

On the Rate of Multivariate Poisson Convergence

Bero Roos

Institut für Mathematische Stochastik, Universität Hamburg, Germany

E-mail: roos@math.uni-hamburg.de

Received March 20, 1998; revised August 31, 1998

The distribution of the sum of independent nonidentically distributed Bernoulli random vectors in \mathbf{R}^k is approximated by a multivariate Poisson distribution. By using a multivariate adaption of Kerstan's (1964, *Z. Wahrsch. verw. Gebiete* 2, 173–179) method, we prove a conjecture of Barbour (1988, *J. Appl. Probab.* 25A, 175–184) on removing a log-term in the upper bound of the total variation distance. Second-order approximations are included. © 1999 Academic Press

AMS 1991 subject classifications: 60F05, 60G50.

Key words and phrases: Bernoulli random vectors; multivariate Poisson approximation; total variation distance.

1. INTRODUCTION

1.1. The General Setting

In this paper, we consider the distribution P^{S_n} of the sum S_n of $n \in \mathbf{N} = \{1, 2, \dots\}$ independent Bernoulli random vectors X_1, \dots, X_n in \mathbf{R}^k ($k \in \mathbf{N}$) with probabilities

$$P(X_i = e_r) = p_{i,r} \in [0, 1], \quad i \in \{1, \dots, n\}, \quad r \in \{1, \dots, k\},$$

$$P(X_i = (0, \dots, 0)) = 1 - \sum_{r=1}^k p_{i,r} \in [0, 1], \quad i \in \{1, \dots, n\}.$$

Here, $e_r \in \mathbf{R}^k$ denotes the vector with entry 1 at position r and entry 0 otherwise. Let $\lambda(r) = \sum_{i=1}^n p_{i,r} > 0$ for $r \in \{1, \dots, k\}$ be the mean of the r th component $S_n(r)$ of S_n .

1.2. The Problem

The task here is to give convenient bounds for the approximation error between P^{S_n} and the multivariate Poisson distribution $\mathcal{P}(\lambda)$ consisting of independent components with mean vector $\lambda = (\lambda(1), \dots, \lambda(k))$, that is,

$$\mathcal{P}(\lambda)(\{m\}) = \prod_{r=1}^k \left(e^{-\lambda(r)} \frac{\lambda(r)^{m_r}}{m_r!} \right), \quad m = (m_1, \dots, m_k) \in \mathbf{Z}_+^k,$$

where $\mathbf{Z}_+ = \{0, 1, 2, \dots\}$. As a measure of accuracy, we consider the total variation distance, that is, we search for bounds for

$$d_\tau := \frac{1}{2} \sum_{m \in \mathbf{Z}_+^k} |P^{S_n}(\{m\}) - \mathcal{P}(\lambda)(\{m\})| = \sup_{A \in \mathbf{Z}_+^k} |P(S_n \in A) - \mathcal{P}(\lambda)(A)|.$$

Our choice of a Poisson distribution with independent components is justified by McDonald [11, Theorem 1], who proved that

$$d_\tau \leq \sum_{i=1}^n \left(\sum_{r=1}^k p_{i,r} \right)^2 = E \left[\sum_{r=1}^k S_n(r) \right] - \text{Var} \left[\sum_{r=1}^k S_n(r) \right]. \quad (1)$$

In what follows, we are interested in sharper bounds.

1.3. What Is Already Known?

First we look at the univariate case $k=1$. Then the problem reduces to the Poisson approximation of the Poisson binomial distribution, a setting studied by many authors (for example, see Barbour and Hall [4], Borovkov [6], Daley and Vere-Jones [7, pp. 297–299], Deheuvels and Pfeifer [8], Kerstan [10], Presman [12], Roos [13, 14], Serfling [16], Witte [17], and the references therein). The most beautiful estimate came from Barbour and Hall [4, Theorems 1 and 2], who used the Stein–Chen method to prove, for $k=1$,

$$\frac{1}{32} \min \left\{ \frac{1}{\lambda(1)}, 1 \right\} \sum_{i=1}^n p_{i,1}^2 \leq d_\tau \leq \frac{1 - e^{-\lambda(1)}}{\lambda(1)} \sum_{i=1}^n p_{i,1}^2 \leq \min \left\{ \frac{1}{\lambda(1)}, 1 \right\} \sum_{i=1}^n p_{i,1}^2. \quad (2)$$

Hence in the case $k=1$, d_τ and $\min\{\lambda(1)^{-1}, 1\} \sum_{i=1}^n p_{i,1}^2$ are of the same order and it easily follows that d_τ tends to zero if and only if $\lambda(1)^{-1} \sum_{i=1}^n p_{i,1}^2 = 1 - \text{Var}[S_n(1)]/E[S_n(1)]$ tends to zero.

