The Annals of Applied Probability
1992, Vol. 2, No. 3, 536–554

STEIN'S METHOD AND MULTINOMIAL APPROXIMATION[1]

BY WEI-LIEM LOH

Purdue University

In this paper Stein's method is considered in the context of approximation by a multinomial distribution. By using a probabilistic argument of Barbour, whereby the essential ingredients necessary for the application of Stein's method are derived, the Stein equation for the multinomial distribution is obtained. Bounds on the smoothness of its solution are derived and are used in three examples to give error bounds for the multinomial approximation to the distribution of a random vector.

1. Introduction. Stein (1970) introduced a powerful and general method for obtaining an explicit bound for the error in the normal approximation to the distribution of a sum of dependent random variables. This method was extended from the normal distribution to the Poisson distribution by Chen (1975). Since then, Stein's method has found considerable applications in combinatorics, probability and statistics. Recent literature pertaining to this method includes Arratia, Goldstein and Gordon (1989, 1990), Baldi and Rinott (1989), Barbour (1988, 1990), Barbour, Chen and Loh (1990), Bolthausen and Götze (1989), Chen (1987), Götze (1991), Green (1989), Holst and Janson (1990), Schneller (1989), Stein (1990) and the references cited therein. Stein (1986) gives an excellent account of this method.

In this paper we consider Stein's method in the context of approximation by means of a multinomial distribution. To obtain the necessary ingredients for the application of Stein's method, we use a probabilistic argument of Barbour (1988) which we shall now sketch. At the heart of Stein's method lies a Stein equation. For example, in the case of the normal approximation we have

$$\frac{df}{dw}(w) - wf(w) = g(w), \qquad \forall\, w \in R,$$

and in the Poisson approximation, we have

$$\lambda f(w+1) - wf(w) = g(w), \qquad \forall\, w \in Z^+.$$

Barbour (1988) observed that we can associate with each of these equations a stochastic process. For the normal approximation, we have the Ornstein–Uhlenbeck process and for the Poisson approximation, we have the immigration–death process with immigration rate λ and unit per capita death rate. One of the basic ingredients of Stein's method lies in the problem of getting

Received January 1991; revised October 1991.

[1]Research supported in part by NSF Grant DMS-89-23071.

AMS 1980 *subject classifications*. Primary 60E15; secondary 60J10.

Key words and phrases. Stein's method, multinomial distribution, rate of convergence, multi-urn Ehrenfest model, total variation distance.

smoothness estimates for the solutions of Stein's equations. By embedding the Stein equation in a stochastic process, bounds on the smoothness estimates may be obtained by probabilistic arguments. In many cases, these arguments are easier to apply than the usual analytic ones. This probabilistic technique has been successfully applied to Poisson process approximations, multivariate Poisson approximations [see Barbour (1988)], diffusion approximations [see Barbour (1990)], compound Poisson approximations [see Barbour, Chen and Loh (1990)] and multivariate normal approximations [see Götze (1991) and Bolthausen and Götze (1989)].

The rest of this paper is organized as follows. Section 2 develops the basic ingredients of Stein's method in the multinomial setting. In particular, the Stein equation for the multinomial distribution is obtained in Theorem 1 and smoothness estimates of its solution are given in Theorems 2 and 3. In Section 3 these results are used in three examples to give error bounds for the multinomial approximation to the distribution of a random vector. The first example involves the base M expansion of a random integer and the second gives a multinomial approximation to the multivariate Poisson binomial distribution. The third example gives a similar multinomial approximation to the multivariate hypergeometric distribution. Unfortunately, the error bounds obtained in these examples are somewhat crude. However, we feel that future research would inevitably sharpen these results. On the more positive side, this paper gives yet another probability distribution, namely the multinomial distribution, for which Stein's method can be applied. This should, at the very least, contribute to a broader and hopefully better understanding of Stein's method.

2. Multinomial approximation. We first consider the following multi-urn version of the Ehrenfest model with continuous time. Let there be M urns and N balls distributed in these urns. The system is said to be in state $n = (n_1, \dots, n_M)$ if there are n_i balls in urn i, $i = 1, \dots, M$. Events occur at random times and the time intervals T between successive events are independent random variables all with the same exponential distribution

$$P(T > t) = \exp(-Nt), \qquad \forall\, t \geq 0.$$

When an event occurs, a ball is chosen uniformly at random, removed from its urn and then placed in urn i with probability p_i, $i = 1, \dots, M$, with $\sum_{i=1}^{M} p_i = 1$.

The state of the system at time t, $Z^{(n)}(t)$, is a stationary Markov process with continuous time having state space

$$\Omega = \left\{ (k_1, \dots, k_M) \colon \sum_{i=1}^{M} k_i = N,\, k_i \geq 0,\, 1 \leq i \leq M \right\},$$

where $Z^{(n)}(0) = n$. It is clear that the stationary distribution of $Z^{(n)}(t)$ is $\mathrm{MULT}(N, p_1, \dots, p_M)$, the multinomial distribution with parameters $N, p_1, \dots, p_M$. Multi-urn versions of the Ehrenfest model were first proposed by Siegert (1950) and a treatment can be found in Karlin and McGregor (1965).

The rest of this section is heavily influenced by the techniques developed in Barbour (1988). For $A \subseteq \Omega$, define

$$f_A(n) = \int_0^\infty \left[P(Z^{(n)}(t) \in A) - P(W \in A) \right] dt, \tag{1}$$

where $W \sim \text{MULT}(N, p_1, \dots, p_M)$ and $\sum_{i=1}^M n_i = N$. Also for simplicity, we define $I_A(\cdot)$ to be the indicator function of A and $e^{(i)}$ to be the M-tuple with the ith component equal to 1 and its remaining $M-1$ components equal to 0. We shall now proceed to derive a bound on f_A.

PROPOSITION 1. *With the above notation*, $\sup_{n \in \Omega} |f_A(n)| \le N$.

PROOF. Let τ_i denote the time taken for ball i to be chosen the first time, $i = 1, \dots, N$. Then it is easy to see that when $t > \max_{1 \le i \le N} \tau_i$, we have $Z^{(n)}(t) \sim MULT(N, p_1, \dots, p_M)$. Thus

$$\begin{aligned} |f_A(n)| &= \left| \int_0^\infty E\left[I_A(Z^{(n)}(t)) - I_A(W) | \max \tau_i > t \right] P(\max \tau_i > t)\, dt \right| \\ &\le N \int_0^\infty P(\tau_1 > t)\, dt \\ &= N. \end{aligned}$$

The last equality uses the fact that τ_1 is a standard exponential random variable. □

Next we define

$$\Delta_{i,j} f_A(n) = f_A(n - e^{(i)} + e^{(j)}) - f_A(n), \tag{2}$$

whenever $n, n - e^{(i)} + e^{(j)} \in \Omega$ and $A \subseteq \Omega$.

