PROBABILITY
AND
MATHEMATICAL STATISTICS
Published online 18.2.2025
doi:10.37190/0208-4147.00208

ON THE POISSON APPROXIMATION OF RANDOM DIAGONAL SUMS OF BERNOULLI MATRICES

BY

BERO ROOS (TRIER)

Abstract. We use the Stein–Chen method to prove new explicit inequalities for the total variation, Wasserstein and local distances between the distribution of a random diagonal sum of a Bernoulli matrix and a Poisson distribution. Approximation results using a finite signed measure of higher order are given as well. Some of our total variation bounds improve existing results in the literature.

2020 Mathematics Subject Classification: Primary 60F05; Secondary 62E17.

Key words and phrases: approximation error, Hoeffding statistic, Poisson approximation.

1. INTRODUCTION AND REVIEW OF SOME KNOWN RESULTS

Let $n \in \mathbb{N} = \{1, 2, 3, \dots\}$ with $n \neq 1$ and $X_{j,r}$ for $j, r \in \underline{n} := \{1, \dots, n\}$ be independent random variables with Bernoulli distributions $P^{X_{j,r}} = \mathrm{Be}(p_{j,r})$ with success probabilities $p_{j,r} \in [0, 1]$. For $k, \ell \in \mathbb{N}$ with $k \leqslant \ell$, we let $\underline{\ell}^k_{\neq} = \{(j_1, \dots, j_k) \mid j_1, \dots, j_k \in \underline{\ell} \text{ pairwise distinct}\}$, where we also write $(j(1), \dots, j(k))$ for $(j_1, \dots, j_k)$. In particular, $\underline{n}^n_{\neq}$ is the set of all permutations of $\underline{n}$. Let $\pi = (\pi(1), \dots, \pi(n))$ be a random permutation uniformly distributed on $\underline{n}^n_{\neq}$. We assume that all the $X_{j,r}$'s and π are independent. Let us call $(X_{1,\pi(1)}, X_{2,\pi(2)}, \dots, X_{n,\pi(n)})$ a *random* (generalized) *diagonal* of the Bernoulli matrix $X = (X_{j,r})$ and

$$S_n = \sum_{j=1}^{n} X_{j,\pi(j)}$$

the corresponding *random diagonal sum*. If the $X_{j,r}$'s are constants, S_n is sometimes also called the Hoeffding permutation statistic (see Barbour et al. [7]) or Hoeffding statistic (see Adamczak et al. [1]). Further let

$$\overline{p}_{j,\cdot} = \frac{1}{n} \sum_{r=1}^{n} p_{j,r} \quad \text{for } j \in \underline{n}, \qquad \overline{p}_{\cdot,r} = \frac{1}{n} \sum_{j=1}^{n} p_{j,r} \quad \text{for } r \in \underline{n},$$

$$\lambda = \mathrm{E}\, S_n = \frac{1}{n}\sum_{j=1}^{n}\sum_{r=1}^{n} p_{j,r} = \sum_{j=1}^{n} \overline{p}_{j,\cdot} = \sum_{r=1}^{n} \overline{p}_{\cdot,r} > 0, \qquad \overline{p} = \frac{\lambda}{n}.$$

In this paper, we consider the approximation of the distribution P^{S_n} of S_n by a Poisson distribution and also by a signed measure of higher order. To measure the accuracy, we use the total variation, Wasserstein and local norms, the definitions of which require some notation. Let $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$. For two sets A and B, let B^A be the set of functions from A to B. For $f \in \mathbb{R}^{\mathbb{Z}_+}$, let $\Delta f \in \mathbb{R}^{\mathbb{Z}_+}$ be defined by $\Delta f(m) = f(m+1) - f(m)$ for $m \in \mathbb{Z}_+$ and set $\|f\|_\infty = \sup_{m\in\mathbb{Z}_+}|f(m)|$ and $\|f\|_1 = \sum_{m\in\mathbb{Z}_+}|f(m)|$. Let $\mathcal{F}_{\mathrm{W}} = \{f \in \mathbb{R}^{\mathbb{Z}_+} \mid \|\Delta f\|_\infty \leqslant 1\}$. Let $\mathcal{M}$ be the vector space of all finite signed measures on the power set of $\mathbb{Z}_+$ and set

$$\mathcal{M}' = \Big\{Q \in \mathcal{M} \,\Big|\, Q(\mathbb{Z}_+) = 0, \sum_{m=0}^{\infty} m|f_Q(m)| < \infty\Big\},$$

where $f_Q \in \mathbb{R}^{\mathbb{Z}_+}$ with $f_Q(m) = Q(\{m\})$ for $m \in \mathbb{Z}_+$ is the counting density of Q. Let

$$\|Q\|_{\mathrm{TV}} = \|f_Q\|_1, \qquad \|Q\|_{\mathrm{loc}} = \|f_Q\|_\infty,$$
$$d_{\mathrm{TV}}(Q_1, Q_2) = \sup_{A\subseteq\mathbb{Z}_+} |Q_1(A) - Q_2(A)|$$

be the *total variation norm* and the *local norm* of $Q \in \mathcal{M}$, and the *total variation distance* between $Q_1, Q_2 \in \mathcal{M}$. The *Wasserstein norm* $\|Q\|_{\mathrm{W}}$ (sometimes also called the Fortet–Mourier norm or Kantorovich norm) of $Q \in \mathcal{M}'$ is defined by

$$\|Q\|_{\mathrm{W}} = \sum_{m=0}^{\infty} |Q(\underline{m} \cup \{0\})|.$$

It is well-known that, for $Q_1, Q_2 \in \mathcal{M}$ and $Q \in \mathcal{M}'$,

$$d_{\mathrm{TV}}(Q_1, Q_2) = \frac{1}{2}\|Q_1 - Q_2\|_{\mathrm{TV}} \qquad \text{if } Q_1(\mathbb{Z}_+) = Q_2(\mathbb{Z}_+),$$

$$\|Q\|_{\mathrm{W}} = \sup_{f\in\mathcal{F}_{\mathrm{W}}}\Big|\int f\,\mathrm{d}Q\Big|, \qquad \|(\delta_1 - \delta_0) * Q\|_{\mathrm{W}} = \|Q\|_{\mathrm{TV}}, \tag{1.1}$$

where $\delta_m \in \mathcal{M}$ is the Dirac measure at $m \in \mathbb{Z}_+$, and $*$ denotes convolution. Let $\mathrm{Po}(t)$ be the Poisson distribution with mean $t \in (0, \infty)$.

The literature contains some explicit inequalities for the total variation distance between P^{S_n} and the Poisson distribution with the same mean. We are not aware of any published explicit bounds concerning the Wasserstein and local distances in this context. However, the local distance was considered by Barbour et al. [7, Theorem 2.10], but they used a translated Poisson distribution to improve the accuracy of approximation and the estimate given there is not explicit, since it contains an O-term.

Let us discuss some total variation bounds. Chen [10, Theorem 2.1] adapted Stein's [22] method to prove that, for $n \geqslant 5$,

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant 7.875 \min\left\{1, \frac{1}{\sqrt{\lambda}}\right\}\Big(\sum_{j=1}^{n} \overline{p}_{j,\cdot}^2 + \sum_{r=1}^{n} \overline{p}_{\cdot,r}^2\Big), \tag{1.2}$$

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant 22.625 \frac{1}{\lambda}\Big(\sum_{j=1}^{n} \overline{p}_{j,\cdot}^2 + \sum_{r=1}^{n} \overline{p}_{\cdot,r}^2\Big). \tag{1.3}$$

Barbour and Holst [4, Theorem 7.1] and Barbour et al. [5, Theorem 4.A, p. 78] used the Stein–Chen method and coupling to show refinements of (1.2) and (1.3) when the matrix $(p_{j,r})$ is in $\{0,1\}^{\underline{n}\times\underline{n}}$. We only state the result in [5], which improves on that in [4]. It says that

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\left(\frac{n-2}{n}(\lambda - \operatorname{Var} S_n) + \frac{2\lambda^2}{n}\right) \tag{1.4}$$

$$\leqslant \frac{3}{2}\,\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\left(\sum_{j=1}^{n} \overline{p}_{j,\cdot}^2 + \sum_{r=1}^{n} \overline{p}_{\cdot,r}^2 - \frac{2\lambda}{3n}\right). \tag{1.5}$$

As mentioned in [5, Remark 4.1.3], the proof of [5, Theorem 4.A] can be adapted to prove (1.3) in the general case $(p_{j,r}) \in [0,1]^{\underline{n}\times\underline{n}}$ with constant $3/2$ in place of 22.625. Our results below imply that (1.4) also holds for $(p_{j,r}) \in [0,1]^{\underline{n}\times\underline{n}}$; on the other hand, (1.5) has to be slightly adapted (see Remark 2.4).

Under the additional assumption that $p_{j,r} = \mathbf{1}_{[-1,a_j]}(r)$ for all $j, r \in \underline{n}$ with $a_1, \ldots, a_n \in \underline{n} \cup \{-1, 0\}$, (1.4) can be improved to

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant (1-\mathrm{e}^{-\lambda})\left(1 - \frac{\operatorname{Var} S_n}{\lambda}\right). \tag{1.6}$$

Here, for a set A, $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise. We note that (1.6) does not follow directly from (1.4) or its proof, but it can be shown again using the Stein–Chen method [5, p. 80]. However, more can be said:

REMARK 1.1. Let $(p_{j,r}) \in [0,1]^{\underline{n}\times\underline{n}}$ have (weakly) decreasing columns, that is, $p_{j,r} \geqslant p_{j+1,r}$ for all $j \in \underline{n-1}$ and $r \in \underline{n}$. Then the following hold:

(1) The probability generating function $\psi_{S_n}(z) = \sum_{k=0}^{n} P(S_n = k) z^k$ $(z \in \mathbb{C})$ of P^{S_n} has only real roots.

(2) The distribution P^{S_n} is a Bernoulli convolution, that is, it is the distribution of the sum of n independent Bernoulli random variables.

(3) Inequality (1.6) holds and

$$\frac{1}{14} \min\left\{1, \frac{1}{\lambda}\right\}(\lambda - \operatorname{Var} S_n) \leqslant d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)). \tag{1.7}$$

Proof. Since $\psi_{S_n}(z)$ is equal to the permanent of the matrix $(1+p_{j,r}(z-1)) \in \mathbb{C}^{\underline{n}\times\underline{n}}$ divided by $n!$, (1) follows directly from the monotone column permanent theorem (see Brändén et al. [8]). It is well-known that (1) implies (2) (see e.g., Pitman [14, Proposition 1]). Part (3) is a consequence of (2), [3, Theorem 1] and [5, Remark 3.2.2]. ■

It is clear that the statement of Remark 1.1 remains valid if it is assumed that the matrix $(p_{j,r})$ has decreasing rows, that is, its transpose has decreasing columns. The same holds for increasing columns (or rows). In particular, if $p_{j,1} = \cdots = p_{j,n}$ for all $j \in \underline{n}$, then (1.6) and (1.7) hold and the distribution of S_n is a Bernoulli convolution with success probabilities $p_{1,1}, \dots, p_{n,1}$.

Unfortunately, an inequality of the form

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant C\Big(1 - \frac{\operatorname{Var} S_n}{\lambda}\Big)$$

with an absolute constant $C \in (0,\infty)$ cannot generally hold. Indeed, if $(p_{j,r}) \in [0,1]^{\underline{n}\times\underline{n}}$ is the identity matrix, then $\operatorname{Var} S_n = 1 = \lambda$ and $P^{S_n} \neq \mathrm{Po}(\lambda)$ (see, e.g., [5, Example 4.2.1]). Therefore, to get a general upper bound of $d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda))$, one has to enlarge $1 - \frac{\operatorname{Var} S_n}{\lambda}$ somewhat. In this context, the following formula for the variance of S_n is useful:

$$\operatorname{Var} S_n = \lambda - \sum_{j=1}^{n} \overline{p}_{j,\cdot}^2 - \gamma, \tag{1.8}$$

where

$$\gamma = \frac{1}{2n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} (p_{j,r} - p_{j,s})(p_{k,r} - p_{k,s}). \tag{1.9}$$

For a proof, see Section 5. Further, we have

$$\begin{aligned} \operatorname{Var} S_n &= \lambda - \frac{n}{n-1}\Big(\sum_{j\in\underline{n}} \overline{p}_{j,\cdot}^2 + \sum_{r\in\underline{n}} \overline{p}_{\cdot,r}^2 - \frac{1}{n^2}\sum_{(j,r)\in\underline{n}^2} p_{j,r}^2 - \frac{\lambda^2}{n}\Big) \\ &= \lambda - \frac{1}{n-1}\sum_{j\in\underline{n}}\sum_{r\in\underline{n}} p_{j,r}\Big(\overline{p}_{j,\cdot} + \overline{p}_{\cdot,r} - \frac{p_{j,r}}{n} - \overline{p}\Big), \end{aligned} \tag{1.10}$$

from which it easily follows that

$$\begin{aligned} &\lambda(1-m) \leqslant \operatorname{Var} S_n \leqslant \lambda, \qquad \text{where} \\ &m = \frac{n}{n-1}\max_{(j,r)\in\underline{n}^2}\Big(\overline{p}_{j,\cdot} + \overline{p}_{\cdot,r} - \frac{p_{j,r}}{n} - \overline{p}\Big) \leqslant \min\Big\{\lambda, \frac{2n}{n-1}\Big\}. \end{aligned} \tag{1.11}$$

Identity (1.10) and the second inequality in (1.11) were proved in [5, Proposition 4.1.1] in the case $(p_{j,r}) \in \{0,1\}^{\underline{n}\times\underline{n}}$ with $\frac{1}{n^2}\sum_{(j,r)\in\underline{n}^2} p_{j,r}^2$ replaced by $\overline{p}$. The proof in the general case is analogous. We note that the first inequality in (1.11) can only be useful if $m < 1$.

