

Poisson Approximations of Multinomial Distributions and Point Processes

PAUL DEHEUVELS

Université Paris VI, Paris, France

AND

DIETMAR PFEIFER

University of Oldenburg, West Germany

We present an extension to the multinomial case of former estimations for univariate Poisson binomial approximation problems and generalize a result obtained by N. K. Arenbaev (*Theory Probab. Appl.* **21** (1976), 805–810). As an application, we evaluate the total variation distance between superpositions of independent Bernoulli point processes and a suitable Poisson process. The main tool will be a multiparameter semigroup approach. © 1988 Academic Press, Inc.

1. INTRODUCTION

Operator methods in connection with Poisson approximation problems have received some attention recently (Shur [18], Presman [15], Barbour and Hall [2], Deheuvels and Pfeifer [5–7], and Pfeifer [13, 14]), extending or improving an approach introduced originally by LeCam [12].

All these papers deal with the univariate case, giving estimations or asymptotic expansions for distances between the distribution of sums of independent Bernoulli summands and a suitable Poisson distribution.

In the following, we give an extension of the semigroup approach developed in Deheuvels and Pfeifer [5, 6] for a general multinomial approximation with respect to the total variation distance. This problem

AMS 1980 subject classifications: 60G50, 60F05, 20M99.

Key words and phrases: Poisson approximations, multinomial distributions, total variation distance, point processes, semigroups.

has been studied by Arenbaev [1] in the case of i.i.d. multinomial summands, and we generalize his results in a wider setting.

Our methods can be applied to the estimation of the total variation distance between the superposition of independent point processes and a suitable Poisson process (see, e.g., Serfozo [17] and the references therein). The estimations we obtain are generally sharper than those obtained by martingale (the papers of Freedman [8] and Serfling [16] are essentially a martingale approach in discrete time) and compensator approaches (the first compensator approach of this problem is given in Brown [3, 4]; see, e.g., Valkheila [19], Kabanov, Liptser, and Shiryaev [10], and the references therein), due to the specific setting tailored to the Poissonian semigroup.

This paper is organized as follows. In Section 1 we give our main results for multinomial approximations. Section 2 is devoted to the semigroup evaluations. In Sections 4 and 5, we compute the leading terms of our expansions. Section 6 contains our results for point processes.

2. MAIN RESULTS

Let $k \geq 1$ be a fixed integer, and let $Z_n = (Z_n(1), \dots, Z_n(k))$, $n = 1, 2, \dots$, be a sequence of independent random vectors of \mathbb{R}^k , such that, for $j = 1, \dots, k$,

$$\begin{aligned} P(Z_n(1) = \dots = Z_n(j-1) = 0, Z_n(j) = 1, Z_n(j+1) = \dots = Z_n(k) = 0) \\ = p_{nj} \geq 0, \end{aligned}$$

$$P(Z_n(1) = \dots = Z_n(k) = 0) = 1 - \sum_{j=1}^k p_{nj} = 1 - P_n \geq 0.$$

Consider $S_n = \sum_{i=1}^n Z_i = (S_n(1), \dots, S_n(k))$, where $S_n(j) = \sum_{i=1}^n Z_i(j)$, $j = 1, \dots, k$.

Let, for $j = 1, \dots, k$, $\lambda_j = \sum_{i=1}^n p_{ij}$, and define $T_n = (T_n(1), \dots, T_n(k))$ as a random vector such that $T_n(1), \dots, T_n(k)$ are independent, and that $T_n(j)$ follows a Poisson distribution with mean λ_j , $j = 1, \dots, k$.

The main purpose of this paper is to provide sharp evaluations for the total variation distance Δ_n between the distribution $L(S_n)$ of S_n and $L(T_n)$ of T_n :

$$\begin{aligned} \Delta_n = d_v(L(S_n), L(T_n)) &= \sup_E |P(S_n \in E) - P(T_n \in E)| \\ &= \frac{1}{2} \sum_m |P(S_n = m) - P(T_n = m)|. \end{aligned} \tag{2.1}$$

Up to now, this problem has received attention in the particular cases listed in the examples below.

EXAMPLE 2.1. For $k = 1$, Z_n follows a Bernoulli $B(P_n)$ distribution, and T_n a Poisson distribution with mean $\sum_{i=1}^n P_i$. This is the classical Poisson approximation problem for sums of independent Bernoulli random variables which has received an extensive treatment (see the references in Section 1).

EXAMPLE 2.2. When $p_{ij} = p_j$ is independent of $i = 1, \dots, n$ for all $j = 1, \dots, k$, S_n follows a multinomial distribution such that

$$P(S_n(1) = r_1, \dots, S_n(k) = r_k) = \frac{n!}{r_1! \dots r_k! (n-R)!} p_1^{r_1} \dots p_k^{r_k} \left(1 - \sum_{j=1}^k p_j\right)^{n-R}, \quad (2.2)$$

where $r_1 \geq 0, \dots, r_k \geq 0$, and $R = \sum_{j=1}^k r_j \leq n$.

Under these assumptions, Arenbaev [1] has proved that, whenever p_1, \dots, p_k are fixed and $\sum_{j=1}^k p_j > 0$, we have, as $n \rightarrow \infty$,

$$\frac{1}{2} \sum_{\substack{r_1 + \dots + r_k \leq n \\ r_j \geq 0, 1 \leq j \leq k}} |P(S_n = r) - P(T_n = r)| \\ = \left\{ \sum_{j=1}^k p_j \right\} \left(\frac{1}{\sqrt{2\pi e}} + O \left(\left\{ n \sum_{j=1}^k p_j \right\}^{-1/2} \right) \right). \quad (2.3)$$

Let $U_n = T_n(1) + \dots + T_n(k)$ and $P = \sum_{j=1}^k p_j$. Routine manipulations show that $P(U_n \geq n+1) \leq (P/(1-P))(2\pi n)^{-1/2} \exp(-n(P-1-\log P)) = o(n^{-1/2})$ as $n \rightarrow \infty$. Hence, Arenbaev has shown that, for any fixed $0 < P < 1$,

$$\Delta_n = \frac{P}{\sqrt{2\pi e}} \left(1 + O \left(\frac{1}{\sqrt{nP}} \right) \right) \quad \text{as } n \rightarrow \infty. \quad (2.4)$$

In our first theorem, we show, among other results, that the validity of (2.4) can be extended to the case where p_1, \dots, p_k vary with n .

THEOREM 2.1. Assume that $p_1 = p_{i1}, \dots, p_k = p_{ik}$ are independent of $i = 1, \dots, n$. Let $P = \sum_{j=1}^k p_j$ and $\theta = nP$. Then

$$\Delta_n = \frac{1}{2} P\theta \left(\frac{\theta^{\alpha-1}(\alpha-\theta)}{\alpha!} - \frac{\theta^{\beta-1}(\beta-\theta)}{\beta!} \right) e^{-\theta} + R_n \\ = \frac{1}{4} d_n + R_n = \frac{P}{4} D(\theta) + R_n, \quad (2.5)$$

where (with $[u]$ denoting the integer part of u)

$$\alpha = [\theta + \frac{1}{2} + (\theta + \frac{1}{4})^{1/2}], \quad \beta = [\theta + \frac{1}{2} - (\theta + \frac{1}{4})^{1/2}],$$

and

$$|R_n| \leq \min\{5nP^3, 16P^2\}.$$

Furthermore, if $n \geq 1$ and P vary in such a way that $P \rightarrow 0$, we always have

$$\Delta_n = \frac{1}{4}d_n(1 + O(P)) \sim \frac{1}{4}d_n. \quad (2.6)$$

In addition, if

$$\theta = nP \rightarrow \infty \quad \text{and} \quad P \rightarrow 0, \quad (2.7)$$

then

$$\Delta_n = \frac{P}{\sqrt{2\pi e}} \left(1 + O \left(\max \left\{ P, \frac{1}{\sqrt{nP}} \right\} \right) \right). \quad (2.8)$$

On the other hand, if

$$\theta = nP \rightarrow 0, \quad (2.9)$$

then

$$\Delta_n = nP^2(1 + O(P)). \quad (2.10)$$

Proof. It follows from Lemma 5.1 in the sequel. Observe that (2.6) and (2.8) follow from (2.5) by straightforward expansions as in Deheuvels and Pfeifer [5].

Remark 2.1. In Theorem 2.1, we do not make any growth assumption on $n \geq 1$. Likewise, $k \geq 1$ need not remain fixed. In particular we may not assume that $n \rightarrow \infty$. On the other hand, the estimations of the error term R_n lack precision when $P \not\rightarrow 0$.

