APPROXIMATION OF SUMS OF LOCALLY DEPENDENT RANDOM
VARIABLES VIA PERTURBATION OF STEIN OPERATOR∗
ZHONGGEN SU†, VLADIMIR V. ULYANOV‡, AND XIAOLIN WANG§

Abstract. Let \((X_i, i \in J)\) be a family of locally dependent nonnegative integer-valued random
variables, and consider the sum \(W = \sum_{i \in J} X_i\). We first establish a general error upper bound
for \(d_{TV}(W, M)\) using Stein’s method, where the target variable \(M\) is either the mixture of Poisson
distribution and binomial or negative binomial distribution. As applications, we attain \(O(|J|^{-1})\)
error bounds for \((k_1,k_2)\)-runs and \(k\)-runs under some special cases. Our results are significant
improvements of the existing results in literature, say \(O(|J|^{-0.5})\) in Peköz [Bernoulli, 19 (2013)]
and \(O(1)\) in Upadhye, et al. [Bernoulli, 23 (2017)].

Key words. Local dependence structure, Stein’s method, Total variation distance, \((k_1,k_2)\)-runs

1. Introduction. Stein [26] first introduced a new method, now referred to as
Stein’s method, to obtain the bound for the departure of the distribution of the sum
of \(n\) terms of a stationary random sequence from a normal distribution. Soon af-

∗Submitted to the editors DATE.
†School of Mathematical Sciences, Zhejiang University, Hangzhou, 310058, China. (suzhonggen@zju.edu.cn)
‡Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University,
Moscow, 119991, Russia
§School of Mathematical Sciences, Zhejiang University, Hangzhou, 310058, China. (12035020@zju.edu.cn)

This manuscript is for review purposes only.
Call \( \{X_i, i \in J\} \) a local dependence structure if for any \( i \in J \), there exist \( A_i, B_i \) and \( C_i \) such that \( i \in A_i \subseteq B_i \subseteq C_i \subseteq J \) and \( X_i \) is independent of \( X_{A_i^c} \). \( X_{A_i^c} \) is independent of \( X_{B_i^c} \) and \( X_{B_i^c} \) is independent of \( X_{C_i^c} \). This local dependence structure has been studied in literature, say Chen and Shao [12], in which it was referred to as Assumption (LD3). The object of our study is defined by

\[
W = \sum_{i \in J} X_i.
\]

We aim to find as good an approximation as possible for \( W \) when \( EW^3 < \infty \). The target distribution is either \( M_1 = B(n, \rho) \ast P(\lambda) \) or \( M_2 = NB(r, \bar{\rho}) \ast P(\lambda) \) with same expectation \( \mu \) and variance \( \sigma^2 \) of \( W \). In fact, there has been a lot of research work on approximating \( W \), say Vellaisamy et al. [29], Upadhye et al. [28].

To state our main results, we need some additional notation. Define

\[
\theta_1 = \frac{\lambda}{n + \lambda/p} q, \quad \theta_2 = \frac{\lambda \bar{q}}{(rq + \lambda \bar{p})}, \quad \Theta_1 = \frac{\theta_1 q}{1 - 2\theta_1 \lambda p}, \quad \Theta_2 = \frac{\theta_2}{1 - 2\theta_2 \lambda q}.
\]

\[
D(W | X_{C_i}) = \left\| \mathcal{L}(W | X_{C_i}) \ast (I_1 - I_0)^2 \right\|_{TV}, \quad \text{for each } i \in J,
\]

where \( I_1 \) and \( I_0 \) denote the degenerate distributions concentrated at 1 and 0 respectively, and \( \mathcal{L}(W | X_{C_i}) \) stands for the conditional distribution of \( W \) conditioned upon \( X_{C_i} \). Denote

\[
\Xi_{i,j} = 9 \sum_{(E)} [(1 - \beta_3)E_{X_1}(E)X_{A_1}(E)X_{B_1}(E)X_{C_1} + \beta_3 E_{X_1}(E)X_{B_1}(E)X_{C_1}] \ast \text{ess } D(W | X_{C_i}),
\]

where \( \beta_3 = -p/q, \beta_1 = \bar{q}, \text{ess } X \) denotes the essential supremum of \( X \) and \( \sum_{(E)} \) denotes the sum over all possible choices of \( E \) in front of \( X \)'s, \((E)X_i \) stands for \( \mathcal{E} | X_i \) or \( \mathcal{E} | X_i \).

We also need to make some basic assumptions about the parameters of the target distributions.

(H1): The triple \( \{n, p, \lambda\} \) satisfies \( \varepsilon_1 := \lambda p/q < p((n + \lambda/p))/2, q = 1 - p \).

(H2): The triple \( \{r, \bar{\rho}, \lambda\} \) satisfies \( \varepsilon_2 := \lambda \bar{q} < (rq + \lambda \bar{p})/2, \bar{q} = 1 - \bar{\rho} \).

Our main results read as follows.

**Theorem 1.1.** With the above notation and assumptions, we have

(i) if \( EW < \text{Var } W \), then

\[
d_{TV}(W, M_2) \leq \Theta_2 \sum_{i \in J} \Xi_{i,2};
\]

(ii) if \( EW > \text{Var } W \), then

\[
d_{TV}(W, M_1) \leq \Theta_1 \left[ \sum_{i \in J} \Xi_{i,1} + \frac{\delta \bar{p}^2}{q} \right].
\]

Note that \( M_1 \) and \( M_2 \) can be represented as sums of independent identical random variables. Indeed, \( M_1 = \sum_{1}^{[J]} \xi_{1} \) with \( \xi_{1} \sim B(n/|J|, p) \ast P(\lambda/|J|) \), and \( M_2 = \sum_{1}^{[J]} \xi_{2} \) with \( \xi_{2} \sim NB(r/|J|, \bar{p}) \ast P(\lambda/|J|) \). A recent result in Bobkov and Ulyanov [7] yields a refined central limit theorem. Denote by \( \Phi \) the standard normal distribution function, \( \phi \) the standard normal density function

\[
l_{3,j} = \frac{1}{\sigma^3} E(M_{j} - EM_{j})^3, \quad \Phi_{3,j}(x) = \Phi(x) - \frac{l_{3,j}}{6}(x^2 - 1)\phi(x), \quad x \in \mathbb{R},
\]

\[
L_{4,j} = \frac{1}{\sigma^4} \sum_{i \in J} E(\xi_{i,j} - EM_{j})^4, \quad V_j = -\sum_{i=1}^{[J]} \sup_{0 \leq t \leq 2n} \frac{Ee^{it\xi_{i,j}}}{1 - \cos t}, \quad j = 1, 2.
\]