THEOREM 1. *Let f_A be defined as in* (1). *Then f_A satisfies the equation*

$$\sum_{i,j=1}^M n_i p_j \Delta_{i,j} f_A(n) = P(W \in A) - I_A(n), \qquad \forall\, n \in \Omega. \tag{3}$$

PROOF. Let $f_A(n, t) = \int_0^t [P(Z^{(n)}(u) \in A) - P(W \in A)]\, du$. By considering the first jump of the process $Z^{(n)}(u)$, we have

$$\begin{aligned} f_A(n,t) = \int_0^t e^{-Nu} \Bigg\{ & uN\left[I_A(n) - P(W \in A) \right] \\ & + \sum_{i,j=1}^M n_i p_j f_A(n - e^{(i)} + e^{(j)}, t - u) \Bigg\}\, du \\ & + te^{-Nt}\left[I_A(n) - P(W \in A) \right]. \end{aligned} \tag{4}$$

We observe as in the proof of Proposition 1 that

$$|f_A(n,t)| \le N, \qquad \forall\, t \ge 0. \tag{5}$$

Thus it follows from (4) and the definitions of $f_A(n)$ and $f_A(n,t)$ that

$$\begin{aligned} f_A(n) &= \lim_{t\to\infty} f_A(n,t) \\ &= N^{-1}[I_A(n) - P(W \in A)] \\ &\quad + \lim_{t\to\infty} \int_0^\infty e^{-Nu} \sum_{i,j=1}^{M} n_i p_j f_A(n - e^{(i)} + e^{(j)}, t-u) I_{[0,t]}(u)\, du \\ &= N^{-1}\left[I_A(n) - P(W \in A) + \sum_{i,j=1}^{M} n_i p_j f_A(n - e^{(i)} + e^{(j)}) \right]. \end{aligned}$$

The last equality uses (5) and the dominated convergence theorem. This completes the proof of Theorem 1. □

Remark. The Stein equation for the multinomial distribution is given by (3). This equation is of crucial importance in our applications in Section 3.

The next theorem gives an estimate of the "smoothness" of $f_A(n)$ by bounding $\Delta_{i,j} f_A(n)$.

Theorem 2. *Let f_A be defined as in* (1). *Then for $N \ge 1$, we have*

$$\sup_{n \in \Omega:\, n_i > 0} |\Delta_{i,j} f_A(n)| \le \begin{cases} 1, & \textit{if } C^{(1)} N^{-1/2} \le 1, \\ 2C^{(1)} N^{-1/2}(1 - C^{(1)} N^{-1/2}/2), & \textit{otherwise}, \end{cases}$$

where $C^{(1)} = \sup_{i<j}[C(i,j) \wedge C(j,i)]$ and

$$C(i,j) = \left[\frac{2}{p_i} + \frac{3}{p_j} + \frac{1}{ep_j(1-p_j)} \right]^{1/2} + \left[\frac{1}{2ep_j(1-p_j)^2} \right]^{1/2}. \tag{6}$$

Proof. Without loss of generality, we shall assume that $i \ne j$. Let $n, n - e^{(i)} + e^{(j)} \in \Omega$. It is convenient to couple $f_A(n - e^{(i)} + e^{(j)})$ and $f_A(n)$ on the same probability space as follows. Let there be M urns and initially (i.e., at time $t = 0$), $N + 1$ balls are placed in these urns such that n_m balls are placed in urn m, $m \ne j$, and $n_j + 1$ balls are placed in urn j. Suppose for convenience that the balls are labeled 1 to $N + 1$. Select a ball from urn i (say ball a) and select a ball from urn j (say ball b).

We assume that events occur at random times and time intervals T between successive events are independent random variables all with the same exponential distribution

$$P(T > t) = \exp(-Nt), \qquad \forall\, t \ge 0.$$

When an event occurs, a ball is chosen uniformly at random from the N balls numbered 1 to $N + 1$ excluding ball b, removed from its urn and then placed in urn m with probability p_m, satisfying $\sum_{m=1}^{M} p_m = 1$. We further impose the condition that the balls a and b have identical jump times and they jump to the same urn. This implies that the processes of these two balls coincide after their first jump.

For each l, $1 \le l \le N + 1$, let τ_l denote the time ball l is picked for the first time. Thus $\tau_a = \tau_b$ and balls a and b are kept in the same urn whenever $t > \tau_a$. Also the law of τ_l is the standard exponential distribution such that the τ_l's, $l \neq b$, are stochastically independent. At time t, we define for $m = 1, \dots, M$,

$$a(t) = e^{(m)} \quad \text{if ball } a \text{ is in urn } m,$$

$$b(t) = e^{(m)} \quad \text{if ball } b \text{ is in urn } m.$$

Writing $n' = (n_1, \dots, n_{i-1}, n_i - 1, n_{i+1}, \dots, n_M)$, it follows from (1) and (2) that

$$\Delta_{i,j} f_A(n) = \int_0^\infty \left[P(Z^{(n')}(t) + b(t) \in A) - P(Z^{(n')}(t) + a(t) \in A) \right] dt.$$

From the above coupling construction, we observe that $a(t) = b(t)$ whenever $t > \tau_a$. Hence

$$\begin{aligned} &\Delta_{i,j} f_A(n) \\ (7) \quad &= \int_0^\infty E\left[I_A(Z^{(n')}(t) + b(t)) - I_A(Z^{(n')}(t) + a(t)) | t < \tau_a \right] P(t < \tau_a)\, dt \\ &= \int_0^\infty \left[P(Z^{(n')}(t) + e^{(j)} \in A) - P(Z^{(n')}(t) + e^{(i)} \in A) \right] e^{-t}\, dt. \end{aligned}$$

The last equality uses the independence of τ_a and $Z^{(n')}(t)$. Since $|P(Z^{(n')}(t) + e^{(j)} \in A) - P(Z^{(n')}(t) + e^{(i)} \in A)| \le 1$, it follows from (7) and Lemma 1 (see the Appendix) that

$$\begin{aligned} |\Delta_{i,j} f_A(n)| &\le \int_0^\infty \left\{ 1 \wedge \frac{C^{(1)}}{[(1 - e^{-t})N]^{1/2}} \right\} e^{-t}\, dt \\ &= \begin{cases} 1, & \text{if } C^{(1)} N^{-1/2} \ge 1, \\ 2C^{(1)} N^{-1/2}(1 - C^{(1)} N^{-1/2}/2), & \text{otherwise.} \end{cases} \end{aligned}$$

This completes the proof. □

COROLLARY 1. *Let f_A be defined as in* (1) *and $p_1 = \cdots = p_M = 1/M$. Then for $N \ge 1$, we have*

$$\sup_{n \in \Omega:\, n_i > 0} |\Delta_{i,j} f_A(n)| \le \begin{cases} 1, & \textit{if } 3.3\sqrt{M/N} \ge 1, \\ 6.6\sqrt{M/N}\left(1 - 3.3\sqrt{M/N}/2\right), & \textit{otherwise}. \end{cases}$$

PROOF. The proof is immediate from Theorem 2 by assuming $M \geq 2$ since the $M = 1$ case is trivial. □

REMARK. The coupling used in the proof of Theorem 2 was suggested by a referee. The coupling that was originally constructed gives, in general, slightly cruder bounds and can be found in Loh (1991).