By choosing an arbitrary $0 < \varepsilon < 1$ and by applying Arenbaev's [1] technique for $P \geq \varepsilon$ and Theorem 2.1 for $0 < P < \varepsilon$, we can easily prove that, if $n \geq 1$ and P vary in such a way that $nP \rightarrow \infty$ and $0 < P < 1 - \delta$ for some fixed $\delta > 0$, then

$$\Delta_n = \frac{P}{\sqrt{2\pi e}} (1 + o(1)). \quad (2.11)$$

The evaluations in (2.5) cover the situation where $nP \rightarrow a \in (0, \infty)$, $n \rightarrow \infty$, $P \rightarrow 0$, in which case we have $\Delta_n = (P/4) D(a)(1 + o(1))$.

Our next theorem deals with the general case. We obtain the following results.

THEOREM 2.2. *Let p_{ij} , $j = 1, \dots, k$, $i = 1, \dots, n$, be arbitrary. We have*

$$A_n = \frac{1}{4}d_n + R_n, \quad (2.12)$$

where

$$d_n = E \left(\left| \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{\tau_j}{\lambda_j} \right) \right\}^2 - \sum_{i=1}^n \sum_{j=1}^k \frac{p_{ij}^2}{\lambda_j^2} \tau_j \right| \right) \\ \leq \min \left(4 \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^2, 2 \sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-1} \right), \quad (2.13)$$

$$|R_n| \leq \exp \left(4K \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^2 \right) \\ \times \left(\frac{8}{3} (K-1)^2 \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^3 + 8K(K-1) \left\{ \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^2 \right\}^2 \right) \\ + 8(K-1) \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^3, \quad (2.14)$$

$$R_n = o(d_n) \quad \text{if} \quad \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^2 = O(1),$$

$$K = 1 + \frac{1}{2} \exp \left(2 \max_{1 \leq i \leq n} \sum_{j=1}^k p_{ij} \right) \leq 1 + \frac{1}{2} e^2 < 4.70,$$

and where τ_1, \dots, τ_k denote independent Poisson random variables with expectations $E(\tau_j) = \sum_{i=1}^n p_{ij}$, $j = 1, \dots, k$. Throughout, we use the convention that $0/0 = 0$.

Proof. It follows from (3.19)–(3.30) and (5.3) in the sequel.

Remark 2.2. By (2.10) we see that the upper bound evaluation of the leading term $\frac{1}{4}d_n$ in (2.13) is sharp in the range where $\sum_{i=1}^n \sum_{j=1}^k p_{ij} \rightarrow 0$. On the other hand, in the case covered by Theorem 2.1, if $\sum_{i=1}^n \sum_{j=1}^k p_{ij} = n \sum_{j=1}^k p_j \rightarrow \infty$, (2.13) yields the upper bound

$$\frac{1}{4}d_n \leq \frac{1}{2} \left\{ \sum_{j=1}^k p_j \right\}, \quad (2.15)$$

to be compared with the exact asymptotic coefficient given by (2.8):

$$\frac{1}{4}d_n \sim \frac{1}{\sqrt{2\pi e}} \left\{ \sum_{j=1}^k p_j \right\}. \quad (2.16)$$

It can be verified that $1/\sqrt{2\pi e} \simeq 0.242 < \frac{1}{2} = 0.5$. It follows that in this range, the upper bound in (2.13) is sharp up to a coefficient only, even though it gives the right order of magnitude (see Theorem 4.2 in the sequel).

In order to make simple evaluations of (2.13) in closed form, complementary assumptions have to be made on the p_{ij} 's. An example is given as follows.

THEOREM 2.3. *Assume that $p_{ij}p_{il} = 0$ for all $1 \leq j \neq l \leq k$ and $1 \leq i \leq n$. Suppose that $k \geq 1$ is fixed and that the p_{ij} 's vary in such a way that*

$$\left\{ \sum_{i=1}^n p_{i1}^2 \right\} / \left\{ \sum_{i=1}^n p_{ij} \right\} \sim \dots \sim \left\{ \sum_{i=1}^n p_{ik}^2 \right\} / \left\{ \sum_{i=1}^n p_{ik} \right\}$$

and (2.17)

$$\sum_{i=1}^n p_{ij} \rightarrow \infty \quad \text{for } j = 1, \dots, k.$$

Let d_n be as in (2.11). Then (with the convention $0/0 = 0$)

$$d_n \sim 2V_k \left(\frac{k}{2\pi e} \right)^{k/2} \sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-1}, \quad (2.18)$$

where

$$V_{2m} = \frac{\pi^m}{m!} \quad \text{and} \quad V_{2m+1} = \frac{2^{2m+1}m!}{(2m+1)!} \pi^m, \quad m = 0, 1, \dots$$

Proof. See Theorem 4.3 in the sequel.

Remark 2.3. Let $k = 1$ and $P_i = p_{i1}$, $i = 1, \dots, n$. Since $V_1 = 2$, we have by (2.16) and (2.18) that

$$\frac{1}{4} d_n \sim \frac{1}{\sqrt{2\pi e}} \left\{ \sum_{i=1}^n P_i^2 \right\} \left\{ \sum_{i=1}^n P_i \right\}^{-1} \quad \text{as } \sum_{i=1}^n P_i \rightarrow \infty. \quad (2.19)$$

By Theorem 2.2, we see that $\Delta_n \sim \frac{1}{4} d_n$ if, in addition, we have $\{\sum_{i=1}^n P_i^2\} / \{\sum_{i=1}^n P_i\} \rightarrow 0$. This corresponds to the Poisson approximation of the sum of independent Bernoulli summands, where it has been shown in Deheuvels and Pfeifer [5] that

$$\Delta_n = \frac{1}{\sqrt{2\pi e}} \left\{ \sum_{i=1}^n P_i^2 \right\} \left\{ \sum_{i=1}^n P_i \right\}^{-1} \quad \text{as } \sum_{i=1}^n P_i \rightarrow \infty \text{ and } \sum_{i=1}^n P_i^2 = O(1). \quad (2.20)$$

Remark 2.4. Let $k \geq 1$ be arbitrary and consider an array $\{X_{lj}, 1 \leq l \leq N, 1 \leq j \leq k\}$ of independent Bernoulli random variables such that

$$P(X_{lj} = 1) = 1 - P(X_{lj} = 0) = P_j, \quad j = 1, \dots, k, l = 1, \dots, N. \quad (2.21)$$

Put $n = Nk$, and let

$$S_n = (S_n(1), \dots, S_n(k)) = \left(\sum_{l=1}^N X_{l1}, \dots, \sum_{l=1}^N X_{lk} \right). \quad (2.22)$$

It is straightforward that an alternative representation for S_n is $S_n = \sum_{i=1}^n Z_i$, where the random vectors $Z_i, i = 1, \dots, n$, collect all k -vectors of the form $(0, \dots, X_{lj}, 0, \dots, 0)$ with X_{lj} in j th position ($j = 1, \dots, k$)($l = 1, \dots, N$). The correspondence between i and (l, j) may be chosen arbitrarily as long as it defines a one-to-one mapping between $\{1, \dots, n\}$ and $\{1, \dots, N\} \times \{1, \dots, k\}$.

In this case, Theorem 2.3 (see also Theorem 4.1 in the sequel) applies and

$$A_n = d_v(L(S_n), L(T_n)) \sim \frac{1}{2} V_k \left(\frac{k}{2\pi e} \right)^{k/2} \sum_{j=1}^k P_j, \quad (2.23)$$

whenever $k \geq 1$ is fixed and $N \geq 1, P_1, \dots, P_N$ vary in such a way that

$$N \sum_{j=1}^k P_j \rightarrow \infty \quad \text{and} \quad N \sum_{j=1}^N P_j^2 = O(1). \quad (2.24)$$

Assume, in addition to (2.24), that

$$NP_j \rightarrow \infty, \quad j = 1, \dots, k. \quad (2.25)$$

Then, we have likewise (or by (2.20))

$$d_v(L(S_n(j)), L(T_n(j))) \sim \frac{P_j}{\sqrt{2\pi e}}, \quad j = 1, \dots, k. \quad (2.26)$$

Since (see Remark 4.2 in the sequel) $2V_k(k/2\pi e)^{k/2} \sim 2/\sqrt{\pi k}$ as $k \rightarrow \infty$, we see that for any $\varepsilon > 0$ there exists a $k = k_\varepsilon \geq 1$ such that, under (2.24) and (2.25), ultimately

$$\begin{aligned} (1 - \varepsilon) \left\{ \frac{e}{2k} \right\}^{1/2} &\leq d_v(L(S_n), L(T_n)) \Big/ \sum_{j=1}^k d_v(L(S_n(j)), L(T_n(j))) \\ &\leq (1 + \varepsilon) \left\{ \frac{e}{2k} \right\}^{1/2}. \end{aligned} \quad (2.27)$$