This manuscript is for review purposes only.
Applying Theorem 4.1 in [7], we obtain

\begin{equation}
\sup_{k \in \mathbb{Z}} \left| P(M_j \leq k) - \Phi_{3,j}\left(\frac{k + 1/2 - \mu}{\sigma}\right) \right| \leq \frac{\sigma^2 L_{4,j}}{V_j}.
\end{equation}

Combining (1.6) and Theorem 1.1 yields

**Theorem 1.2.** Denote \( K_j = \sup_{i \in J} |A_i||B_i||C_i| \). Assume further that \( D(W \mid X_{C_i}) \leq C/\sigma^2 \). Then for \( j = 1, 2 \), we have

\begin{equation}
\sup_{k \in \mathbb{Z}} \left| P(W \leq k) - \Phi_{3,j}\left(\frac{k + 1/2 - \mu}{\sigma}\right) \right| \leq C \left[ \frac{\Theta_j K_j \sum_{i \in J} E Y_i^3}{\sigma^2} + \frac{\sigma^2 L_{4,j}}{V_j} \right].
\end{equation}

Moreover, if \( \lim_{|J| \to \infty} K_j \) and all \( E Y_i^3 \) are bounded by a constant, then

\begin{equation}
\sup_{k \in \mathbb{Z}} \left| P(W \leq k) - \Phi_{3,j}\left(\frac{k + 1/2 - \mu}{\sigma}\right) \right| \leq C \frac{|J|}{\sigma^4 \wedge (\sigma^2 \mu)}.
\end{equation}

The rest of the paper is organized as follows. Section 2 first obtain the Stein operators for \( M_1 \) and \( M_2 \), and then express them as the perturbation of some classic Stein operators. We also determine the parameters of \( M_1 \) and \( M_2 \) by matching their first three factorial cumulants and those of \( W \). In Section 3, we prove Theorem 1.1. A key ingredient is to control \( |E g(W) - E g(M_j)| \) by using the third-order difference expansion of \( g \). Section 4 is devoted to the application of main results to the runs of independent Bernoulli trials. Sharper upper error bounds are obtained by explicitly computing the factorial cumulants and estimating the measure of smoothness.

**2. Stochastic perturbation tricks.** Let us start with the Stein operator associated with a specific nonnegative integer-valued random variable. It is sometimes easy to figure out explicitly. Indeed, let \( Y \) be a nonnegative random variable that takes values in \( \mathbb{N} \) with \( \mu_k = P(Y = k) > 0 \). Then the Stein operator \( \mathcal{A}_Y \) can be defined as \( \mathcal{A}_Y g(k) = (k + 1)\mu_{k+1}g(k+1)/\mu_k - kg(k), k \in \mathbb{Z}_+, \) namely \( E \mathcal{A}_Y g(Y) = 0 \). Below are three basic examples, which are frequently used throughout the paper.

- Let \( \xi_1 \sim B(n, p) \), \( n \in \mathbb{Z}_+, p \in (0, 1) \). Namely,
  \[ P(\xi_1 = k) = \binom{n}{k} p^k q^{n-k}, k = 0, \ldots, n, q = 1 - p. \]
  The Stein operator for \( \xi_1 \) is \( \mathcal{A}_{\xi_1} g(k) = (n-k)pg(k+1)/q - kg(k), k \in \mathbb{Z}_+ \).

- Let \( \xi_2 \sim NB(r, \bar{p}) \), \( r > 0, \bar{p} \in (0, 1) \). Namely,
  \[ P(\xi_2 = k) = \binom{r+k-1}{k} \bar{p}^k \bar{q}^{r-k}, k \in \mathbb{N}, \bar{q} = 1 - \bar{p}. \]
  The Stein operator for \( \xi_2 \) is \( \mathcal{A}_{\xi_2} g(k) = \bar{q}(r+k)g(k+1) - kg(k), k \in \mathbb{Z}_+ \).

- Fix \( N \in (1, +\infty) \) (not necessarily an integer), \( 0 < p < 1 \). Let \( \xi_3 \sim P(B(N, p) \) the pseudo-binomial distribution (see Čekanavičius and Roos [8], p. 370). Namely,
  \[ P(\xi_3 = k) = \frac{1}{C_{N,p}} \binom{N}{k} p^k q^{N-k}, \quad k = 0, \ldots, [N] \]
  where \( q = 1 - p \), \( C_{N,p} \) is the normalization constant. The Stein operator for \( \xi_3 \) is
  \[ \mathcal{A}_{\xi_3} g(k) = (Np/q - pk/q)g(k+1) - kg(k), \quad k = 1, 2, \ldots, [N]. \]
As the reader notice, these three operators can be expressed in a unified way,

\[(2.1) \quad A_{\xi_i} g(k) = (\alpha_i + \beta_i k)g(k+1) - kg(k), \quad k \in \mathbb{Z}_+, \]

where

\[(2.2) \quad \alpha_1 = np/q, \quad \beta_1 = -p/q; \quad \alpha_2 = r\bar{q}, \quad \beta_2 = \bar{q}; \quad \alpha_3 = Np/q, \quad \beta_3 = -p/q. \]

Next, we turn to the Stein operator for the sum of two independent random variables. The basic tool is the probability generating function approach. Lemma 2.1 can be shown in a completely parallel way to that of Proposition 2.2 in [28].