Using a multivariate Charlier expansion, Roos [15, Corollary 1] gave the following bound for the multivariate case $k \in \mathbf{N}$:

$$\begin{aligned} d_\tau &\leq \frac{1}{2 - \sqrt{3}} \left[\sum_{r=1}^k \sqrt{\min \left\{ \frac{1}{\lambda(r)}, 2e \right\} \sum_{i=1}^n p_{i,r}^2} \right]^2 \\ &\leq \frac{2ke}{2 - \sqrt{3}} \sum_{r=1}^k \left[\min \left\{ \frac{1}{\lambda(r)}, 1 \right\} \sum_{i=1}^n p_{i,r}^2 \right]. \end{aligned} \quad (3)$$

This estimate generalizes the one-dimensional upper bound (2) if one considers the order only. It was also shown that an inequality $d_\tau \leq Mk^\alpha \sum_{r=1}^k [\min\{\lambda(r)^{-1}, 1\} \sum_{i=1}^n p_{i,r}^2]$ with absolute constants $M \in (0, \infty)$ and $\alpha \in [0, 1)$ cannot hold; further, (3) and the lower bound in (2)

were used to prove that, in case of bounded or fixed dimension k , the distance d_τ tends to zero if and only if $\sum_{r=1}^k [\lambda(r)^{-1} \sum_{i=1}^n p_{i,r}^2] = \sum_{r=1}^k [1 - \text{Var}[S_n(r)]/E[S_n(r)]]$ tends to zero.

An interesting refinement of McDonald's bound (1) was derived by Barbour [3, Theorem 1], with the help of the Stein–Chen method, giving

$$d_\tau \leq \sum_{i=1}^n \min \left\{ c_A \sum_{r=1}^k \frac{p_{i,r}^2}{\lambda(r)}, \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\}, \quad (4)$$

where $c_A = \frac{1}{2} + \log^+(2A)$ and $A = \sum_{r=1}^k \lambda(r)$. He conjectured that it would be possible that (4) remains valid if c_A is replaced by an absolute constant $c \in (0, \infty)$. In Theorem 1, we prove as (6) a slightly weaker form of this conjecture: We replace c_A by $c = 8.8$ but put it outside the “min”. For practical purposes, we provide a much sharper bound, which has a somewhat different type [see (5)]. A further result of this paper gives two bounds in a second-order approximation of P^{S_n} (see Theorem 2). Considering the order, then, as (3), Barbour's conjectured bound leads to a generalization of the one-dimensional upper bound (2); further, for not necessarily bounded or fixed dimension k , the distance d_τ tends to zero if $\sum_{r=1}^k [\lambda(r)^{-1} \sum_{i=1}^n p_{i,r}^2]$ tends to zero.

For a further comparison of (3) and (4), let us assume that $k \in \mathbf{N}$ is arbitrary and that the vectors X_1, \dots, X_n are identically distributed, that is, $p_{1,r} = \dots = p_{n,r}$ for all $r \in \{1, \dots, k\}$. Then P^{S_n} is a multinomial distribution, for which Deheuvels and Pfeifer [9, Lemma 5.1] (see also Arenbaev [1, (9')]) proved that $d_\tau = \tilde{d}_\tau$, where \tilde{d}_τ is the total variation distance between the univariate binomial distribution with parameter $n \in \mathbf{N}$ and success probability $\sum_{r=1}^k p_{1,r}$ and the univariate Poisson distribution with mean $n \sum_{r=1}^k p_{1,r}$. Applying the upper bound in (2) to this setting, we get $d_\tau = \tilde{d}_\tau \leq \min \{ \sum_{r=1}^k p_{1,r}, n(\sum_{r=1}^k p_{1,r})^2 \}$, being a bound of correct order; in this case, the estimates (3) and (4) (using here the fact that c_A can indeed be replaced by an absolute constant c and put outside the “min”) lead to

$$d_\tau \leq \frac{1}{2 - \sqrt{3}} \left[\sum_{r=1}^k \sqrt{\min \{ p_{1,r}, 2en p_{1,r}^2 \}} \right]^2 \quad (3')$$

and

$$d_\tau \leq c \min \left\{ \sum_{r=1}^k p_{1,r}, n \left(\sum_{r=1}^k p_{1,r} \right)^2 \right\}. \quad (4')$$

Hence, Barbour's conjectured bound has also correct order in this case. By easy examples, one can show that (3') sometimes leads to weaker estimates:

If $p_{i,r} = p$ for all i and r then (3') and (4') give $d_\tau \leq 1/(2 - \sqrt{3}) \min\{k^2 p, 2enk^2 p^2\}$ and $d_\tau \leq c \min\{kp, nk^2 p^2\}$. The difference in the order is the factor k .

For other papers concerning the asymptotic behavior of d_τ , see Arenbaev [1] for the multinomial case and Deheuvels and Pfeifer [9] for the general case. Consult the references in Roos [15] for further papers on the multivariate problem.

1.4. The Method

We use a method originally due to Kerstan [10], who gave results in the univariate case. He treated independent Bernoulli summands as well as general independent summands, which also seem to be possible in the multivariate case. A refinement of the method, in the presented form, is the use of Cauchy's inequality for sequences (see the proof of Lemma 1). Indeed, in the univariate case, the resulting norm estimates and therefore the bounds for the total variation distance are sharper than those obtained by Witte [17], who improved Kerstan's method in this case and gave considerable sharp bounds (see Remark 6 for Theorem 1). The method used has advantages over the Stein–Chen method because of the quality of the second-order results (see Theorem 2); even in the univariate case, the Stein–Chen method does not give the order in (11) (see Remark 6 for Theorem 2). The presented method and that in Roos [15] are entirely different: Our main arguments are an expansion of the difference of generating functions with a somewhat different grouping of terms as in Kerstan [10] and Witte [17], the polynomial theorem, and Cauchy's inequality; we do not need any integrals. See Daley and Vere–Jones [7, pp. 297–299] and Roos [13, Kapitel 8] for further insights with respect to Kerstan's method in the univariate case.