It is noteworthy that $S_n(1), \dots, S_n(k)$ are independent and that $S_n(j)$ follows a binomial $B(N, P_j)$ distribution. Recall that $T_n(1), \dots, T_n(j)$ are also independent and such that $T_n(j)$ follows a Poisson distribution with expectation NP_j . The evaluation in (2.27) shows that, for large k 's, the upper bound

$$\begin{aligned} d_v(L(S_n), L(T_n)) &\leq 1 - \prod_{j=1}^k \{1 - d_v(L(S_n(j)), L(T_n(j)))\} \\ &\leq \sum_{j=1}^k d_v(L(S_n(j)), L(T_n(j))), \end{aligned} \quad (2.28)$$

which can be obtained by maximal independent couplings of $S_n(j)$ and $T_n(j)$, $j=1, \dots, k$, is far from optimal (Recall that a maximal coupling of ξ and ζ is a construction of ξ and ζ on the same probability space such that $P(\xi \neq \zeta) = d_v(L(\xi), L(\zeta))$. Such a construction always exists and can be made here with $\xi = S_n(j)$ and $\zeta = T_n(j)$ in such a way that $(S_n(1), T_n(1)), \dots, (S_n(k), T_n(k))$ are independent 2-vectors.)

Consider now the general situation described in Example 2.2 and Theorem 2.1, corresponding to the multinomial distribution. In this case $S_n(1), \dots, S_n(k)$ are dependent so that (2.28) does not hold. However, it is remarkable that, under the assumptions of Theorem 2.1, if (2.7) holds, we have

$$\begin{aligned} d_v(L(S_n), L(T_n)) &\sim \sum_{j=1}^k d_v(L(S_n(j)), L(T_n(j))) \\ &\sim 1 - \prod_{j=1}^k \{1 - d_v(L(S_n(j)), L(T_n(j)))\}. \end{aligned} \quad (2.29)$$

On the other hand, if (2.9) holds, and with the particular choice of $p_j = P/k$, $j=1, \dots, k$, we see that

$$d_v(L(S_n), L(T_n)) \Big/ \sum_{j=1}^k d_v(L(S_n(j)), L(T_n(j))) \rightarrow k, \quad (2.30)$$

which can be rendered as great as desired by a suitable choice of $k \geq 1$.

These examples show that there is no hope to obtain sharp evaluation of $d_v(L(S_n), L(T_n))$ in terms of $\sum_{j=1}^k d_v(L(S_n(j)), L(T_n(j)))$ without specific assumptions on the p_{ij} 's.

Remark 2.5. It is straightforward that $d_v(L(S_n), L(T_n)) \geq d_v(L(\sum_{j=1}^k S_n(j)), L(\sum_{j=1}^k T_n(j)))$. A simple proof of this statement uses a maximal coupling between S_n and T_n and the inequalities $P(S_n \neq T_n) \geq$

$P(\sum_{j=1}^k S_n(j) \neq \sum_{j=1}^k T_n(j)) \geq d_v(L(\sum_{j=1}^k S_n(j)), L(\sum_{j=1}^k T_n(j)))$. This enables one to obtain lower bounds for \mathcal{A}_n by using classical Poisson approximation arguments (see, e.g., Example 2.1).

3. THE SEMIGROUP SETTING

Let $k \geq 1$ be a fixed integer, and consider the Banach space $l_1^{(k)}$ if all sequences $f = f(m)$, $m = (m_1, \dots, m_k) \in \mathbb{N}^k = \{0, 1, \dots\}^k$, such that

$$\|f\| = \sum_{m \in \mathbb{N}^k} |f(m)| < \infty. \quad (3.1)$$

For $f, g \in l_1^{(k)}$, the convolution $f * g = g * f$ is defined by

$$f * g(m_1, \dots, m_k) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} f(r_1, \dots, r_k) g(m_1 - r_1, \dots, m_k - r_k), \quad m \in \mathbb{N}^k. \quad (3.2)$$

Note for further use that if, $f, g \in l_1^{(k)}$,

$$\|f * g\| \leq \|f\| \|g\|, \quad (3.3)$$

with equality whenever $f(r)g(s) \geq 0$ for all $r \in \mathbb{N}^k$ and $s \in \mathbb{N}^k$.

In the sequel, we shall identify a bounded measure μ on \mathbb{N}^k with the sequence $f \in l_1^{(k)}$ via the equivalence

$$\mu \approx f \Leftrightarrow f(m) = \mu(\{m\}), \quad \text{all } m \in \mathbb{N}^k. \quad (3.4)$$

In particular, the set \mathbb{M}^k of all probability measures on \mathbb{N}^k will be identified with the subset of $l_1^{(k)}$ composed of all nonnegative sequences f such that $\|f\| = 1$.

Any sequence $f \in l_1^{(k)}$, or equivalently, any bounded measure $\mu \approx f$ on \mathbb{N}^k , defines a bounded linear operator on $l_1^{(k)}$ by

$$g \in l_1^{(k)} \rightarrow \mu g = f * g \in l_1^{(k)}, \quad (3.5)$$

where μ and f are related via (3.4).

Let ε_i^j denote the unit mass at point (m_1, \dots, m_k) , where $m_i = 0$, $i \neq j$, and $m_j = 1$. Then the identity operator I on $l_1^{(k)}$ corresponds via (3.5) to ε_0^0 , since we have

$$I g = g = \varepsilon_0^0 * g, \quad \text{all } g \in l_1^{(k)}. \quad (3.6)$$

By (3.3) and (3.6), we have, for any $f, g \in l_1^{(k)}$ and $\mu \approx f$,

$$\|f * g\| = \|\mu g\| = \|(f * \varepsilon_0^0) * g\| \leq \|f\| \|g\| = \|(f * \varepsilon_0^0)\| \|g\|, \quad (3.7)$$

with equality when $g = \varepsilon_0^0$. This shows that the operator norm $\sup\{\|f * g\|/\|g\|: g \neq 0\}$ of the operator defined by (3.5) coincides with $\|f\| = \|f * \varepsilon_0^0\|$. In the sequel, both norms will be denoted by $\|f\|$.

Consider now, for $j = 1, \dots, k$, the operator B_j defined by

$$g \in I_1^{(k)} \rightarrow B_j g = \varepsilon_1^j * g. \quad (3.8)$$

It is straightforward that B_1, \dots, B_k commute and that the operator in (3.5) corresponds to

$$g \in I_1^{(k)} \rightarrow f * g = \left(\sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} f(r_1, \dots, r_k) B_1^{r_1} \cdots B_k^{r_k} \right) g. \quad (3.9)$$

For $j = 1, \dots, k$, $A_j = B_j - I$ defines the generator of the contraction semigroup $\{\exp(t_j A_j), t_j \geq 0\}$. The operator $\exp(t_j A_j)$ corresponds via (3.5) and (3.9) to a probability measure which is a product of unit masses at the origin for the coordinates $1, \dots, j-1, j+1, \dots, k$, and of a Poisson distribution with mean t_j for the j th coordinate.

Likewise, the multiparameter semigroup

$$\exp\left(\sum_{j=1}^k t_j A_j\right) = \prod_{j=1}^k \exp(t_j A_j), \quad t_1, \dots, t_k \geq 0, \quad (3.10)$$

corresponds to products of probability measures having mean t_j on the j th coordinate, $j = 1, \dots, k$.

Let E_1, \dots, E_k denote disjoint random events with $p_j = P(E_j)$, $j = 1, \dots, k$, and $\sum_{j=1}^k p_j \leq 1$. Let $N_j = 1_{E_j}$ denote the number of outcomes of E_j , $j = 1, \dots, k$, and let ν stand for the probability distribution of the random vector (N_1, \dots, N_k) . It will be convenient to denote such a distribution by $B(p_1, \dots, p_k)$. By (3.9), we see that ν corresponds by (3.5) to the operator $I + \sum_{j=1}^k p_j A_j$.

