, p. 3–4]) the accuracy is worse than in Poisson approximation.

P r o o f o f T h e o r e m 3.1. Using (9), we obtain

$$\begin{aligned}
\mathbf{E}\,(\mathscr{A}g)(W) &= \mathbf{E}\,\{rqg(W+1) + qWg(W+1) - Wg(W)\} \\
&= \sum_{i=1}^{n} \Big\{p\mu_i\,\mathbf{E}\,g(W+1) + q\,\mathbf{E}\,X_i g(W+1) - \mathbf{E}\,X_i g(W)\Big\} \\
&= \sum_{i=1}^{n} \Big\{p\mu_i \sum_{k=0}^{\infty} p_{ik}\,\mathbf{E}\,g(W_i+k+1) + q\sum_{k=0}^{\infty} kp_{ik}\,\mathbf{E}\,g(W_i+k+1) \\
&\qquad - \sum_{k=0}^{\infty} kp_{ik}\,\mathbf{E}\,g(W_i+k)\Big\} \\
&= \sum_{i=1}^{n} \Big\{p\mu_i\Big(1 - \sum_{k=1}^{\infty} p_{ik}\Big)\mathbf{E}\,g(W_i+1) \\
&\qquad + p\mu_i \sum_{k=1}^{\infty} p_{ik}\,\mathbf{E}\,g(W_i+k+1) \\
&\qquad + q\sum_{k=1}^{\infty} kp_{ik}\,\mathbf{E}\,g(W_i+k+1) - \sum_{k=1}^{\infty} kp_{ik}\,\mathbf{E}\,g(W_i+k)\Big\}
\end{aligned}$$

$$= \sum_{i=1}^{n} \Big\{ p\mu_i \sum_{k=1}^{\infty} p_{ik}(\mathbf{E}\, g(W_i + k + 1) - \mathbf{E}\, g(W_i + 1))$$
$$+ q \sum_{k=1}^{\infty} k p_{ik}(\mathbf{E}\, g(W_i + k + 1) - \mathbf{E} g(W_i + 1))$$
$$- \sum_{k=2}^{\infty} k p_{ik}(\mathbf{E}\, g(W_i + k) - \mathbf{E}\, g(W_i + 1)) \Big\}$$
$$= \sum_{i=1}^{n} \sum_{k=1}^{\infty} \{p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}\} \sum_{s=1}^{k} \mathbf{E}\, \Delta g(W_i + s). \tag{11}$$

Thus,

$$|\mathbf{E}\, (\mathscr{A} g)(W)| \leqslant \|\Delta g\| \sum_{i=1}^{n} \sum_{k=1}^{\infty} k |p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}|.$$

Applying (5), the proof follows.

We next compute the bound for the case when X_i are independent geometric $\mathrm{Ge}(p_i)$ random variables and compare it with known bounds in the literature.

Corollary 3.1. *Let X_i be independent* $\mathrm{Ge}(p_i)$ *random variables,* $u_i = q_i/p_i$, $q_i = 1 - p_i$, *and* $v_i = \lfloor q_i/p_i \rfloor + 1$. *Then,*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant \frac{1}{rq} \sum_{i=1}^{n} \frac{|p - p_i|}{p_i}\, \kappa(i), \tag{12}$$

where $\kappa(i) = u_i[2(v_i - 1)v_i q_i^{v_i+1} - 2(v_i - 1)(2v_i + 1)q_i^{v_i} + 2v_i^2 q_i^{v_i-1} - 1]$ *and* $\lfloor x \rfloor$ *denotes the integer part of* x.

P r o o f. Using $p_{ik} = p_i q_i^k$ and $\mu_i = q_i/p_i$, we get

$$\sum_{k=1}^{\infty} k|p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}| = \sum_{k=1}^{\infty} kq_i^k |pq_i + qkp_i - (k+1)p_i q_i|$$
$$= \sum_{k=1}^{\infty} kq_i^k |(k - (k+1)q_i)(p_i - p)| = \frac{|p - p_i|}{p_i}\, \kappa(i), \tag{13}$$

where the last equality follows from the fact that $k - (k+1)q_i$ is positive for $k \geqslant u_i$ and negative for $k < u_i$ and then splitting the sum with respect to v_i (since k is nonnegative integer). This proves the result.