We end this section with a bound for $\Delta_{j,i}[\Delta_{k,j} f_A(n)]$.

THEOREM 3. *Let f_A be defined as in* (1). *Then for $N \geq 2$, we have*

$$\sup_{n \in \Omega:\, n_j > 0,\, n_k > 0} \left|\Delta_{j,i}\left[\Delta_{k,j} f_A(n)\right]\right| \leq \begin{cases} 1, & \text{if } C^{(2)}[2(N-1)]^{-1} \geq 1, \\ \dfrac{C^{(2)}}{N-1} \log\left[\dfrac{2(N-1)}{C^{(2)}}\right] + \left[\dfrac{C^{(2)}}{2(N-1)}\right]^2, & \text{otherwise}, \end{cases}$$

where $C^{(2)} = \sup_{i,j,k:\, i \neq j,\, k \neq j} C(j,i,k)$ and

$$C(j,i,i) = \frac{2}{p_i} + \frac{2}{p_j},$$

$$\begin{aligned} C(j,i,k) = {} & \frac{2}{ep_i(1-p_i)} + \frac{2}{ep_k(1-p_k)} + \frac{1}{p_j} \\ & + \left[\frac{2}{p_j} + \frac{3}{p_i} + \frac{1}{ep_i(1-p_i)}\right]^{1/2} \left[\frac{2}{p_j} + \frac{3}{p_k} + \frac{1}{ep_k(1-p_k)}\right]^{1/2}, \end{aligned} \tag{8}$$

for $i \neq k$.

PROOF. Without loss of generality, we can assume that $i \neq j$ and $k \neq j$, since otherwise we have $\Delta_{j,i}[\Delta_{k,j} f_A(n)] = 0$.

Let $n \in \Omega$ with n_j, n_k nonzero. We shall couple $f_A(n)$, $f_A(n - e^{(k)} + e^{(j)})$, $f_A(n - e^{(k)} + e^{(i)})$ and $f_A(n - e^{(j)} + e^{(i)})$ on the same probability space as follows. Let there be M urns and $N + 2$ balls. Initially at time $t = 0$, these balls are placed in the urns such that n_m balls are placed in urn m whenever $m \neq i, j$ and $n_m + 1$ balls are placed in urn m whenever $m = i, j$. For convenience, we suppose that the $N + 2$ balls are numbered from 1 to $N + 2$.

Case I. Suppose $i \neq k$. Select a ball from urn i (say ball a), select two balls from urn j (say balls b' and b'') and select a ball from urn k (say ball c).

Case II. Suppose $i = k$. Select two balls from urn i (say balls a and c) and select two balls from urn j (say balls b' and b'').

We assume that events occur at random times and time intervals T between successive events are independent random variables all with the same exponential distribution

$$P(T > t) = \exp(-Nt), \qquad \forall\, t \geq 0.$$

When an event occurs, a ball is chosen uniformly at random from the N balls numbered 1 to $N+2$ excluding balls b' and b'', removed from its urn and then placed in urn m with probability p_m, $m = 1,\dots,M$. We further impose the condition that balls a and b' have identical jump times and they jump to the same urn. Thus the processes of these two balls coincide after their first jump. Likewise we assume that balls b'' and c have identical jump times and they jump to the same urn.

For each l, $1 \le l \le N+2$, let τ_l denote the time ball l is picked for the first time. Thus $\tau_a = \tau_{b'}$ and $\tau_{b''} = \tau_c$. A consequence of this is that once picked, balls a and b' shall always be kept in the same urn and balls b'' and c shall be kept in the same urn. Also the law of τ_l is the standard exponential distribution such that the τ_l's, $l \neq b', b''$, are stochastically independent. Next we write $n' = (n'_1,\dots,n'_M)$ with

$$n'_m = \begin{cases} n_m, & \text{if } m \neq j, k, \\ n_m - 1, & \text{if } m = j, k. \end{cases}$$

Furthermore at time t, we define for $m = 1,\dots,M$,

$$\begin{aligned} a(t) &= e^{(m)} \quad \text{if ball } a \text{ is in urn } m, \\ b'(t) &= e^{(m)} \quad \text{if ball } b' \text{ is in urn } m, \\ b''(t) &= e^{(m)} \quad \text{if ball } b'' \text{ is in urn } m, \\ c(t) &= e^{(m)} \quad \text{if ball } c \text{ is in urn } m. \end{aligned}$$

For simplicity, we write

$$\begin{aligned} D = {} & I_A\big(Z^{(n')}(t) + e^{(i)} + e^{(j)}\big) - I_A\big(Z^{(n')}(t) + 2e^{(j)}\big) \\ & + I_A\big(Z^{(n')}(t) + e^{(j)} + e^{(k)}\big) - I_A\big(Z^{(n')}(t) + e^{(i)} + e^{(k)}\big). \end{aligned} \tag{9}$$

Since $a(t) = b'(t)$ if $t > \tau_a$, and $b''(t) = c(t)$ if $t > \tau_c$, we have $D = 0$ whenever $t > \tau_a \wedge \tau_c$. Hence it follows from (1) and (2) that

$$\begin{aligned} \Delta_{j,i}\big[\Delta_{k,j} f_A(n)\big] &= f_A(n - e^{(k)} + e^{(i)}) - f_A(n - e^{(k)} + e^{(j)}) \\ &\quad - f_A(n - e^{(j)} + e^{(i)}) + f_A(n) \\ &= \int_0^\infty E(D) P(t < \tau_a \wedge \tau_c)\, dt \\ &= \int_0^\infty E(D) e^{-2t}\, dt. \end{aligned} \tag{10}$$

The last equality uses the observation that τ_a and τ_c are independent standard exponential random variables and that $\tau_a \wedge \tau_c$ and D are stochastically independent. Since $|D| \le 2$, it follows from Lemma 2 (see the Appendix) and (10) that

$$
\begin{aligned}
&\left|\Delta_{j,i}\left[\Delta_{k,j} f_A(n)\right]\right| \\
&\quad \le \int_0^\infty \left[2 \wedge \frac{C^{(2)}}{(1-e^{-t})(N-1)}\right] e^{-2t}\,dt \\
&\quad = \begin{cases} 1, & \text{if } C^{(2)}/[2(N-1)] \ge 1, \\ \dfrac{C^{(2)}}{N-1} \log\left[\dfrac{2(N-1)}{C^{(2)}}\right] + \left[\dfrac{C^{(2)}}{2(N-1)}\right]^2, & \text{otherwise.} \end{cases}
\end{aligned}
$$