Assume now that $Z_i = (Z_i(1), \dots, Z_i(k))$, $i = 1, \dots, n$, are independent random vectors such that for each $i = 1, \dots, n$, Z_i follows a $B(p_{i1}, \dots, p_{ik})$ distribution. Set $\lambda_j = \sum_{i=1}^n p_{ij}$, $j = 1, \dots, k$. If $S_n = (S_n(1), \dots, S_n(k)) = \sum_{i=1}^n Z_i$, and if $T_n = (T_n(1), \dots, T_n(k))$, where $T_n(1), \dots, T_n(k)$ are independent Poisson random variables with means $\lambda_1, \dots, \lambda_k$, then the total variation distance between the distributions $L(S_n)$ of S_n and $L(T_n)$ of T_n , namely

$$\begin{aligned} d_v(L(S_n), L(T_n)) &= \sup_{A \subset \mathbb{N}^k} |P(S_n \in A) - P(T_n \in A)| \\ &= \frac{1}{2} \sum_{m \in \mathbb{N}^k} |P(S_n = m) - P(T_n = m)|, \end{aligned} \quad (3.11)$$

is nothing else but the corresponding operator norm halved:

$$\begin{aligned} d_v(L(S_n), L(T_n)) &= \frac{1}{2} \|L(S_n) - L(T_n)\| \\ &= \frac{1}{2} \left\| \exp\left(\sum_{j=1}^k \lambda_j A_j\right) - \prod_{i=1}^n \left(I + \sum_{j=1}^k p_{ij} A_j\right) \right\|. \end{aligned} \quad (3.12)$$

In the sequel, we shall evaluate this expression by suitable Taylor expansions for semigroups as in Deheuvels and Pfeifer [5, 6]. We begin with general evaluations dealing with linear operators in Banach spaces.

THEOREM 3.1. *Let D_1, \dots, D_n be bounded linear operators on a Banach space \mathbb{X} , with values in \mathbb{X} , and such that $I + D_i$ are contractions for $i = 1, \dots, n$, where I denotes the identity operator on \mathbb{X} . Let $\|\cdot\|$ denote the norm (and the operator norm) on \mathbb{X} . Then also the operators $\exp(D_i)$ are contractions for $i = 1, \dots, n$, and we have*

$$\left\| \exp\left(\sum_{i=1}^n D_i\right) - \prod_{i=1}^n (I + D_i) \right\| = \frac{1}{2} \left\| \sum_{i=1}^n D_i^2 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\| + R'_n, \quad (3.13)$$

where

$$\begin{aligned} |R'_n| &\leq \sum_{i=1}^n \left(\frac{(K-1)}{3} \left\| D_i^3 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\| + K \sum_{m=1}^{i-1} \left\| D_i^2 D_m^2 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\| \right) \\ &\quad \times \exp\left(K \sum_{m=1}^{i-1} \|D_m\|^2\right), \end{aligned} \quad (3.14)$$

and

$$K = 1 + \frac{1}{2} \exp(\max\{\|D_1\|, \dots, \|D_n\|\}).$$

Proof. Recall that $u: \mathbb{X} \rightarrow \mathbb{X}$ is a contraction iff $\|u\| \leq 1$. We have by (3.3) that

$$\|\exp(D_i)\| \leq \|e^{-I}\| \|e^{I+D_i}\| \leq e^{-1} \exp(\|I+D_i\|) \leq 1, \quad (3.15)$$

where we have used the fact that $\{e^{\lambda I}\} x = e^\lambda x$ for all $x \in \mathbb{X}$ and $\lambda \in \mathbb{R}$, jointly with the inequality $\|e^C\| \leq e^{\|C\|}$ valid for all bounded operators C on \mathbb{X} . Hence $\exp(D_i)$ is a contraction for $i = 1, \dots, n$.

Next, by the same factorization technique for differences of products as in Deheuvels and Pfeifer [5, proof of Theorem 2.1] (see also LeCam [12]), we have

$$\begin{aligned}
& \exp\left(\sum_{i=1}^n D_i\right) - \prod_{i=1}^n (I + D_i) - \frac{1}{2} \sum_{i=1}^n D_i^2 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \\
&= \sum_{i=1}^n \left\{ \prod_{m=1}^{i-1} (I + D_m) e^{-D_m} \right\} \left\{ \left(e^{D_i} - I - D_i - \frac{1}{2} D_i^2 \right) \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\} \\
&+ \frac{1}{2} \sum_{i=1}^n D_i^2 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \left\{ \prod_{m=1}^{i-1} (I + D_m) e^{-D_m} - I \right\}. \quad (3.16)
\end{aligned}$$

Consider in (3.16):

$$\begin{aligned}
(I + D_m) e^{-D_m} &= (I + D_m) \left\{ I - D_m + \int_0^1 (1-t) e^{-tD_m} D_m^2 dt \right\} \\
&= I - D_m^2 + (I + D_m) \int_0^1 (1-t) e^{-tD_m} D_m^2 dt.
\end{aligned}$$

By taking norms and making use of the fact that $(I + D_m)$ is a contraction, we get

$$\|(I + D_m) e^{-D_m}\| \leq 1 + \|D_m\|^2 + \frac{1}{2} \|D_m\|^2 e^{\|D_m\|} \leq \exp(K \|D_m\|^2),$$

where $K = 1 + \frac{1}{2} \exp(\max\{\|D_1\|, \dots, \|D_m\|\})$.

Similar arguments show that

$$\begin{aligned}
& \left\| \left(e^{D_i} - I - D_i - \frac{1}{2} D_i^2 \right) \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\| \\
&\leq \left\| D_i^3 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\| \left\| \int_0^1 \frac{(1-t)^2}{2} e^{tD_i} dt \right\| \\
&\leq \frac{(K-1)}{3} \left\| D_i^3 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\|.
\end{aligned}$$

By all this, we have proved that

$$\begin{aligned}
& \left\| \left\{ \prod_{m=1}^{i-1} (I + D_m) e^{-D_m} \right\} \left\{ \left(e^{D_i} - I - D_i - \frac{1}{2} D_i^2 \right) \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\} \right\| \\
&\leq \frac{(K-1)}{3} \left\| D_i^3 \exp\left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l\right) \right\| \exp\left(K \sum_{m=1}^{i-1} \|D_m\|^2\right). \quad (3.17)
\end{aligned}$$

Next, we have

$$\prod_{m=1}^{i-1} (I + D_m) e^{-D_m} - I = \sum_{m=1}^{i-1} \left(\prod_{j=m+1}^{i-1} \{(I + D_j) e^{-D_j}\} \right) ((I + D_m) e^{-D_m} - I),$$

and

$$(I + D_m) e^{-D_m} - I = D_m^2 \left(-I + (I + D_m) \int_0^1 (1-t) e^{-tD_m} dt \right).$$

It follows that

$$\begin{aligned} & \left\| D_i^2 \exp \left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l \right) \left\{ \prod_{m=1}^{i-1} (I + D_m) e^{-D_m} - I \right\} \right\| \\ & \leq \sum_{m=1}^{i-1} \left\| D_i^2 D_m^2 \exp \left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l \right) \right\| K \exp \left(K \sum_{j=m+1}^{i-1} \|D_j\|^2 \right). \end{aligned} \quad (3.18)$$

Here, we have used the fact that

$$\left\| -I + (I + D_m) \int_0^1 (1-t) e^{-tD_m} dt \right\| \leq 1 + \frac{1}{2} e^{\|D_m\|} \leq K.$$

By (3.17) and (3.18), we have

$$\begin{aligned} |R'_n| & \leq \sum_{i=1}^n \left(\frac{(K-1)}{3} \left\| D_i^3 \exp \left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l \right) \right\| + K \sum_{m=1}^{i-1} \left\| D_i^2 D_m^2 \exp \left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l \right) \right\| \right) \\ & \quad \times \exp \left(K \sum_{m=1}^{i-1} \|D_m\|^2 \right), \end{aligned}$$

which completes the proof of Theorem 3.1.

THEOREM 3.2. *Under the assumptions of Theorem 3.1, we have*

$$\left\| \exp \left(\sum_{i=1}^n D_i \right) - \prod_{i=1}^n (I + D_i) \right\| = \frac{1}{2} \left\| \sum_{i=1}^n D_i^2 \exp \left(\sum_{l=1}^n D_l \right) \right\| + R''_n, \quad (3.19)$$

where

$$\begin{aligned} |R''_n| & \leq 2(K-1) \sum_{i=1}^n \left\| D_i^3 \exp \left(\sum_{l=1}^n D_l \right) \right\| \left(1 + \frac{(K-1)}{3} \exp \left(K \sum_{m=1}^{i-1} \|D_m\|^2 \right) \right) \\ & \quad + 2K(K-1) \sum_{i=1}^n \sum_{m=1}^{i-1} \left\| D_i^2 D_m^2 \exp \left(\sum_{l=1}^n D_l \right) \right\| \exp \left(K \sum_{m=1}^{i-1} \|D_m\|^2 \right). \end{aligned}$$

Proof. We have

$$\begin{aligned} & \left\| \sum_{i=1}^n D_i^2 \left(\exp \left(\sum_{\substack{l=1 \\ l \neq i}}^n D_l \right) - \exp \left(\sum_{l=1}^n D_l \right) \right) \right\| \\ & \leq \sum_{i=1}^n \left\| D_i^3 \exp \left(\sum_{l=1}^n D_l \right) \right\| \left\| \int_0^1 e^{-tD_i} dt \right\|, \end{aligned}$$

where we have used the fact that $I - e^{-D_i} = D_i \int_0^1 e^{-tD_i} dt$. It is also straightforward that $\left\| \int_0^1 e^{-tD_i} dt \right\| \leq e^{\|D_i\|} \leq 2(K-1)$, and likewise that $\|e^{-D_i}\| \leq 2(K-1)$. This, jointly with Theorem 3.1, suffices for proof of Theorem 3.2.