2. RESULTS

THEOREM 1. Let $g(x) = 2 \sum_{s=2}^{\infty} x^{s-2}(s-1)/s!$ for $x \in \mathbf{R}$ and

$$\alpha = \sum_{i=1}^n g \left(2 \sum_{r=1}^k p_{i,r} \right) \min \left\{ 2^{-3/2} \sum_{r=1}^k \frac{p_{i,r}^2}{\lambda(r)}, \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\},$$

$$\beta = \sum_{i=1}^n \min \left\{ \sum_{r=1}^k \frac{p_{i,r}^2}{\lambda(r)}, \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\}.$$

If $\alpha < (2e)^{-1}$ then

$$d_\tau \leq \frac{\alpha}{1 - 2\alpha e}. \quad (5)$$

The estimate

$$d_\tau \leq c\beta, \quad (6)$$

is valid without any restrictions, where $c = 8.8$.

Remarks. (1) We have $g(x) \leq e^x$, for $x \geq 0$, and $g(x) = 2e^x(e^{-x} - 1 + x) \times x^{-2}$, for $x \neq 0$. Since $\sum_{r=1}^k p_{i,r} \in [0, 1]$ for $i \in \{1, \dots, n\}$, we get $\max_{1 \leq i \leq n} g(2 \sum_{r=1}^k p_{i,r}) \leq g(2) \leq 4.1946$.

(2) By (5), we can choose c nearly one if we assume that β and $\max_{1 \leq i \leq n} \sum_{r=1}^k p_{i,r}$ are small.

(3) An inequality $d_\tau \leq c\beta$ with an absolute constant $c < 1$ cannot hold, because, for $k = 1$, this would lead to $d_\tau \leq c\lambda(1)^{-1} \sum_{i=1}^n p_{i,1}^2 \leq c < 1$; but if we choose $p_{1,1} = \dots = p_{n,1} = 1$ then $d_\tau = 1 - e^{-n^n/n!}$ would tend to one, for $n \rightarrow \infty$, giving the contradiction.

(4) Inequality (6) has theoretical value. For practical usage, we prefer the sharper (5).

(5) Simple worst case considerations give the inequality $\beta \leq \min\{k, n\}$, which is sharp in the following sense: If $n \leq k$ and $p_{i,r}$ is 1 for $i = r \in \{1, \dots, n\}$ and 0 otherwise then we have $\beta = n$; further, if $k < n$ and $p_{i,r}$ is 1 for $i = r \in \{1, \dots, k\}$ and 0 otherwise then we get $\beta = k$. In the important cases of identically distributed random vectors X_1, \dots, X_n or $k = 1$, we obtain the bound $\beta \leq 1$ in a better correspondence with the trivial inequality $d_\tau \leq 1$.

(6) Using Kerstan's method, Witte [17, (1.10)] proved in the case $k = 1$ that

$$d_\tau \leq \frac{e^{2p_{0,1}} \theta(1)}{\sqrt{2\pi} [1 - 2\theta(1) e^{2p_{0,1}}]} \quad \text{if } \theta(1) := \lambda(1)^{-1} \sum_{i=1}^n p_{i,1}^2 < \frac{1}{2} e^{-2p_{0,1}}, \quad (7)$$

where $p_{0,1} = \max_{1 \leq i \leq n} p_{i,1}$; for a comparison, we derive from (5) the bound

$$d_\tau \leq \frac{g(2p_{0,1}) \theta(1)}{2^{3/2} [1 - 2^{-1/2} eg(2p_{0,1}) \theta(1)]} \quad \text{if } \theta(1) < \frac{\sqrt{2}}{eg(2p_{0,1})}, \quad (8)$$

which is always better than (7). However, Witte [17, (1.11)] gave a further, more complicated two-term bound, which is better than (8).

In the next theorem, we give results for the (second order) approximation of P^{S_n} by the finite signed measure Q , concentrated on \mathbf{Z}_+^k with counting density

$$Q(\{m\}) = \mathcal{P}(\lambda)(\{m\}) \left(1 - \frac{1}{2} \sum_{i=1}^n \left[\left(\sum_{r=1}^k \frac{p_{i,r}(m_r - \lambda(r))}{\lambda(r)} \right)^2 - \sum_{r=1}^k \frac{m_r p_{i,r}^2}{\lambda(r)^2} \right] \right), \quad (9)$$

for $m = (m_1, \dots, m_k) \in \mathbf{Z}_+^k$. Indeed, Q has finite total variation measure, because, as is shown in the proof of Theorem 2, we have $\sum_{m \in \mathbf{Z}_+^k} |Q(\{m\})| \leq 1 + 2\beta < \infty$ with β as in Theorem 1. Note that $Q(\mathbf{Z}_+^k) = 1$, being a necessary condition for a successful approximation. Consult (28) for the generating function of Q . Let

$$d'_\tau = \frac{1}{2} \sum_{m \in \mathbf{Z}_+^k} |P^{S_n}(\{m\}) - Q(\{m\})| = \sup_{A \subseteq \mathbf{Z}_+^k} |P(S_n \in A) - Q(A)|$$

be the total variation distance between P^{S_n} and Q .