This proves the theorem. □

COROLLARY 2. *Let f_A be defined as in* (1) *and $p_1 = \cdots = p_M = 1/M$. Then for $N \ge 2$, we have*

$$
\sup_{n \in \Omega:\, n_j > 0,\, n_k > 0} \left|\Delta_{j,i}\left[\Delta_{k,j} f_A(n)\right]\right| \le \begin{cases} 1, & \textit{if } 5M(N-1)^{-1} \ge 1, \\ \dfrac{10M}{N-1} \log\left(\dfrac{N-1}{5M}\right) + \left(\dfrac{5M}{N-1}\right)^2, & \textit{otherwise}. \end{cases}
$$

PROOF. The proof follows from Theorem 3 by assuming $M \ge 2$ since the $M = 1$ case is trivial. □

3. Applications.

3.1. *On the base M expansion of a random integer.* Let k and M be natural numbers, with $M \ge 2$, and X a random variable uniformly distributed over the set $\{0, \dots, k-1\}$. Define N to satisfy $M^{N-1} < k \le M^N$. Then the base M expansion of $a = k - 1$ and X can be written as

$$
a = \sum_{i=1}^{N} a_i M^{N-i}, \qquad X = \sum_{i=1}^{N} X_i M^{N-i},
$$

where $a_i, X_i \in \{0, \dots, M-1\}$. Also define for $i = 1, \dots, M$,

$$
U_i = \sum_{j=1}^{N} I_{\{X_j = i-1\}}, \qquad U = (U_1, \dots, U_M).
$$

We are interested in approximating the distribution of U by a multinomial distribution. We note that the distribution of U is exactly MULT$(N, 1/M, \dots, 1/M)$ if $k = M^N$.

Variations of this problem when $M = 2$ have been studied by Delange (1975), Diaconis (1977) and Stein (1986). In particular, the expected value of the number of ones, U_2, in the binary expansion of a random integer was studied as a function of k by Delange. Diaconis (1977), jointly with Stein, exhibited a central limit theorem for U_2. Stein (1986) showed that for large k, U_2 has approximately a binomial distribution. His argument is more analytic, rather than probabilistic, and it is not immediately clear how it could be extended to treat the general multinomial situation.

We shall use the total variation distance as a means of measuring how close the distribution of U is to $\mathrm{MULT}(N, 1/M, \dots, 1/M)$.

DEFINITION. The total variation distance between two probability measures F and G on Ω is defined by

$$d(F,G) = \sup_A |F(A) - G(A)|,$$

where the supremum is taken over all subsets A of Ω. Also for simplicity, we denote the law of a random vector S by $\mathscr{L}(S)$.

THEOREM 4. *For $M \geq 2$, we have*

$$d(\mathscr{L}(U), MULT(N, 1/M, \dots, 1/M)) \leq 3.3(M-1)M\sqrt{M/N}\,.$$

PROOF. First we construct an exchangeable pair of random vectors (U, U^*) on the same probability space as follows. Let I be a random variable uniformly distributed over $\{0, \dots, M-1\}$ and let J be a random variable uniformly distributed over $\{1, \dots, N\}$ with I, J, X mutually independent. Define

$$X^* = \sum_{i=1}^{N} X_i^* M^{N-i},$$

where

$$X_i^* = \begin{cases} X_i, & \text{if } i \neq J, \\ X_i, & \text{if } i = J \text{ and } X - X_J M^{N-J} + (M-1)M^{N-J} \geq k, \\ I, & \text{if } i = J \text{ and } X - X_J M^{N-J} + (M-1)M^{N-J} < k. \end{cases}$$

Define for $i = 1, \dots, M$,

$$U_i^* = \sum_{j=1}^{N} I_{\{X_j^* = i-1\}}, \qquad U^* = (U_1^*, \dots, U_M^*).$$

The following alternative construction shows that (U, U^*) is exchangeable. First choose (I', I'', J) uniformly at random from the set $\{(i', i'', j)\colon 0 \leq i', i'' \leq M-1,\ 1 \leq j \leq N\}$. Next choose $(X_1, \dots, X_{J-1}, X_{J+1}, \dots, X_N)$ uniformly at random from $\{(x_1, \dots, x_{J-1}, x_{J+1}, \dots, x_N)\colon \sum_{i \neq j} x_i M^{N-i} < k\}$. If $(M-1)M^{N-J} + \sum_{i \neq J} X_i M^{N-i} < k$, we define

$$X = I' M^{N-J} + \sum_{i \neq J} X_i M^{N-i}, \qquad X^* = I'' M^{N-J} + \sum_{i \neq J} X_i M^{N-i}.$$

On the other hand, if $(M-1)M^{N-J} + \sum_{i \neq J} X_i M^{N-i} \geq k$, let M_0 denote the greatest integer satisfying $M_0 M^{N-J} + \sum_{i \neq J} X_i M^{N-i} < k$. Choose I''' uniformly at random from $\{i\colon 0 \leq i \leq M_0\}$. We define

$$X = X^* = I''' M^{N-J} + \sum_{i \neq J} X_i M^{N-i}.$$

The symmetry of (X, X^*) in the above construction immediately shows that (U, U^*) is an exchangeable pair of random variables.

Next we consider, as in Stein (1986), the antisymmetric function

$$(U, U^*) \mapsto f_A(U) I_{\{U = U^* - e^{(i)} + e^{(j)}\}} - f_A(U^*) I_{\{U^* = U - e^{(i)} + e^{(j)}\}},$$

with f_A defined as in (1). Then

$$\begin{aligned} 0 &= E\left[f_A(U) I_{\{U = U^* - e^{(i)} + e^{(j)}\}} - f_A(U^*) I_{\{U^* = U - e^{(i)} + e^{(j)}\}} \right] \\ &= E\big[f_A(U) P(U^* = U + e^{(i)} - e^{(j)} | X) \\ &\qquad - f_A(U - e^{(i)} + e^{(j)}) P(U^* = U - e^{(i)} + e^{(j)} | X) \big] \\ &= E\left[f_A(U) \frac{U_j - R_j}{MN} - f_A(U - e^{(i)} + e^{(j)}) \frac{U_i - R_i}{MN} \right], \end{aligned}$$

where $R_l = |\{m\colon\ X_m = l - 1,\ X - X_m M^{N-m} + (M-1) M^{N-m} \geq k\}|$, $l = 1, \dots, M$. Now it follows from Theorem 1 that

$$\begin{aligned} &|P(U \in A) - P(W \in A)| \\ &\quad = \left| E \sum_{i,j=1}^{M} \left[f_A(U - e^{(i)} + e^{(j)}) - f_A(U) \right] (R_i / M) \right| \qquad (11) \\ &\quad \leq \sup_{n, n - e^{(i)} + e^{(j)} \in \Omega} |f_A(n - e^{(i)} + e^{(j)}) - f_A(n)| E \sum_{l=1}^{M} (1 - 1/M) R_l, \end{aligned}$$

whenever $A \subseteq \Omega$ with $W \sim \mathrm{MULT}(N, 1/M, \dots, 1/M)$. We further observe from the definition of R_l that

$$\begin{aligned} ER_l &= \sum_{m=1}^{N} P\left[X \geq k - (M - l) M^{N-m} \right] \\ &= \sum_{m=1}^{N} (M - l) M^{N-m} / k \\ &\leq (M - l) M / (M - 1), \end{aligned}$$

and hence

$$(1 - 1/M) \sum_{l=1}^{M} ER_l \leq (M - 1) M / 2. \qquad (12)$$

We conclude from (11), (12) and Corollary 1 that

$$|P(U \in A) - P(W \in A)| \leq 3.3 (M - 1) M \sqrt{M/N}.$$

This completes the proof. □

REMARK. We do not know if the rate given by Theorem 4 is optimal. A referee conjectured the right rate to be $\sqrt{M/N}$ at worst.