EXAMPLE 3.1. Put $k=1$, $A=A_1$, $B=B_1$, and $D_i=P_i A$, $0 \leq P_i \leq 1$, $i=1, \dots, n$. Let $A_m = \sum_{i=1}^n P_i^m$, and assume, as in Example 2.1, that S_n is the sum of independent Bernoulli $B(P_1), \dots, B(P_n)$ random variables while T_n follows a Poisson distribution with mean $A=A_1$. Then, a direct application of Theorem 3.2 yields $\|D_i\| = 2P_i$, $K \leq 1 + \frac{1}{2}e^2 < 4.70$, and

$$d_v(L(S_n), L(T_n)) = \frac{1}{4}A_2 \|A^2 e^{AA}\| + R_n, \quad (3.20)$$

where

$$\begin{aligned} |R_n| & \leq (K-1) A_3 \|A^3 e^{AA}\| \left(1 + \frac{(K-1)}{3} e^{4KA_2} \right) \\ & \quad + \frac{K(K-1)}{2} A_2^2 \|A^4 e^{AA}\| e^{4KA_2} \\ & \leq (8.25 A_3 \|A^3 e^{AA}\| + 8.68 A_2^2 \|A^4 e^{AA}\|) e^{18.78 A_2}. \end{aligned}$$

It can be verified (see, e.g., Deheuvels and Pfeifer [7]) that $A_3 \leq A_2$, $\|A^4 e^{AA}\| \leq 16A^{-2}$, $\|A^3 e^{AA}\| \leq 8A^{-3/2}$, and $\|A^2 e^{AA}\| \sim 4A^{-1}(2\pi e)^{-1/2}$ as $A \rightarrow \infty$. Hence a direct application of (3.20) shows that, whenever $A \rightarrow \infty$ and $A_2 = O(1)$, we have

$$d_v(L(S_n), L(T_n)) \sim A_2 A^{-1} (2\pi e)^{-1/2}. \quad (3.21)$$

This result has been proved by similar arguments in Deheuvels and Pfeifer [7].

We turn back now to (3.12), which corresponds to the particular case where $D_i = \sum_{j=1}^k p_{ij} A_j$, $i=1, \dots, n$. It is here straightforward that $I + D_i$ is a contraction for all i since $I + D_i$ corresponds to a probability distribution.

Hence Theorems 3.1 and 3.2 apply. Note that, for $i = 1, \dots, n$,

$$\|D_i\| = 2 \sum_{j=1}^k p_{ij} = 2P_i \leq 2$$

and

$$D_i = -P_i I + \sum_{j=1}^k p_{ij} B_j,$$

$$\|D_i^2\| = 4 \left\{ \sum_{j=1}^k p_{ij} \right\} = 4P_i^2$$

and

$$D_i^2 = P_i^2 I - 2P_i \sum_{j=1}^k p_{ij} B_j + \sum_{j=1}^k \sum_{l=1}^k p_{ij} p_{ijk} B_j B_l.$$

Likewise, we have, for all $m \geq 1$,

$$\|D_i^m\| \leq \|D_i\|^m = 2^m \left\{ \sum_{j=1}^k p_{ij} \right\}^m = 2^m P_i^m. \quad (3.24)$$

In the sequel, let

$$A_m = \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^m = \sum_{i=1}^n P_i^m \quad \text{and} \quad A = A_1 = \sum_{j=1}^k \lambda_j. \quad (3.25)$$

By (3.12) and (3.19), we have

$$d_v(L(S_n), L(T_n)) = \frac{1}{4} \left\| \sum_{i=1}^n D_i^2 \exp \left(\sum_{l=1}^n D_l \right) \right\| + R_n = \frac{1}{4} d_n + R_n, \quad (3.26)$$

where

$$\begin{aligned} |R_n| &\leq \sum_{i=1}^n \left(L_1 \left\| D_i^3 \exp \left(\sum_{l=1}^n D_l \right) \right\| \right. \\ &\quad \left. + L_2 \sum_{m=1}^{i-1} \left\| D_i^2 D_m^2 \exp \left(\sum_{l=1}^n D_l \right) \right\| \right) \exp(L_3 A_2) \\ &\leq (8L_1 A_3 + 16L_2 A_2^2) \exp(L_3 A_2), \end{aligned} \quad (3.27)$$

and where $L_1 = (K-1)(1 + (K-1)/3) < 8.25$, $L_2 = K(K-1)/2 < 8.68$, $L_3 = 4K < 18.78$, and

$$K = 1 + \frac{1}{2} \exp(2 \max\{P_1, \dots, P_n\}) \leq 1 + \frac{1}{2} e^2 < 4.70.$$

We will now evaluate $d_n = \|\sum_{i=1}^n D_i^2 \exp(\sum_{i=1}^n D_i)\|$. By straightforward calculations, paralleling the procedure in Deheuvels and Pfeifer [5], we obtain

$$\begin{aligned} d_n &= \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} \frac{\lambda_1^{r_1} \cdots \lambda_k^{r_k}}{r_1! \cdots r_k!} e^{-A} \left| \sum_{i=1}^n \left(\left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{r_j}{\lambda_j}\right) \right\}^2 - \sum_{j=1}^k \frac{p_{ij}^2 r_j}{\lambda_j^2} \right) \right| \\ &= E \left(\left| \sum_{i=1}^n \left(\left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{\tau_j}{\lambda_j}\right) \right\}^2 - \sum_{j=1}^k \frac{p_{ij}^2 \tau_j}{\lambda_j^2} \right) \right| \right), \end{aligned} \quad (3.28)$$

where τ_1, \dots, τ_k are independent Poisson random variables with expectations $\lambda_1, \dots, \lambda_k$.

Routine computations show also that

$$E \left(\sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{\tau_j}{\lambda_j}\right) \right\}^2 \right) = E \left(\sum_{i=1}^n \sum_{j=1}^k \frac{p_{ij}^2 \tau_j}{\lambda_j^2} \right) = \sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-1}. \quad (3.29)$$

It follows from (3.29) and the triangle inequality that we have in (3.28)

$$d_n \leq 2 \sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-1}. \quad (3.30)$$

In the next section, we shall detail these evaluations. Note here that the same arguments as in (3.22) show that $R_n = o(d_n)$ whenever $\sum_{i=1}^n P_i^2 = O(1)$.

4. THE MAIN TERM

Consider d_n defined as in (3.28):

$$d_n = E \left(\left| \sum_{i=1}^n \left(\left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{\tau_j}{\lambda_j}\right) \right\}^2 - \sum_{j=1}^k \frac{p_{ij}^2 \tau_j}{\lambda_j^2} \right) \right| \right). \quad (4.1)$$

In the sequel, we shall show that, in general, d_n is close to D_n , where

$$D_n = E \left(\left| \sum_{i=1}^n \left(\left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{\tau_j}{\lambda_j}\right) \right\}^2 - \sum_{j=1}^k \frac{p_{ij}^2}{\lambda_j} \right) \right| \right). \quad (4.2)$$

It is straightforward that

$$|d_n - D_n| \leq \sum_{j=1}^k \left\{ \lambda_j^{-2} \sum_{i=1}^n p_{ij}^2 \right\} E(|\tau_j - \lambda_j|). \quad (4.3)$$

LEMMA 4.1. *Let τ follow a Poisson distribution with mean λ . Then, there exists an absolute constant C such that, for any $\lambda > 0$,*

$$E(|\tau - \lambda|) \leq C\lambda^{1/2}. \quad (4.4)$$

Proof. We have (see, e.g., Johnson and Kotz [9, p. 91]) $E(|\tau - \lambda|) = 2e^{-\lambda}\lambda^{[\lambda]+1}/[\lambda]! \sim \lambda$ as $\lambda \rightarrow 0$, while $E(|\tau - \lambda|) \sim \{2\lambda/\pi\}^{1/2}$ as $\lambda \rightarrow \infty$. This suffices for (4.4).