THEOREM 2. *Let $h(x) = 3 \sum_{s=3}^{\infty} x^{s-3}(s-1)/s!$ for $x \in \mathbf{R}$ and $g(x)$ and β be defined as in Theorem 1. Further, let*

$$\gamma = \sum_{i=1}^n h \left(2 \sum_{r=1}^k p_{i,r} \right) \min \left\{ \frac{3}{2^{7/3}} \sum_{r=1}^k \frac{p_{i,r}^2}{\lambda(r)}, \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\}^{3/2},$$

$$\delta = \sum_{i=1}^n g \left(2 \sum_{r=1}^k p_{i,r} \right) \min \left\{ \sum_{r=1}^k \frac{p_{i,r}^2}{\lambda(r)}, \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\}.$$

If $\delta < 2^{1/2} e^{-1}$ then

$$d'_\tau \leq \frac{4}{3} \gamma + \delta^2 \left(1 + \frac{0.82 \delta}{1 - 2^{-1/2} \delta e} \right). \quad (10)$$

The estimate

$$d'_\tau \leq c \beta^{3/2}, \quad (11)$$

is valid without any restrictions, where $c = 28.3$.

Remarks. (1) We have $h(x) \leq e^x$, for $x \geq 0$, and $h(x) = 3(e^x(e^{-x} - 1 + x) - 2^{-1}x^2)x^{-3}$, for $x \neq 0$. As above for $g(x)$, $\max_{1 \leq i \leq n} h(2 \sum_{r=1}^k p_{i,r}) \leq h(2) \leq 2.3959$.

(2) The bound in (11) has weaker order than that in (10) but it seems to have best possible order as a function of β and does not have a singularity.

(3) By use of (10), we can choose c in (11) nearly $4/3$ if we assume that β and $\max_{1 \leq i \leq n} \sum_{r=1}^k p_{i,r}$ are small.

(4) An inequality $d'_\tau < c \beta^{3/2}$ with an absolute constant $c < 1$ cannot hold, because, in the case of $k=1$, we would then have $d'_\tau \leq c \beta \leq c < 1$; but if we choose $p_{1,1} = \dots = p_{n,1} = 1$ we get the contradiction by $d'_\tau \geq |P(S_n = n) - Q(\{n\})| = |1 - 3e^{-n}n^n/(2n!)|$, since the lower bound tends to one for $n \rightarrow \infty$.

(5) Roos [15, Theorem 1 with the choice $t = \lambda = (\lambda(1), \dots, \lambda(k))$, $u = 2$] derived the bound

$$d'_\tau \leq \sqrt{2} \frac{(\sum_{r=1}^k \sqrt{\kappa(r)})^3}{1 - \sum_{r=1}^k \sqrt{2\kappa(r)}} \quad \text{if} \quad \sum_{r=1}^k \sqrt{\kappa(r)} < 2^{-1/2}, \quad (12)$$

where $\kappa(r) = \min\{\lambda(r)^{-1}, 2e\} \sum_{i=1}^n p_{i,r}^2$. Indeed, the multivariate Charlier expansion was used for the approximation of P^{S_n} by signed measures of higher order related to $\mathcal{P}(t)$ with respect to an arbitrary mean vector $t = (t_1, \dots, t_k)$. The signed measure $Q(2, \lambda)$ of that paper is our Q . For a comparison with (11), we argue as in 1.3 with respect to (3) and (4) and see that (12) sometimes leads to weaker estimates than (11).

(6) Barbour [2, Corollary 2.4] (see also Barbour and Jensen [5, Theorem 1]) proved that, for $k=1$,

$$d'_\tau \leq 4 \frac{1 - e^{-\lambda(1)}}{\lambda(1)} \sum_{r=1}^n p_{i,1}^3 \leq 4 \min \left\{ \frac{1}{\lambda(1)}, 1 \right\} \sum_{i=1}^n p_{i,1}^3. \quad (13)$$

He used the Stein–Chen method and a Poisson type expansion similar to the Edgeworth expansion in the normal approximation for higher order approximations. His signed measure Q_2 is our Q for $k=1$. It should be mentioned that his signed measures of third and higher order do not coincide with those of the univariate Charlier expansion (see Roos [14, Theorem 2]). In the case $k=1$, the inequalities (11) and (12) yield an estimate of type $d'_\tau \leq c [\min\{\lambda(1)^{-1}, 1\} \sum_{i=1}^n p_{i,1}^2]^{3/2}$ with an absolute constant c (for (12), argue with $d'_\tau \leq 1 + \beta \leq 2$, ($k=1$), as is shown in the proof of Theorem 2). In case of $\lambda(1)$ being large, this inequality is often sharper than (13).

3. PROOFS

14. The Crux of the Method

The method used is originally due to Kerstan [10], who treated the univariate case. Its adaption to the multivariate problem is based on the

following expansion of the difference of the probability generating functions $\phi(z)$ and $\psi(z)$ of P^{S_n} and $\mathcal{P}(\lambda)$, respectively. (We take $z \in \mathbf{C}^k$, where \mathbf{C} denotes the set of complex numbers, and write $m = (m_1, \dots, m_k)$ whenever m has k components.) By independence,

$$\begin{aligned}
 \phi(z) - \psi(z) &= \sum_{m \in \mathbf{Z}_+^k} P(S_n = m) z_1^{m_1} \cdots z_k^{m_k} - \sum_{m \in \mathbf{Z}_+^k} \mathcal{P}(\lambda)(\{m\}) z_1^{m_1} \cdots z_k^{m_k} \\
 &= \prod_{j=1}^n \left(1 + \sum_{r=1}^k p_{j,r}(z_r - 1) \right) - \exp \left(\sum_{r=1}^k \lambda(r)(z_r - 1) \right) \\
 &= \left(\prod_{j=1}^n \left[\left(1 + \sum_{r=1}^k p_{j,r}(z_r - 1) \right) \exp \left(- \sum_{r=1}^k p_{j,r}(z_r - 1) \right) \right] - 1 \right) \\
 &\quad \times \exp \left(\sum_{r=1}^k \lambda(r)(z_r - 1) \right) \\
 &= \sum_{j=1}^n \sum_{1 \leq i(1) < \cdots < i(j) \leq n} \prod_{s=1}^j \left[L_{i(s)}(z) \exp \left(\frac{1}{j} \sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right], \tag{14}
 \end{aligned}$$