3.2. *On the multivariate Poisson binomial distribution.* Let $X^{(1)}, \dots, X^{(N)}$ be independent M-variate random vectors such that $P(X^{(l)} = e^{(m)}) = p_m^{(l)}$, whenever $1 \le m \le M$, $1 \le l \le N$, and $\sum_{m=1}^M p_m^{(l)} = 1$ for each l. We define

$$S = (S_1, \dots, S_M) = \sum_{l=1}^N X^{(l)}, \qquad p_m = N^{-1} \sum_{l=1}^N p_m^{(l)}.$$

The random vector S is said to have the multivariate Poisson binomial distribution. In this example, we are interested in approximating the distribution of S by a multinomial distribution. It is clear that if $p_m^{(l)} = p_m$ for all $1 \le m \le M$ and $1 \le l \le N$, then the law of S is exactly $\mathrm{MULT}(N, p_1, \dots, p_M)$.

For the special case of $M = 2$, this problem has been given a comprehensive treatment by Ehm (1991) using also Stein's method. However, his solution rests on solving the Stein equation, namely (3) explicitly. While this is possible for $M = 2$, it is unclear how that argument could be extended to arbitrary M.

THEOREM 5. *Let $C^{(1)}$ and $C^{(2)}$ be defined as in Theorem 2 and 3 respectively. Then for $N \ge 2$, we have*

$$\begin{aligned} &d(\mathscr{L}(S), MULT(N, p_1, \dots, p_M)) \\ &\quad \le \sum_{l=1}^N \sum_{i<j} |p_i^{(l)} p_j - p_j^{(l)} p_i| \left\{ \frac{C^{(2)}}{N-1} \log\left[\frac{2(N-1)}{C^{(2)}}\right] + \left[\frac{C^{(2)}}{2(N-1)}\right]^2 \right. \\ &\qquad \left. + 2C^{(1)} N^{-1/2} \min\left[\prod_{l'=1}^N (1 - p_i^{(l')}), \prod_{l''=1}^N (1 - p_j^{(l'')})\right]\right\}, \end{aligned}$$

whenever $\max\{C^{(1)} N^{-1/2}, C^{(2)}[2(N-1)]^{-1}\} \le 1$.

PROOF. For each $1 \le l \le N$, we write $S^{(l)} = \sum_{l' \ne l} X^{(l')}$ and $X^{(l)} = (X_1^{(l)}, \dots, X_M^{(l)})$. Letting $W \sim \mathrm{MULT}(N, p_1, \dots, p_M)$, it follows from Theorem 1 that for $A \subseteq \Omega$, we have

$$\begin{aligned} &P(W \in A) - P(S \in A) \\ &\quad = E \sum_{l=1}^N \sum_{i,j=1}^M X_i^{(l)} p_j [f_A(S - e^{(i)} + e^{(j)}) - f_A(S)] \\ &\quad = E \sum_{l=1}^N \sum_{i,j=1}^M p_i^{(l)} p_j [f_A(S^{(l)} + e^{(j)}) - f_A(S^{(l)} + e^{(i)})] \\ &\quad = E \sum_{l=1}^N \sum_{i<j} (p_i^{(l)} p_j - p_j^{(l)} p_i)[f_A(S^{(l)} + e^{(j)}) - f_A(S^{(l)} + e^{(i)})] \\ &\quad = \sum_{l=1}^N \sum_{i<j} (p_i^{(l)} p_j - p_j^{(l)} p_i) \Big\{ E\big[f_A(S - X^{(l)} + e^{(j)}) \\ &\qquad - f_A(S - X^{(l)} + e^{(i)}) + f_A(S - e^{(j)} + e^{(i)}) - f_A(S) | S_j > 0\big] P(S_j > 0) \\ &\qquad + E\big[f_A(S^{(l)} + e^{(j)}) - f_A(S^{(l)} + e^{(i)}) | S_j = 0\big] P(S_j = 0) \Big\}. \end{aligned}$$

Hence it follows from Theorems 2 and 3 that

$$
\begin{aligned}
&|P(W \in A) - P(S \in A)| \\
&\quad \le \sup_{n\in\Omega:\, n_{j'}>0,\, n_{k'}>0} \left|\Delta_{j',i'}\left[\Delta_{k',j'} f_A(n)\right]\right| \sum_{l=1}^{N} \sum_{i<j} |p_i^{(l)} p_j - p_j^{(l)} p_i| \\
&\qquad + \sup_{n\in\Omega,\, n_{i'}>0} |\Delta_{i',j'} f_A(n)| \sum_{l=1}^{N} \sum_{i<j} |p_i^{(l)} p_j - p_j^{(l)} p_i| \prod_{l'=1}^{N} \left(1 - p_j^{(l')}\right) \\
&\quad \le \left\{ \frac{C^{(2)}}{N-1} \log\left[\frac{2(N-1)}{C^{(2)}}\right] + \left[\frac{C^{(2)}}{2(N-1)}\right]^2 \right\} \sum_{l=1}^{N} \sum_{i<j} |p_i^{(l)} p_j - p_j^{(l)} p_i| \\
&\qquad + 2C^{(1)} N^{-1/2} \sum_{l=1}^{N} \sum_{i<j} |p_i^{(l)} p_j - p_j^{(l)} p_i| \prod_{l'=1}^{N} \left(1 - p_j^{(l')}\right),
\end{aligned}
$$

whenever $\max\{C^{(1)}N^{-1/2}, C^{(2)}[2(N-1)]^{-1}\} \le 1$. The theorem now follows from the symmetry of i and j in the above argument. □

In the special case of the binomial approximation to the Poisson binomial distribution, Theorem 5 reduces to the following corollary.

COROLLARY 3. *Let* $C^{(1)}$ *and* $C^{(2)}$ *be defined as in Theorems* 2 *and* 3 *respectively. Then for* $N \ge 2$, *we have*

$$
\begin{aligned}
&d(\mathscr{L}(S), B(N, p_1)) \\
&\quad \le \left\{ \frac{C^{(2)}}{N-1} \log\left[\frac{2(N-1)}{C^{(2)}}\right] + \left[\frac{C^{(2)}}{2(N-1)}\right]^2 + 2^{1-N} \right\} \sum_{l=1}^{N} |p_1^{(l)} - p_1|, \,.
\end{aligned}
$$

whenever $\max\{C^{(1)}N^{-1/2}, C^{(2)}[2(N-1)]^{-1}\} \le 1$.