Let C be as in (4.4). We have, by (4.3),

$$|d_n - D_n| \leq C \sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-3/2}. \quad (4.5)$$

We will now evaluate D_n . First, we introduce some notation. Let

$$M_n = \sum_{i=1}^n \begin{pmatrix} p_{i1} \\ \vdots \\ p_{ik} \end{pmatrix} (p_{i1} \cdots p_{ik}), \quad R_n = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{pmatrix},$$

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix}, \quad \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix}.$$

Consider $\Gamma_n = R_n^{-1/2} M_n R_n^{-1/2}$ and the sets $\Theta_n = \{u \in R^k: u' \Gamma_n u \leq \text{Tr}(\Gamma_n)\}$ and

$$Q_n = \lambda + R_n^{1/2} \Theta_n$$

$$= \left\{ t = (t_1, \dots, t_k) \in R^k: \sum_{i=1}^n \left(\left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{t_j}{\lambda_j} \right) \right\}^2 - \sum_{j=1}^k \frac{p_{ij}^2}{\lambda_j} \right) \leq 0 \right\}.$$

Note for further use that $\text{Tr}(\Gamma_n) = \sum_{j=1}^k \{ \sum_{i=1}^n p_{ij}^2 \} \{ \sum_{i=1}^n p_{ij} \}^{-1} \leq k$ and that

$$D_n = 2E \left(\mathbf{1}_{\{t \in Q_n\}} \sum_{i=1}^n \left(\sum_{j=1}^k \frac{p_{ij}^2}{\lambda_j} - \left\{ \sum_{j=1}^k p_{ij} \left(1 - \frac{\tau_j}{\lambda_j} \right) \right\}^2 \right) \right). \quad (4.6)$$

Our next result describes the limiting behaviour of d_n .

THEOREM 4.1. *Assume that*

$$\min\{\lambda_1, \dots, \lambda_k\} \rightarrow \infty. \quad (4.7)$$

Then

$$d_n \sim 2(2\pi)^{-k/2} \int_{\{u \in R^k: u' \Gamma_n u \leq \text{Tr}(\Gamma_n)\}} (\text{Tr}(\Gamma_n) - u' \Gamma_n u) \exp\left(-\frac{1}{2} u' u\right) du. \quad (4.8)$$

Proof. First, we remark that $\{u \in R^k: u'u = 1\} \subset \Theta_n = \{u \in R^k: u' \Gamma_n u \leq \text{Tr}(\Gamma_n)\}$. This follows from the inequality $u' \Gamma_n u \leq (u'u) \beta_n$, where $\beta_n \leq \text{Tr}(\Gamma_n)$ is the greatest eigenvalue of Γ_n . We have therefore the inequalities

$$(1/k)(1 - u'u) \text{Tr}(\Gamma_n) \leq \beta_n(1 - u'u) \leq \text{Tr}(\Gamma_n) - u' \Gamma_n u \leq \text{Tr}(\Gamma_n) \leq k\beta_n. \quad (4.9)$$

Next, we use the central limit theorem in local form. Here, we can see by Stirling's formula that

$$P(\tau_j = \lambda_j + v\lambda_j^{1/2}) = \frac{e^{-v^2/2}}{\sqrt{2\pi\lambda_j}} \left\{ 1 + O\left(\frac{1+v^2}{\lambda_j^{1/2}}\right) \right\}, \quad (4.10)$$

for $\lambda_j + v\lambda_j^{1/2}$ integer and $\lambda_j \rightarrow \infty$.

Let $v_n = (\min\{\lambda_1, \dots, \lambda_k\})^{1/5}$ and set $\Delta_n = \{u \in R^k: u'u \leq v_n^2\}$. Put $\xi = R_n^{-1/2}(\tau - \lambda)$. We have

$$\begin{aligned} D_n &= 2E(1_{\{\xi \in \Theta_n \cap \Delta_n\}}(\text{Tr}(\Gamma_n) - \xi' \Gamma_n \xi)) + 2E(1_{\{\xi \in \Theta_n \cap \Delta_n^c\}}(\text{Tr}(\Gamma_n) - \xi' \Gamma_n \xi)) \\ &= D_{n1} + D_{n2}. \end{aligned}$$

It is now straightforward that $D_{n2} \leq k\beta_n P(\xi \notin \Delta_n) = o(\beta_n)$, while ultimately

$$D_{n1} \geq 2\beta_n E(1_{\{\xi' \xi \leq 1\}}(1 - \xi' \xi)) \sim \gamma \beta_n$$

for some constant γ (here we have used the central limit theorem for ξ). It follows that $D_{n2} = o(D_{n1})$. The conclusion follows by (4.5) and (4.10).

In the course of our proof, we have shown the following result.

THEOREM 4.2. *Assume that (4.7) holds. Then, for any $\varepsilon > 0$, we have*

$$C_0(1 - \varepsilon) \leq d_n \left/ \left(\sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-1} \right) \right. \leq 2, \quad (4.11)$$

for all n sufficiently large, where

$$C_0 = \frac{2}{k} (2\pi)^{k/2} \int_{\{u'u \leq 1\}} (1 - u'u) \exp\left(-\frac{1}{2}u'u\right) du. \quad (4.12)$$

Proof. The upper bound in (4.11) is given by (3.30), while the lower bound follows from (4.9).

Theorem 4.2 shows that, in the range where (4.7) holds, the rate given by (3.30) is sharp and can be determined only up to a fixed constant.

We apply now the preceding results to specific examples.

THEOREM 4.3. Assume that there exists a sequence $\{\rho_n\}$ such that uniformly in $1 \leq j \neq l \leq k$, as $n \rightarrow \infty$,

$$\frac{\sum_{i=1}^n P_{ij}^2}{\sum_{i=1}^n P_{ij}} = \rho_n(1 + o(1)) \quad \text{and} \quad \frac{\sum_{i=1}^n P_{ij} P_{il}}{\{\sum_{i=1}^n P_{ij}\}^{1/2} \{\sum_{i=1}^n P_{il}\}^{1/2}} = o(\rho_n). \quad (4.13)$$

Suppose, in addition, that for all $1 \leq j \leq k$, as $n \rightarrow \infty$,

$$\sum_{i=1}^n p_{ij} \rightarrow \infty. \quad (4.14)$$

Then

$$d_n \sim 2V_k \left(\frac{k}{2\pi e} \right)^{k/2} \sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-1} \quad \text{as } n \rightarrow \infty, \quad (4.15)$$

where

$$V_{2m} = \frac{\pi^m}{m!}, \quad V_{2m+1} = \frac{2^{2m+1} m!}{(2m+1)!} \pi^m, \quad m = 0, 1, \dots \quad (4.16)$$

Proof. By our assumptions, we have $\Gamma_n \sim \rho_n I$ and $\text{Tr}(\Gamma_n) \sim k\rho_n$ (here I denotes the $(k \times k)$ identity matrix). By Theorem 4.1, it follows that

$$d_n \sim 2\rho_n (2\pi)^{-k/2} \int_{\{u'u \leq k\}} (k - u'u) \exp\left(-\frac{1}{2} u'u\right) du.$$

Put $u'u = s^2$, so that $du = kV_k s^{k-1} ds$, where V_k is the Lebesgue measure of the unit ball in \mathbb{R}^k . We have now

$$d_n \sim 2\rho_n (2\pi)^{-k/2} kV_k \int_0^{\sqrt{k}} (k - s^2) s^{k-1} e^{-s^2/2} ds = 2\rho_n (2\pi)^{-k/2} kV_k k^{k/2} e^{-k/2},$$

which proves (4.15). The proof of (4.16) will be omitted.

Remark 4.1. Let $k = 1$ in (4.15). We have then $V_1 = 2$ and

$$d_n \sim \frac{4}{\sqrt{2\pi e}} \left\{ \sum_{i=1}^n P_i^2 \right\} \left\{ \sum_{i=1}^n P_i \right\}^{-1} \quad \text{as } n \rightarrow \infty. \quad (4.17)$$

This result gives a new proof of Theorems 1.2 and 2.2 (2.13) in Deheuvels and Pfeifer [5].