R e m a r k 3.2. (i) If $q_i < 1/2$, then

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant \frac{1}{rq} \sum_{i=1}^{n} |p - p_i|\, \frac{q_i}{p_i^2} = \frac{1}{rq} \sum_{i=1}^{n} |p - p_i| \mathbf{D} X_i. \tag{14}$$

(ii) The order of the above result improves upon Theorem 2.2 in [11].
(iii) Roos [10, Theorem 1] proved for this case the following estimate:

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant 8.8 \sum_{i=1}^{n} p_i^2 \min\left\{ \sum_{j=1}^{\infty} \frac{p_i^2 q_i^{2j}}{\sum_{i=1}^{n} p_i^2 q_i^j}, 1 \right\}. \tag{15}$$

It can be seen that the bound given in (12) is comparable and improves the constant. Moreover, estimate in (12) takes into account the closeness of p_i and is exact, when $p_i \equiv p$.

(iv) It is not difficult to note that

$$\kappa(i) = u_i(2v_i^2 p_i^2 q_i^{v_i - 1} + 2(v_i p_i + 1) q_i^{v_i} - 1) \leqslant 5u_i,$$

since $v_i p_i \leqslant 1$. Therefore, for $1/2 \leqslant q_i < 1$, a rougher version of (14) holds, where $\mathbf{D} X_i$ is replaced by $5\mathbf{D} X_i$.

4. Two-parametric approximation. If the random variables have three finite moments, we can utilize both the parameters of the NB law to fit the mean and the variance of W. Recall that $W = X_1 + X_2 + \cdots + X_n$, and $W_i = W - X_i$, where the X_i are independent nonnegative integer-valued random variables. Choose now

$$r = \frac{(\mathbf{E}\, W)^2}{\mathbf{D} W - \mathbf{E}\, W}, \qquad p = \frac{\mathbf{E}\, W}{\mathbf{D} W}, \tag{16}$$

so that $\mathbf{E}\, W = rq/p$ and $\mathbf{D} W = rq/p^2$. Let

$$\tau := 2 \max_{1 \leqslant i \leqslant n} d_{\mathrm{TV}}(\mathscr{L}(W_i), \mathscr{L}(W_i + 1)) = \max_{1 \leqslant i \leqslant n} \sum_{k=0}^{\infty} |\mathbf{P}(W_i = k) - \mathbf{P}(W_i = k-1)|.$$

Theorem 4.1. *Let $\mathbf{E}\, X_i^3 < \infty$, $(i = 1, 2, \ldots, n)$ and let $Y \sim \mathrm{NB}(r,p)$ with r and p defined by* (16). *If $\mathbf{D} W > \mathbf{E}\, W$, then the following estimate holds*:

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant \frac{\tau}{rq} \sum_{i=1}^{n} \sum_{k=1}^{\infty} k\left(\frac{k-1}{2} + \mu_i\right) |p\mu_i p_{ik} + qkp_{ik} - (k+1) p_{i,k+1}|. \tag{17}$$

R e m a r k 4.1. Let $\tau_i = \min\{1/2, 1 - d_{\mathrm{TV}}(\mathscr{L}(X_i), \mathscr{L}(X_i + 1))\}$ and $\tau^* = \max_{1 \leqslant i \leqslant n} \tau_i$. Then

$$\tau \leqslant \sqrt{\frac{2}{\pi}} \left(\frac{1}{4} + \sum_{j=1}^{n} \tau_j - \tau^* \right)^{-1/2}, \tag{18}$$

see [8, Corollary 1.6].

(ii) The right-hand side of (17) is less than

$$\frac{\tau}{rq} \left\{ \left(\frac{3+q}{2} \mu_i + q \right) \mathbf{E}\, X_i(X_i - 1) + \frac{q+1}{2} \mathbf{E}\, X_i(X_i - 1)(X_i - 2) + p\mu_i^3 + q\mu_i^2 \right\}.$$

(iii) Note here that the bound given in Theorem 4.1 is not applicable to the case of Bernoulli variables, since the choice p given in (16) is not less than unity.

Observe also that (18) significantly improves the order of accuracy. Indeed, let us assume that all p_{ik} are uniformly bounded from below by some absolute positive constant and that the maximum lattice span for each X_i is unity. Then $\tau_i \geqslant C > 0$, and the estimate (17) is of the order $O(n^{-1/2})$. The same order in weaker Kolmogorov metric can be obtained by the normal approximation. In this case, the estimate (10) is of the trivial order $O(1)$.