**Lemma 2.1.** Fix \(n, p, \lambda, r, \bar{p}\). Let \(\xi_1, \eta \sim B(n, p)\) or \(\xi_2 \sim NB(r, \bar{p})\). Then \(M_i := \xi_i + \eta\) has a Stein operator of the form

\[(2.3) \quad A_{M_i} g(k) = (\alpha_i + \beta_i k)g(k+1) - kg(k) - \lambda \beta_i \Delta g(k+1), \quad i = 1, 2 \]

where \(\alpha_i = \alpha_i + \lambda \bar{\beta}_i\) and \(\alpha_i, \beta_i\) were given by (2.2) accordingly.

Comparing (2.1) and (2.3), it is easy to give the Stein operators of \(M_1\) and \(M_2\).

**Proposition 2.1.** Let \(U_1(g)(k) = \lambda \Delta g(k+1)p/q\) and \(U_2(g)(k) = -\lambda \bar{q}\Delta g(k+1)\).

1. Fix \(n \geq 1, 0 < p < 1, \lambda > 0\). Let \(N = n + \lambda/p, \zeta_1 \sim PB(N, p)\), then

\[(2.4) \quad A_{M_1} = A_{\zeta_1} + U_1 \]

2. Fix \(r > 0, 0 < \bar{p} < 1, \lambda > 0\). Let \(\zeta_2 \sim NB(r + \lambda \bar{p}/\bar{q}, \bar{p})\), then

\[(2.5) \quad A_{M_2} = A_{\zeta_2} + U_2. \]

Having the Stein operator \(A_{M_i}\), we take a close look at the properties of solution \(g_f\) to the following equation

\[(2.6) \quad A_{M_i} g(k) = f(k) - \mathbb{E} f(M_i), \quad f \in \mathcal{G}. \]

It follows from Lemma 2.2 in [2], Lemma 9.2.1 in [5] and (57) of [8] that

\[(2.7) \quad \|\Delta g_f^\xi\| \leq \frac{2\|f\|}{np}, \quad \|\Delta g_f^{\bar{\zeta}}\| \leq \frac{2\|f\|}{r\bar{q}} \quad \text{and} \quad \|\Delta g_f^\zeta\| \leq \frac{2\|f\|}{[N]p}, \]

where \(g_f^\xi\) is the solution to the corresponding Stein equation.

In our context, we are mainly concerned with the Stein operators in (2.4) and (2.5). Note that \(\|U_1 g\| \leq \varepsilon_1 \|\Delta g\|\) and \(\|U_2 g\| \leq \varepsilon_2 \|\Delta g\|\) where \(\varepsilon_1 = \lambda p/q\) and \(\varepsilon_2 = \lambda \bar{q}\). The following lemma due to Barbour et al.[4] offers an upper bound for \(\|\Delta g_h^M\|\) which is the solution of (2.6) when we restrict the domain of Stein operator \(A_{M_i}\) to \(\mathcal{H}\), the set of indicator functions.

**Lemma 2.2.** With \((H1)\) and \((H2)\), we have

\[(2.8) \quad \|\Delta g_h^{M_1}\| \leq \frac{1}{n + \lambda p/q - 2\varepsilon_1}, \quad h \in \mathcal{H}; \quad \|\Delta g_h^{M_2}\| \leq \frac{1}{r\bar{q} + \lambda \bar{p} - 2\varepsilon_2}, \quad h \in \mathcal{H}. \]

The proof is similar to Lemma 3.1 of [28] with some minor modifications. The interested reader is referred to [4] for a general framework.

Next turn to the primary object of study, \(d_{TV}(W, M_i), i = 1, 2\). Note that

\[(2.9) \quad d_{TV}(W, M_i) = \sup_{h \in \mathcal{H}} \mathbb{E} \left| A_{M_i} g_h^{M_i} (W) \right|. \]

Applying Lemma 2.2, we have

---

This manuscript is for review purposes only.
PROPOSITION 2.2. (i) Assume \( \textbf{(H1)} \) and \( \left| \mathbb{E} \left( A_{M_1} g_{h_1}^{M_1} \right) (W) \right| \leq \varepsilon \| \Delta g_{h_1}^{M_1} \|\), then

\begin{equation}
\left( 2.8 \right) \quad d_{TV}(W, M_1) \leq \frac{\varepsilon q}{n + \lambda/p |pq - 2\lambda p|}.
\end{equation}

(ii) Assume \( \textbf{(H2)} \) and \( \left| \mathbb{E} \left( A_{M_2} g_{h_2}^{M_2} \right) (W) \right| \leq \varepsilon \| \Delta g_{h_2}^{M_2} \|\), then

\begin{equation}
\left( 2.9 \right) \quad d_{TV}(W, M_2) \leq \frac{\varepsilon}{(rq + \lambda \bar{p}) - 2q}.
\end{equation}

To conclude this section, we would like briefly explain how to decide the parameters \((\lambda, n, p)\) and \((\lambda, r, \bar{p})\). Its basic principle is as follows.

Denote by \( \Gamma_j \) the \( j \)-th factorial cumulant, namely, \( \Gamma_j(\cdot) = \frac{d^j(\log \phi_X(z))}{dz^j} \big|_{z=1} \), where \( \phi_X(z) \) is the generating function of some random variable \( X \). Then by some basic algebra, we obtain

\[ \Gamma_1(M_i) = \frac{\alpha_i}{1 - \beta_i} + \lambda, \quad \Gamma_2(M_i) = \frac{\alpha_i \beta_i}{(1 - \beta_i)^2}, \quad \Gamma_3(M_i) = 2 \frac{\alpha_i \beta_i^2}{(1 - \beta_i)^3}, \quad i = 1, 2. \]

Besides, it is well known that \( \Gamma_1(W) = m_1, \Gamma_2(W) = (m_2 - m_1^2 - m_1), \Gamma_3(W) = m_3 - 3m_1m_2 + 2m_1^3 - 3m_2 + 3m_1^2 + 2m_1 \), where \( m_1 \) be the \( j \)-th origin moment of \( W \).