where

$$L_i(z) = \left(1 + \sum_{r=1}^k p_{i,r}(z_r - 1) \right) \exp \left(- \sum_{r=1}^k p_{i,r}(z_r - 1) \right) - 1 \tag{15}$$

$$= - \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right)^2 \sum_{s=2}^{\infty} \left(- \sum_{r=1}^k p_{i,r}(z_r - 1) \right)^{s-2} \frac{s-1}{s!}, \tag{16}$$

for $i \in \{1, \dots, n\}$. It was used that $(1+x)e^{-x} = 1 - x^2 \sum_{s=2}^{\infty} (-x)^{s-2} \times (s-1)/s!$ for $x \in \mathbf{C}$. Note that the expansion (14) has a somewhat different form as in Kerstan [10] and Witte [17]. The connection to the total variation distance is given by the identity $d_{\tau} = \frac{1}{2} \sum_{m \in \mathbf{Z}_+^k} |a_m|$, where the a_m , $m \in \mathbf{Z}_+^k$, are the coefficients of the power series $\phi(z) - \psi(z)$. During this paper, all of our power series take the form $\sum_{m \in \mathbf{Z}_+^k} b_m z_1^{m_1} \cdots z_k^{m_k}$ with real coefficients b_m , $m \in \mathbf{Z}_+^k$, and converge absolutely for all $z \in \mathbf{C}^k$. In particular, the order of summation may be chosen arbitrarily. For any such power series $f(z)$ with coefficients b_m , $m \in \mathbf{Z}_+^k$, we define the norm $\|f(z)\| = \sum_{m \in \mathbf{Z}_+^k} |b_m|$ and use the easy fact that $\|f_1(z) f_2(z)\| \leq \|f_1(z)\| \|f_2(z)\|$ for power series $f_1(z)$ and $f_2(z)$. The application of the triangle inequality and the polynomial theorem leads to

$$\begin{aligned}
d_\tau &= \frac{1}{2} \left\| \phi(z) - \psi(z) \right\| \\
&\leq \frac{1}{2} \sum_{j=1}^n \sum_{1 \leq i(1) < \dots < i(j) \leq n} \prod_{s=1}^j \left\| L_{i(s)}(z) \exp \left(\frac{1}{j} \sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \\
&\leq \frac{1}{2} \sum_{j=1}^n \frac{1}{j!} \left[\sum_{i=1}^n \left\| L_i(z) \exp \left(\frac{1}{j} \sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \right]^j. \tag{17}
\end{aligned}$$

We call (17) the fundamental estimate. Inequality (5) can immediately be shown after estimating the norm term in (17). For a small constant in (6) and the second-order approximation in Theorem 2, we consider the first terms in (14) and (17) separately. This will be done in the proofs of Theorems 1 and 2.

3.2. Norm Estimates and Remaining Proofs

LEMMA 1. *Let $t = (t_1, \dots, t_k) \in (0, \infty)^k$, $i \in \{1, \dots, n\}$, $g(x) = 2 \sum_{s=2}^{\infty} x^{s-2} \times (s-1)/s!$ for $x \in \mathbf{R}$. Then*

$$\begin{aligned}
&\left\| \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right) \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
&\leq \min \left\{ \sum_{r=1}^k \frac{p_{i,r}^2}{t_r}, \left(2 \sum_{r=1}^k p_{i,r} \right)^2 \right\}^{1/2}, \tag{18}
\end{aligned}$$

$$\begin{aligned}
&\left\| \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right)^2 \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
&\leq \min \left\{ \sqrt{2} \sum_{r=1}^k \frac{p_{i,r}^2}{t_r}, \left(2 \sum_{r=1}^k p_{i,r} \right)^2 \right\}, \tag{19}
\end{aligned}$$

$$\begin{aligned}
&\left\| L_i(z) \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
&\leq g \left(2 \sum_{r=1}^k p_{i,r} \right) \min \left\{ \frac{1}{\sqrt{2}} \sum_{r=1}^k \frac{p_{i,r}^2}{t_r}, 2 \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\}. \tag{20}
\end{aligned}$$

Proof. For the easy part of (18), note that

$$\left\| \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right) \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \leq \left\| \sum_{r=1}^k p_{i,r}(z_r - 1) \right\| = 2 \sum_{r=1}^k p_{i,r}.$$

We now prove the rest of (18). Simple calculus shows that

$$\begin{aligned} & \left(\sum_{r=1}^k p_{i,r}(z_r-1) \right) \exp \left(\sum_{r=1}^k t_r(z_r-1) \right) \\ &= \sum_{m \in \mathbf{Z}_+^k} \mathcal{P}(t)(\{m\}) \left[\sum_{r=1}^k \frac{p_{i,r}(m_r-t_r)}{t_r} \right] z_1^{m_1} \dots z_k^{m_k}, \end{aligned}$$

and the application of Cauchy's inequality leads to the following bound for the left-hand side of (18):

$$\left(\sum_{m \in \mathbf{Z}_+^k} \mathcal{P}(t)(\{m\}) \left[\sum_{r=1}^k \frac{p_{i,r}(m_r-t_r)}{t_r} \right]^2 \right)^{1/2} = \left(\sum_{r=1}^k \frac{p_{i,r}^2}{t_r} \right)^{1/2}.$$