The rest of the paper is structured as follows. Sections 2 and 3 contain our main approximation inequalities for the total variation, Wasserstein and local distances. The proofs are given in Sections 4 and 5. In what follows, the assumptions of Section 1 are supposed to hold unless otherwise stated.

2. RESULTS FOR THE TOTAL VARIATION DISTANCE

To state our first result, we need further quantities related to γ. Let

$$\begin{aligned}
\gamma' &= \frac{2}{n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} (p_{j,r}-p_{j,s})_+(p_{k,s}-p_{k,r})_+,\\
\gamma'' &= \frac{1}{2n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} |p_{j,r}-p_{j,s}|\,|p_{k,r}-p_{k,s}|,\\
\gamma''' &= \frac{1}{4n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} (|p_{j,r}-p_{j,s}|-|p_{k,r}-p_{k,s}|)^2.
\end{aligned}$$

Here and below, we set $x_+ = \max\{0, x\}$ for $x \in \mathbb{R}$. Our first result improves on (1.4).

THEOREM 2.1. *We have*

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant \frac{1-\mathrm{e}^{-\lambda}}{\lambda}(\lambda - \operatorname{Var} S_n + \gamma') \tag{2.1}$$

$$= \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{j=1}^{n} \overline{p}_{j,\cdot}^2 + \gamma''\Big) \tag{2.2}$$

$$= \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\frac{1}{n} \sum_{(j,r)\in\underline{n}^2} p_{j,r}^2 - \gamma'''\Big). \tag{2.3}$$

Let us compare the inequality in Theorem 2.1 with (1.4). We first note that the upper bound in Theorem 2.1 is always smaller than or equal to the right-hand side of (1.4). However, both upper bounds are of the same order. This follows from the next lemma.

LEMMA 2.1. *Let $A = \lambda - \operatorname{Var} S_n + \gamma'$ and $B = \frac{n-2}{n}(\lambda - \operatorname{Var} S_n) + \frac{2\lambda^2}{n}$. Then*

$$A \leqslant B \leqslant \Big(3 - \frac{2}{n}\Big)A. \tag{2.4}$$

REMARK 2.1. In view of (2.3), we see that the bound in Theorem 2.1 is always at most $1 - \mathrm{e}^{-\lambda}$.

EXAMPLE 2.1. The right-hand side in (1.4) is not always bounded by 1 and the second inequality in (2.4) may be an equality. To show this, let $p_{j,r} = p \in [0,1]$ for all $j, r \in \underline{n}$. Then P^{S_n} is the binomial distribution with parameters n and p. Further, $\lambda = np$ and inequality (1.4) states that

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(np)) \leqslant \left(3 - \frac{2}{n}\right)(1 - \mathrm{e}^{-np})p. \tag{2.5}$$

On the other hand, Theorem 2.1 gives $d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(np)) \leqslant (1 - \mathrm{e}^{-np})p$. In particular, in the second inequality in (2.4) equality holds. If $p > 1/3$, then the upper bound in (2.5) is greater than 1 for n large enough.

REMARK 2.2. If the matrix $(p_{j,r})$ has decreasing rows, then $\gamma' = 0$. In this case, the inequalities in Theorem 2.1 and (1.6) are identical. We recall that (1.4) does not imply (1.6).

REMARK 2.3. The distribution of S_n remains unchanged if we replace the matrix $(p_{j,r})$ with its transpose. In the case $(p_{j,r}) \in \{0,1\}^{\underline{n} \times \underline{n}}$, both upper bounds in (1.4) and Theorem 2.1 remain unchanged as well; see Lemma 2.2(1) below. However, if we consider the general case $(p_{j,r}) \in [0,1]^{\underline{n} \times \underline{n}}$, the bound in Theorem 2.1 can indeed change. So, in this case, we possibly get two different inequalities, the better of which should be used. For instance, let $n = 2$ and $(p_{j,r}) = \left(\begin{smallmatrix} 1 & 1/4 \\ 3/4 & 1/2 \end{smallmatrix}\right)$. Then we have $\gamma' = 0$, but the analogous term for the transpose of $(p_{j,r})$ is equal to

$$\frac{2}{n^2(n-1)} \sum_{(j,k) \in \underline{n}^2_{\neq}} \sum_{(r,s) \in \underline{n}^2_{\neq}} (p_{j,r} - p_{k,r})_+ (p_{k,s} - p_{j,s})_+ = \frac{1}{16}.$$

Let us now collect some properties of γ'.

LEMMA 2.2.

(1) *We have*

$$\gamma' \leqslant \frac{2}{n^2(n-1)} \sum_{(j,k) \in \underline{n}^2_{\neq}} \sum_{(r,s) \in \underline{n}^2_{\neq}} p_{j,r}(1 - p_{j,s}) p_{k,s}(1 - p_{k,r}). \tag{2.6}$$

Here, equality holds if and only if for all $j \in \underline{n}$ and $(r,s) \in \underline{n}^2_{\neq}$ such that $(p_{j,r}, p_{j,s}) \in (0,1)^2$, we have $p_{k,r} = p_{k,s} \in \{0,1\}$ for all $k \in \underline{n} \setminus \{j\}$. In particular, equality holds if in every row of the matrix $(p_{j,r})$ there is at most one entry in $(0,1)$.

The right-hand side of (2.6) does not change if we replace the matrix $(p_{j,r})$ with its transpose.

(2) *We have*

$$\gamma' \leqslant \min\left\{\mathrm{Var}\, S_n - \frac{1}{n} \sum_{(j,r) \in \underline{n}^2} p_{j,r}(1 - p_{j,r}), \frac{2}{n}(\mathrm{Var}\, S_n - \lambda + \lambda^2)\right\}. \tag{2.7}$$

(3) *If* $(p_{j,r}) \in \{0,1\}^{\underline{n}\times\underline{n}}$, *then*

$$\gamma' = \frac{2}{n-1} \sum_{(r,s)\in\underline{n}^2_{\neq}} \Big(\overline{p}_{\bullet,r} - \frac{1}{n}\sum_{j\in\underline{n}} p_{j,r}p_{j,s}\Big)\Big(\overline{p}_{\bullet,s} - \frac{1}{n}\sum_{j\in\underline{n}} p_{j,r}p_{j,s}\Big).$$

REMARK 2.4. We note that (2.1) together with the second entry in the right-hand side of (2.7) can be used to show (1.4) in the general case $p_{j,r} \in [0,1]^{\underline{n}\times\underline{n}}$. Further, an analogue of (1.5) can be shown: in view of (1.4), (1.10) and (1.11), we see that

$$\frac{3}{2}\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{j=1}^{n}\overline{p}_{j,\bullet}^{2} + \sum_{r=1}^{n}\overline{p}_{\bullet,r}^{2} - \frac{2}{3n^2}\sum_{(j,r)\in\underline{n}^2} p_{j,r}^2\Big)$$

is larger than or equal to the right-hand side of (1.4).

EXAMPLE 2.2. The left entry L and the right entry R in the minimum term in (2.7) are not comparable in general:

If $p_{j,r} = 1$ for all $j, r \in \underline{n}$, then $\lambda = n$ and $\operatorname{Var} S_n = 0$, that is $L = 0 < 2(n-1) = R$.

On the other hand, if $(p_{j,r}) \in [0,1]^{\underline{n}\times\underline{n}}$ is the identity matrix, then $\operatorname{Var} S_n = 1 = \lambda$; see e.g. [5, Example 4.2.1]. In this case, we have $R = 2/n \leqslant 1 = L$.

REMARK 2.5. Let us discuss conditions for the smallness of the total variation distance between P^{S_n} and $\mathrm{Po}(\lambda)$. For this, we consider a triangular scheme, where n, all $X_{j,r}$ and $p_{j,r}$, π, and in turn λ, γ', and S_n depend on a further variable $k \in \mathbb{N}$, which we let go to infinity later. In order to simplify the notation, we do not indicate the dependence on k. Barbour et al. [5, Corollary 4.A.1] showed that, under the assumptions

$$(2.8) \qquad (p_{j,r}) \in \{0,1\}^{n\times n}, \qquad \lim_{k\to\infty}\frac{\lambda}{n} = 0 \quad \text{and} \quad \liminf_{k\to\infty}\lambda > 0,$$

we have

$$(2.9) \qquad d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \to 0 \quad \text{if and only if} \quad \frac{\operatorname{Var} S_n}{\lambda} \to 1.$$

However, since (1.4) also holds in the case $(p_{j,r}) \in [0,1]^{n\times n}$ according to Remark 2.4, the condition $(p_{j,r}) \in \{0,1\}^{n\times n}$ in (2.8) can be dropped. Further, a slight refinement of the proof in [5] shows that the condition $\liminf_{k\to\infty}\lambda > 0$ in (2.8) can be dropped as well. In fact, sufficiency follows from (1.4), and Theorem 3.A in [5] shows that if $d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \to 0$, then $\min\{1,\lambda\}\big(1 - \frac{\operatorname{Var} S_n}{\lambda}\big) \to 0$. The first and third inequalities in (1.11) imply that $\lambda - \operatorname{Var} S_n \leqslant \lambda^2$, leading to

$$\Big(1 - \frac{\operatorname{Var} S_n}{\lambda}\Big)^2 \leqslant \min\{1,\lambda\}\Big(1 - \frac{\operatorname{Var} S_n}{\lambda}\Big) \to 0.$$

So, the condition $\liminf_{k\to\infty} \lambda > 0$ is not needed here. If we use (2.1), instead of (1.4), in the sufficiency part, we obtain the following corollary without using the assumptions in (2.8).

COROLLARY 2.1. *Consider the triangular scheme above. If $\gamma'/\lambda \to 0$, then* (2.9) *holds.*

REMARK 2.6. In the situation of Corollary 2.1, it follows from (2.7) and (1.11) that $\gamma'/\lambda \leqslant 2\lambda/n$. This implies that the assumption $\gamma'/\lambda \to 0$ is weaker than $\lambda/n \to 0$. In particular, if the matrices $(p_{j,r})$ have decreasing rows, we have $\gamma' = 0$ for all k, so that (2.9) holds.

In the next theorem, we consider the approximation of P^{S_n} by the finite signed measure Q_2 concentrated on $\mathbb{Z}_+$ with

$$\begin{aligned} Q_2(A) &= \mathrm{Po}(\lambda)(A) - \frac{1}{2}(\lambda - \mathrm{Var}\, S_n)((\delta_1-\delta_0)^{*2} * \mathrm{Po}(\lambda))(A) \\ &= \sum_{k\in A} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\Big(1 - \frac{1}{2\lambda^2}(\lambda - \mathrm{Var}\, S_n)(\lambda^2 - 2k\lambda + k(k-1))\Big) \end{aligned} \tag{2.10}$$

for $A \subseteq \mathbb{Z}_+$. Comparable approximations in the case of independent summands were considered by Kerstan [13], Chen [9], Shorgin [21], Barbour and Hall [3], Roos [16] and others.

THEOREM 2.2. *Let $n \geqslant 4$. Then*

$$d_{\mathrm{TV}}(P^{S_n}, Q_2) \leqslant \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\varepsilon, \tag{2.11}$$

where

$$\begin{aligned} \varepsilon &= \min\Big\{1, \sqrt{\frac{2}{\lambda \mathrm{e}}}\Big\}(\varepsilon_1+\varepsilon_2) + \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\varepsilon_3, \\ \varepsilon_1 &= \frac{2}{n}\sum_{j=1}^n\sum_{r=1}^n \overline{p}_{j,\cdot}\, p_{j,r}|\lambda - \lambda'_{j,r}|, \\ \varepsilon_2 &= \frac{1}{n^2(n-1)}\sum_{(j,k)\in\underline{n}^2_{\neq}}\ \sum_{(r,s)\in\underline{n}^2_{\neq}} |p_{j,r}-p_{j,s}|\,|p_{k,r}-p_{k,s}|\,|\lambda - \lambda''_{j,k,r,s}|, \\ \varepsilon_3 &= \frac{2n^2}{(n-2)^2}\Big(\frac{n-2}{n-1}\sum_{j=1}^n \overline{p}^2_{j,\cdot} + \sqrt{\frac{n-1}{n-3}}\gamma''\Big)^2 \end{aligned}$$

and, for $(j,k),(r,s) \in \underline{n}^2_{\neq}$,

$$\lambda'_{j,r} = \frac{1}{n-1}\sum_{u\in\underline{n}\setminus\{j\}}\ \sum_{v\in\underline{n}\setminus\{r\}} p_{u,v} = \frac{1}{n-1}\big(n(\lambda - \overline{p}_{j,\cdot} - \overline{p}_{\cdot,r}) + p_{j,r}\big),$$

$$\lambda''_{j,k,r,s} = \frac{1}{n-2} \sum_{u \in \underline{n} \setminus \{j,k\}} \sum_{v \in \underline{n} \setminus \{r,s\}} p_{u,v}$$
$$= \frac{1}{n-2} \big(n(\lambda - \overline{p}_{j,\cdot} - \overline{p}_{k,\cdot} - \overline{p}_{\cdot,r} - \overline{p}_{\cdot,s}) + p_{j,r} + p_{k,r} + p_{j,s} + p_{k,s} \big).$$

In what follows, let

$$\varepsilon_0 = \frac{1 - \mathrm{e}^{-\lambda}}{\lambda} (\lambda - \operatorname{Var} S_n + \gamma') = \frac{1 - \mathrm{e}^{-\lambda}}{\lambda} \Big(\sum_{j=1}^{n} \overline{p}_{j,\cdot}^2 + \gamma'' \Big)$$

be the upper bound in Theorem 2.1.