Remark 4.2. In (4.16), we have $\gamma_k = 2V_k (k/2\pi e)^{k/2} < 2$ for $k = 1, 2, \dots$, which is in agreement with (2.30). We have here

$$\gamma_1 = 4/\sqrt{2\pi e} \simeq 0.97, \quad \gamma_2 = 2/e \simeq 0.74, \dots, \quad \gamma_k \sim 2/\sqrt{\pi k} \quad \text{as } k \rightarrow \infty.$$

Remark 4.3. In Theorem 4.3, we suppose that $\Gamma_n \sim \rho_n I$ as $n \rightarrow \infty$. If we assume more generally that

$$\Gamma_n / \text{Tr}(\Gamma_n) \rightarrow (1/k) \Gamma \quad \text{as } n \rightarrow \infty, \quad (4.18)$$

then the same arguments as above show that (whenever (4.7) holds)

$$d_n \sim C \sum_{j=1}^k \left\{ \sum_{i=1}^n p_{ij}^2 \right\} \left\{ \sum_{i=1}^n p_{ij} \right\}^{-1} \quad \text{as } n \rightarrow \infty, \quad (4.19)$$

where

$$C = \frac{2}{k} (2\pi)^{-k/2} \int_{\{u' \Gamma u \leq k\}} (k - u' \Gamma u) \exp\left(-\frac{1}{2} u' u\right) du. \quad (4.20)$$

Note here that $\text{Tr}(\Gamma) = k$, and that (4.19) and (4.20) hold even when $\Gamma \geq 0$ is a singular matrix. In particular, we may treat the i.i.d. case with the notations of Example 2.2, which gives $\lambda_j = np_j$, $j = 1, \dots, k$,

$$M_n = n \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} (p_1 \cdots p_k), \quad \Gamma_n = \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_k} \end{pmatrix} (\sqrt{p_1} \cdots \sqrt{p_k}).$$

Here, we may see directly in (4.8) that

$$\begin{aligned} & 2(2\pi)^{-k/2} \int_{\{u' \Gamma_n u \leq \text{Tr}(\Gamma_n)\}} (\text{Tr}(\Gamma_n) - u' \Gamma_n u) \exp\left(-\frac{1}{2} u' u\right) du \\ &= \left(2(2\pi)^{-1/2} \int_{-1}^1 (1 - u^2) e^{-u^2/2} du \right) \text{Tr}(\Gamma_n) = \frac{4}{\sqrt{2\pi e}} \sum_{j=1}^k p_j. \end{aligned} \quad (4.21)$$

In the following section, we shall give a direct proof of (4.21).

5. THE SPECIAL CASE OF IDENTICAL SUMMANDS

We assume here that $p_1 = p_{i1}, \dots, p_k = p_{ik}$ are independent of $i = 1, \dots, n$. Our main result is as follows.

LEMMA 5.1. *We have*

$$d_v(L(S_n), L(T_n)) = d_v\left(L\left(\sum_{j=1}^k S_n(j)\right), L\left(\sum_{j=1}^k T_n(j)\right)\right). \quad (5.1)$$

Proof. By (2.2), if $R = \sum_{j=1}^k r_j$, $P = \sum_{j=1}^k p_j$,

$$\begin{aligned} P(S_n = r)/P(T_n = r) &= P\left(\sum_{j=1}^k S_n(j) = R\right) / P\left(\sum_{j=1}^k T_n(j) = R\right) \\ &= \frac{n!}{(n-R)!} (1-P)^{n-R} e^P. \end{aligned}$$

It follows that

$$\begin{aligned} d_v(L(S_n), L(T_n)) &= \frac{1}{2} \sum_{R=0}^{\infty} \sum_{\substack{r_1 + \dots + r_k = R \\ r_1, \dots, r_k \geq 0}} \left| \frac{P(S_n = r)}{P(T_n = r)} - 1 \right| P(T = r) \\ &= \frac{1}{2} \sum_{R=0}^{\infty} \left| \frac{P(\sum_{j=1}^k S_n(j) = R)}{P(\sum_{j=1}^k T_n(j) = R)} - 1 \right| P\left(\sum_{j=1}^k T_n(j) = R\right) \\ &= d_v\left(L\left(\sum_{j=1}^k S_n(j)\right), L\left(\sum_{j=1}^k T_n(j)\right)\right), \end{aligned}$$

as requested.

Lemma 5.1 shows that the i.i.d. case yields dimension-free results which can be treated by classical methods. It turns out that Theorem 2.1 is therefore a direct consequence of Theorem 2.1 in Deheuvels and Pfeifer [5]. It may also be seen that (4.21) follows directly from (2.8).

Remark 5.1. Consider again the general case where p_{i1}, \dots, p_{ik} depend upon $i = 1, 2, \dots$. By the coupling inequality applied as in (2.28), we have

$$\begin{aligned} d_v(L(S_n), L(T_n)) &\leq \sum_{i=1}^n d_v\left(L\left(\sum_{j=1}^k Z_i(j)\right), \prod\left(\sum_{j=1}^k p_{ij}\right)\right) \\ &\leq \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^2, \end{aligned} \tag{5.2}$$

where $\prod(\lambda)$ denotes a Poisson distribution with expectation λ . It is remarkable that the same upper bound as in (5.2) holds for the leading term $\frac{1}{4}d_n$, since in (3.26) and (2.5),

$$\begin{aligned} d_n &\leq \sum_{i=1}^n \|D_i \exp(D_i)\| \\ &= \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\} D\left(\sum_{j=1}^k p_{ij}\right) \leq 4 \sum_{i=1}^n \left\{ \sum_{j=1}^k p_{ij} \right\}^2, \end{aligned} \tag{5.3}$$

where the last inequality follows from $D(\theta) \leq 4\theta$.

6. APPLICATIONS TO POINT PROCESSES

Consider a Polish space (E, δ) , where δ is a metric which renders E separable and complete (see, e.g., Kallenberg [11, p. 93]). Let ξ_1, ξ_2, \dots be independent E -valued random variables, whose distributions will be denoted by μ_1, μ_2, \dots , e.g.,

$$\mu_i(B) = P(\xi_i \in B), \quad (6.1)$$

where here and in the sequel B, B_1, \dots, B_k denote arbitrary Borel subsets of E .

For $n = 1, 2, \dots$, the point process $\{\xi_i, 1 \leq i \leq n\}$ defines a random measure

$$N_n(B) = \sum_{i=1}^n 1_{\{\xi_i \in B\}}, \quad (6.2)$$

with intensity

$$M_n(B) = \sum_{i=1}^n \mu_i(B). \quad (6.3)$$

Consider on E the Poisson process $\Pi_n(\cdot)$ with intensity $M_n(\cdot)$ and let C denote an arbitrary closed subset of E . Denote by $N_n^C(\cdot)$ and $\Pi_n^C(\cdot)$ the point processes induced by $M_n(\cdot)$ on C , and by $L(M_n^C)$ and $L(\Pi_n^C)$ the corresponding probability distributions.

Observe that $C = E \cap C$ is a Polish space with the metric induced by δ , and that $N_n^C(\cdot)$ and $\Pi_n^C(\cdot)$ are random variables with values in $(\mathfrak{N}, \mathcal{E})$, where \mathfrak{N} denotes the set of all locally finite integer-valued nonnegative Radon measures on C , and where \mathcal{E} is the Borel ring of subsets of \mathfrak{N} induced by the vague topology. It noteworthy that \mathfrak{N} is Polish (see, e.g., Kallenberg [11, p. 95]) in the vague topology.

In the sequel, we shall consider the distance in variation between $L(N_n^C)$ and $L(\Pi_n^C)$, namely

$$\begin{aligned} d_v(L(N_n^C), L(\Pi_n^C)) &= \sup_{F \in \mathcal{E}} |P(N_n^C \in F) - P(\Pi_n^C \in F)| \\ &= \sup_{0 \leq h \leq 1} |E(N_n^C(h)) - E(\Pi_n^C(h))|, \end{aligned} \quad (6.4)$$

where the supremum is taken over all \mathcal{E} -measurable functions h such that $0 \leq h \leq 1$.

Because of the fact that $(\mathfrak{N}, \mathcal{E})$ is Polish, any $F \in \mathcal{E}$ has the property that there exists a sequence $\{G_i, i \geq 1\}$ of compact sets in \mathfrak{N} such that $F \supset \bigcup_{i=1}^{\infty} G_i$ and $P(N_n^C \in F - \bigcup_{i=1}^{\infty} G_i) = P(\Pi_n^C \in F - \bigcup_{i=1}^{\infty} G_i) = 0$. It follows that in (6.4) we may replace \mathcal{E} by the set of finite unions of elements of \mathcal{H} ,

where \mathcal{H} is a basis for the vague topology (by this, we mean that, for any $\nu \in \mathfrak{N}$ and for any neighborhood V of ν , there exists a neighborhood W of ν such that $W \subset V$ and $W \in \mathcal{H}$).