At first glance, it seems that two-parametric approximation is always preferable to one-parametric approximation. It is easy to construct an example showing that, as far as our results are concerned, this is not the case. Let $X_i \sim \mathrm{Ge}(1/3)$ for $1 \leq i \leq (n-1)$, $a = \sum_{k=1}^{\infty}(1/k^4)$ and $\mathbf{P}(X_n = k) = 1/(ak^4)$, $k = 1, 2, \ldots$. Then we cannot apply (17). Meanwhile, applying Theorem 3.1 with $p = 1/3$ and using (13), we easily obtain $d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r, 1/3)) = O(n^{-1})$.

Corollary 4.1. *Let X_i be independent* $\mathrm{Ge}(p_i)$ *random variables with* $q_i < 1/2$, $i = 1, 2, \ldots, n$. *Then*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant 3\sqrt{\frac{2}{\pi}}\left(\sum_{j=1}^{n} q_j - \frac{1}{4}\right)^{-1/2}\left(\sum_{k=1}^{n}\frac{q_k}{p_k}\right)^{-1} \times \sum_{i=1}^{n}\left|\frac{1}{p_i} - \frac{1}{p}\right|\left(\frac{q_i}{p_i}\right)^2. \tag{19}$$

R e m a r k 4.2. When $p_i = p$, the approximation is exact, as expected. Moreover, if $\sum_{i=1}^{n} q_i \geqslant 1$, then

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(r,p)) \leqslant C\frac{\sum_{i=1}^{n} q_i^2}{(\sum_{j=1}^{n} q_j)^{3/2}},$$

where $C > 0$. Note that the estimate has much better order than the ones given in (14) and (15).

The rest of this section is devoted to the proofs.

P r o o f o f T h e o r e m 4.1. We have

$$\begin{aligned}
\mathbf{E}\,\Delta g(W+1) &= \sum_{j=0}^{\infty} p_{ij}\,\mathbf{E}\,\Delta g(W_i + j + 1) \\
&= \sum_{j=0}^{\infty} p_{ij}\Big(\mathbf{E}\,\Delta g(W_i+1) + \sum_{l=1}^{j}\mathbf{E}\,\Delta^2 g(W_i + l)\Big) \\
&= \mathbf{E}\,\Delta g(W_i+1) + \sum_{j=0}^{\infty} p_{ij}\sum_{l=1}^{j}\mathbf{E}\,\Delta^2 g(W_i+l).
\end{aligned}$$

Consequently,

$$\mathbf{E}\,\Delta g(W_i+1) = \mathbf{E}\,\Delta g(W+1) - \sum_{j=0}^{\infty} p_{ij} \sum_{l=1}^{j} \mathbf{E}\,\Delta^2 g(W_i+l)$$

and

$$\begin{aligned}\mathbf{E}\,\Delta g(W_i+s) &= \mathbf{E}\,\Delta g(W_i+1) + \sum_{m=1}^{s-1} \mathbf{E}\,\Delta^2 g(W_i+m)\\ &= \mathbf{E}\,\Delta g(W+1) - \sum_{j=0}^{\infty} p_{ij} \sum_{l=1}^{j} \mathbf{E}\,\Delta^2 g(W_i+l) + \sum_{m=1}^{s-1} \mathbf{E}\,\Delta^2 g(W_i+m). \quad (20)\end{aligned}$$

Due to (16), we have

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{k=1}^{\infty} k\{p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}\}\\ &\quad = \sum_{i=1}^{n} [p\mu_i^2 + q(\sigma_i^2+\mu_i^2) - \mathbf{E}X_i(X_i-1)] = \sum_{i=1}^{n}(\mu_i - p\sigma_i^2) = 0. \quad (21)\end{aligned}$$

Therefore, substituting (20) into (11), we obtain

$$\begin{aligned}\mathbf{E}(\mathscr{A}g)(W) &= \sum_{i=1}^{n}\sum_{k=1}^{\infty} \{p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}\}\\ &\times \sum_{s=1}^{k} \Big[-\sum_{j=0}^{\infty} p_{ij} \sum_{l=1}^{j} \mathbf{E}\,\Delta^2 g(W_i+l) + \sum_{m=1}^{s-1} \mathbf{E}\,\Delta^2 g(W_i+m)\Big]. \quad (22)\end{aligned}$$

It is shown in [4] (see also [3, p. 517]) that $|\mathbf{E}\,\Delta^2 g(W_i+m)| \leqslant \tau\|\Delta g\|$. Therefore, the assertion of the theorem follows from (22) and (5).