REMARK 2.7. Let the assumptions of Theorem 2.2 hold. Then

$$\varepsilon_1 \leqslant 2(\overline{p}_{\max,\cdot} + \overline{p}_{\cdot,\max}) \sum_{j=1}^{n} \overline{p}_{j,\cdot}^2, \tag{2.12}$$

$$\varepsilon_2 \leqslant 4(\overline{p}_{\max,\cdot} + \overline{p}_{\cdot,\max})\gamma'', \qquad \Big(\frac{1 - \mathrm{e}^{-\lambda}}{\lambda} \Big)^2 \varepsilon_3 \leqslant \frac{2n^2(n-1)}{(n-2)^2(n-3)} \varepsilon_0^2, \tag{2.13}$$

where $\overline{p}_{\max,\cdot} = \max_{j \in \underline{n}} \overline{p}_{j,\cdot}$ and $\overline{p}_{\cdot,\max} = \max_{r \in \underline{n}} \overline{p}_{\cdot,r}$. These inequalities together with (2.11) lead to the somewhat crude bound

$$d_{\mathrm{TV}}(P^{S_n}, Q_2) \leqslant 4\varepsilon_0 \bigg(\min\Big\{1, \sqrt{\frac{2}{\lambda \mathrm{e}}}\Big\} (\overline{p}_{\max,\cdot} + \overline{p}_{\cdot,\max}) + \frac{n^2(n-1)}{2(n-2)^2(n-3)} \varepsilon_0 \bigg),$$

which shows that the right-hand side of (2.11) is of a better order than ε_0. The inequalities above are easily proved by using the fact that, for $(j,k), (r,s) \in \underline{n}^2_{\neq}$,

$$\lambda'_{j,r} - \lambda = \frac{\lambda'_{j,r}}{n} - \Big(\overline{p}_{\cdot,r} + \overline{p}_{j,\cdot} - \frac{p_{j,r}}{n} \Big),$$
$$\lambda''_{j,k,r,s} - \lambda = \frac{2}{n} \lambda''_{j,k,r,s} - \Big(\overline{p}_{j,\cdot} + \overline{p}_{k,\cdot} + \frac{1}{n} \sum_{u \in \underline{n} \setminus \{j,k\}} p_{u,r} + \frac{1}{n} \sum_{u \in \underline{n} \setminus \{j,k\}} p_{u,s} \Big),$$

giving

$$|\lambda - \lambda'_{j,r}| \leqslant \max\Big\{ \frac{\lambda'_{j,r}}{n},\ \overline{p}_{\cdot,r} + \overline{p}_{j,\cdot} - \frac{p_{j,r}}{n} \Big\} \leqslant \overline{p}_{\max,\cdot} + \overline{p}_{\cdot,\max}$$

and

$$|\lambda - \lambda''_{j,k,r,s}| \leqslant \max\Big\{ \frac{2}{n} \lambda''_{j,k,r,s},\ \overline{p}_{j,\cdot} + \overline{p}_{k,\cdot} + \frac{1}{n} \sum_{u \in \underline{n} \setminus \{j,k\}} p_{u,r} + \frac{1}{n} \sum_{u \in \underline{n} \setminus \{j,k\}} p_{u,s} \Big\}$$
$$\leqslant 2(\overline{p}_{\max,\cdot} + \overline{p}_{\cdot,\max}).$$

In what follows, C denotes a positive absolute constant, the value of which may change from line to line. Further, let $\lfloor x \rfloor$ for $x \in \mathbb{R}$ be the largest integer $\leqslant x$.

COROLLARY 2.2. *Let the assumptions of Theorem 2.2 hold. Then*

$$\Big|d_{\mathrm{TV}}(P^{S_n},\mathrm{Po}(\lambda))-\frac{\lambda-\mathrm{Var}\,S_n}{\sqrt{2\pi\mathrm{e}}\,\lambda}\Big| \leqslant C\min\Big\{1,\frac{1}{\lambda}\Big\}\Big(\frac{\lambda-\mathrm{Var}\,S_n}{\lambda}+\varepsilon\Big), \tag{2.14}$$

$$d_{\mathrm{TV}}(P^{S_n},\mathrm{Po}(\lambda)) \leqslant \min\Big\{1,\frac{3}{4\lambda\mathrm{e}}\Big\}(\lambda-\mathrm{Var}\,S_n)+\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\varepsilon. \tag{2.15}$$

The constant $\frac{3}{4\mathrm{e}}$ *in* (2.15) *is the best possible.*

REMARK 2.8. Let the assumptions of Theorem 2.2 hold.

(1) Let the matrix $(p_{j,r})$ have decreasing rows. Then P^{S_n} is a Bernoulli convolution (see Remark 1.1). Let $\theta=\frac{\lambda-\mathrm{Var}\,S_n}{\lambda}$. Since $\gamma'=0$, we have

$$\begin{aligned}\varepsilon_0&=(1-\mathrm{e}^{-\lambda})\theta=\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{j=1}^n\overline{p}_{j,\cdot}^2+\gamma''\Big),\\ \theta&=\frac{1}{\lambda}\Big(\sum_{j=1}^n\overline{p}_{j,\cdot}^2+\gamma''\Big).\end{aligned} \tag{2.16}$$

Further, (2.12) and (2.13) imply that

$$\begin{aligned}&\varepsilon_1\leqslant 4\sum_{j=1}^n\overline{p}_{j,\cdot}^2,\qquad \varepsilon_2\leqslant 8\gamma'',\qquad \varepsilon_3\leqslant C(\lambda\theta)^2,\\ &\varepsilon\leqslant C\lambda\theta\min\Big\{1,\frac{1}{\sqrt{\lambda}}+\theta\Big\},\end{aligned} \tag{2.17}$$

and (2.14) yields

$$\Big|d_{\mathrm{TV}}(P^{S_n},\mathrm{Po}(\lambda))-\frac{\theta}{\sqrt{2\pi\mathrm{e}}}\Big|\leqslant C\theta\min\Big\{1,\frac{1}{\sqrt{\lambda}}+\theta\Big\}.$$

The latter also follows from Roos [16, formula (32)] and is a generalization, resp. refinement, of the results of Prokhorov [15, Theorem 2] (see also Barbour et al. [5, p. 2]) and Deheuvels and Pfeifer [11, Theorem 1.2].

(2) Inequality (2.15) is a refinement of (2.1). Further, the optimality of the constant $\frac{3}{4\mathrm{e}}$ on the right hand side of (2.15) can be verified by using the special example of $p_{j,r}=p$ for all $j,r\in\underline{n}$. Here, S_n has a binomial distribution with parameters n and p. Further, we have $\lambda=np$ and $\mathrm{Var}\,S_n=np(1-p)$. In view of (2.15) and the definition of ε, we see that

$$d_{\mathrm{TV}}(P^{S_n},\mathrm{Po}(\lambda))\leqslant\frac{3}{4\mathrm{e}}p+Cp^2. \tag{2.18}$$

From [18, Theorem 2], it follows that, in the present situation,

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \sim \frac{3}{4\mathrm{e}} p$$

as $p \to 0$ and $np \to 1$, where we use the triangular scheme as in Remark 2.5. Here $\sim$ means that the quotient of both sides tends to 1. This shows the optimality of the constant $\frac{3}{4\mathrm{e}}$ in (2.18) and (2.15).

We now consider the setting in the general matching problem (see [5, pp. 82–83]).

EXAMPLE 2.3. (1) Let $m \in \underline{n}$, $a_1, \dots, a_m, b_1, \dots, b_m \in \underline{n} \cup \{0\}$ with $n = \sum_{\ell=1}^{m} a_\ell = \sum_{\ell=1}^{m} b_\ell$. For $\ell \in \{0, \dots, m\}$, let $A_\ell = \sum_{j=1}^{\ell} a_j$ and $B_\ell = \sum_{j=1}^{\ell} b_j$, where $\sum_{j=1}^{0} a_j = 0$ denotes the empty sum. Define $p_{j,r} = 1$ for all $(j, r) \in \bigcup_{\ell=1}^{m} ((A_{\ell-1}, A_\ell] \cap \underline{n}) \times ((B_{\ell-1}, B_\ell] \cap \underline{n})$ and $p_{j,r} = 0$ otherwise. Then

$$\begin{aligned}
\overline{p}_{j,\cdot} &= \frac{1}{n} \sum_{r=1}^{n} p_{j,r} = \frac{b_\ell}{n} \quad \text{if } \ell \in \underline{m} \text{ and } j \in (A_{\ell-1}, A_\ell] \cap \underline{n}, \\
\overline{p}_{\cdot,r} &= \frac{1}{n} \sum_{j=1}^{n} p_{j,r} = \frac{a_\ell}{n} \quad \text{if } \ell \in \underline{m} \text{ and } r \in (B_{\ell-1}, B_\ell] \cap \underline{n}, \\
\lambda &= \frac{1}{n} \sum_{\ell=1}^{m} a_\ell b_\ell, \qquad \sum_{j=1}^{n} \overline{p}_{j,\cdot}^2 = \frac{1}{n^2} \sum_{\ell=1}^{m} a_\ell b_\ell^2, \qquad \sum_{r=1}^{n} \overline{p}_{\cdot,r}^2 = \frac{1}{n^2} \sum_{\ell=1}^{m} a_\ell^2 b_\ell, \\
\operatorname{Var} S_n &= \lambda - \frac{1}{n-1} \left(\frac{1}{n} \sum_{\ell=1}^{m} a_\ell b_\ell (a_\ell + b_\ell) - \lambda^2 - \lambda \right), \\
\gamma' &= \frac{2}{n^2(n-1)} \sum_{\ell \in \underline{m}^2_{\neq}} a_{\ell(1)} b_{\ell(1)} a_{\ell(2)} b_{\ell(2)}.
\end{aligned}$$

Therefore, (2.1) gives

$$\begin{aligned}
&d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \\
&\quad = \frac{1 - \mathrm{e}^{-\lambda}}{\lambda} \left(\frac{1}{n-1} \left(\frac{1}{n} \sum_{\ell=1}^{m} a_\ell b_\ell (a_\ell + b_\ell) - \lambda^2 - \lambda \right) + \gamma' \right),
\end{aligned} \tag{2.19}$$

slightly improving (1.4) in this case, which reads as follows:

$$\begin{aligned}
&d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \\
&\quad \leqslant (1 - \mathrm{e}^{-\lambda}) \left(\frac{n-2}{\lambda n(n-1)} \left(\frac{1}{n} \sum_{\ell=1}^{m} a_\ell b_\ell (a_\ell + b_\ell) - \lambda^2 - \lambda \right) + \frac{2\lambda}{n} \right)
\end{aligned}$$

(see also [5, p. 83]).

(2) Let the assumptions in (1) hold and $a_1 = \cdots = a_m = b_1 = \cdots = b_m = d$. Then $\overline{p}_{j,\cdot} = d/n$ for $j \in \underline{n}$, $\overline{p}_{\cdot,r} = d/n$ for $r \in \underline{n}$, $md = n$, $\lambda = d$, and $\operatorname{Var} S_n = d - \frac{d(d-1)}{n-1}$. Further, (2.19) reduces to

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant (1-\mathrm{e}^{-d})\Big(\frac{3d-1}{n} - \frac{(d-1)(2d-1)}{n(n-1)}\Big), \tag{2.20}$$

which coincides with an inequality in [5, p. 82] proved by different arguments.

In the case $n \geqslant 4$, it follows from (2.15), (2.12) and (2.13) that

$$\begin{aligned} d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) &\leqslant \frac{3}{4\mathrm{e}}\,\frac{d-1}{n-1} + 4\varepsilon_0\Big(\frac{2^{3/2}\sqrt{d}}{n\sqrt{\mathrm{e}}} + \frac{n^2(n-1)}{2(n-2)^2(n-3)}\varepsilon_0\Big) \\ &\leqslant \frac{3}{4\mathrm{e}}\,\frac{d-1}{n-1} + C\Big(\frac{d}{n}\Big)^2, \end{aligned} \tag{2.21}$$

where ε_0 is the upper bound in (2.20). We note that (2.21) is better than (2.20) if d/n is small. In particular, if $d = 1$, we obtain a bound of order $(d/n)^2$.