We shall consider the case where \mathcal{H} collects all sets of the form $\{m \in \mathfrak{N}: m(B_i) = n_i, i = 1, \dots, k\}$, where B_1, \dots, B_k are Borel subsets of E , $k \geq 1$, and n_1, \dots, n_k are nonnegative integers. Recall that $\nu_n \rightarrow \nu$ vaguely iff for any B such that $\nu(\partial B) = 0$, we have $\nu_n(B) \rightarrow \nu(B)$, which implies $\nu_n(B) = \nu(B)$ for n large enough. We have evidently

$$\begin{aligned} & d_\nu(L(N_n^C), L(\Pi_n^C)) \\ &= \sup_{F \in \mathcal{H}} |P(N_n^C \in F) - P(\Pi_n^C \in F)| \\ &= \sup_{k \geq 1} \sup_{B_1, \dots, B_k} d_\nu(L\{N_n^C(B_1), \dots, N_n^C(B_k)\}, L\{\Pi_n^C(B_1), \dots, \Pi_n^C(B_k)\}). \end{aligned} \quad (6.5)$$

Here we can assume, without loss of generality, that B_1, \dots, B_k are disjoint Borel subsets of E .

By (6.5), the evaluation of $d_\nu(L(N_n^C), L(\Pi_n^C))$ can be reduced to the problem treated in the preceding section. As a direct application, we obtain:

THEOREM 6.1. *Assume that $\mu = \mu_i$ is independent of $i = 1, 2, \dots, n$. Let $P = \mu(C)$ and $\theta = nP$. Then*

$$d_\nu(L(N_n^C), L(\theta_n^C)) = \frac{1}{2} P\theta \left(\frac{\theta^{\alpha-1}(\alpha-\theta)}{\alpha!} - \frac{\theta^{\beta-1}(\beta-\theta)}{\theta!} \right) e^{-\theta} + R_n, \quad (6.6)$$

where α and β and R_n are as in Theorem 2.1.

In addition, if n, μ , and C vary in such a way that, for a fixed $\varepsilon > 0$,

$$n\mu(C) \rightarrow \infty \quad \text{and} \quad \mu(C) < 1 - \varepsilon, \quad (6.7)$$

then

$$d_\nu(L(N_n^C), L(\Pi_n^C)) \sim \frac{\mu(C)}{\sqrt{2\pi\varepsilon}}. \quad (6.8)$$

On the other hand, if

$$n\mu(C) \rightarrow 0, \quad (6.9)$$

then

$$d_\nu(L(N_n^C), L(\Pi_n^C)) \sim n\mu^2(C). \quad (6.10)$$

Proof. By (6.5) and Theorem 2.1, we have $d_\nu = \sup(\frac{1}{4}d_n + R_n)$, where the supremum is taken over all B_1, \dots, B_k disjoint Borel subsets of C . It suffices

now to use the fact that $\frac{1}{4}d_n$ and $|R_n|$ are increasing functions of $\mu(\bigcup_{i=1}^k B_i)$, and maximal when $\bigcup_{i=1}^k B_i = C$. The conclusion follows by (2.11).

In order to obtain similar evaluations in the non-i.i.d. case, assume that there exists a Radon measure ν on E such that $\mu_i \ll \nu$ for $i = 1, \dots, n$. Such a measure always exists, since we may take $\nu = \sum_{i=1}^n \mu_i$.

Let $f_i = d\mu_i/d\nu$ denote the Radon–Nikodym derivative of μ_i respectively to ν , and consider the Radon measure σ_n defined (with the convention $0/0 = 0$) by

$$d\sigma_n(\omega) = \left\{ \sum_{i=1}^n f_i^2(\omega) \right\} \left\{ \sum_{i=1}^n f_i(\omega) \right\}^{-1} d\nu(\omega). \quad (6.11)$$

Observe that σ_n is independent of ν . Furthermore, since the choice $\nu = \sum_{i=1}^n \mu_i$ implies $f_i \leq 1$ (ν -a.e.), we have evidently $\sigma_n \leq \sum_{i=1}^n \mu_i$. Note also that, whenever $\mu_1 = \dots = \mu_n = \mu$, we have $\sigma_n = \mu$.

Let us now use again Theorem 2.2 and (6.5). It is straightforward that if we use the upper bound (3.30) for d_n , we obtain the upper bound

$$d_n \leq 2\sigma_n(C). \quad (6.12)$$

We obtain therefore the following result.

THEOREM 6.2. *Let σ_n be as in (6.11). We have*

$$d_v(L(N_n^C), L(\Pi_n^C)) = \frac{1}{4}d_n + R_n,$$

where

$$\begin{aligned} d_n &\leq \min(2\sigma_n(C), 4 \sum_{i=1}^n \mu_i^2(C)), \\ |R_n| &\leq \exp\left(4K \sum_{i=1}^n \mu_i^2(C)\right) \\ &\quad \times \left(\frac{8}{3}(K-1)^2 \sum_{i=1}^n \mu_i^3(C) + 8K(K-1) \left\{ \sum_{i=1}^n \mu_i^2(C) \right\}^2\right) \\ &\quad + 8K(K-1) \sum_{i=1}^n \mu_i^3(C), \end{aligned} \quad (6.13)$$

and

$$K = 1 + \frac{1}{2}e^2 < 4.70.$$

Furthermore,

$$R_n = o(d_n) \quad \text{whenever} \quad \sum_{i=1}^n \mu_i^2(C) = O(1).$$

Remark 6.1. The condition $\sum_{i=1}^n \mu_i^2(C) = O(1)$ is probably not necessary for the validity of (6.13) (see, e.g., Theorem 6.1).

ACKNOWLEDGMENTS

We thank the referee for some helpful comments and remarks. The second author is indebted to the Deutsche Forschungsgemeinschaft for financial support by a Heisenberg research grant.

REFERENCES

- [1] ARENBAEV, N. K. (1976). Asymptotic behavior of the multinomial distribution. *Theory Probab. Appl.* **21**, 805–810.
- [2] BARBOUR, A. D., AND HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95**, 473–480.
- [3] BROWN, T. C. (1981). Poisson approximations and exchangeable random variables. In *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, Eds.), pp. 177–184, North-Holland, Amsterdam.
- [4] BROWN, T. C. (1983). Some Poisson approximations using compensators. *Ann. Probab.* **11**, 726–744.
- [5] DEHEUVELS, P., AND PFEIFER, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* **14** 663–676.
- [6] DEHEUVELS, P., AND PFEIFER, D. (1986). Operator semigroups and Poisson convergence in selected metrics. *Semigroup Forum* **34** 203–224.
- [7] DEHEUVELS, P., AND PFEIFER, D. (1987). Semigroups and Poisson approximation. In *New Perspectives in Theoretical and Applied Statistics* (M. Puri, J. P. Vilaplana and W. Wertz, Eds.), pp. 439–448, Wiley, New York.
- [8] FREEDMAN, D. (1974). The Poisson approximation for dependent events. *Ann. Probab.* **2** 256–269.
- [9] JOHNSON, N. L., AND KOTZ, S. (1969). *Discrete Distributions*, Houghton Mifflin, Boston.
- [10] KABANOV, Y. M., LIPTSER, R. S., AND SHIRYAEV, A. N. (1983). Weak and strong convergence of the distributions of counting processes. *Theory Probab. Appl.* **28**, 303–336.
- [11] KALLENBERG, O. (1976). *Random Measures*. Academic Press, New York.
- [12] LECAM, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197.
- [13] PFEIFER, D. (1985). A semigroup setting for distance measures in connection with Poisson approximation. *Semigroup Forum* **31**, 199–203.
- [14] PFEIFER, D. (1985). *On the Distance between Mixed Poisson and Poisson Distributions*. Technical Report 115, University of North Carolina, Chapel Hill.
- [15] PRESMAN, E. L. (1984). Approximation of binomial distributions by infinitely divisible ones. *Theory Probab. Appl.* **29**, 393–403.
- [16] SERFLING, R. J. (1975). A general Poisson approximation theorem. *Ann. Probab.* **3**, 726–731.
- [17] SERFOZO, R. F. (1985). Partitions of point processes: Multivariate Poisson approximations. *Stochastic Process. Appl.* **20**, 281–294.
- [18] SHUR, M. G. (1982). The Poisson theorem and Markov chains. *Theory Probab. Appl.* **27**, 124–125.
- [19] VALKEILA, E. (1982). A general Poisson approximation theorem. *Stochastics* **7**, 159–171.