P r o o f o f C o r o l l a r y 4.1. Note that

$$\begin{aligned}d_{\mathrm{TV}}(\mathscr{L}(X_i), \mathscr{L}(X_i+1)) &= \frac{1}{2}\sum_{k=0}^{\infty} |\mathbf{P}(X_i=k) - \mathbf{P}(X_i=k-1)|\\ &= \frac{1}{2}\Big(p_i + \sum_{k=1}^{\infty} |p_i q_i^k - p_i q_i^{k-1}|\Big) = p_i.\end{aligned}$$

Thus, $\tau_i = q_i$, $\tau^* \leqslant 1/2$. Similar to the proof of Corollary 3.1, we get

$$\begin{aligned}&|p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}| = q_i^k |pq_i + qkp_i - (k+1)p_i q_i|\\ &\quad = q_i^k |k - (k+1)q_i|\,|p_i - p| = p_i q_i^k \left|k - \frac{q_i}{p_i}\right| |p_i - p|. \quad (23)\end{aligned}$$

Note that $q_i/p_i < 1$, since $q_i < 1/2$ and hence the right-hand side of (17) is less than

$$\begin{aligned}
&\frac{\tau}{rq}\sum_{i=1}^{n} k\left(\frac{k-1}{2}+\mu_i\right)|p_i-p|p_i q_i^k\left(k-\frac{q_i}{p_i}\right)\\
&=\frac{\tau}{rq}\sum_{i=1}^{n}|p_i-p|\left(\frac{1}{2}\,\mathbf{E}\,X_i(X_i-1)(X_i-\mu_i)+\mu_i\,\mathbf{E}\,X_i(X_i-\mu_i)\right)\\
&=\frac{\tau}{rq}\sum_{i=1}^{n}|p_i-p|\left(\frac{1}{2}\,\mathbf{E}X_i(X_i-1)(X_i-2)+\frac{1}{2}\,(2-\mu_i)\,\mathbf{E}\,X_i(X_i-1)+\mu_i\sigma_i^2\right)\\
&=\frac{\tau}{rq}\sum_{i=1}^{n}|p_i-p|\left(3\,\frac{q_i^3}{p_i^3}+\left(2-\frac{q_i}{p_i}\right)\frac{q_i^2}{p_i^2}+\frac{q_i^2}{p_i^3}\right)=\frac{3\tau}{rq}\sum_{i=1}^{n}|p_i-p|\,\frac{q_i^2}{p_i^3}\\
&=\frac{3\tau}{p\,\mathbf{E}\,W}\sum_{i=1}^{n}|p_i-p|\,\frac{q_i^2}{p_i^3}=\frac{3\tau}{\mathbf{E}\,W}\sum_{i=1}^{n}\left|\frac{1}{p_i}-\frac{1}{p}\right|\,\frac{q_i^2}{p_i^2}.
\end{aligned}$$

Collecting all estimates and using (18), the bound in (19) follows.

5. Negative binomial perturbation. For an improvement of the accuracy of approximation, we need more than just two-parametric distribution. However, we want to retain the NB law as the main part of approximation. Of course, there are many choices for such approximations. In this section, we consider the convolution of NB and generalized Poisson distribution $\mathrm{Pois}(\lambda)$, where λ is real. Note that $\mathrm{Pois}(0)$ is degenerate at zero. More precisely, let $\mathrm{NB}(N,p)*\mathrm{Pois}(\lambda)$ be (signed) measure with the following generating function:

$$\begin{aligned}
\left(\frac{p}{1-qz}\right)^{N}\exp\{\lambda(z-1)\} &= \exp\bigg\{\left(N\frac{q}{p}+\lambda\right)(z-1)\\
&\qquad+\frac{N}{2}\left(\frac{q}{p}\right)^{2}(z-1)^2+\frac{N}{3}\left(\frac{q}{p}\right)^{3}(z-1)^3+\cdots\bigg\}.
\end{aligned}$$

Let us denote $M(k)=\mathrm{NB}(N,p)*\mathrm{Pois}(\lambda)\{k\}$ and write the generating function

$$\left(\frac{p}{1-qz}\right)^{N}\exp\{\lambda(z-1)\}=\sum_{k=0}^{\infty}M(k)z^k.$$

Taking derivative with respect to z, we obtain

$$\frac{Nq}{1-qz}\sum_{k=0}^{\infty}M(k)z^k+\lambda\sum_{k=0}^{\infty}M(k)z^k=\sum_{k=0}^{\infty}kM(k)z^{k-1}.$$

Consequently,

$$\sum_{k=0}^{\infty}M(k)(Nqz^k+\lambda z^k-\lambda q z^{k+1}-kz^{k-1}+kqz^k)=0$$

and $M(k)(Nq+\lambda+kq)-\lambda qM(k-1)-(k+1)M(k+1)=0$.