REMARK. Ehm (1991) showed that for $M = 2$ and $p_1 \in (0,1)$,

$$
d(\mathscr{L}(S), B(N, p_1)) \ge \frac{1}{124}\left[1 \wedge \frac{1}{Np_1(1-p_1)}\right] \sum_{l=1}^{N} \left(p_1^{(l)} - p_1\right)^2,
$$

$$
d(\mathscr{L}(S), B(N, p_1)) \le \frac{1}{(N+1)p_1(1-p_1)} \sum_{l=1}^{N} \left(p_1^{(l)} - p_1\right)^2.
$$

Comparing this result with that of Corollary 3, we infer that the bound given by Theorem 5 is unfortunately somewhat crude.

3.3. *On the multivariate hypergeometric distribution*. Consider a population of N_0 individuals, of which α_1 are of type 1, α_2 are of type $2, \dots, \alpha_M$ are of type M, with $\sum_{i=1}^{M} \alpha_i = N_0$. Suppose a sample of size N is chosen without replacement from among these N_0 individuals. For each $i = 1, \dots, M$, let V_i denote the number of individuals of type i found in the sample. Then the random vector $V = (V_1, \dots, V_M)$ is said to have the multivariate hypergeometric distribution with parameters $N, \alpha_1, \dots, \alpha_M$ [see, e.g., Johnson and Kotz (1969)]. When $M = 2$, it reduces to the usual hypergeometric distribution. In this subsection we are interested in approximating the distribution of V by MULT$(N, p_1, \dots, p_M)$, where $p_i = \alpha_i/N_0$, $i = 1, \dots, M$.

THEOREM 6. *With the above notation,*

$$d(\mathscr{L}(V), MULT(N, p_1, \dots, p_M)) \le \min\left\{(N-1)N/(2N_0), (1 \wedge 2C^{(1)}/\sqrt{N})\left(1 - \sum_{i=1}^{M} \alpha_i^2/N_0^2\right)N^2/N_0\right\},$$

where $C^{(1)}$ is defined as in Theorem 2.

PROOF. Let $W \sim$ MULT$(N, p_1, \dots, p_M)$. We couple V and W on the same probability space in the following way. Choose the sample of N individuals with replacement from the population of N_0 individuals. This determines W. If there are no repetitions, set $V = W$. Otherwise, replace those repeated individuals in the sample by individuals chosen at random uniformly from the remaining population without replacement so that the eventual sample has no repetitions. This determines V. Consequently, we have

$$\begin{aligned} d(\mathscr{L}(V), \mathscr{L}(W)) &\le P(V \ne W) \\ &\le 1 - \prod_{i=1}^{N-1}(1 - i/N_0) \qquad (13)\\ &\le (N-1)N/(2N_0). \end{aligned}$$

The above argument was suggested to us by Professor Herman Rubin.

As in the first example, we now construct an exchangeable pair of random vectors (V, V^*). Suppose we have the sample of N individuals obtained in the manner described in the first paragraph of this section. This determines V. Now choose an individual, call it a, uniformly at random from the sample and independently choose another individual, call that b, uniformly at random from the population of N_0 individuals. If b is already in the original sample, define $V^* = V$. Otherwise, replace a by b in the sample. Define for each $i = 1, \dots, M$, V_i^* to be the number of individuals of type i found in the revised sample. Write $V^* = (V_1^*, \dots, V_M^*)$.

By considering the antisymmetric function

$$(V, V^*) \mapsto f_A(V) I_{\{V = V^* - e^{(i)} + e^{(j)}\}} - f_A(V^*) I_{\{V^* = V - e^{(i)} + e^{(j)}\}},$$

with f_A defined as in (1), we have

$$\begin{aligned} 0 &= E\left[f_A(V) I_{\{V = V^* - e^{(i)} + e^{(j)}\}} - f_A(V^*) I_{\{V^* = V - e^{(i)} + e^{(j)}\}}\right] \\ &= E\left[f_A(V) P(V^* = V + e^{(i)} - e^{(j)} | V) \right.\\ &\qquad \left. - f_A(V - e^{(i)} + e^{(j)}) P(V^* = V - e^{(i)} + e^{(j)} | V)\right] \\ &= E\left[f_A(V)\frac{V_j p_i}{N}\left(1 - \frac{V_i}{p_i N_0}\right) - f_A(V - e^{(i)} + e^{(j)})\frac{V_i p_j}{N}\left(1 - \frac{V_j}{p_j N_0}\right)\right]. \end{aligned}$$

From Theorems 1 and 2, we have

$$
\begin{aligned}
&|P(V \in A) - P(W \in A)| \\
&\quad \leq \sup_{n, n - e^{(i)} + e^{(j)} \in \Omega} |f_A(n - e^{(i)} + e^{(j)}) - f_A(n)| E \sum_{i,j:\, i \neq j} V_i V_j / N_0 \\
(14)\quad &\quad \leq (1 \wedge 2C^{(1)}/\sqrt{N}) \sum_{i,j:\, i \neq j} (EV_i)(EV_j)/N_0 \\
&\quad \leq (1 \wedge 2C^{(1)}/\sqrt{N}) \left(1 - \sum_{i=1}^{M} \alpha_i^2 / N_0^2\right) N^2/N_0.
\end{aligned}
$$

In the second to last inequality, we have used the fact that for $i \neq j$, V_i and V_j are negatively correlated. Now the result follows from (13) and (14). □

COROLLARY 4. *Let* $p_1 = \cdots = p_M = 1/M$. *Then*

$$
\begin{aligned}
&d(\mathscr{L}(V), MULT(N, 1/M, \dots, 1/M)) \\
&\quad \leq \min\{(N-1)N/(2N_0), 6.6(1 - 1/M)\sqrt{M} N^{3/2}/N_0\}.
\end{aligned}
$$

PROOF. This is immediate from Corollary 1 and Theorem 6. □

REMARK. This example is probably the least satisfactory of the three examples considered in this section. It is known [see Ehm (1991)] that in the case $M = 2$,

$$C(N-1)/(N_0 - 1) \leq d(\mathscr{L}(V), B(N, p_1)) \leq (N-1)/(N_0 - 1),$$

whenever $1 \leq N \leq \alpha_1 \wedge \alpha_2$, for some universal positive constant C provided that $Np_1(1 - p_1) \geq 1$. This indicates that the bound given by Theorem 6 is not optimal for $M = 2$ (for large N, the bound is essentially a factor $N^{1/2}$ too large) and also probably not for $M > 2$.