The basic principle is to guarantee the first three moments of both \( W \) and \( M_i \) match each other, namely, that is

\begin{align}
\left( 2.10 \right) & \quad \Gamma_1(M_i) = \Gamma_1(W), \quad \Gamma_2(M_i) = \Gamma_2(W), \quad \Gamma_3(M_i) = \Gamma_3(W), \quad i = 1, 2.
\end{align}

In particular, the parameters \( \{n, p, \lambda\} \) in \( M_1 \) and \( \{r, \bar{p}, \lambda\} \) in \( M_2 \) are delicately chosen to satisfy the requirements

\begin{align}
\left( 2.11 \right) & \quad n = \left[ \frac{4 \Gamma_2(W)^3}{\Gamma_3(W)^2} \right], \quad p = -\frac{\Gamma_3(W)}{2 \Gamma_2(W)}, \quad \lambda = \mathbb{E}W - np, \quad \delta = \frac{4 \Gamma_2(W)^3}{\Gamma_3(W)^2} - n;
\end{align}

\begin{align}
\left( 2.12 \right) & \quad r = \frac{4 \Gamma_2(W)^3}{\Gamma_3(W)^2}, \quad p = \frac{2 \Gamma_2(W)}{2 \Gamma_2(W) + \Gamma_3(W)}, \quad \lambda = \mu - \frac{rq}{p}.
\end{align}

3. Proofs of Main Results.

Proof of Theorem 1.1. Denote \( \bar{\beta}_1 = 1 - \beta_1 \) and \( \bar{\beta}_2 = 1 - \beta_2 \). Let us begin with the proof of (i). From Lemma 2.1, we have

\[ A_{M_2} g(k) = (a_2 + \beta_2 k) g(k + 1) - k g(k) - \lambda_2 \Delta g(k + 1), \]

where \( a_2 = \alpha_2 + \lambda_2 \bar{\beta}_2 \). Taking expectation with respect to \( W \) yields

\[ \mathbb{E} A_{M_2} g(W) = \mathbb{E}(a_2 + \beta_2 W) g(W + 1) - \mathbb{E}W g(W) - \lambda_2 \mathbb{E} \Delta g(W + 1) \]

\[ = \bar{\beta}_2 \left[ \frac{a}{\bar{\beta}_2} \mathbb{E}[g(W + 1)] - \mathbb{E}[W g(W)] \right] + \beta_2 \mathbb{E}[W \Delta g(W)] - \lambda_2 \mathbb{E} \Delta g(W + 1). \]

Since \( a_2/\bar{\beta}_2 = \mathbb{E}W = \sum_{i \in J} \mathbb{E}X_i \), then

\[ \mathbb{E} [A_{M_2} g(W)] = \bar{\beta}_2 \left[ \sum_{i \in J} \mathbb{E}X_i [g(W + 1)] - \sum_{i \in J} \mathbb{E}[X_i g(W)] \right] \]

\[ + \beta_2 \mathbb{E}[W \Delta g(W)] - \lambda_2 \mathbb{E} \Delta g(W + 1). \]
Set
\[ W_i = W - X_{A_i}, \quad W_i^* = W - X_{B_i}^*, \quad W_i^{**} = W - X_{C_i}^*. \]

Using the independence of \( X_i \) and \( W_i \), we obtain
\[
\mathbb{E}[A_{M_2}g(W)] = \beta_2 \sum_{i \in J} \mathbb{E}X_i \mathbb{E} \left[ \sum_{j=1}^{X_{A_i}} \Delta g(W_i + j) \right] - \lambda \beta_2 \mathbb{E}\Delta g(W + 1)
\]
\[
- \tilde{\beta}_2 \sum_{i \in J} \mathbb{E} \left[ X_i \sum_{j=1}^{X_{A_i}} \Delta g(W_i + j) \right] + \beta_2 \sum_{i \in J} \mathbb{E}[X_i \Delta g(W)].
\]

(3.1)

A crucial observation is
\[
\beta_2 \left\{ \sum_{i \in J} \mathbb{E}X_i \mathbb{E} \left[ \sum_{j=1}^{X_{A_i}} 1 \right] - \sum_{i \in J} \mathbb{E} \left[ X_i \sum_{j=1}^{X_{A_i}} 1 \right] \right\} + \beta_2 \sum_{i \in J} \mathbb{E}X_i
\]
\[
= \sum_{i \in J} \mathbb{E}X_i - \tilde{\beta}_2 \sum_{i \in J, j \in A_i} (\mathbb{E}X_iX_j - \mathbb{E}X_i\mathbb{E}X_j) = \mathbb{E}W - \tilde{\beta}_2 \text{Var} W.
\]

Note that for \( M_2, \mathbb{E}W = \alpha_2/\beta_2 + \lambda, \text{Var} W = \alpha_2/\beta_2^2 + \lambda \). So (3.2) is equal to say
\[
\lambda \beta_2 = \beta_2 \left\{ \sum_{i \in J} \mathbb{E}X_i \mathbb{E} \left[ X_{A_i} \right] - \sum_{i \in J} \mathbb{E} \left[ X_i (X_{A_i} - 1) \right] \right\} + \beta_2 \sum_{i \in J} \mathbb{E}X_i.
\]

Substituting (3.3) into (3.1), noting the independence of \( X_{A_i}^* \) and \( W_i^* \) and the following elementary second-order difference identity,
\[
\Delta g(W_i^* + m) - \Delta g(W_i^* + 1) = \sum_{t=1}^{X_{B_i} - X_{A_i}^* - m} \Delta^2 g(W_i^* + t), \quad m \in \mathbb{Z},
\]
we obtain
\[
\mathbb{E}[A_{M_2}g(W)] = - \sum_{i \in J} \left\{ \beta_2 \left[ \mathbb{E}X_i \mathbb{E}X_{A_i}^* - \mathbb{E}X_i (X_{A_i}^* - 1) \right] + \beta_2 \mathbb{E}X_i \right\}
\]
\[
\times \mathbb{E} \sum_{t=1}^{X_{B_i}^*} \Delta^2 g(W_i^* + t) + \beta_2 \sum_{i \in J} \mathbb{E}X_i \left[ X_i \sum_{t=1}^{X_{B_i}^*} \Delta^2 g(W_i^* + t) \right]
\]
\[
+ \tilde{\beta}_2 \sum_{i \in J} \mathbb{E}X_i \mathbb{E} \left[ \sum_{j=1}^{X_{A_i}^*} \sum_{t=1}^{X_{B_i}^* + j} \Delta^2 g(W_i^* + t) \right]
\]
\[
- \tilde{\beta}_2 \sum_{i \in J} \mathbb{E}X_i \left[ X_i \sum_{j=1}^{X_{A_i}^* - 1} \sum_{t=1}^{X_{B_i}^* + j - 1} \Delta^2 g(W_i^* + t) \right] = \Upsilon_2.
\]

A nice coincidence is that \( \Upsilon_2 \) above with \( \Delta^2 g(W_i^* + t) \) replaced by 1 vanishes. For clarity, we state it as a lemma.