Therefore, the corresponding Stein's operator is

$$(\mathscr{A}_1 g)(k) := q\Big(N+\frac{\lambda p}{q}+k\Big)g(k+1)-kg(k)-\lambda q\Delta g(k+1). \tag{24}$$

Comparing (24) with (1) and (6), we see that $r=N+\lambda p/q$, $\tilde{\varepsilon}=|\lambda|q$ and the sufficient condition for (8) to hold is $|\lambda|q<rq=p(Nq/p+\lambda)=p\sum_{i=1}^n \mu_i$.

Further, we discuss the choice of parameters. We can write the following formal expression for the generating function of $W=\sum_{i=1}^n X_i$ as

$$\mathbf{E}\, z^W = \exp\left\{\Gamma_1(z-1)+\frac{\Gamma_2}{2}\,(z-1)^2+\frac{\Gamma_3}{3!}\,(z-1)^3+\cdots\right\}.$$

Here Γ_j is j-th factorial cumulant of W. We can choose real N, p, and λ to match the first three factorial cumulants of W and $\mathrm{NB}(N,p)*\mathrm{Pois}(\lambda)$. Consequently, the first three moments will be matched as well. It is obvious that Γ_j can be expressed through moments of X_i. However, for our purposes it is more convenient to use factorial moments. Therefore, for $i=1,2,\dots,n$, let

$$\nu_{2i}=\mathbf{E}\,X_i(X_i-1),\quad \nu_{3i}=\mathbf{E}\,X_i(X_i-1)(X_i-2),$$
$$\nu_{4i}=\mathbf{E}\,X_i(X_i-1)(X_i-2)(X_i-3).$$

Using the relations between factorial cumulants and factorial moments, given by $\Gamma_1=\sum_{i=1}^n\mu_i$, $\Gamma_2=\sum_{i=1}^n(\nu_{2i}-\mu_i^2)$, $\Gamma_3=\sum_{i=1}^n(\nu_{3i}-3\mu_i\nu_{2i}+2\mu_i^3)$, we see that the parameters N, p and λ must satisfy the following equations:

$$N\,\frac{q}{p}+\lambda = \sum_{i=1}^n \mu_i, \tag{25}$$

$$N\Big(\frac{q}{p}\Big)^2 = \sum_{i=1}^n(\nu_{2i}-\mu_i^2), \tag{26}$$

$$N\Big(\frac{q}{p}\Big)^3 = \frac{1}{2}\sum_{i=1}^n(\nu_{3i}-3\mu_i\nu_{2i}+2\mu_i^3). \tag{27}$$

We want $N>0$ and $0<p<1$, which imposes additional assumptions on the X_i. For example, (26) requires $\mathbf{D}W>\mathbf{E}\,W$, since $\nu_{2i}-\mu_i^2=\sigma_i^2-\mu_i$.

Let

$$\widetilde{\tau}=\sup_{1\leqslant i\leqslant n}\sum_{k=0}^{\infty}|\mathbf{P}(W_i=k-2)-2\mathbf{P}(W_i=k-1)+\mathbf{P}(W_i=k)|.$$

Theorem 5.1. *Let* $\mathbf{E}\,X_i^4<\infty$, $i=1,2,\dots,n$, *and assume that* (25)–(27) *can be solved for* $N>0$ *and* $p\in(0,1)$. *If* $|\lambda|q<p\sum_{i=1}^n\mu_i$, *then*

$$d_{\mathrm{TV}}(\mathscr{L}(W),\mathrm{NB}(N,p)*\mathrm{Pois}(\lambda))\leqslant\frac{\widetilde{\tau}}{p\sum_{j=1}^n\mu_j-|\lambda|q}$$

$$\times \sum_{i=1}^{n} \sum_{k=1}^{\infty} |p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}|$$
$$\times \left(\frac{\nu_{2i}}{2} + k\mu_i^2 + \frac{k(k-1)}{2}\,\mu_i + \frac{k(k-1)(k-2)}{6} \right). \tag{28}$$

R e m a r k 5.1. (i) Let $\tau_i = \min\{1/2, 1 - d_{\mathrm{TV}}(\mathscr{L}(X_i), \mathscr{L}(X_i+1))\}$, $\tau^* = \max_{1 \leqslant i \leqslant n} \tau_i$, and $V = \sum_{i=1}^n \tau_i$. Then

$$\widetilde{\tau} \leqslant 4\left(1 \wedge \frac{2}{(V - 4\tau^*)_+}\right) \leqslant \frac{16}{V}, \tag{29}$$

using [3, equation (4.10)].