The proof of (19) is analogous by using the identity

$$\begin{aligned} & \left(\sum_{r=1}^k p_{i,r}(z_r-1) \right)^2 \exp \left(\sum_{r=1}^k t_r(z_r-1) \right) \\ &= \sum_{m \in \mathbf{Z}_+^k} \mathcal{P}(t)(\{m\}) \left[\left(\sum_{r=1}^k \frac{p_{i,r}(m_r-t_r)}{t_r} \right)^2 - \sum_{r=1}^k \frac{m_r p_{i,r}^2}{t_r^2} \right] z_1^{m_1} \dots z_k^{m_k} \end{aligned} \quad (21)$$

and some straightforward calculus. In order to prove (20), we use (16) and obtain

$$\left\| L_i(z) \exp \left(\sum_{r=1}^k t_r(z_r-1) \right) \right\| \leq T \left\| \left(\sum_{r=1}^k p_{i,r}(z_r-1) \right)^2 \exp \left(\sum_{r=1}^k t_r(z_r-1) \right) \right\|,$$

where the latter term on the right-hand side can be estimated by (19) and

$$\begin{aligned} T &:= \left\| \sum_{s=2}^{\infty} \left(- \sum_{r=1}^k p_{i,r}(z_r-1) \right)^{s-2} \frac{s-1}{s!} \right\| \leq \sum_{s=2}^{\infty} \left\| \sum_{r=1}^k p_{i,r}(z_r-1) \right\|^{s-2} \frac{s-1}{s!} \\ &= \sum_{s=2}^{\infty} \left(2 \sum_{r=1}^k p_{i,r} \right)^{s-2} \frac{s-1}{s!} = \frac{1}{2} g \left(2 \sum_{r=1}^k p_{i,r} \right). \end{aligned}$$

Combining these inequalities, (20) is shown. \blacksquare

Remarks. (1) In the case $t_r \geq p_{i,r}$ for all $r \in \{1, \dots, k\}$ and fixed $i \in \{1, \dots, n\}$ (for example, if $t_r = \lambda(r)$ for all r), we have the estimate

$$\begin{aligned}
& \left\| L_i(z) \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
& \stackrel{(15)}{=} \left\| \left[1 + \sum_{r=1}^k p_{i,r}(z_r - 1) - \exp \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right) \right] \right. \\
& \quad \times \exp \left(\sum_{r=1}^k (t_r - p_{i,r})(z_r - 1) \right) \Big\| \\
& \leq \left\| 1 + \sum_{r=1}^k p_{i,r}(z_r - 1) - \exp \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right) \right\| \\
& = 2 \left(\sum_{r=1}^k p_{i,r} \right) \left[1 - \exp \left(- \sum_{r=1}^k p_{i,r} \right) \right] \tag{22}
\end{aligned}$$

$$\leq 2 \left(\sum_{r=1}^k p_{i,r} \right)^2. \tag{23}$$

Equality (22) indicates the value of $2d_\tau$ in the case $n=1$ and is easy to verify; for $k=1$, it can also be found in Serfling [16, Lemma 4.1]. Inequality (23) is the same as (1) in the case of $n=1$. Hence, under the above assumptions,

$$\begin{aligned}
& \left\| L_i(z) \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
& \leq \min \left\{ 2^{-1/2} g \left(2 \sum_{r=1}^k p_{i,r} \right) \sum_{r=1}^k \frac{p_{i,r}^2}{t_r}, 2 \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\}, \tag{24}
\end{aligned}$$

which is better than (20). Inequality (24) can be applied to the first summand in (17) (that is, in the case $j=1$).

(2) A weaker bound for the left-hand side of (19) can be obtained by (18) in the following easy way:

$$\begin{aligned}
& \left\| \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right)^2 \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
& \leq \left\| \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right) \exp \left(\frac{1}{2} \sum_{r=1}^k t_r(z_r - 1) \right) \right\|^2 \\
& \leq \min \left\{ 2 \sum_{r=1}^k \frac{p_{i,r}^2}{t_r}, \left(2 \sum_{r=1}^k p_{i,r} \right)^2 \right\}.
\end{aligned}$$

Proof of Theorem 1. We use the fundamental inequality (17), the bound (20), and Stirling's formula to obtain (5):

$$\begin{aligned}
d_\tau &\leq \frac{1}{2} \sum_{j=1}^n \frac{1}{j!} \left[\sum_{i=1}^n g \left(2 \sum_{r=1}^k p_{i,r} \right) \min \left\{ \frac{j}{\sqrt{2}} \sum_{r=1}^k \frac{p_{i,r}^2}{\lambda(r)}, 2 \left(\sum_{r=1}^k p_{i,r} \right)^2 \right\} \right]^j \\
&\leq \alpha + \frac{1}{2} \sum_{j=2}^{\infty} \frac{(2\alpha e)^j}{\sqrt{2\pi j}} \leq \alpha + \frac{(\alpha e)^2}{\sqrt{\pi} (1 - 2\alpha e)} \leq \frac{\alpha}{1 - 2\alpha e} \quad \text{if } \alpha < (2e)^{-1}.
\end{aligned}$$