**Lemma 3.1.** Let
\[
\Upsilon_{20} := \sum_{i \in J} \left\{ \beta_2 \mathbb{E}X_i \mathbb{E}(X_{A_i}^* (X_{B_i}^* - X_{A_i}^* - j) - \mathbb{E}X_i (X_{A_i}^* - 1) (X_{B_i}^* - X_{A_i}^* - j + 1)) + \beta_2 \mathbb{E}X_i (X_{B_i}^* - 1) - \beta_2 \left[ \mathbb{E}X_i \mathbb{E}X_{A_i}^* - \mathbb{E}X_i (X_{A_i}^* - 1) \right] + \beta_2 \mathbb{E}X_i \right\} \mathbb{E}X_{B_i}^*,
\]
then \( \Upsilon_{20} = 0 \).

**Proof.** It is easy to see
\[
\Upsilon_{20} = \sum_{i \in J} \left\{ \left[ \mathbb{E}X_i \mathbb{E}X_{A_i}^* X_{B_i}^* - \mathbb{E}X_i X_{A_i}^* X_{B_i}^* - \mathbb{E}X_i \mathbb{E}X_{A_i}^* \mathbb{E}X_{B_i}^* + \mathbb{E}X_i X_{A_i}^* \mathbb{E}X_{B_i}^* \right]
\]
\[
+ \frac{1}{2} \left[ \mathbb{E}X_i (X_{A_i}^*)^2 - \mathbb{E}X_i (X_{A_i}^*)^2 + \mathbb{E}X_i X_{A_i}^* - \mathbb{E}X_i \mathbb{E}X_{A_i}^* \right] \tilde{\beta}_2
\]
\[
+ \mathbb{E}X_i X_{B_i}^* - \mathbb{E}X_i \mathbb{E}X_{B_i}^* - \mathbb{E}X_i \right\}.
\]

(3.5)
By local independence, we have the following two equality:

\[(3.6) \quad \sum_{i \in J} (E_i X_{B_i} - E_i E X_{B_i}) = \sum_{i \in J} (E_i X_{A_i} - E_i E X_{A_i}) = m_2 - m_1^*;\]

\[(3.7) \quad E_i E X_{A_i} W - E_i E X_{A_i} W - E_i E X_{A_i} E W + E_i E X_{A_i} E W\]

Applying (3.6) and (3.7), (3.5) becomes

\[(3.10) \quad (\hat{\beta}_2/2 + 1)(m_2 - m_1^*) - m_1 + \hat{\beta}_2(m_2 - m_1^*)m_1.\]

In addition, noting \(W = X_{A_i} + X_{A_i}^*\), it follows

\[(3.9) \quad E_i E (X_{A_i}^*)^2 - E_i E (X_{A_i}^*)^2 + 2(E_i E X_{A_i} W - E_i E X_{A_i} W)\]

Combining equations (3.8) and (3.9), we end up with that (3.5) becomes

\[(3.10) \quad (\hat{\beta}_2/2 + 1)(m_2 - m_1^*) - m_1 + \hat{\beta}_2(m_2 - m_1^*)m_1 - \hat{\beta}_2(m_3 - m_1m_2)/2.\]

Solving (2.10) yields

\[(3.11) \quad \hat{\beta}_2 = \frac{m_3 - 3m_1m_2 - 3m_2 + 2m_1^* + 3m_1^* + 2m_1}{m_3 - 3m_1m_2 - m_2 + 2m_1^* + m_1^*}.\]

Substituting (3.11) into (3.10), we conclude that \(\Upsilon_0 = 0\), as desired.

Proceed with the proof of Theorem 1.1. Applying Lemma 3.1, noting the independence of \(X_{B_i}\) and \(W_{i}^*\) and the following third-order difference identity

\[\Delta^2 g (W_i^* + m) - \Delta^2 g (W_i^{**} + 1) = \sum_{k=1}^{\xi_1} X_{C_i} \sum_{k=1}^{m-1} \Delta^3 g (W_i^* + k), \quad m \in \mathbb{Z},\]

some simple algebra lead to

\[|E.A_M g(W)| \leq \sum_{(E)} \left[ \sup_{k \in \mathbb{Z}} E[\Delta^3 g(W + k) \mid X_{C_i}] \right] \times \sum_{k=1}^{\xi_1} \left[ \delta \beta E_i E X_{A_i} X_{B_i} (E) X_{B_i} (E) X_{C_i} + \beta \delta E_i E X_{A_i} X_{B_i} (E) X_{C_i} \right].\]

It is easy to verify that \(E[\Delta^3 g(W + k) \mid X_{C_i}] \leq ||\Delta g|| D(W \mid X_{C_i})\), where \(D(W \mid X_{C_i})\) was defined in (1.3). Using the definition of \(\xi_{i,2}\) and some apparent comparison, the last inequation becomes \(|E.A_M g(W)| \leq ||\Delta g|| \sum_{i \in J} \xi_{i,2}\), which, together with (2.9) implies (1.4) holds.

Turn to the proof of (ii). Following the proof of (i), the R.H.S of (3.2) is equal to \(E W - \hat{\beta}_1 \text{Var} W = \lambda \beta_1 + \delta \hat{\beta}_1 / \hat{\beta}_1\). Thus we immediately have

\[(3.12) \quad E.A_M g(W) \leq \Upsilon_1 + \frac{\delta \hat{\beta}_1^2}{\beta_1} E \Delta g(W + 1),\]
where Υ_1 is obtained by replacing β_2 with β_1 from Υ_2. Noting (2.10), an analog of the proof of (1.4) yields

\[ |E_A M_t g(W)| \leq ||\Delta g|| \left( \sum_{i \in J} \Xi_{i,1} + \frac{\delta p^2}{q} \right). \]

Finally, using (2.8) and (3.13), the proof is completed.