Further, (2.12)–(2.14) imply that

$$\Big| d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) - \frac{d-1}{\sqrt{2\pi\mathrm{e}}(n-1)} \Big| \leqslant C\Big(\frac{d-1}{d(n-1)} + \Big(\frac{d}{n}\Big)^2\Big).$$

In particular, it follows that

$$d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \sim \frac{d}{\sqrt{2\pi\mathrm{e}}\,n} \quad \text{as } d \to \infty,\ n \to \infty, \text{ and } d/n \to 0.$$

3. RESULTS FOR WASSERSTEIN AND LOCAL DISTANCES

Let the one-point concentration of a $\mathbb{Z}_+$-valued random variable Z be defined by

$$c(Z) = \sup_{m \in \mathbb{Z}_+} P(Z = m). \tag{3.1}$$

THEOREM 3.1. *For* $(j,k) \in \underline{n}^2_{\neq}$, *let*

$$S_n^{(j)} = \sum_{i \in \underline{n}\setminus\{j\}} X_{i,\pi(i)} \quad \text{and} \quad S_n^{(j,k)} = \sum_{i \in \underline{n}\setminus\{j,k\}} X_{i,\pi(i)}.$$

Set $\eta_1 = \max_{j\in\underline{n}} c(S_n^{(j)})$ *and* $\eta_2 = \max_{(j,k)\in\underline{n}^2_{\neq}} c(S_n^{(j,k)})$. *Then*

$$\|P^{S_n} - \mathrm{Po}(\lambda)\|_{\mathrm{W}} \leqslant \min\Big\{1, \frac{4}{3}\sqrt{\frac{2}{\lambda\mathrm{e}}}\Big\}\Big(\sum_{j=1}^n \overline{p}_{j,\cdot}^2 + \gamma''\Big), \tag{3.2}$$

$$\|P^{S_n} - \mathrm{Po}(\lambda)\|_{\mathrm{loc}} \leqslant 2\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\eta_1 \sum_{j=1}^n \overline{p}_{j,\cdot}^2 + \eta_2\gamma''\Big). \tag{3.3}$$

REMARK 3.1. Let the assumptions of Theorem 3.1 hold.

(1) The right-hand side of (3.2) is equal to

$$\min\Big\{1, \frac{4}{3}\sqrt{\frac{2}{\lambda \mathrm{e}}}\Big\}(\lambda - \mathrm{Var}\, S_n + \gamma') = \min\Big\{1, \frac{4}{3}\sqrt{\frac{2}{\lambda \mathrm{e}}}\Big\}\Big(\frac{1}{n}\sum_{(j,r)\in \underline{n}^2} p_{j,r}^2 - \gamma'''\Big)$$

(see Theorem 2.1). A similar statement holds with respect to (3.3) after estimating η_1 and η_2 by their maximum. In comparison to (3.2), the bound in Theorem 2.1 contains an additional factor $\lambda^{-1/2}$ as expected; for instance, see [18, Theorem 2] in the case of Bernoulli convolutions.

(2) For practical applications of (3.3), it is necessary to use an explicit upper bound of η_1 and η_2, for which we refer the reader to Roos [20, Theorem 1.7]. In many cases, the bounds are of order $\lambda^{-1/2}$. This would lead to an upper bound of $\|P^{S_n} - \mathrm{Po}(\lambda)\|_{\mathrm{loc}}$ with an additional factor $\lambda^{-1/2}$ in comparison with the bound in Theorem 2.1. However, a detailed discussion is omitted.

(3) For $(j,k) \in \underline{n}^2_{\neq}$, we have

$$c(S_n^{(j)}) \leqslant 2c(S_n^{(j,k)}),$$

since $P(S_n^{(j)} = m) \leqslant P(S_n^{(j,k)} \in \{m-1, m\}) \leqslant 2c(S_n^{(j,k)})$ for $m \in \mathbb{Z}_+$. This implies that $\eta_1 \leqslant 2\eta_2$, which can be used to estimate the right-hand side of (3.3).

For a finite set B, let $|B|$ denote its cardinality.

THEOREM 3.2. *Under the assumptions of Theorem 2.2, we have*

$$\|P^{S_n} - Q_2\|_{\mathrm{W}} \leqslant \min\Big\{1, \frac{4}{3}\sqrt{\frac{2}{\lambda \mathrm{e}}}\Big\}\varepsilon, \tag{3.4}$$

$$\|P^{S_n} - Q_2\|_{\mathrm{loc}} \leqslant 2\Big(\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big)^2(\varepsilon_1 + \varepsilon_2 + \kappa\varepsilon_3), \tag{3.5}$$

where, for $(j,k), (r,s) \in \underline{n}^2_{\neq}$,

$$(T'_{j,r})^B := \sum_{i\in \underline{n}\setminus(\{j\}\cup B)} X_{i,\pi_{j,r}(i)} \qquad \text{for } B \subseteq \underline{n}\setminus\{j\},$$
$$(T''_{j,k,r,s})^B := \sum_{i\in \underline{n}\setminus(\{j,k\}\cup B)} X_{i,\pi_{j,k,r,s}(i)} \qquad \text{for } B \subseteq \underline{n}\setminus\{j,k\},$$

the random variable $\pi_{j,r}$*, resp.* $\pi_{j,k,r,s}$*, is independent of* X *and has uniform distribution on* $(\underline{n}\setminus\{r\})^{\underline{n}\setminus\{j\}}_{\neq}$*, resp.* $(\underline{n}\setminus\{r,s\})^{\underline{n}\setminus\{j,k\}}_{\neq}$*, and*

$$\kappa = \max\Big\{\max_{B\subseteq \underline{n}\setminus\{j\}:\, 1\leqslant |B|\leqslant 2} c((T'_{j,r})^B), \max_{B\subseteq \underline{n}\setminus\{j,k\}:\, 1\leqslant |B|\leqslant 2} c((T''_{j,k,r,s})^B)\Big\}.$$

COROLLARY 3.1. *Let the assumptions of Theorem 3.2 hold. Then*

$$\|P^{S_n}-\mathrm{Po}(\lambda)\|_{\mathrm{W}} \leqslant \min\Big\{1, \frac{1}{\sqrt{2\lambda \mathrm{e}}}\Big\}(\lambda-\mathrm{Var}\, S_n)+\min\Big\{1, \frac{4}{3}\sqrt{\frac{2}{\lambda \mathrm{e}}}\Big\}\varepsilon, \tag{3.6}$$

$$\Big|\|P^{S_n}-\mathrm{Po}(\lambda)\|_{\mathrm{W}}-\frac{\lambda-\mathrm{Var}\, S_n}{\sqrt{2\pi\lambda}}\Big| \leqslant C\min\Big\{1, \frac{1}{\sqrt{\lambda}}\Big\}\Big(\frac{\lambda-\mathrm{Var}\, S_n}{\sqrt{\lambda}}+\varepsilon\Big), \tag{3.7}$$

$$\begin{aligned}\|P^{S_n}-\mathrm{Po}(\lambda)\|_{\mathrm{loc}} &\leqslant \min\Big\{1, \frac{1}{2}\Big(\frac{3}{2\lambda\mathrm{e}}\Big)^{3/2}\Big\}(\lambda-\mathrm{Var}\, S_n)\\ &\quad+2\Big(\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big)^2(\varepsilon_1+\varepsilon_2+\kappa\varepsilon_3),\end{aligned} \tag{3.8}$$

$$\begin{aligned}&\Big|\|P^{S_n}-\mathrm{Po}(\lambda)\|_{\mathrm{loc}}-\frac{\lambda-\mathrm{Var}\, S_n}{2\sqrt{2\pi}\lambda^{3/2}}\Big|\\ &\qquad\leqslant C\min\Big\{1, \frac{1}{\lambda}\Big\}\Big(\frac{\lambda-\mathrm{Var}\, S_n}{\lambda^{3/2}}+\min\Big\{1, \frac{1}{\lambda}\Big\}(\varepsilon_1+\varepsilon_2+\kappa\varepsilon_3)\Big).\end{aligned} \tag{3.9}$$

The constants $\frac{1}{\sqrt{2\mathrm{e}}}$ *and* $(\frac{3}{2\mathrm{e}})^{3/2}$ *in* (3.6) *and* (3.8) *are the best possible.*

REMARK 3.2. Let the assumptions of Theorem 3.2 hold.

(1) Let the matrix $(p_{j,r})$ have decreasing rows. Then P^{S_n} is a Bernoulli convolution (see Remark 1.1). Let $\theta=\frac{\lambda-\mathrm{Var}\, S_n}{\lambda}$. Then (3.7), (2.16) and (2.17) imply that

$$\Big|\|P^{S_n}-\mathrm{Po}(\lambda)\|_{\mathrm{W}}-\frac{\theta\sqrt{\lambda}}{\sqrt{2\pi}}\Big| \leqslant C\theta\sqrt{\lambda}\min\Big\{1, \frac{1}{\sqrt{\lambda}}+\theta\Big\}. \tag{3.10}$$

Further, in many cases, we have $\kappa \leqslant C\lambda^{-1/2}$ (see Remark 3.1(2)). In this case, (3.9), (2.16) and (2.17) yield

$$\Big|\|P^{S_n}-\mathrm{Po}(\lambda)\|_{\mathrm{loc}}-\frac{\theta}{2\sqrt{2\pi\lambda}}\Big| \leqslant C\frac{\theta}{\sqrt{\lambda}}\min\Big\{1, \frac{1}{\sqrt{\lambda}}+\theta\Big\}. \tag{3.11}$$

For (3.10) and (3.11) in the present situation, see also Roos [16, (32)].

(2) The optimality of the constants $\frac{1}{\sqrt{2\mathrm{e}}}$ and $(\frac{3}{2\mathrm{e}})^{3/2}$ in (3.6) and (3.8) can be verified by using the special example of $p_{j,r}=p$ for all $j, r \in \underline{n}$. This follows from arguments similar to those given in Remark 2.8(2).

4. PROOFS OF MAIN RESULTS

Let $t \in (0,\infty)$ and Y be a $\mathrm{Po}(t)$-distributed random variable. In what follows, we use the classical Stein–Chen approach without coupling (see [9] or [5]). This is based on the following idea: If $f \in \mathbb{R}^{\mathbb{Z}_+}$, then there exists a function $g := g_{t,f} \in \mathbb{R}^{\mathbb{Z}_+}$ such that $g(0)=0$ and the *Stein equation*

$$f(m)=tg(m+1)-mg(m) \quad \text{for all } m\in\mathbb{Z}_+ \tag{4.1}$$

holds. It turns out that g is unique on $\mathbb{N}$ and satisfies

$$g(m+1) = \frac{1}{t\mathrm{po}(m,t)} \sum_{j=0}^{m} \mathrm{po}(j,t) f(j) \qquad \text{for all } m \in \mathbb{Z}_+,$$

where $\mathrm{po}(m,t) = \mathrm{e}^{-t}\frac{t^m}{m!}$ for $m \in \mathbb{Z}_+$. If additionally $\mathrm{E}\,|f(Y)| < \infty$ and $\mathrm{E}\,f(N) = 0$, then also

$$g(m+1) = -\frac{1}{t\mathrm{po}(m,t)} \sum_{j=m+1}^{\infty} \mathrm{po}(j,t) f(j) \qquad \text{for all } m \in \mathbb{Z}_+. \tag{4.2}$$

Now suppose that W is a $\mathbb{Z}_+$-valued random variable, whose distribution is to be approximated by $\mathrm{Po}(t)$. Suppose further that we want to measure the approximation error in terms of differences like $|\mathrm{E}\,h(W) - \mathrm{E}\,h(Y)|$ for certain functions $h \in \mathbb{R}^{\mathbb{Z}_+}$, where we assume that the expectations are finite. Letting $f = h - \mathrm{E}\,h(Y)$ and $g := g_{t,f}$, we then find from the Stein equation that

$$|\mathrm{E}\,h(W) - \mathrm{E}\,h(Y)| = |\mathrm{E}\,f(W)| = |\mathrm{E}(tg(W+1) - Wg(W))|, \tag{4.3}$$

where the right-hand side has to be further estimated. To achieve this, estimates for g are necessary, some of which are given in the following lemma.