(ii) The right-hand side of (28) is less than

$$\frac{\widetilde{\tau}}{p\sum_{j=1}^n \mu_j - |\lambda| q} \sum_{i=1}^{n} \left\{ p\mu_i^2\nu_{2i} + \frac{1+q}{2}\,\nu_{2i}^2 + \frac{3q}{2}\,\mu_i\nu_{2i} + p\mu_i^4 \right.$$
$$\left. + (1+q)\mu_i^2\nu_{2i} + q\mu_i^3 + \frac{2+q}{3}\,\mu_i\nu_{3i} + \frac{1+q}{6}\,\nu_{4i} + \frac{q}{2}\,\nu_{3i} \right\}.$$

Distributions satisfying (25)–(27) have large probability mass at zero. This condition is natural for the so-called aggregate claim distribution for the individual model in insurance mathematics. More precisely, it is assumed that $X_i = \xi_i\eta_i$, where ξ_i and η_i are independent, $\xi_i \sim \mathrm{Be}(\alpha_i)$ and η_i is a positive random variable. One can give the following interpretation: ξ_i reflects the possibility of occurrence of claim with the small probability α_i and η_i denotes the distribution of the claim amount. Note that, if η_i is integer-valued random variable, then Theorem 5.1 can easily be applied, since $\nu_{ki} = \alpha_i \mathbf{E}\,\eta_i(\eta_i - 1)\cdots(\eta_i - k + 1)$. Note also that Theorem 5.1 cannot be applied to the Poisson binomial distribution (i.e., to the case $\mathbf{P}(\eta_i = 1) = 1$), where $\mathbf{D}W < \mathbf{E}\,W$.

As in previous sections, we reformulate Theorem 5.1 for the sum of geometric random variables so that $X_i \sim \mathrm{Ge}(p_i)$, $1 \leqslant i \leqslant n$. In this case,

$$\frac{q}{p} = \frac{\sum_{i=1}^n (q_i/p_i)^3}{\sum_{i=1}^n (q_i/p_i)^2}, \quad N\left(\frac{q}{p}\right)^2 = \sum_{i=1}^{n}\left(\frac{q_i}{p_i}\right)^2, \quad \lambda = \sum_{i=1}^{n} \frac{q_i}{p_i} - N\left(\frac{q}{p}\right).$$

It is easy to check that $\lambda \geqslant 0$.

R e m a r k 5.2. Note that the bound obtained in Theorem 5.1 has more flexibility due to the parameter λ as compared to the one in Theorem 4.1.

Corollary 5.1. *Let X_i be independent* $\mathrm{Ge}(p_i)$ *random variables*, $q_i < 1/2$ $(i = 1, 2, \ldots, n)$ *and* $\sum_{i=1}^n q_i \geqslant 3$. *Then*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(N,p) * \mathrm{Pois}(\lambda))$$
$$\leqslant 56\left(\sum_{j=1}^{n} q_j - 2\right)^{-1} \left(\sum_{j=1}^{n} \frac{q_j}{p_j}\left(1 - \frac{q}{p} + \frac{q_j}{p_j}\right)\right)^{-1} \sum_{i=1}^{n} \left|\frac{1}{p_i} - \frac{1}{p}\right| \left(\frac{q_i}{p_i}\right)^3.$$

R e m a r k 5.3. (i) Note that if $p_i = p$, then the approximation is exact, as expected. Moreover, if $q_i < 1/2$ and $\sum_{i=1}^n q_i \geqslant 4$, then, for some positive constant C,

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{NB}(N,p) * \mathrm{Pois}(\lambda)) \leqslant C \frac{\sum_{i=1}^n q_i^3}{(\sum_{j=1}^n q_j)^2}.$$

(ii) The above bound is an improvement over the better bound obtained for NB approximation in Remark 4.2.