4. Applications.

4.1. (k_1, k_2)-runs. Fix k_1, k_2 \geq 1, let m = k_1 + k_2 - 1 and J = \{1, 2, \cdots, Nm\}. Suppose that \xi_1, \xi_2, \cdots, \xi_{Nm} are a sequence of independent Bernoulli random variables with \( P(\xi_j = 1) = p \) for \( j \in J \). We say that a \((k_1, k_2)\)-event if there occurs \( k_1 \) consecutive 0's followed by \( k_2 \) consecutive 1's. To avoid edge effects, we identify \( i + Nm_j \) as \( i \in J, j \in Z \). Define

\[ X_j = (1 - \xi_j) \cdots (1 - \xi_{j+k_1-1}) \xi_{j+k_1} \cdots \xi_{j+k_1+k_2-1}. \]

It is easy to see that the local dependence structure is satisfied with

\[ A_i = \{ j \in J : |j - i| \leq m \}; B_i = \{ j : |j - i| \leq 2m \}; C_i = \{ j : |j - i| \leq 3m \}. \]

Denote by \( W_N = \sum_{j=1}^{Nm} X_j \) the number of occurrences of \((k_1, k_2)\)-events in \( Nm \) trials, which is often called a modified binomial distribution. For ease of notation, we henceforth suppress the dependence of quantities on \( N \) when it is clear from the context. Let \( b = E X_j = (1 - p)^{k_1} p^{k_2} \). It follows from some simple but tedious calculations,

\[ \Gamma_1(W) = Nmb, \quad \Gamma_2(W) = -Nm(2m + 1)b^2, \quad \Gamma_3(W) = Nm[9m^2 + 9m + 2]b^3. \]

Since \( \Gamma_2(W) \) is negative, we use \( M_1 = B(n, p) * \mathcal{P}(\lambda) \) to approximate \( W \) by matching the first three moments. Following the argument in Section 3, we get

\[ n = \left\lceil \frac{Nm(2m + 1)^3}{(2m + 1)^2 + \frac{m(m+1)}{2}} \right\rceil, \quad p = \left[ 2m + 1 + \frac{m(m+1)}{2(2m + 1)} \right] b, \quad \lambda = Nmb - np. \]

**Theorem 4.1.** Let \( m \geq 1 \) and \( \{p_i, i \in J\} \) are identical to \( p \), assume

\[ b := (1 - p)^{k_1} p^{k_2} < \frac{2(2m + 1)}{2(2m + 1)^2 + 3m(m+1)} := c_m. \]

Then we have

\[ d_{TV}(W, M_1) \leq O(N^{-1}); \]

and

\[ \sup_{k \in Z} \left| P(W \leq k) - \Phi_3\left( \frac{k + 1/2 - \mu}{\sigma} \right) \right| \leq O(N^{-1}). \]

We claim that (4.3) is a significant improvement over the bound given in the literature. In fact, Theorem 5.2 of Barbour [6] and Example 2.1 of Kumar [18] only attain \( O(N^{-1/2}) \), where they used the two-parameter compound Poisson distribution and pseudo-binomial distribution to approximate \( W \) under \( m = 1 \), respectively; and Theorem 3.1 of [27] by using Poisson approximation, which is of order \( O(N^{-1/2}) \) under \( m \geq 2 \).
Define $T_i = \sum_{j=0}^{2m+1} X_j$ for $i = 1, 2, \ldots, N$. Let $J' = \{1, 2, \ldots, N\}$. Hence $T_1, T_2, \ldots, T_N$ are $1$-dependent random variables, which means the local dependence structure is satisfied with

$$A_i' = \{j : |j - i| \leq 1\} \cap J', \quad B_i' = \{j : |j - i| \leq 2\} \cap J', \quad C_i' = \{j : |j - i| \leq 3\} \cap J'.$$

Define $Z := \{T_{2k}, 1 \leq k \leq [N/2]\}, N(i, Z) = |T_{J' \cap C_i}' \setminus Z|$ where $|A|$ stands for the cardinality of $A$ and the specific value of $N(i, Z)$ is $O(N)$. Set $F(i, Z) = \{1, \ldots, N(i, Z)\}$. Note $T_{J' \cap C_i}' \setminus Z$ can be written as $\{T^Z_{ij}, j \in F(i, Z), l_1 < l_2 < \cdots < l_{N(i, Z)}\}$ where $T^Z_{ij}$ satisfies $\mathcal{L}(T^Z_{ij}) = \mathcal{L}(T_{ij} \mid Z, X_{C_i})$. Hence these ($T^Z_{ij}, j \in F(i, Z)$) are independent of each other. Fixing $Z = z \in \{0, 1\}^{N/2}$, $\mathcal{L}(W_i^* \mid Z = z, X_{C_i})$ can be represented as the sum of independent random variables $\{T^Z_{ij}, j \in F(i, z)\}$. Thus it follows

$$((W_i)^* \mid Z, X_{C_i}) \overset{d}{=} \left(\sum_{j \in J'} T^Z_{ij} \mid Z, X_{C_i}\right) \overset{d}{=} \left(\sum_{j \in J} T^Z_{ij}\right).$$

Noting that $D(W_i^* \mid T_{C_i}') = D(W \mid X_{C_i})$, it suffices to control the $D(W_i^* \mid T_{C_i}')$. We define

$$D(X) = \|\mathcal{L}(X) \ast (I_1 - I_0)\|_{TV} = \sum_{k=0}^{\infty} |P(X = k + 1) - P(X = k)|,$$

$$V_2 = \sum_{j \in F(i, Z)} \left[1/2 \wedge \left(1 - D(T^Z_{ij})/2\right)\right], \quad v_2^* = \max_{j \in F(i, Z)} \left\{1/2 \wedge \left(1 - D(T^Z_{ij})/2\right)\right\}.$$