LEMMA 4.1. *Let $t \in (0,\infty)$ and Y be a $\mathrm{Po}(t)$-distributed random variable.*

(1) *Let $A \subseteq \mathbb{Z}_+$, $h = \mathbf{1}_A \in \mathbb{R}^{\mathbb{Z}_+}$, $f = h - \mathrm{Po}(t)(A) \in \mathbb{R}^{\mathbb{Z}_+}$ and $g_{t,A} := g_{t,f}$. Then*

$$\|g_{t,A}\|_\infty \leqslant \min\left\{1, \sqrt{\frac{2}{t\mathrm{e}}}\right\}, \quad \|\Delta g_{t,A}\|_\infty \leqslant \frac{1-\mathrm{e}^{-t}}{t} \leqslant \min\left\{1, \frac{1}{t}\right\}. \tag{4.4}$$

(2) *Let $h \in \mathcal{F}_{\mathrm{W}}$ and $f = h - \mathrm{E}\,h(Y) \in \mathbb{R}^{\mathbb{Z}_+}$. Then*

$$\|g_{t,f}\|_\infty \leqslant 1, \quad \|\Delta g_{t,f}\|_\infty \leqslant \min\left\{1, \frac{4}{3}\sqrt{\frac{2}{t\mathrm{e}}}\right\}. \tag{4.5}$$

(3) *Let $a \in \mathbb{Z}_+$, $h = \mathbf{1}_{\{a\}} \in \mathbb{R}^{\mathbb{Z}_+}$, $f = h - \mathrm{po}(a,t) \in \mathbb{R}^{\mathbb{Z}_+}$ and $g_{t,\{a\}}$ be defined as in* (1). *Let Z be an arbitrary $\mathbb{Z}_+$-valued random variable and $c(Z)$ be defined as in* (3.1). *Then*

$$\|g_{t,\{a\}}\|_\infty \leqslant 2\frac{1-\mathrm{e}^{-t}}{t}, \quad \mathrm{E}\,|\Delta g_{t,\{a\}}(Z)| \leqslant \min\{1, 2c(Z)\}\,\frac{1-\mathrm{e}^{-t}}{t}.$$

Proof. For the proof of the second inequality in (1), see Barbour and Eagleson [2, Lemma 4(ii)]. The remaining inequalities in (1) and (2) can be found in [5, pp. 7, 14, 15]. The inequalities in (3) follow from [5, Lemma 9.1.5, p. 176]. Here it is used that

$$\sum_{m=0}^{\infty} |\Delta g_{t,\{a\}}(m)| = 2\Delta g_{t,\{a\}}(a) \tag{4.6}$$

(see [2, proof of Lemma 4(ii)] and also Barbour and Jensen [6, p. 79]). ∎

The following proposition will be useful in the proofs of Corollary 4.1 and Theorems 2.2 and 3.2.

PROPOSITION 4.1. *Let F, G be two finite sets with cardinality $m = |F| = |G| \geqslant 2$, $(p_{j,r}) \in [0,1]^{F\times G}$, $X_{j,r}$ be a $\mathrm{Be}(p_{j,r})$-distributed random variable for $(j,r) \in F\times G$, and ρ be a random variable uniformly distributed over $G^F_{\neq}$. Assume that all the random variables ρ, $X_{j,r}$ ($j \in F$, $r \in G$) are independent. Let*

$$W = \sum_{j\in F} X_{j,\rho(j)}, \quad \overline{p}'_{j,\cdot} = \mathrm{E}\, X_{j,\rho(j)} = \frac{1}{m}\sum_{r\in G} p_{j,r} \quad (j \in F),$$
$$\mu = \mathrm{E}\, W = \sum_{j\in F} \overline{p}'_{j,\cdot}.$$

For $j, k \in F$ with $j \neq k$, let $W_j = W - X_{j,\rho(j)}$ and $W_{j,k} = W - X_{j,\rho(j)} - X_{k,\rho(k)}$. For an arbitrary function $h \in \mathbb{R}^{\mathbb{Z}_+}$, we have

$$\mathrm{E}(\mu h(W+1) - W h(W)) = D_1 + D_2, \tag{4.7}$$

where

$$D_1 = \sum_{j\in F} \mathrm{E}(\overline{p}'_{j,\cdot} p_{j,\rho(j)} \Delta h(W_j + 1)),$$
$$D_2 = \frac{1}{2m} \sum_{(j,k)\in F^2_{\neq}} \mathrm{E}\big((p_{j,\rho(j)} - p_{j,\rho(k)})(p_{k,\rho(j)} - p_{k,\rho(k)})\Delta h(W_{j,k}+1)\big).$$

Proof. Using Fubini's theorem, we get, for $j \in F$,

$$\begin{aligned}\mathrm{E}(X_{j,\rho(j)} h(W)) &= \sum_{\ell\in G^F_{\neq}} \mathrm{E}\big(\mathbf{1}_{\{\rho=\ell\}} X_{j,\ell(j)} h(W_j + X_{j,\ell(j)})\big)\\ &= \mathrm{E}(p_{j,\rho(j)} h(W_j+1))\end{aligned}$$

and similarly

$$\mathrm{E}\, h(W+1) = \mathrm{E}\big(p_{j,\rho(j)} h(W_j+2) + (1-p_{j,\rho(j)}) h(W_j+1)\big).$$

Hence

$$
\begin{aligned}
&\mathrm{E}(\mu h(W+1) - Wh(W))\\
&= \sum_{j\in F} \mathrm{E}(\overline{p}'_{j,\cdot} h(W+1) - X_{j,\rho(j)} h(W))\\
&= \sum_{j\in F} \mathrm{E}\big(\overline{p}'_{j,\cdot} p_{j,\rho(j)} h(W_j+2) + \overline{p}'_{j,\cdot}(1-p_{j,\rho(j)}) h(W_j+1) - p_{j,\rho(j)} h(W_j+1)\big)\\
&= D_1 + D'_2,
\end{aligned}
$$

where

$$D'_2 = \sum_{j\in F} \mathrm{E}((\overline{p}'_{j,\cdot} - p_{j,\rho(j)}) h(W_j+1)).$$

We have to show that $D'_2 = D_2$. Similarly to the above, for $(j,k)\in F^2_{\neq}$ we obtain

$$
\begin{aligned}
&\mathrm{E}((p_{j,\rho(k)} - p_{j,\rho(j)}) h(W_j+1))\\
&\qquad = \mathrm{E}\big((p_{j,\rho(k)} - p_{j,\rho(j)})(p_{k,\rho(k)} h(W_{j,k}+2) + (1-p_{k,\rho(k)}) h(W_{j,k}+1))\big).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
D'_2 &= \sum_{j\in F} \mathrm{E}((\overline{p}'_{j,\cdot} - p_{j,\rho(j)}) h(W_j+1))\\
&= \frac{1}{m} \sum_{j\in F}\sum_{k\in F} \mathrm{E}((p_{j,\rho(k)} - p_{j,\rho(j)}) h(W_j+1))\\
&= \frac{1}{m} \sum_{(j,k)\in F^2_{\neq}} \mathrm{E}\big((p_{j,\rho(k)} - p_{j,\rho(j)})\\
&\qquad\qquad \times (p_{k,\rho(k)} h(W_{j,k}+2) + (1-p_{k,\rho(k)}) h(W_{j,k}+1))\big)\\
&= \frac{1}{m} \sum_{(j,k)\in F^2_{\neq}} \mathrm{E}\big((p_{j,\rho(k)} - p_{j,\rho(j)}) p_{k,\rho(k)} \Delta h(W_{j,k}+1)\big).
\end{aligned}
\tag{4.8}
$$

The latter equality follows from the observation that, for $(j,k)\in F^2_{\neq}$, we have

$$\mathrm{E}(p_{j,\rho(k)} h(W_{j,k}+1)) = \mathrm{E}(p_{j,\rho(j)} h(W_{j,k}+1)).$$

In fact, the left-hand side does not change if we replace ρ with the composition $\rho\circ\tau_{j,k}$, where $\tau_{j,k}\in F^F_{\neq}$ is the transposition which interchanges j and k. Similarly,

$$
\begin{aligned}
&\mathrm{E}\big((p_{j,\rho(k)} - p_{j,\rho(j)}) p_{k,\rho(k)} \Delta h(W_{j,k}+1)\big)\\
&\qquad\qquad = \mathrm{E}\big((p_{j,\rho(j)} - p_{j,\rho(k)}) p_{k,\rho(j)} \Delta h(W_{j,k}+1)\big).
\end{aligned}
$$

Hence

$$D_2' = -\frac{1}{m} \sum_{(j,k)\in F^2_{\neq}} \mathrm{E}\big((p_{j,\rho(k)} - p_{j,\rho(j)})p_{k,\rho(j)}\Delta h(W_{j,k}+1)\big). \tag{4.9}$$

By adding the right-hand sides in (4.8) and (4.9) and dividing by 2 we obtain $D_2' = D_2$. This proves the assertion. ∎

REMARK 4.1. The combination of (2.10) and (2.11) in [10] leads to an identity which is similar but not identical to (4.7). It may be possible to get our results by using that identity. However, we prefer (4.7), since it does not require additional notation like that in [10, (2.3)–(2.9)].

COROLLARY 4.1. *Let the assumptions of Proposition* 4.1 *hold. For a* $\mathbb{Z}_+$-*valued random variable* Z, *let* $c(Z)$ *be defined as in* (3.1). *Set*

$$\alpha = \sum_{j\in F} (\overline{p}_{j,\cdot}')^2, \qquad \beta = \frac{1}{2m^2(m-1)} \sum_{(j,k)\in F^2_{\neq}} \sum_{(r,s)\in G^2_{\neq}} |p_{j,r} - p_{j,s}|\,|p_{k,r} - p_{k,s}|.$$

Then, for $t \in (0,\infty)$,

$$d_{\mathrm{TV}}(P^W, \mathrm{Po}(t)) \leqslant |t-\mu| \min\left\{1, \sqrt{\frac{2}{t\mathrm{e}}}\right\} + \frac{1-\mathrm{e}^{-t}}{t}(\alpha+\beta),$$
$$\|P^W - \mathrm{Po}(t)\|_{\mathrm{W}} \leqslant |t-\mu| + \min\left\{1, \frac{4}{3}\sqrt{\frac{2}{t\mathrm{e}}}\right\}(\alpha+\beta),$$

and

$$\|P^W - \mathrm{Po}(t)\|_{\mathrm{loc}} \leqslant 2\frac{1-\mathrm{e}^{-t}}{t}\Big(|t-\mu| + \Big(\max_{j\in F} c(W_j)\Big)\alpha + \Big(\max_{(j,k)\in F^2_{\neq}} c(W_{j,k})\Big)\beta\Big). \tag{4.10}$$

Proof. Let $t \in (0,\infty)$ and Y be a $\mathrm{Po}(t)$-distributed random variable. Let $h \in \mathbb{R}^{\mathbb{Z}_+}$ be such that $h = \mathbf{1}_A$ for $A \subseteq \mathbb{Z}_+$ or $h \in \mathcal{F}_{\mathrm{W}}$ or $h = \mathbf{1}_{\{a\}}$ for $a \in \mathbb{Z}_+$. Let $f = h - \mathrm{E}\,h(Y)$ and $g := g_{t,f}$. By using (4.3), Proposition 4.1 and (4.4), we obtain

$$\begin{aligned}|\mathrm{E}\,h(W) - \mathrm{E}\,h(Y)| &= |\mathrm{E}(tg(W+1) - Wg(W))| \\ &= |(t-\mu)\,\mathrm{E}\,g(W+1)| + |\mathrm{E}(\mu g(W+1) - Wg(W))|,\end{aligned}$$

where

$$
\begin{aligned}
&|\mathrm{E}(\mu g(W+1) - W g(W))| \\
&\leqslant \sum_{j\in F} \mathrm{E}(\overline{p}'_{j,\cdot} p_{j,\rho(j)} |\Delta g(W_j+1)|) \\
&\quad + \frac{1}{2m} \sum_{(j,k)\in F^2_{\neq}} \mathrm{E}\,|(p_{j,\rho(j)} - p_{j,\rho(k)})(p_{k,\rho(j)} - p_{k,\rho(k)}) \Delta g(W_{j,k}+1)| \\
&\leqslant \Big(\max_{j\in F} \mathrm{E}\,|\Delta g(W_j+1)|\Big) \frac{1}{m!} \sum_{j\in F} \sum_{\ell\in G^F_{\neq}} \overline{p}'_{j,\cdot} p_{j,\ell(j)} \\
&+ \Big(\max_{(j,k)\in F^2_{\neq}} \mathrm{E}\,|\Delta g(W_{j,k}+1)|\Big) \frac{1}{2m\, m!} \sum_{(j,k)\in F^2_{\neq}} \sum_{\ell\in G^F_{\neq}} |p_{j,\ell(j)} - p_{j,\ell(k)}|\, |p_{k,\ell(j)} - p_{k,\ell(k)}| \\
&= \Big(\max_{j\in F} \mathrm{E}\,|\Delta g(W_j+1)|\Big)\alpha + \Big(\max_{(j,k)\in F^2_{\neq}} \mathrm{E}\,|\Delta g(W_{j,k}+1)|\Big)\beta.
\end{aligned}
$$

The proof is easily completed by using Lemma 4.1. ■

LEMMA 4.2. *We have*

$$
\gamma + \gamma' = \gamma'', \qquad \gamma'' + \gamma''' = \frac{1}{n} \sum_{(j,r)\in \underline{n}^2} p^2_{j,r} - \sum_{j\in\underline{n}} \overline{p}^2_{j,\cdot}\,. \tag{4.11}
$$

Proof. For $a, b \in \mathbb{R}$, we have $|ab| - ab = 2(a_+(-b)_+ + (-a)_+ b_+)$, and therefore

$$
\begin{aligned}
\gamma'' - \gamma &= \frac{1}{n^2(n-1)} \sum_{(j,k)\in \underline{n}^2_{\neq}} \sum_{(r,s)\in \underline{n}^2_{\neq}} \big((p_{j,r} - p_{j,s})_+ (p_{k,s} - p_{k,r})_+ \\
&\qquad + (p_{j,s} - p_{j,r})_+ (p_{k,r} - p_{k,s})_+\big) \\
&= \frac{2}{n^2(n-1)} \sum_{(j,k)\in \underline{n}^2_{\neq}} \sum_{(r,s)\in \underline{n}^2_{\neq}} (p_{j,r} - p_{j,s})_+ (p_{k,s} - p_{k,r})_+ = \gamma'.
\end{aligned}
$$

Further,

$$
\begin{aligned}
\gamma'' + \gamma''' &= \frac{1}{4n^2(n-1)} \sum_{(j,k)\in \underline{n}^2_{\neq}} \sum_{(r,s)\in \underline{n}^2_{\neq}} ((p_{j,r} - p_{j,s})^2 + (p_{k,r} - p_{k,s})^2) \\
&= \frac{1}{n} \sum_{(j,r)\in \underline{n}^2} p^2_{j,r} - \sum_{j\in\underline{n}} \overline{p}^2_{j,\cdot}\,. \quad ■
\end{aligned}
$$

Proof of Theorem 2.1. Using Corollary 4.1 with $F = G = \underline{n}$, $m = n$, $\rho = \pi$, $W = S_n$, $\overline{p}'_{j,\cdot} = \overline{p}_{j,\cdot}$, and $\mu = t = \lambda$, we obtain

$$
d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) \leqslant \frac{1 - \mathrm{e}^{-\lambda}}{\lambda} \Big(\sum_{j=1}^{n} \overline{p}^2_{j,\cdot} + \gamma'' \Big).
$$

The proof is easily completed by using (1.8) and (4.11). ■

The proof of Theorem 3.1 is analogous and therefore omitted. The following lemma is needed in the proof of Theorems 2.2 and 3.2.