APPENDIX

LEMMA 1. *With the notation of Theorem* 2, *we have*

$$|P(Z^{(n')}(t) + e^{(j)} \in A) - P(Z^{(n')}(t) + e^{(i)} \in A)| \leq C^{(1)}[(1 - e^{-t})N]^{-1/2}.$$

PROOF. We shall use the same coupling and notation as in the proof of Theorem 2. We write

$$Z^{(n')}(t) = W(t) + Y(t),$$

where $W(t) = (W_1(t), \dots, W_M(t))$ and $W_m(t)$ denotes the number of balls in urn

m (neglecting balls a and b) at time t which have not been picked even once. It is easily seen that

$$W_m(t) \sim \begin{cases} B(n_m, e^{-t}), & \text{if } m \neq i, \\ B(n_i - 1, e^{-t}), & \text{if } m = i. \end{cases}$$

$B(n_m, e^{-t})$ denotes the binomial distribution with parameters (n_m, e^{-t}). Furthermore, given that $W(t) = k$, with $k = (k_1, \dots, k_M)$ and $K = \sum_m k_m$,

$$Y(t) = (Y_1(t), \dots, Y_M(t)) \sim \text{MULT}(N - K - 1, p_1, \dots, p_M).$$

Thus

$$\begin{aligned} &P(Z^{(n')}(t) + e^{(j)} \in A) - P(Z^{(n')}(t) + e^{(i)} \in A) \\ &\quad = \sum_k P(W(t) = k) \sum_{l:\, l+k \in A} \{P[Y(t) = l - e^{(j)} | W(t) = k] \\ &\qquad\qquad -P[Y(t) = l - e^{(i)} | W(t) = k]\}. \end{aligned} \tag{15}$$

We observe that

$$\begin{aligned} &\left| \sum_{l:\, l+k \in A} \{P[Y(t) = l - e^{(j)} | W(t) = k] - P[Y(t) = l - e^{(i)} | W(t) = k]\} \right| \\ &\quad \leq \sum_{l:\, l_j \neq 0} P[Y(t) = l - e^{(j)} | W(t) = k] \left| 1 - \frac{l_i p_j}{l_j p_i} \right| \\ &\qquad + P[Y_j(t) = 0 | W(t) = k] \\ &\quad \leq \left\{ E\left[1 - \frac{L_i p_j}{(L_j + 1) p_i}\right]^2 \right\}^{1/2} + (1 - p_j)^{N-K-1}, \end{aligned} \tag{16}$$

where $L = (L_1, \dots, L_M) \sim \text{MULT}(N - K - 1, p_1, \dots, p_M)$. The last inequality uses the Cauchy–Schwarz inequality. To bound the right-hand side of (16), we first write

$$\begin{aligned} E\left[1 - \frac{L_i p_j}{(L_j + 1) p_i}\right]^2 = E\Bigg\{ 1 - \left(\frac{p_j}{p_i}\right) \frac{2L_i}{L_j + 1} \\ + \left(\frac{p_j}{p_i}\right)^2 \Bigg[\frac{L_i(L_i - 1)}{(L_j + 1)(L_j + 2)} + \frac{L_i}{(L_j + 1)(L_j + 2)} \\ + \frac{L_i(L_i - 1)}{(L_j + 1)^2 (L_j + 2)} + \frac{L_i}{(L_j + 1)^2 (L_j + 2)} \Bigg] \Bigg\}. \end{aligned} \tag{17}$$

Some straightforward algebra reveals that

$$\frac{p_j}{p_i} E \frac{L_i}{L_j + 1} = 1 - P(L_j = 0),$$

$$\left(\frac{p_j}{p_i}\right)^2 E \frac{L_i(L_i - 1)}{(L_j + 1)(L_j + 2)} \le 1 - P(L_j = 0),$$

$$\left(\frac{p_j}{p_i}\right)^2 E \frac{L_i}{(L_j + 1)(L_j + 2)} \le \frac{1}{p_i(N - K)},$$

$$\left(\frac{p_j}{p_i}\right)^2 E \frac{L_i(L_i - 1)}{(L_j + 1)^2(L_j + 2)} \le \frac{3}{p_j(N - K)},$$

$$\left(\frac{p_j}{p_i}\right)^2 E \frac{L_i}{(L_j + 1)^2(L_j + 2)} \le \frac{1}{p_i(N - K)}.$$

Hence it follows from (17) that

$$E\left[1 - \frac{L_i p_j}{(L_j + 1)p_i}\right]^2 \le P(L_j = 0) + \frac{2}{p_i(N - K)} + \frac{3}{p_j(N - K)} \le \frac{C_{i,j}}{N - K},$$

where

$$C_{i,j} = \frac{2}{p_i} + \frac{3}{p_j} + \frac{1}{ep_j(1 - p_j)}. \tag{18}$$

Now from (15) and (16), we get

$$\begin{aligned}
&|P(Z^{(n')}(t) + e^{(j)} \in A) - P(Z^{(n')}(t) + e^{(i)} \in A)| \\
&\quad \le \sum_k P(W(t) = k)\left\{\left[E\left(1 - \frac{L_i p_j}{(L_j + 1)p_i}\right)^2\right]^{1/2} + (1 - p_j)^{N-K-1}\right\} \\
&\quad \le \left\{C_{i,j}^{1/2} + \left[2ep_j(1 - p_j)^2\right]^{-1/2}\right\}\left(E\frac{1}{N - \Sigma_n W_m(t)}\right)^{1/2} \\
&\quad \le C(i,j)\left[(1 - e^{-t})N\right]^{-1/2},
\end{aligned}$$

with $C(i,j)$ defined as in (6). The second to last inequality follows from Jensen's inequality and the last inequality uses the fact that $\Sigma_m W_m(t)$ is distributed as $B(N - 1, e^{-t})$. From the symmetry of i and j in the above argument, we conclude that

$$\begin{aligned}
&|P(Z^{(n')}(t) + e^{(j)} \in A) - P(Z^{(n')}(t) + e^{(i)} \in A)| \\
&\quad \le [C(i,j) \wedge C(j,i)]\left[(1 - e^{-t})N\right]^{-1/2}.
\end{aligned}$$

Taking the supremum over all i, j satisfying $i < j$ proves the lemma. □

LEMMA 2. *With the notation of Theorem* 3, *we have*

$$|E(D)| \le C^{(2)}\left[(1-e^{-t})(N-1)\right]^{-1}.$$

PROOF. We shall use the same coupling and notation as in the proof of Theorem 3. We write

$$Z^{(n')}(t) = W(t) + Y(t),$$

where $W(t) = (W_1(t), \dots, W_M(t))$ and $W_m(t)$ denotes the number of balls in urn m, $m = 1, \dots, M$ (neglecting balls a, b', b'' and c), which have not been picked even once up to time t. It is easy to see that

$$W_m(t) \sim \begin{cases} B(n_m, e^{-t}), & \text{if } m \ne j, k, \\ B(n_m - 1, e^{-t}), & \text{if } m = j, k. \end{cases}$$