According to (5.12) of [24] and (4.9) of [3], we have

$$D(W_i^* \mid T_{C_i}') \leq \mathbb{E} [D(W_i^*) \mid Z, T_{C_i}'] \leq 4\mathbb{E} \left\{1 \wedge \frac{2}{(V_2 - 4v_2^*)^+}\right\} = O(N^{-1}).$$

**Proof of Theorem 4.1.** As for (4.3), in this case we get

$$\lim_{N \to \infty} \theta_1 = \frac{m(m+1)b}{2(2m+1) - [2(2m+1)^2 + m(m+1)]b}$$

from (1.2) and (4.1). We claim $\lim_{N \to \infty} \theta_1 < 1/2$ under the assumption (4.2). In fact,

$$\lim_{N \to \infty} \theta_1 = \frac{m(m+1)b}{2(2m+1) - [2(2m+1)^2 + m(m+1)]b}.$$

Notice that $f(x) = m(m+1)x \left\{2(2m+1) - [2(2m+1)^2 + m(m+1)]x\right\}^{-1}$ is monotonic increasing and $f(c_m) = 1/2$. So, $\lim_{N \to \infty} \theta_1 = f(b) < f(c_m) = 1/2$.

Observeably, $\delta \in [0, 1]$ and (4.5) implies $\sum_{i \in J} \Xi_{i, 1} \leq O(N)$. Then substituting (4.5) into Theorem 1.1 directly yields (4.3). The proof of (4.4) is an immediate result of (4.3) and Theorem 1.2.

**4.2. k-runs.** In this subsection, we turn to another special case with $k_1 = 0$ and $k_2 = k > 0$. It is easy to see that the local dependence structure is satisfied with

$$A_i = \{j : |j - i| \leq k\} \cap J, \quad B_i = \{j : |j - i| \leq 2k\} \cap J, \quad C_i = \{j : |j - i| \leq 3k\} \cap J.$$

Denote $W_{N, k} = \sum_{j=1}^{N} X_j$, which is often termed as $k$-runs. As before, we suppress the dependence of quantities on $N$ and $k$ in the sequel. We focus on the case that $p$ and $k$ depend on $N$ with $k \log p \rightarrow -\infty$. The expectation of $W$ is $Np^k$, so is $\Gamma_1(W)$.
We omit the cumbersome calculations and give asymptotic expressions of the second and third factorial cumulants:

\(\Gamma_2(W) = \left[\frac{2p}{1-p} + O(p^k)\right]Np^k, \quad \Gamma_3(W) = \left[\frac{6p^2}{(1-p)^2} + O(p^k)\right]Np^k.\)

From (2.12) and (4.6), it is easy to obtain

\(r = O(Np^k), \quad \bar{p} = O(1), \quad \lambda = O(Np^k)\)

Moreover, by (4.6), and the definition (1.2) of \(\theta_2\), \(\lim_{N \to \infty} 2(1-p)\theta_2/p = 1\). Hence we assume further \(p < 1/2\) to guarantee that \(\theta_2 < 1/2\). Our result reads.

**Theorem 4.2.** Assume \(p < 1/2\), \(N > 10k\) and \(k \log p \to -\infty\) as \(N \to \infty\). Then

\[d_{TV}(W, M_2) \leq \left(9 \wedge \frac{95.22(1-2p)}{(N-10k+8)p^k(1-p)^2}\right) \left(2k-1\right) \left(4k-3\right)(6k-5)p^3.\]

Assume \(p\) and \(k\) are numeric constants (do not depend on \(N\)), the gap between \(W\) and \(M_2\) using Theorem 4.2 is actually \(O(N^{-1})\), which is an improvement of Corollary 1.1 of [30] that attains \(O(N^{-1/2})\). Moreover, in the case that \(p\) and \(k\) depend on \(N\), Corollary 1.1 of Wang and Xia [30] showed

\[d_{TV}(W, NB(r', p')) \leq 4.5(4k-3)(2k-1)p^2 \left(2 \wedge \frac{4.6 \sqrt{N}}{\sqrt{(N-4k+2)p^k(1-p)^3}}\right),\]

where \(r' = \Gamma_1(W)^2/\Gamma_2(W), \quad p' = \Gamma_1(W)/[\Gamma_1(W) + \Gamma_2(W)].\) The error upper bound (4.8) is in general better than (4.9).

Before proving Theorem 4.2, we introduce some additional notation and give a lemma. Let \(N_1 = \lfloor (N-k)/2 \rfloor, \quad W_1 = \sum_{j=1}^{N_1} X_j, \quad W_2 = \sum_{j=N_1+k}^{N} X_j, \quad \) Denote \(C_{1,i} = C_i \cap \{1, \ldots, N_1\}, \quad C_{2,i} = C_i \cap \{N_1+k, \ldots, N\}.\) Without loss of generality, we assume \(\{N_1+1, \ldots, N_1+k-1\} \subseteq C_1\) and \(|C_{1,i}| = 2.5k-3, |C_{2,i}| = 2.5k-2.6\).

**Lemma 4.1.** For \(N > 10k\), we have

\[D(W_j|X_{C_{j,i}}) \leq 1 \wedge \frac{2.3}{\sqrt{(0.5N-5k+4)(1-p)^3} p^k}, \quad j = 1, 2.\]

**Proof.** We start with the proof of (4.10) with \(j = 1\). Note that \(X_{C_{1,i}}\) contains \((2.5k-3)\) consecutive elements from \(Y_{N_1+1-2.5k+2}\) to \(Y_{N_1}\) which determined by \(\{[N_1]-2.5k+2, [N_1]-2.5k+3, \ldots, [N_1]+k-1\}\).