LEMMA 4.3. *Let $t \in (0,\infty)$, let Y be a $\mathrm{Po}(t)$-distributed random variable and $h \in \mathbb{R}^{\mathbb{Z}_+}$ such that $\mathrm{E}\,|h(Y+3)| < \infty$. Set $f = h - \mathrm{E}\,h(Y)$ and $g = g_{t,f}$. Then*

$$\int h\,\mathrm{d}((\delta_1-\delta_0)^{*2} * \mathrm{Po}(t)) = -2\,\mathrm{E}(\Delta g(Y+1)) \tag{4.12}$$

is finite.

Proof. If $h_0 : \mathbb{Z}_+ \to \mathbb{R}$ is an arbitrary function, then $\mathrm{E}\,|Yh_0(Y)| < \infty$ if and only if $\mathrm{E}\,|h_0(Y+1)| < \infty$. In this case, we have $\mathrm{E}(th_0(Y+1)) = \mathrm{E}(Yh_0(Y))$. This implies that $\mathrm{E}\,|h(Y+j)| < \infty$ for $j \in \{0,1,2,3\}$. Since

$$(\delta_1-\delta_0)^{*2} * \mathrm{Po}(t) = (\delta_2 - 2\delta_1 + \delta_0) * \mathrm{Po}(t) = P^Y - 2P^{Y+1} + P^{Y+2},$$

the left-hand side of (4.12) is finite. Using (4.1), we obtain

$$\begin{aligned}
&\int h\,\mathrm{d}((\delta_1-\delta_0)^{*2} * \mathrm{Po}(t))\\
&\quad = \mathrm{E}(h(Y) - 2h(Y+1) + h(Y+2))\\
&\quad = \mathrm{E}(-2f(Y+1) + f(Y+2))\\
&\quad = \mathrm{E}(-2(tg(Y+2) - (Y+1)g(Y+1)) + tg(Y+3) - (Y+2)g(Y+2)).
\end{aligned}$$

Since $\mathrm{E}\,f(Y) = 0$, we deduce from (4.2) that, for $j \in \{0,1,2\}$,

$$\begin{aligned}
\mathrm{E}\,|g(Y+j+1)| &\leqslant \sum_{m=0}^{\infty} \frac{\mathrm{po}(m,t)}{t\mathrm{po}(m+j,t)} \sum_{k=m+j+1}^{\infty} \mathrm{po}(k,t)|f(k)|\\
&= \frac{1}{(j+1)}\,\mathrm{E}\,|f(Y+j+1)| < \infty.
\end{aligned}$$

Hence $\mathrm{E}(tg(Y+j+1)) = \mathrm{E}(Yg(Y+j))$ for $j \in \{0,1,2\}$, which implies that

$$\begin{aligned}
&\int h\,\mathrm{d}((\delta_1-\delta_0)^{*2} * \mathrm{Po}(t))\\
&\quad = \mathrm{E}(-2(Yg(Y+1) - (Y+1)g(Y+1)) + Yg(Y+2) - (Y+2)g(Y+2))\\
&\quad = -2\,\mathrm{E}(\Delta g(Y+1)). \quad \blacksquare
\end{aligned}$$

Proof of Theorems 2.2 and 3.2. For $\ell \in \underline{n}^n_{\neq}$ and a real-valued random variable Z, we write $\mathrm{E}_\ell\,Z = \mathrm{E}_\ell(Z) = \mathrm{E}(\mathbf{1}_{\{\pi=\ell\}}Z)$ whenever this exists. Let Y be a $\mathrm{Po}(\lambda)$-distributed random variable independent of π and X and let $h = \mathbf{1}_A \in \mathbb{R}^{\mathbb{Z}_+}$ for a set $A \subseteq \mathbb{Z}_+$ or $h \in \mathcal{F}_{\mathrm{W}}$. Let $f = h - \mathrm{E}\,h(Y)$ and $g = g_{\lambda,f}$. Using (4.1) and

Proposition 4.1 with $F = G = \underline{n}$, $\rho = \pi$, $W = S_n$, and $\mu = \lambda$, we obtain

$$
\begin{aligned}
(4.13)\quad & \mathrm{E}\, h(S_n) - \mathrm{E}\, h(Y) = \mathrm{E}\, f(S_n) = \mathrm{E}(\lambda g(S_n+1) - S_n g(S_n)) \\
&= \sum_{j=1}^{n} \mathrm{E}(\overline{p}_{j,\cdot} p_{j,\pi(j)} \Delta g(T_j+1)) \\
&\quad + \frac{1}{2n} \sum_{(j,k)\in \underline{n}^2_{\neq}} \mathrm{E}\big((p_{j,\pi(j)} - p_{j,\pi(k)})(p_{k,\pi(j)} - p_{k,\pi(k)})\Delta g(T_{j,k}+1)\big) \\
&= \sum_{j=1}^{n}\sum_{r=1}^{n} \overline{p}_{j,\cdot} p_{j,r} \sum_{\ell\in \underline{n}^n_{\neq}:\, \ell(j)=r} \mathrm{E}_\ell\, \Delta g(T_j+1) \\
&\quad + \frac{1}{2n} \sum_{(j,k)\in \underline{n}^2_{\neq}} \sum_{(r,s)\in \underline{n}^2_{\neq}} (p_{j,r} - p_{j,s})(p_{k,r} - p_{k,s}) \sum_{\substack{\ell\in \underline{n}^n_{\neq} \\ \ell(j)=r,\, \ell(k)=s}} \mathrm{E}_\ell\, \Delta g(T_{j,k}+1),
\end{aligned}
$$

where $T_j = S_n - X_{j,\pi(j)}$ and $T_{j,k} = S_n - X_{j,\pi(j)} - X_{k,\pi(k)}$ for $(j,k) \in \underline{n}^2_{\neq}$. Combining (2.10), (1.8), (4.12), (4.13) and (1.9), we obtain

$$
\begin{aligned}
(4.14)\quad & \mathrm{E}\, h(S_n) - \int h \,\mathrm{d}Q_2 \\
&= \mathrm{E}\, f(S_n) - \Big(\sum_{j=1}^{n} \overline{p}^2_{j,\cdot} + \gamma\Big) \mathrm{E}(\Delta g(Y+1)) = D_1 + D_2,
\end{aligned}
$$

where

$$
D_1 = \sum_{j=1}^{n}\sum_{r=1}^{n} \overline{p}_{j,\cdot} p_{j,r} \sum_{\ell\in \underline{n}^n_{\neq}:\, \ell(j)=r} \mathrm{E}_\ell(\Delta g(T_j+1) - \Delta g(Y+1))
$$

and

$$
\begin{aligned}
D_2 = \frac{1}{2n} & \sum_{(j,k)\in \underline{n}^2_{\neq}} \sum_{(r,s)\in \underline{n}^2_{\neq}} (p_{j,r} - p_{j,s})(p_{k,r} - p_{k,s}) \\
\times & \sum_{\substack{\ell\in \underline{n}^n_{\neq} \\ \ell(j)=r,\, \ell(k)=s}} \mathrm{E}_\ell(\Delta g(T_{j,k}+1) - \Delta g(Y+1)).
\end{aligned}
$$

We note that, for $j, r \in \underline{n}$ and $m \in \mathbb{Z}_+$,

$$
\begin{aligned}
(4.15)\quad & \sum_{\ell\in \underline{n}^n_{\neq}:\, \ell(j)=r} P(\pi = \ell, T_j = m) \\
&= \frac{1}{n!} \sum_{\ell\in \underline{n}^n_{\neq}:\, \ell(j)=r} P\Big(\sum_{i\in \underline{n}\setminus\{j\}} X_{i,\ell(i)} = m\Big) \\
&= \frac{1}{n} \sum_{\ell\in (\underline{n}\setminus\{r\})^{\underline{n}\setminus\{j\}}_{\neq}} P(\pi_{j,r} = \ell, T'_{j,r} = m) = \frac{1}{n} P(T'_{j,r} = m),
\end{aligned}
$$

where $T'_{j,r} := \sum_{i\in\underline{n}\setminus\{j\}} X_{i,\pi_{j,r}(i)}$ and the random variable $\pi_{j,r}$ is independent of X and has uniform distribution on $(\underline{n}\setminus\{r\})_{\neq}^{\underline{n}\setminus\{j\}}$. Hence

$$
\begin{aligned}
(4.16)\quad &\Big|\sum_{\substack{\ell\in\underline{n}^n_{\neq}\\ \ell(j)=r}} \mathrm{E}_\ell(\Delta g(T_j+1)-\Delta g(Y+1))\Big|\\
&=\Big|\sum_{m\in\mathbb{Z}_+}\sum_{\substack{\ell\in\underline{n}^n_{\neq}\\ \ell(j)=r}} (P(\pi=\ell,T_j=m)-P(\pi=\ell,Y=m))\Delta g(m+1)\Big|\\
&=\Big|\frac{1}{n}\sum_{m\in\mathbb{Z}_+}(P(T'_{j,r}=m)-P(Y=m))\Delta g(m+1)\Big|\\
&\leqslant \frac{2\|\Delta g\|_\infty}{n} d_{\mathrm{TV}}(P^{T'_{j,r}},\mathrm{Po}(\lambda)).
\end{aligned}
$$

We have $\mathrm{E}\,T'_{j,r}=\lambda'_{j,r}$ for $j,r\in\underline{n}$. Using Corollary 4.1 with $F=\underline{n}\setminus\{j\}$, $G=\underline{n}\setminus\{r\}$, $m=n-1$, $\rho=\pi_{j,r}$, $W=T'_{j,r}$, $\overline{p}'_{u,\cdot}=\frac{1}{n-1}\sum_{v\in\underline{n}\setminus\{r\}}p_{u,v}$ for $u\in\underline{n}\setminus\{j\}$, $\mu=\lambda'_{j,r}$ and $t=\lambda$, we obtain

$$
\begin{aligned}
&d_{\mathrm{TV}}(P^{T'_{j,r}},\mathrm{Po}(\lambda))\\
&\leqslant|\lambda-\lambda'_{j,r}|\min\Big\{1,\sqrt{\frac{2}{\lambda\mathrm{e}}}\Big\}+\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{u\in\underline{n}\setminus\{j\}}(\overline{p}'_{u,\cdot})^2\\
&\quad+\frac{1}{2(n-1)^2(n-2)}\sum_{(u,u')\in(\underline{n}\setminus\{j\})^2_{\neq}}\sum_{(v,v')\in(\underline{n}\setminus\{r\})^2_{\neq}}|p_{u,v}-p_{u,v'}|\,|p_{u',v}-p_{u',v'}|\Big)\\
&\leqslant|\lambda-\lambda'_{j,r}|\min\Big\{1,\sqrt{\frac{2}{\lambda\mathrm{e}}}\Big\}+\frac{n^2}{(n-1)^2}\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{u\in\underline{n}}\overline{p}^2_{u,\cdot}+\frac{n-1}{n-2}\gamma''\Big).
\end{aligned}
$$

Similarly to (4.15), we infer that, for $(j,k),(r,s)\in\underline{n}^2_{\neq}$ and $m\in\mathbb{Z}_+$,

$$
\sum_{\substack{\ell\in\underline{n}^n_{\neq}\\ \ell(j)=r,\,\ell(k)=s}} P(\pi=\ell,T_{j,k}=m)=\frac{1}{n(n-1)}P(T''_{j,k,r,s}=m),
$$

where $T''_{j,k,r,s} := \sum_{i\in\underline{n}\setminus\{j,k\}} X_{i,\pi_{j,k,r,s}(i)}$ and the random variable $\pi_{j,k,r,s}$ is independent of X and has uniform distribution on $(\underline{n}\setminus\{r,s\})_{\neq}^{\underline{n}\setminus\{j,k\}}$. Similarly to (4.16), we get

$$\text{(4.17)}\qquad \Big|\sum_{\substack{\ell\in\underline{n}^n_{\neq}\\ \ell(j)=r,\,\ell(k)=s}} \mathrm{E}_\ell(\Delta g(T_{j,k}+1)-\Delta g(Y+1))\Big|$$
$$= \frac{1}{n(n-1)}\Big|\sum_{m\in\mathbb{Z}_+}(P(T''_{j,k,r,s}=m)-P(Y=m))\Delta g(m+1)\Big|$$
$$\leqslant \frac{2\|\Delta g\|_\infty}{n(n-1)} d_{\mathrm{TV}}(P^{T''_{j,k,r,s}},\mathrm{Po}(\lambda)).$$