P r o o f o f T h e o r e m 5.1. First note that $Nq + \lambda p = \sum_{i=1}^n \mu_i$. Also, from (25)–(27) and (21), we get

$$\lambda q = \sum_{i=1}^n \sum_{k=1}^\infty k\{p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}\}. \tag{30}$$

Proceeding now exactly as in the proof of (11), and using (30), we obtain

$$\begin{aligned} \mathbf{E}\,(\mathscr{A}_1 g)(W) &= \sum_{i=1}^n \sum_{k=1}^\infty \{p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}\} \\ &\quad \times \sum_{s=1}^k \mathbf{E}\,\Delta g(W_i + s) - \lambda q\, \mathbf{E}\,\Delta g(W+1) \\ &= \sum_{i=1}^n \sum_{k=1}^\infty \{p\mu_i p_{ik} + qkp_{ik} - (k+1)p_{i,k+1}\} \\ &\quad \times \sum_{s=1}^k (\mathbf{E}\,\Delta g(W_i + s) - \mathbf{E}\,\Delta g(W+1)). \end{aligned} \tag{31}$$

For $s = 1, 2, \dots$, we can write Newton's expansion in the form

$$g(\omega + s) = g(\omega+1) + (s-1)\Delta g(\omega+1) + \sum_{l=1}^{s-2}(s-1-l)\Delta^2 g(\omega+l). \tag{32}$$

Using (32), we obtain

$$\begin{aligned} \mathbf{E}\,\Delta g(W_i+s) &= \mathbf{E}\,\Delta g(W_i+1) + (s-1)\,\mathbf{E}\,\Delta^2 g(W_i+1) + R_{1i}, \\ \mathbf{E}\,\Delta g(W+1) &= \mathbf{E}\,\Delta g(W_i+1) + \mu_i\,\mathbf{E}\,\Delta^2 g(W_i+1) + R_{2i}, \\ \mathbf{E}\,\Delta^2 g(W_i+1) &= \mathbf{E}\,\Delta^2 g(W+1) - R_{3i}, \end{aligned}$$

where

$$\begin{aligned} R_{1i} &= \sum_{m=1}^{s-2} (s-1-m)\,\mathbf{E}\,\Delta^3 g(W_i+m), \\ R_{2i} &= \sum_{j=0}^\infty p_{ij} \sum_{m=1}^{j-1} (j-m)\,\mathbf{E}\,\Delta^3 g(W_i+m), \\ R_{3i} &= \sum_{j=0}^\infty p_{ij} \sum_{l=1}^{j} \mathbf{E}\,\Delta^3 g(W_i+l). \end{aligned}$$

Applying the above results to (31) gives us

$$\begin{aligned}&\mathbf{E}\,\Delta g(W_i+s)-\mathbf{E}\,\Delta g(W+1)\\&\qquad=(s-1-\mu_i)\,\mathbf{E}\,\Delta^2 g(W+1)-(s-1-\mu_i)R_{3i}+R_{1i}-R_{2i}.\end{aligned}$$

It follows easily from (21) and (30)

$$\sum_{i=1}^{n}\sum_{k=1}^{\infty}\Big\{p\mu_i p_{ik}+qkp_{ik}-(k+1)p_{i,k+1}\Big\}\sum_{s=1}^{k}(s-1-\mu_i)=0.$$

Consequently,

$$\begin{aligned}|\mathbf{E}\,(\mathscr{A}_1 g)(W)|\leqslant{}&\sum_{i=1}^{n}\sum_{k=1}^{\infty}|p\mu_i p_{ik}+qkp_{ik}-(k+1)p_{i,k+1}|\\&\times\sum_{s=1}^{k}[(s-1+\mu_i)|R_{3i}|+|R_{1i}|+|R_{2i}|].\end{aligned}\tag{33}$$

It is known that $|\mathbf{E}\,\Delta^3 g(W_i+s)|\leqslant\widetilde{\tau}\|\Delta g\|$, see [3, equation (4.12)]. Therefore,

$$|R_{1i}|\leqslant\widetilde{\tau}\|\Delta g\|\,\frac{(s-1)(s-2)}{2},\quad |R_{2i}|\leqslant\widetilde{\tau}\|\Delta g\|\,\frac{\nu_{2i}}{2},\quad |R_{3i}|\leqslant\widetilde{\tau}\|\Delta g\|\mu_i.$$

Putting the last estimates into (33), the proof follows.