For the proof of (6), we use the fundamental inequality but write it down in this way:

$$\begin{aligned}
d_\tau &\leq \frac{1}{2} \sum_{i=1}^n \left\| L_i(z) \exp \left(\sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \\
&\quad + \frac{1}{4} \left[\sum_{i=1}^n \left\| L_i(z) \exp \left(\frac{1}{2} \sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \right]^2 \\
&\quad + \frac{1}{2} \sum_{j=3}^{\infty} \frac{1}{j!} \left[\sum_{i=1}^n \left\| L_i(z) \exp \left(\frac{1}{j} \sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \right]^j.
\end{aligned}$$

Applying (24) and (20) to the first and the remaining terms, respectively, we get $d_\tau \leq \min\{1, f(g(2)\beta)\}$, where $f(x) = 2^{-3/2}x + x^2 + \frac{1}{2} \sum_{j=3}^{\infty} (jx/\sqrt{2})^j \times (1/j!)$, for $x \geq 0$. If $x_0 \in (0, \infty)$ denotes the unique positive solution of the equation $f(x) = 1$ then $d_\tau \leq g(2)\beta/x_0$. Numerical computations give $0.477 < x_0 < 0.478$ and hence $d_\tau \leq 8.8\beta$. ■

In order to prove Theorem 2, we define

$$M_i(z) = \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right)^3 \sum_{s=3}^{\infty} \left(- \sum_{r=1}^k p_{i,r}(z_r - 1) \right)^{s-3} \frac{s-1}{s!} \quad (25)$$

for $i \in \{1, \dots, n\}$ and $z \in \mathbf{C}^k$, which can be obtained if we remove the first summand in the representation (16) of $L_i(z)$. We need the following norm estimates.

LEMMA 2. *Let $t = (t_1, \dots, t_k) \in (0, \infty)^k$, $i \in \{1, \dots, n\}$, and $h(x) = 3 \sum_{s=3}^{\infty} x^{s-3}(s-1)/s!$ for $x \in \mathbf{R}$. Then*

$$\begin{aligned}
&\left\| \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right)^3 \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
&\leq \min \left\{ \frac{3}{2^{1/3}} \sum_{r=1}^k \frac{p_{i,r}^2}{t_r}, \left(2 \sum_{r=1}^k p_{i,r} \right)^2 \right\}^{3/2}, \quad (26)
\end{aligned}$$

$$\begin{aligned}
&\left\| M_i(z) \exp \left(\sum_{r=1}^k t_r(z_r - 1) \right) \right\| \\
&\leq \frac{1}{3} h \left(2 \sum_{r=1}^k p_{i,r} \right) \min \left\{ \frac{3}{2^{1/3}} \sum_{r=1}^k \frac{p_{i,r}^2}{t_r}, \left(2 \sum_{r=1}^k p_{i,r} \right)^2 \right\}^{3/2}. \quad (27)
\end{aligned}$$

Proof. We estimate the left-hand side of (26) and receive the product of terms as in (18) and (19), using the weights $1/3$ and $2/3$ for the vector t . By application of (18) and (19), the inequality (26) is shown. The proof of (27) uses (26) and is similar to that of (20). ■

Proof of Theorem 2. Let $\tilde{\psi}(z) = \sum_{m \in \mathbf{Z}_+^k} Q(\{m\}) z_1^{m_1} \cdots z_k^{m_k}$, $z \in \mathbf{C}^k$, be the generating function of Q . Remembering (21), it easy is to see that

$$\tilde{\psi}(z) = \left[1 - \frac{1}{2} \sum_{i=1}^n \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right)^2 \right] \exp \left(\sum_{r=1}^k \lambda(r)(z_r - 1) \right). \quad (28)$$

By (14) and (16), we get

$$\begin{aligned} d'_\tau &= \frac{1}{2} \|\phi(z) - \psi(z) + \psi(z) - \tilde{\psi}(z)\| \\ &= \frac{1}{2} \left\| \sum_{i=1}^n M_i(z) \exp \left(\sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right. \\ &\quad \left. + \sum_{j=2}^n \sum_{1 \leq i(1) < \cdots < i(j) \leq n} \prod_{s=1}^j \left[L_{i(s)}(z) \exp \left(\frac{1}{j} \sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right] \right\| \\ &\leq \frac{1}{2} \sum_{i=1}^n \left\| M_i(z) \exp \left(\sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \\ &\quad + \frac{1}{2} \sum_{j=2}^\infty \frac{1}{j!} \left[\sum_{i=1}^n \left\| L_i(z) \exp \left(\frac{1}{j} \sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \right]^j. \end{aligned}$$

For (10), we proceed as in the proof of Theorem 1: We consider the summand for $j=2$ separately and apply (20) and (27). The proof of (11) is similar: By (28) and (19),

$$\begin{aligned} \sum_{m \in \mathbf{Z}_+^k} |Q(\{m\})| &\leq 1 + \frac{1}{2} \sum_{i=1}^n \left\| \left(\sum_{r=1}^k p_{i,r}(z_r - 1) \right)^2 \exp \left(\sum_{r=1}^k \lambda(r)(z_r - 1) \right) \right\| \\ &\leq 1 + 2\beta. \end{aligned}$$