We note that $\lambda''_{j,k,r,s} = \mathrm{E}\,T''_{j,k,r,s}$ for $(j,k),(r,s)\in\underline{n}^2_{\neq}$. Using Corollary 4.1 with $F=\underline{n}\setminus\{j,k\}$, $G=\underline{n}\setminus\{r,s\}$, $m=n-2$, $\rho=\pi_{j,k,r,s}$, $W=T''_{j,k,r,s}$, $\overline{p}'_{u,\cdot}=\frac{1}{n-2}\sum_{v\in\underline{n}\setminus\{r,s\}}p_{u,v}$ for $u\in\underline{n}\setminus\{j,k\}$, $\mu=\lambda''_{j,k,r,s}$ and $t=\lambda$, we obtain

$$\begin{aligned}
&d_{\mathrm{TV}}(P^{T''_{j,k,r,s}},\mathrm{Po}(\lambda))\\
&\leqslant |\lambda-\lambda''_{j,k,r,s}|\min\Big\{1,\sqrt{\frac{2}{\lambda\mathrm{e}}}\Big\}+\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{u\in\underline{n}\setminus\{j,k\}}(\overline{p}'_{u,\cdot})^2\\
&+\frac{1}{2(n-2)^2(n-3)}\sum_{(u,u')\in(\underline{n}\setminus\{j,k\})^2_{\neq}}\ \sum_{(v,v')\in(\underline{n}\setminus\{r,s\})^2_{\neq}}|p_{u,v}-p_{u,v'}|\,|p_{u',v}-p_{u',v'}|\Big)\\
&\leqslant |\lambda-\lambda''_{j,k,r,s}|\min\Big\{1,\sqrt{\frac{2}{\lambda\mathrm{e}}}\Big\}+\frac{n^2}{(n-2)^2}\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{u\in\underline{n}}\overline{p}^2_{u,\cdot}+\frac{n-1}{n-3}\gamma''\Big).
\end{aligned}$$

Combining the inequalities above, we see that

$$\begin{aligned}
|D_1| &\leqslant \frac{2\|\Delta g\|_\infty}{n}\sum_{j=1}^n\sum_{r=1}^n \overline{p}_{j,\cdot}p_{j,r}d_{\mathrm{TV}}(P^{T'_{j,r}},\mathrm{Po}(\lambda))\\
&\leqslant \|\Delta g\|_\infty\min\Big\{1,\sqrt{\frac{2}{\lambda\mathrm{e}}}\Big\}\varepsilon_1\\
&\quad+2\|\Delta g\|_\infty\frac{n^2}{(n-1)^2}\,\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\sum_{j=1}^n\overline{p}^2_{j,\cdot}\Big(\sum_{u\in\underline{n}}\overline{p}^2_{u,\cdot}+\frac{n-1}{n-2}\gamma''\Big)
\end{aligned}$$

and

$$\begin{aligned}
|D_2| &\leqslant \frac{\|\Delta g\|_\infty}{n^2(n-1)}\sum_{(j,k)\in\underline{n}^2_{\neq}}\ \sum_{(r,s)\in\underline{n}^2_{\neq}}|p_{j,r}-p_{j,s}|\,|p_{k,r}-p_{k,s}|d_{\mathrm{TV}}(P^{T''_{j,k,r,s}},\mathrm{Po}(\lambda))\\
&\leqslant \|\Delta g\|_\infty\varepsilon_2\min\Big\{1,\sqrt{\frac{2}{\lambda\mathrm{e}}}\Big\}\\
&\quad+2\|\Delta g\|_\infty\frac{n^2}{(n-2)^2}\,\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{u\in\underline{n}}\overline{p}^2_{u,\cdot}+\frac{n-1}{n-3}\gamma''\Big)\gamma''.
\end{aligned}$$

Therefore, using (4.14),

$$\begin{aligned}\Big|\mathrm{E}\,h(S_n) - \textstyle\int h\,\mathrm{d}Q_2\Big| &\leqslant |D_1| + |D_2| \\ &\leqslant \|\Delta g\|_\infty\bigg(\min\bigg\{1, \sqrt{\frac{2}{\lambda\mathrm{e}}}\bigg\}(\varepsilon_1+\varepsilon_2) + \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\varepsilon_3\bigg) \\ &= \|\Delta g\|_\infty\varepsilon.\end{aligned}$$

In view of (4.4) and (4.5), we see that (2.11) and (3.4) hold.

We now give a proof of (3.5). Here, we consider the special case that $a \in \mathbb{Z}_+$, $h = \mathbf{1}_{\{a\}} \in \mathbb{R}^{\mathbb{Z}_+}$, $f = h - \mathrm{po}(a,\lambda)$ and $g = g_{\lambda,\{a\}}$. It follows from (4.6) that, for an arbitrary $\mathbb{Z}_+$-valued random variable Z,

$$\begin{aligned}\Big|\sum_{m=0}^{\infty}(P(Z=m)-P(Y=m))\Delta g(m+1)\Big| &\leqslant \|P^Z - \mathrm{Po}(\lambda)\|_{\mathrm{loc}}\sum_{m=0}^{\infty}|\Delta g(m)| \\ &\leqslant 2\|\Delta g\|_\infty\|P^Z - \mathrm{Po}(\lambda)\|_{\mathrm{loc}}.\end{aligned}$$

Under the present assumptions, this leads to the following improvements of (4.16) and (4.17) for $(j,k),(r,s) \in \underline{n}^2_{\neq}$:

$$\Big|\sum_{\ell\in\underline{n}^n_{\neq}:\,\ell(j)=r}\mathrm{E}_\ell(\Delta g(T_j+1)-\Delta g(Y+1))\Big| \leqslant \frac{2\|\Delta g\|_\infty}{n}\|P^{T'_{j,r}} - \mathrm{Po}(\lambda)\|_{\mathrm{loc}},$$

and

$$\begin{aligned}&\Big|\sum_{\substack{\ell\in\underline{n}^n_{\neq}:\\ \ell(j)=r,\ell(k)=s}}\mathrm{E}_\ell(\Delta g(T_{j,k}+1)-\Delta g(Y+1))\Big| \\ &\qquad\qquad \leqslant \frac{2\|\Delta g\|_\infty}{n(n-1)}\|P^{T''_{j,k,r,s}} - \mathrm{Po}(\lambda)\|_{\mathrm{loc}}.\end{aligned}$$

Further, from (4.10), we get

$$\|P^{T'_{j,r}} - \mathrm{Po}(\lambda)\|_{\mathrm{loc}} \leqslant 2\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\bigg(|\lambda-\lambda'_{j,r}| + \frac{\kappa n^2}{(n-1)^2}\Big(\sum_{u\in\underline{n}}\overline{p}^2_{u,\cdot} + \frac{n-1}{n-2}\gamma''\Big)\bigg),$$

and

$$\begin{aligned}&\|P^{T''_{j,k,r,s}} - \mathrm{Po}(\lambda)\|_{\mathrm{loc}} \\ &\qquad \leqslant 2\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\bigg(|\lambda-\lambda''_{j,k,r,s}| + \frac{\kappa n^2}{(n-2)^2}\Big(\sum_{u\in\underline{n}}\overline{p}^2_{u,\cdot} + \frac{n-1}{n-3}\gamma''\Big)\bigg).\end{aligned}$$

In view of (4.14), we see that

$$\mathrm{E}\,h(S_n) - \textstyle\int h\,\mathrm{d}Q_2 = D_1 + D_2,$$

where

$$|D_1| \leqslant 2\|\Delta g\|_\infty \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\varepsilon_1 + 4\kappa\|\Delta g\|_\infty \frac{n^2}{(n-1)^2}\,\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\sum_{j=1}^n \overline{p}_{j,\cdot}^2\Big(\sum_{u\in\underline{n}}\overline{p}_{u,\cdot}^2 + \frac{n-1}{n-2}\gamma''\Big)$$

and

$$|D_2| \leqslant 2\|\Delta g\|_\infty \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\varepsilon_2 + 4\kappa\|\Delta g\|_\infty \frac{n^2}{(n-2)^2}\,\frac{1-\mathrm{e}^{-\lambda}}{\lambda}\Big(\sum_{u\in\underline{n}}\overline{p}_{u,\cdot}^2 + \frac{n-1}{n-3}\gamma''\Big)\gamma''.$$

Therefore

$$\Big|\mathrm{E}\,h(S_n) - \textstyle\int h\,\mathrm{d}Q_2\Big| \leqslant |D_1| + |D_2| \leqslant 2\|\Delta g\|_\infty \frac{1-\mathrm{e}^{-\lambda}}{\lambda}(\varepsilon_1+\varepsilon_2+\kappa\varepsilon_3),$$

which together with (4.4) implies (3.5). ∎

LEMMA 4.4. *Let* $t \in (0,\infty)$. *Then*

(4.18) $$\|(\delta_1-\delta_0)^{*2} * \mathrm{Po}(t)\|_{\mathrm{TV}} \leqslant \min\Big\{4, \frac{3}{t\mathrm{e}}\Big\},$$

(4.19) $$\Big|\|(\delta_1-\delta_0)^{*2} * \mathrm{Po}(t)\|_{\mathrm{TV}} - \frac{4}{t\sqrt{2\pi\mathrm{e}}}\Big| \leqslant \frac{C}{t}\min\Big\{1, \frac{1}{t}\Big\},$$

(4.20) $$\|(\delta_1-\delta_0)^{*2} * \mathrm{Po}(t)\|_{\mathrm{W}} = \|(\delta_1-\delta_0) * \mathrm{Po}(t)\|_{\mathrm{TV}} \leqslant \min\Big\{2, \sqrt{\frac{2}{t\mathrm{e}}}\Big\},$$

(4.21) $$\Big|\|(\delta_1-\delta_0)^{*2} * \mathrm{Po}(t)\|_{\mathrm{W}} - \sqrt{\frac{2}{\pi t}}\Big| \leqslant \frac{C}{\sqrt{t}}\min\Big\{1, \frac{1}{\sqrt{t}}\Big\},$$

(4.22) $$\|(\delta_1-\delta_0)^{*2} * \mathrm{Po}(t)\|_{\mathrm{loc}} \leqslant \min\Big\{2, \Big(\frac{3}{2t\mathrm{e}}\Big)^{3/2}\Big\},$$

(4.23) $$\Big|\|(\delta_1-\delta_0)^{*2} * \mathrm{Po}(t)\|_{\mathrm{loc}} - \frac{1}{\sqrt{2\pi}t^{3/2}}\Big| \leqslant \frac{C}{t^{3/2}}\min\Big\{1, \frac{1}{t}\Big\}.$$

The constants $\frac{3}{\mathrm{e}}$, $\sqrt{\frac{2}{\mathrm{e}}}$ *and* $\big(\frac{3}{2\mathrm{e}}\big)^{3/2}$ *on the right-hand sides of* (4.18), (4.20) *and* (4.22) *are the best possible.*

Proof. For (4.18), (4.22) and the optimality of the constants $\frac{3}{\mathrm{e}}$ and $\big(\frac{3}{2\mathrm{e}}\big)^{3/2}$, see [18, Lemma 3]. Inequality (4.19) was proved in [19, Lemma 5]; the proof of (4.23) is analogous using [17, (45)]. The equality in (4.20) follows from (1.1); the inequality and the optimality of the constant $\sqrt{2/\mathrm{e}}$ are contained in [12, (3.8)]. Inequality (4.21) follows from the more general Proposition 4 in [16]. ∎

REMARK 4.2. We do not know if $\frac{1}{\sqrt{t}}$ in the minimum term in (4.21) can be replaced by $\frac{1}{t}$. The norm terms in Lemma 4.4 can be evaluated using the zeros of some Charlier polynomials. The details are omitted here; more general identities can be found in [16, Corollaries 1, 2].