Proof of Corollary 5.1. In the previous section, we proved that $\tau_i=q_i$, and $\tau^*<1/2$. Therefore, it follows from (29) that

$$\widetilde{\tau}\leqslant 8\bigg(\sum_{i=1}^{n}q_i-2\bigg)^{-1}.$$

From the definition of λ, p, and N given in (25)–(27), we have

$$\lambda\,\frac{q}{p}=\frac{q}{p}\sum_{i=1}^{n}\frac{q_i}{p_i}-N\Big(\frac{q}{p}\Big)^2=\frac{q}{p}\sum_{i=1}^{n}\frac{q_i}{p_i}-\sum_{i=1}^{n}\Big(\frac{q_i}{p_i}\Big)^2.$$

Therefore,

$$p\sum_{i=1}^{n}\mu_i-\lambda q=p\bigg(\sum_{i=1}^{n}\frac{q_i}{p_i}-\lambda\,\frac{q}{p}\bigg)=p\sum_{i=1}^{n}\frac{q_i}{p_i}\Big(1-\frac{q}{p}+\frac{q_i}{p_i}\Big).$$

Applying (23) and performing some standard calculations, we get

$$\begin{aligned}&\sum_{k=1}^{\infty}|p\mu_i p_{ik}+qkp_{ik}-(k+1)p_{i,k+1}|\\&\qquad\times\Big(\frac{\nu_{2i}}{2}+k\mu_i^2+\frac{k(k-1)}{2}\,\mu_i+\frac{k(k-1)(k-2)}{6}\Big)\\&\quad=|p_i-p|\bigg(\Big(\frac{\nu_{2i}}{2}+\mu_i^2\Big)\mathbf{E}\,X_i(X_i-\mu_i)+\frac{\mu_i}{2}\,\mathbf{E}\,X_i(X_i-1)(X_i-\mu_i)\\&\qquad\qquad+\frac{1}{6}\,\mathbf{E}\,X_i(X_i-1)(X_i-2)(X_i-\mu_i)\bigg)=\frac{7}{p_i}\,|p_i-p|\Big(\frac{q_i}{p_i}\Big)^3.\end{aligned}$$

The proof of Corollary 5.1 follows by using the last three expressions.

Acknowledgments. A part of this work was completed during the visit of V. Čekanavičius and N. S. Upadhye to the National University of Singapore, in January 2009. We wish to express our gratitude for generous hospitality and support from the Institute for Mathematical Sciences of the National University of Singapore. We are grateful to the referee for some valuable suggestions and remarks.

REFERENCES

1. *Barbour A. D., Čekanavičius V., Xia A.* On Stein's method and perturbations. — Lat. Am. J. Probab. Math. Statist., 2007, v. 3, p. 31–53.
2. *Barbour A. D., Holst L., Janson S.* Poisson Approximation. New York: Oxford University Press, 1992, 277 p.
3. *Barbour A. D., Čekanavičius V.* Total variation asymptotics for sums of independent integer random variables. — Ann. Probab., 2002, v. 30, № 2, p. 509–545.
4. *Barbour A. D., Xia A.* Poisson perturbations. — ESAIM Probab. Statist., 1999, v. 3, p. 131–150.
5. *Barbour A. D., Xia A.* Estimating Stein's constants for compound Poisson approximation. — Bernoulli, 2000, v. 6, № 4, p. 581–590.
6. *Brown T. C., Phillips M. J.* Negative binomial approximation with Stein's method. — Methodol. Comput. Appl. Probab., 1999, v. 1, № 4, p. 407–421.
7. *Brown T. C., Xia A.* Stein's method and birth-death processes. — Ann. Probab., 2001, v. 29, № 3, p. 1373–1403.
8. *Mattner L., Roos B.* A shorter proof of Kanters Bessel function concentration bound. — Probab. Theory Related Fields, 2007, v. 139, № 1–2, p. 191–205.
9. *Roos B.* Binomial approximation to the Poisson binomial distribution: The Krawtchouk expansion. — Теория вероятн. и ее примен., 2000, т. 45, в. 2, с. 328–344.
10. *Roos B.* Kerstan's method for compound Poisson approximation. — Ann. Probab., 2003, v. 31, № 4, p. 1754–1771.
11. *Vellaisamy P., Upadhye N. S.* Compound negative binomial approximations for sums of random variables. — Probab. Math. Statist., 2009, v. 29, № 2, p. 205–226.
12. *Wang X., Xia A.* On negative binomial approximation to k-runs. — J. Appl. Probab., 2008, v. 45, № 2, p. 456–471.

Поступила в редакцию
3.II.2011