Hence, we have $d'_\tau \leq \frac{1}{2} \sum_{m \in \mathbf{Z}_+^k} [P(S_n = m) + |Q(\{m\})|] \leq 1 + \beta$. Therefore, as in the proof of Theorem 1,

$$d'_\tau \leq \min\{1 + \beta, f(\beta)\} \leq \left(\frac{\beta}{x_0} \right)^{3/2} (1 + x_0),$$

where

$$f(x) = \frac{4}{3} h(2) x^{3/2} + (g(2) x)^2 + \frac{1}{2} \sum_{j=3}^{\infty} \left(\frac{g(2) j x}{\sqrt{2}} \right)^j \frac{1}{j!}, \quad x \geq 0,$$

and x_0 is the unique positive solution of the equation $f(x) = 1 + x$. Numerical computations give $0.1159 < x_0 < 0.1160$, leading to $d'_\tau \leq 28.3 \beta^{3/2}$. ■

4. OPEN PROBLEMS

Theorems 1 and 2 and the method used give rise to the following open problems:

1. Find an expansion, which gives the error bounds as in (6) and (11) for the first and the second order approximations. What are the signed measures of higher order? Give the accompanying error bounds. Of course, we could consider the first summands in (14) but the resulting signed measures are not easy to handle and are therefore unfavorable. Further, the multivariate Charlier expansion (see Roos [15]) seems to be no candidate.

2. Can the method of this paper be applied to other metrics such as the Kolmogorov metric and the point metric? In the univariate case, Witte [17] has shown that this is possible. He treated the total variation distance, the Kolmogorov metric, and the Fortet–Mourier metric. In Roos [13, Kapitel 8], Kerstan's method was used for results with respect to the point metric in the univariate case. In the multivariate case, some technical difficulties occur. It should be mentioned that the method in Roos [15] allows the treatment of several probability metrics.

3. What are the best possible constants in (6) and (11)? As stated in Section 2, they must lie in the intervals $[1, 8.8]$ and $[1, 28.3]$, respectively. The upper bound in (2) shows that, in the case of $k = 1$, the best possible constant in (6) is $c = 1$.

4. Is it true that, for arbitrary behavior of the dimension k , the distance d_τ tends to zero if and only if $\sum_{r=1}^k [\lambda(r)^{-1} \sum_{i=1}^n p_{i,r}^2]$ tends to zero? As mentioned in 1.3, this statement is true for bounded k and, further, the “if”-part is true for arbitrary behavior of k .

5. Can this method be applied to arbitrary independent X_1, \dots, X_n ? In the case $k = 1$, Kerstan's [10] answer is yes.

REFERENCES

1. N. K. Arenbaev, Asymptotic behaviour of the multinomial distribution, *Teor. Veroyatnost. i Primenen.* **21** (1976), 826–831 [Russian]; *Theory Probab. Appl.* **21** (1976), 805–810. [Engl. transl.]
2. A. D. Barbour, Asymptotic expansions in the Poisson limit theorem, *Ann. Probab.* **15** (1987), 748–766.
3. A. D. Barbour, Stein's method and Poisson process convergence, *J. Appl. Probab.* **25A** (1988), 175–184.
4. A. D. Barbour and P. Hall, On the rate of Poisson convergence, *Math. Proc. Cambridge Philos. Soc.* **95** (1984), 473–480.
5. A. D. Barbour and J. L. Jensen, Local and tail approximations near the Poisson limit, *Scand. J. Statist.* **16** (1989), 75–87.
6. K. A. Borovkov, Refinement of Poisson approximation, *Teor. Veroyatnost. i Primenen.* **33** (1988), 364–368 [Russian]; *Theory Probab. Appl.* **33** (1988), 343–347. [Engl. transl.]
7. D. J. Daley and D. Vere-Jones, "An Introduction to the Theory of Point Processes," Springer-Verlag, New York, 1988.
8. P. Deheuvels and D. Pfeifer, A semigroup approach to Poisson approximation, *Ann. Probab.* **14** (1986), 663–676.
9. P. Deheuvels and D. Pfeifer, Poisson approximations of multinomial distributions and point processes, *J. Multivariate Anal.* **25** (1988), 65–89.
10. J. Kerstan, Verallgemeinerung eines Satzes von Prochorow und Le Cam, *Z. Wahrsch. verw. Gebiete* **2** (1964), 173–179.
11. D. R. McDonald, On the Poisson approximation to the multinomial distribution, *Canad. J. Statist.* **8** (1980), 115–118.
12. É. L. Presman, Approximation in variation of the distribution of a sum of independent Bernoulli variables with a Poisson law, *Teor. Veroyatnost. i Primenen.* **30** (1985), 391–396 [Russian]; *Theory Probab. Appl.* **30** (1985), 417–422. [Engl. transl.]
13. B. Roos, "Metrische Poisson Approximation," Ph.D. dissertation, Fachbereich Mathematik, Universität Oldenburg, June 1996a.
14. B. Roos, Asymptotics and sharp bounds in the Poisson approximation to the Poisson binomial distribution, Preprint 96–14, Institut für Mathematische Stochastik, Universität Hamburg, September 1996b, submitted for publication.
15. B. Roos, Metric multivariate Poisson approximation of the generalized multinomial distribution, *Teor. Veroyatnost. i Primenen.* **43** (1998), 404–413.
16. R. J. Serfling, Some elementary results on Poisson approximation in a sequence of Bernoulli trials, *SIAM Rev.* **20** (1978), 567–579.
17. H.-J. Witte, A unification of some approaches to Poisson approximation, *J. Appl. Probab.* **27** (1990), 611–621.