Proof of Corollary 2.2. Using (2.10), Theorem 2.2, (4.18) and (4.19), we obtain

$$\begin{aligned} d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) &= \frac{1}{2}\|P^{S_n} - \mathrm{Po}(\lambda)\|_{\mathrm{TV}} \\ &\leqslant \frac{1}{4}(\lambda - \mathrm{Var}\, S_n)\|(\delta_1 - \delta_0)^{*2} * \mathrm{Po}(\lambda)\|_{\mathrm{TV}} + d_{\mathrm{TV}}(P^{S_n}, Q_2) \\ &\leqslant \min\left\{1, \frac{3}{4\lambda \mathrm{e}}\right\}(\lambda - \mathrm{Var}\, S_n) + \frac{1 - \mathrm{e}^{-\lambda}}{\lambda}\varepsilon \end{aligned}$$

and

$$\begin{aligned} &\left| d_{\mathrm{TV}}(P^{S_n}, \mathrm{Po}(\lambda)) - \frac{\lambda - \mathrm{Var}\, S_n}{\sqrt{2\pi \mathrm{e}}\,\lambda} \right| = \left| \frac{1}{2}\|P^{S_n} - \mathrm{Po}(\lambda)\|_{\mathrm{TV}} - \frac{\lambda - \mathrm{Var}\, S_n}{\sqrt{2\pi \mathrm{e}}\,\lambda} \right| \\ &\quad \leqslant \frac{1}{2}\left\| P^{S_n} - \mathrm{Po}(\lambda) + \frac{1}{2}(\lambda - \mathrm{Var}\, S_n)(\delta_1 - \delta_0)^{*2} * \mathrm{Po}(\lambda) \right\|_{\mathrm{TV}} \\ &\qquad + \frac{1}{4}(\lambda - \mathrm{Var}\, S_n)\left| \|(\delta_1 - \delta_0)^{*2} * \mathrm{Po}(\lambda)\|_{\mathrm{TV}} - \frac{4}{\sqrt{2\pi \mathrm{e}}\,\lambda} \right| \\ &\quad \leqslant \frac{1}{4}(\lambda - \mathrm{Var}\, S_n)\left| \|(\delta_1 - \delta_0)^{*2} * \mathrm{Po}(\lambda)\|_{\mathrm{TV}} - \frac{4}{\lambda\sqrt{2\pi \mathrm{e}}} \right| + \frac{1}{2}\|P^{S_n} - Q_2\|_{\mathrm{TV}} \\ &\quad \leqslant \frac{C}{\lambda}\min\left\{1, \frac{1}{\lambda}\right\}(\lambda - \mathrm{Var}\, S_n) + \frac{1 - \mathrm{e}^{-\lambda}}{\lambda}\varepsilon \\ &\quad \leqslant C\min\left\{1, \frac{1}{\lambda}\right\}\left(\frac{\lambda - \mathrm{Var}\, S_n}{\lambda} + \varepsilon\right). \quad \blacksquare \end{aligned}$$

The proof of Corollary 3.1 is analogous and therefore omitted.

5. REMAINING PROOFS

Proof of (1.8). For $(j, k) \in \underline{n}^2_{\neq}$, we have

$$\mathrm{Var}\, X_{j,\pi(j)} = \overline{p}_{j,\cdot}(1 - \overline{p}_{j,\cdot}), \qquad \mathrm{E}(X_{j,\pi(j)}X_{k,\pi(k)}) = \frac{1}{n(n-1)}\sum_{(r,s)\in\underline{n}^2_{\neq}} p_{j,r}p_{k,s},$$

$$\mathrm{Cov}(X_{j,\pi(j)}, X_{k,\pi(k)}) = \frac{1}{n-1}\left(\overline{p}_{j,\cdot}\overline{p}_{k,\cdot} - \frac{1}{n}\sum_{r\in\underline{n}} p_{j,r}p_{k,r}\right).$$

Therefore

$$\sum_{(j,k)\in\underline{n}^2_{\neq}} \operatorname{Cov}(X_{j,\pi(j)}, X_{k,\pi(k)}) = \frac{1}{n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} p_{j,r}(p_{k,s} - p_{k,r})$$
$$= -\frac{1}{2n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} (p_{j,r} - p_{j,s})(p_{k,r} - p_{k,s}) = -\gamma.$$

Hence

$$\operatorname{Var} S_n = \sum_{j=1}^n \operatorname{Var} X_{j,\pi(j)} + \sum_{(j,k)\in\underline{n}^2_{\neq}} \operatorname{Cov}(X_{j,\pi(j)}, X_{k,\pi(k)}) = \lambda - \sum_{j=1}^n \overline{p}^2_{j,\cdot} - \gamma. \quad \blacksquare$$

Proof of Lemma 2.2. (1) It is easily shown that for $a, b \in [0,1]$ we have $(a-b)_+ \leqslant a(1-b)$, with equality if and only if $(a,b) \notin (0,1)^2$. In particular, this implies (2.6). To prove the second assertion, we note that for $a,b,c,d \in [0,1]$ we have $(a-b)_+(c-d)_+ \leqslant a(1-b)c(1-d)$, with equality if and only if one of the following three conditions is true:

- $(a,b) \notin (0,1)^2$ and $(c,d) \notin (0,1)^2$,
- $(a,b) \in (0,1)^2$ and ($c = 0$ or $d = 1$),
- $(c,d) \in (0,1)^2$ and ($a = 0$ or $b = 1$).

The proof of the second assertion is now easily completed: Let us first show necessity and suppose that equality holds. For all $(j,k), (r,s) \in \underline{n}^2_{\neq}$, we then have

$$(p_{j,r} - p_{j,s})_+(p_{k,s} - p_{k,r})_+ = p_{j,r}(1-p_{j,s})p_{k,s}(1-p_{k,r}),$$
$$(p_{j,s} - p_{j,r})_+(p_{k,r} - p_{k,s})_+ = p_{j,s}(1-p_{j,r})p_{k,r}(1-p_{k,s}).$$

Now, if $j \in \underline{n}$ and $(r,s) \in \underline{n}^2_{\neq}$ are such that $(p_{j,r}, p_{j,s}) \in (0,1)^2$, then $(p_{j,r} - p_{j,s})_+ < p_{j,r}(1-p_{j,s})$ and $(p_{j,s} - p_{j,r})_+ < p_{j,s}(1-p_{j,r})$. This together with the equalities above implies that for all $k \in \underline{n} \setminus \{j\}$ we have $p_{k,s}(1-p_{k,r}) = 0$ and $p_{k,r}(1-p_{k,s}) = 0$, that is, ($p_{k,r} = 1$ or $p_{k,s} = 0$) and ($p_{k,r} = 0$ or $p_{k,s} = 1$), which is equivalent to $p_{k,r} = p_{k,s} \in \{0,1\}$. This proves necessity. Sufficiency is shown similarly.

(2) We have

$$\gamma' \leqslant \frac{2}{n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} p_{j,r}p_{k,s}$$
$$= \frac{2}{n-1}\Big(\lambda^2 - \sum_{j\in\underline{n}} \overline{p}^2_{j,\cdot} - \sum_{r\in\underline{n}} \overline{p}^2_{\cdot,r} + \frac{1}{n^2} \sum_{(j,r)\in\underline{n}^2} p^2_{j,r}\Big)$$
$$= \frac{2}{n}(\operatorname{Var} S_n - \lambda + \lambda^2).$$

On the other hand, using (4.11), we get

$$\gamma' = \gamma'' - \gamma \leqslant \frac{1}{n} \sum_{(j,r)\in\underline{n}^2} p_{j,r}^2 - \sum_{j\in\underline{n}} \overline{p}_{j,\cdot}^2 - \gamma = \frac{1}{n} \sum_{(j,r)\in\underline{n}^2} p_{j,r}^2 + \operatorname{Var} S_n - \lambda$$
$$= \operatorname{Var} S_n - \frac{1}{n} \sum_{(j,r)\in\underline{n}^2} p_{j,r}(1-p_{j,r}).$$

(3) For $(p_{j,r}) \in \{0,1\}^{\underline{n}\times\underline{n}}$, we have

$$\begin{aligned}\gamma' &= \frac{2}{n^2(n-1)} \sum_{(j,k)\in\underline{n}^2_{\neq}} \sum_{(r,s)\in\underline{n}^2_{\neq}} p_{j,r}(1-p_{j,s})p_{k,s}(1-p_{k,r}) \\ &= \frac{2}{n^2(n-1)} \sum_{(j,k)\in\underline{n}^2} \sum_{(r,s)\in\underline{n}^2_{\neq}} p_{j,r}(1-p_{j,s})p_{k,s}(1-p_{k,r}) \\ &\quad - \frac{2}{n^2(n-1)} \sum_{j\in\underline{n}} \sum_{(r,s)\in\underline{n}^2_{\neq}} p_{j,r}(1-p_{j,s})p_{j,s}(1-p_{j,r}) \\ &= \frac{2}{n^2(n-1)} \sum_{(r,s)\in\underline{n}^2_{\neq}} \Big(\sum_{j\in\underline{n}} p_{j,r}(1-p_{j,s})\Big)\Big(\sum_{k\in\underline{n}} p_{k,s}(1-p_{k,r})\Big) \\ &= \frac{2}{n-1} \sum_{(r,s)\in\underline{n}^2_{\neq}} \Big(\overline{p}_{\cdot,r} - \frac{1}{n}\sum_{j\in\underline{n}} p_{j,r}p_{j,s}\Big)\Big(\overline{p}_{\cdot,s} - \frac{1}{n}\sum_{j\in\underline{n}} p_{j,r}p_{j,s}\Big). \quad \blacksquare\end{aligned}$$

Proof of Lemma 2.1. From Lemma 2.2(2), it follows that

$$\gamma' \leqslant \frac{2}{n}(\operatorname{Var} S_n - \lambda + \lambda^2).$$

Therefore $A \leqslant \lambda - \operatorname{Var} S_n + \frac{2}{n}(\operatorname{Var} S_n - \lambda + \lambda^2) = B$. From (4.11), we get $\gamma + \gamma' = \gamma'' \geqslant 0$. Further, the Cauchy–Schwarz inequality implies that $\lambda^2 = (\sum_{j=1}^n \overline{p}_{j,\cdot})^2 \leqslant n \sum_{j=1}^n \overline{p}_{j,\cdot}^2$. Therefore, using (1.8), we obtain

$$\begin{aligned}B &= \frac{n-2}{n}\Big(\sum_{j=1}^n \overline{p}_{j,\cdot}^2 + \gamma\Big) + \frac{2\lambda^2}{n} \leqslant \frac{n-2}{n}\Big(\sum_{j=1}^n \overline{p}_{j,\cdot}^2 + \gamma\Big) + 2\sum_{j=1}^n \overline{p}_{j,\cdot}^2 \\ &\leqslant \Big(3 - \frac{2}{n}\Big)(\lambda - \operatorname{Var} S_n) + 2\gamma' \leqslant \Big(3 - \frac{2}{n}\Big)A. \quad \blacksquare\end{aligned}$$

REFERENCES

[1] R. Adamczak, B. Polaczyk and M. Strzelecki, *Modified log-Sobolev inequalities, Beckner inequalities and moment estimates*, J. Funct. Anal. 282 (2022), art. 109349, 76 pp.

[2] A. D. Barbour and G. K. Eagleson, *Poisson approximation for some statistics based on exchangeable trials*, Adv. Appl. Probab. 15 (1983), 585–600.

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

[4] A. D. Barbour and L. Holst, *Some applications of the Stein–Chen method for proving Poisson convergence*, Adv. Appl. Probab. 21 (1989), 74–90.

[5] A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Clarendon Press, Oxford, 1992.

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

[7] A. D. Barbour, A. Röllin and N. Ross, *Error bounds in local limit theorems using Stein's method*, Bernoulli 25 (2019), 1076–1104.

[8] P. Brändén, J. Haglund, M. Visontai and D. G. Wagner, *Proof of the monotone column permanent conjecture*, in: Notions of Positivity and the Geometry of Polynomials, Springer, Basel, 2011, 63–78.

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

[10] L. H. Y. Chen, *An approximation theorem for sums of certain randomly selected indicators*, Z. Wahrsch. Verw. Gebiete 33 (1975/76), 69–74.

[11] P. Deheuvels and D. Pfeifer, *A semigroup approach to Poisson approximation*, Ann. Probab. 14 (1986), 663–676.

[12] P. Deheuvels and D. Pfeifer, *On a relationship between Uspensky's theorem and Poisson approximations*, Ann. Inst. Statist. Math., 40 (1988), 671–681.

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

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

[15] Yu. V. Prokhorov, *Asymptotic behavior of the binomial distribution*, Uspekhi Mat. Nauk (N.S.) 8 (1953), 135–142 (in Russian); English transl.: Selected Translations in Mathematical Statistics and Probability 1, Amer. Math. Soc., Providence, RI, 1961, 87–95.

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

[17] B. Roos, *Binomial approximation to the Poisson binomial distribution: the Krawtchouk expansion*, Teor. Veroyatnost. i Primenen. 45 (2000), 328–344; also Theory Probab. Appl. 45 (2001), 258–272.

[18] B. Roos, *Sharp constants in the Poisson approximation*, Statist. Probab. Lett. 52 (2001), 155–168.

[19] B. Roos, *Improvements in the Poisson approximation of mixed Poisson distributions*, J. Statist. Plann. Inference 113 (2003), 467–483.

[20] B. Roos, *Smoothness and Lévy concentration function inequalities for distributions of random diagonal sums*, Theory Probab. Math. Statist. 111 (2024), 137–151.

[21] S. Y. Shorgin, *Approximation of a generalized binomial distribution*, Teor. Veroyatnost. i Primenen. 22 (1977), 867–871 (in Russian); English transl.: Theory Probab. Appl. 22 (1977), 846–850.

[22] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, in: Proc. 6th Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, CA, 1970/1971), Vol. II: Probability Theory, Univ. of California Press, Berkeley, CA, 1972, 583–602.

Bero Roos
FB IV–Mathematics
Trier University
54286 Trier, Germany
E-mail: bero.roos@uni-trier.de

Received 4.9.2024;
accepted 3.2.2025