Furthermore, given $W(t) = \theta$, with $\theta = (\theta_1, \dots, \theta_M)$ and $\Theta = \sum_{m=1}^{M} \theta_m$, we have

$$Y(t) = (Y_1(t), \dots, Y_M(t)) \sim \text{MULT}(N - \Theta - 2, p_1, \dots, p_M).$$

Then with D as in (9), we have

$$\begin{aligned} E(D) &= P(Z^{(n')}(t) + e^{(i)} + e^{(j)} \in A) - P(Z^{(n')}(t) + 2e^{(j)} \in A) \\ &\quad + P(Z^{(n')}(t) + e^{(j)} + e^{(k)} \in A) - P(Z^{(n')}(t) + e^{(i)} + e^{(k)} \in A) \\ &= \sum_{\theta} P(W(t) = \theta) \sum_{l:\, l+\theta \in A} \{P(L = l - e^{(i)} - e^{(j)}) - P(L = l - 2e^{(j)}) \\ &\qquad + P(L = l - e^{(j)} - e^{(k)}) - P(L = l - e^{(i)} - e^{(k)})\}, \end{aligned} \tag{19}$$

where $L = (L_1, \dots, L_M) \sim \text{MULT}(N - \Theta - 2, p_1, \dots, p_M)$.

Case I. Suppose $i \ne k$. Writing $l = (l_1, \dots, l_M)$, we observe that

$$\begin{aligned} &\Bigg| \sum_{l:\, l+\theta \in A} \{P(L = l - e^{(i)} - e^{(j)}) - P(L = l - 2e^{(j)}) \\ &\qquad + P(L = l - e^{(j)} - e^{(k)}) - P(L = l - e^{(i)} - e^{(k)})\} \Bigg| \\ &\quad \le 2P(L_i = 0) + 2P(L_k = 0) + \sum_{l:\, l_i > 0,\, l_k > 0} P(L = l - e^{(i)} - e^{(k)}) \\ &\qquad \times \left| 1 - \frac{l_j p_i}{l_i p_j} - \frac{l_j p_k}{l_k p_j} + \frac{l_j (l_j - 1) p_i p_k}{l_i l_k p_j^2} \right| \\ &\quad \le 2P(L_i = 0) + 2P(L_k = 0) + E\left[\frac{L_j p_i p_k}{(L_i + 1)(L_k + 1) p_j^2} \right] \\ &\qquad + \left\{ E\left[1 - \frac{L_j p_i}{(L_i + 1) p_j} \right]^2 \right\}^{1/2} \left\{ E\left[1 - \frac{L_j p_k}{(L_k + 1) p_j} \right]^2 \right\}^{1/2}. \end{aligned} \tag{20}$$

With $C_{i,j}$ defined as in (18), it can be seen on further simplication that the right-hand side of (20) is bounded by

$$2(1-p_i)^{N-\Theta-2} + 2(1-p_k)^{N-\Theta-2} + \frac{\sqrt{C_{j,i}C_{j,k}} + (1/p_j)}{N-\Theta-1}$$
$$\le C(j,i,k)/(N-\Theta-1),$$

where $C(j,i,k)$ is as in (8).

Case II. Suppose $i = k$. Then

$$\left|\sum_{l:\, l+\theta\in A}\{P(L=l-2e^{(j)}) - 2P(L=l-e^{(i)}-e^{(j)}) + P(L=l-2e^{(i)})\}\right|$$
$$\le P(L_i=0) + P(L_j=0)$$
$$+ \sum_{l:\, l_i>0,\, l_j>0} P(L=l-e^{(i)}-e^{(j)})\left|\frac{(l_j-1)p_i}{l_ip_j} + \frac{(l_i-1)p_j}{l_jp_i} - 2\right|$$
$$\le P(L_i=0) + P(L_j=0)$$
$$+ E\left[\frac{L_jp_i}{(L_i+1)p_j} + \frac{L_ip_j}{(L_j+1)p_i} - 2 + \frac{2p_i}{(L_i+1)p_j} + \frac{2p_j}{(L_j+1)p_i}\right]$$
$$\le \left(\frac{2}{p_i} + \frac{2}{p_j}\right)/(N-\Theta-1).$$

Now it follows from both cases, (19) and the definition of $C^{(2)}$ that

$$|E(D)| \le C^{(2)}\sum_\theta P(W(t)=\theta)/(N-\Theta-1)$$
$$\le C^{(2)}E\left[1/\left(N-\sum_m W_m(t) - 1\right)\right]$$
$$\le C^{(2)}\left[(1-e^{-t})(N-1)\right]^{-1},$$

since $\sum_m W_m(t) \sim B(N-2, e^{-t})$. □

Acknowledgments. I would like to thank Professors Charles Stein, Louis Chen and Andrew Barbour for introducing this subject to me. Also I am grateful to Professor J. Michael Steele, an Associate Editor and anonymous referees for their insightful comments and suggestions which led to the sharpening of a number of results.

REFERENCES

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25.

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen-Stein method. *Statist. Sci.* **5** 403–424.

BALDI, P. and RINOTT, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650.

BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A** 175–184.

BARBOUR, A. D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322.

BARBOUR, A. D., CHEN, L. H. Y. and LOH, W. L. (1990). Compound Poisson approximation for nonnegative random variables via Stein's method. Preprint.

BOLTHAUSEN, E. and GÖTZE, F. (1989). The rate of convergence for multivariate sampling statistics. Preprint.

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

CHEN, L. H. Y. (1987). The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. Preprint.

DELANGE, H. (1975). Sur la fonction sommatoire de la fonction "somme des chiffres." *Enseign. Math.* (2) **21** 31–47.

DIACONIS, P. (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5** 72–81.

EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16.

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739.

GREEN, T. A. (1989). Asymptotic enumeration of generalized Latin rectangles. *J. Combin. Theory Ser. A* **51** 149–160.

HOLST, L. and JANSON, S. (1990). Poisson approximation using the Chen–Stein method and coupling: Number of exceedances of Gaussian random variables. *Ann. Probab.* **18** 713–723.

JOHNSON, N.L. and KOTZ, S. (1969). *Discrete Distributions*. Wiley, New York.

KARLIN, S. and MCGREGOR, J. (1965). Ehrenfest urn models. *J. Appl. Probab.* **2** 352–376.

LOH, W. L. (1991). Stein's method and multinomial approximation. Technical report, Dept. Statistics, Purdue Univ.

SCHNELLER, W. (1989). Edgeworth expansions for linear rank statistics. *Ann. Statist.* **17** 1103–1123.

SIEGERT, A. J. F. (1950). On the approach to statistical equilibrium. *Phys. Rev.* **76** 1708–1714.

STEIN, C. M. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 583–602. Univ. California Press, Berkeley.

STEIN, C. M. (1986). *Approximate Computation of Expectations*. IMS, Hayward, Calif.

STEIN, C. M. (1990). Application of Newton's identities to a generalized birthday problem and to the Poisson binomial distribution. Technical report, Dept. Statistics, Stanford Univ.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907