Denote \(G = \{1, 2, \ldots, N_1\}, \quad H = \{j : |N_1| - 2.5k + 2 \leq j \leq |N_1| + k - 1\} \cap J.\) Define \(\gamma_l := P(\xi_l = 1 \mid X_{C_{1,i}}), l \in J.\) It is easy to find for \(l \notin H, \gamma_l = p.\) Using Lemma 2.1 of [30], we obtain

\[D(W_1|X_{C_{1,i}}) \leq 1 \wedge \frac{2.3}{\sqrt{(0.5N-5k+4)(1-p)^3} p^k}\]

By the same token, we complete the proof of (4.10) with \(j = 2.\)

**Proof of Theorem 4.2.** We start with the calculation of \(D(W|X_{C_{1,i}}).\) Notice that \(D(W|X_{C_{1,i}}) = D(W_1 + W_2|X_{C_{1,i}})\) and \(\mathcal{L}(W_1 + W_2|X_{C_{1,i}}) = \mathcal{L}(W_1|X_{C_{1,i}}) * \mathcal{L}(W_2|X_{C_{2,i}}).\) So by (1.2) of [24] and Lemma 4.1 we obtain

\[D(W|X_{C_{1,i}}) \leq D(W_1|X_{C_{1,i}})D(W_2|X_{C_{2,i}}) \leq 1 \wedge \frac{10.58}{(N-10k+8)p^k(1-p)^3}.\]

This manuscript is for review purposes only.
Next, using (1.2), (4.7), the definition of $\Xi_{i,2}'$ and observing $\bar{q} \leq 3p\bar{p}/2$ when $p < 1/2$, we get
\[
\sum_{i=1}^{n} \Theta_{2}\Xi_{i,2}' \leq 9(2k-1)(4k-3)(6k-5)(1-p)p^3 \frac{(1-p)p^3}{1-2p}.
\]
Finally, we use (i) of Theorem 1.1 to conclude the proof.

REFERENCES

[1] R. Arratia, L. Goldstein, and L. Gordon, Two moments suffice for Poisson approximations: the Chen-Stein method, The Annals of Probability, (1989), pp. 9–25.
[2] A. Barbour, Asymptotic expansions in the Poisson limit theorem, The Annals of Probability, (1987), pp. 748–766.
[3] A. D. Barbour and V. Čekanavičius, Total variation asymptotics for sums of independent integer random variables, The Annals of Probability, 30 (2002), pp. 509–545.
[4] A. D. Barbour, V. Čekanavičius, and A. Xia, On Stein’s method and perturbations, arXiv preprint math/0702008, (2007).
[5] A. D. Barbour, L. Holst, and S. Janson, Poisson approximation, vol. 2, The Clarendon Press Oxford University Press, 1992.
[6] A. D. Barbour and A. Xia, Poisson perturbations, ESAIM: Probability and Statistics, 3 (1999), pp. 131–150.
[7] S. G. Bobkov and V. V. Ulyanov, The chebyshev–edgeworth correction in the central limit theorem for integer-valued independent summands, Theory of Probability & Its Applications, 66 (2022), pp. 537–549.
[8] V. Čekanavičius and B. Roos, Two-parametric compound binomial approximations, Lithuanian Mathematical Journal, 44 (2004), pp. 354–373.
[9] S. Chatterjee, J. Fulman, and A. Röllin, Exponential approximation by Stein’s method and spectral graph theory, ALEA Lat. Am. J. Probab. Math. Stat, 8 (2011), pp. 197–223.
[10] L. H. Chen, Poisson approximation for dependent trials, The Annals of Probability, 3 (1975), pp. 534–545.
[11] L. H. Chen, L. Goldstein, and Q.-M. Shao, Normal approximation by Stein’s method, vol. 2, Springer, 2011.
[12] L. H. Chen and Q.-M. Shao, Normal approximation under local dependence, The Annals of Probability, 32 (2004), pp. 1985–2028.
[13] F. Daly, C. Lefèvre, and S. Utev, Stein’s method and stochastic orderings, Advances in Applied Probability, 44 (2012), pp. 343–372.
[14] C. Ley, G. Reiner, Y. Swan, et al., Approximate computation of expectations: the canonical Stein operator, Citeseer, 2014.
[15] C. Ley and Y. Swan, Local pinsher inequalities via Stein’s discrete density approach, IEEE Transactions on Information Theory, 59 (2013), pp. 5584–5591.
[16] C. Ley and Y. Swan, Stein’s density approach and information inequalities, Electronic Communications in Probability, 18 (2013), pp. 1–14.
[17] H. M. Luk, Stein’s method for the gamma distribution and related statistical applications, PhD thesis, University of Southern California, 1994.
[18] B. Mann, Stein’s method for $\chi^2$ of a multinomial, Unpublished manuscript, (1997).
[19] I. Nourdin and G. Peccati, Normal approximations with Malliavin calculus: from Stein’s method to universality, no. 192, Cambridge University Press, 2012.
[20] E. A. Peköz, Stein’s method for geometric approximation, Journal of applied probability, 33 (1996), pp. 707–713.
[21] E. A. Peköz and A. Röllin, New rates for exponential approximation and the theorems of Rényi and Yaglom, The Annals of Probability, 39 (2011), pp. 587–608.
[22] E. A. Peköz, A. Röllin, and N. Ross, Total variation error bounds for geometric approximation, Bernoulli, 19 (2013), pp. 610–632.
[23] E. A. Peköz, A. Röllin, and N. Ross, Generalized gamma approximation with rates for urns, walks and trees, The Annals of Probability, 44 (2016), pp. 1776–1816.
[24] A. Röllin, Symmetric and centered binomial approximation of sums of locally dependent random variables, Electronic Journal of Probability, 13 (2008), pp. 756–776.
[25] N. Ross, Fundamentals of Stein’s method, Probability Surveys, 8 (2011), pp. 210–293.
[26] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, in Proceedings of the sixth Berkeley symposium on mathemat-
ical statistics and probability, volume 2: Probability theory, vol. 6, University of California Press, 1972, pp. 583–603.

[27] N. Upadhye and A. Kumar, *Pseudo-binomial approximation to $(k_1, k_2)$-runs*, Statistics & Probability Letters, 141 (2018), pp. 19–30.

[28] N. S. Upadhye, V. Čekanavicius, and P. Vellaisamy, *On Stein operators for discrete approximations*, Bernoulli, 23 (2017), pp. 2828–2859.

[29] P. Vellaisamy, N. Upadhye, and V. Čekanavicius, *On negative binomial approximation*, Theory of Probability & Its Applications, 57 (2013), pp. 97–109.

[30] X. Wang and A. Xia, *On negative binomial approximation to $k$-runs*, Journal of applied probability, 45 (2008), pp. 456–471.