STEIN'S METHOD AND MULTINOMIAL APPROXIMATION

by

Wei–Liem Loh
Purdue University

Technical Report #91-02

Department of Statistics
Purdue University

January 1991

*

STEIN'S METHOD AND MULTINOMIAL APPROXIMATION[1]

by
Wei-Liem Loh
Department of Statistics
Purdue University
West Lafayette
IN 47907

[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.

STEIN'S METHOD AND MULTINOMIAL APPROXIMATION

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 two 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 (1989), Green (1989), Holst and Janson (1990), Schneller (1989), Stein (1990) and the references cited in them. 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 probabilitic 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

$$f'(w) - wf(w) = g(w), \qquad w \in R,$$

and in the Poisson approximation, we have

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

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 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 (1989) 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 particular, the Stein equation for the multinomial distribution is obtained together with smoothness estimates of its solution. In Section 3, these results are used in two 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 hypergeometric distribution.

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 are distributed in these urns. The system is said to be in state $n = (n_1, \ldots, n_M)$ if there are n_i balls in urn i, $i = 1, \ldots, M$. Events occur at random times and the time intervals T between successive events are independent random variables all with the same exponential distribution

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

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, $i = 1, \ldots, M$.

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

$$\Omega = \{(k_1, \ldots, k_M) : \sum_{i=1}^{M} k_i = N, k_i \geq 0, 1 \leq i \leq M\}$$

where $Z^{(n)}(0) = n$. It is clear that the stationary distribution of $Z^{(n)}(t)$ is $\mathrm{MULT}(N, p_1, \ldots, p_M)$, the multinomial distribution with parameters $N, p_1, \ldots, p_M$. 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

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

where $W \sim \mathrm{MULT}(N, p_1, \ldots, p_M)$ and $\sum_{i=1}^M n_i = N$. Also for simplicity, we define $I_A(.)$ 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 zero. We shall now proceed to derive a bound on f_A.

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

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

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

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

Theorem 1 *Let f_A be defined as in (1). Then f_A satisfies the equations*

$$\sum_{i,j=1}^{M} n_i p_j [f_A(n - e^{(i)} + e^{(j)}) - f_A(n)] = P(W \in A) - I_A(n), \quad \forall n \in \Omega.$$

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

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

Since $f_A(n,t) \to f_A(n)$ as $t \to \infty$, the theorem follows by letting t tend to infinity. □

Now we shall give two bounds on the 'smoothness' of f_A.

Theorem 2 *Let f_A be defined as in (1). Then*

$$\sup_{n, n-e^{(i)}+e^{(j)} \in \Omega} |f_A(n - e^{(i)} + e^{(j)}) - f_A(n)| \leq 3/2.$$

PROOF. 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 $N+1$ balls are placed in these urns such that n_m balls are placed in urn m, $m \neq j$, and $n_j + 1$ balls are placed in urn j. Again, we assume that events occur at random times and the time intervals T' between successive events are independent random variables all having the same exponential distribution

$$P(T' > t) = \exp[-(N+1)t], \quad t \geq 0.$$

When an event occurs, a ball is chosen uniformly at random from among the $N+1$ balls, removed from its urn, and the placed in urn m with probability p_m, $m = 1, \ldots, M$. At time $t = 0$, select a ball from urn i, call that ball a, and select a ball from urn j, call that ball b. At time t, define $a(t) = e^{(m)}$ $[b(t) = e^{(m)}]$ if ball a [ball b] is in urn m respectively, $m = 1, \ldots, M$. Writing $n' = (n_1, \ldots, n_{i-1}, n_i - 1, n_{i+1}, \ldots, n_M)$, we have

$$\begin{aligned} & f_A(n - e^{(i)} + e^{(j)}) - f_A(n) \\ = & \int_0^\infty [P(Z^{(n')}(t) + b(t) \in A) - P(Z^{(n')}(t) + a(t) \in A)]dt. \end{aligned}$$

Let τ_a and τ_b denote the times when balls a and b are first picked up respectively. By symmetry, we observe that $Z^{(n')}(t) + a(t)$ has the same distribution as $Z^{(n')}(t)+b(t)$ whenever $t>\tau_a \vee \tau_b$. Hence

$$\begin{aligned}
& f_A(n-e^{(i)}+e^{(j)})-f_A(n) \\
= & \int_0^\infty \{E[I_A(Z^{(n')}(t)+b(t))-I_A(Z^{(n')}(t)+a(t))|\tau_a<t<\tau_b] \\
& \times P(\tau_a<t<\tau_b) \\
& +E[I_A(Z^{(n')}(t)+b(t))-I_A(Z^{(n')}(t)+a(t))|\tau_a>t>\tau_b] \\
& \times P(\tau_a>t>\tau_b) \\
& +E[I_A(Z^{(n')}(t)+b(t))-I_A(Z^{(n')}(t)+a(t))|t<\tau_a\wedge\tau_b] \\
& \times P(t<\tau_a\wedge\tau_b)\}dt \qquad (2)
\end{aligned}$$

and

$$\begin{aligned}
|f_A(n-e^{(i)}+e^{(j)})-f_A(n)| & \le \int_0^\infty [2P(\tau_a<t<\tau_b)+P(t<\tau_a\wedge\tau_b)]dt \\
& = \int_0^\infty [2e^{-t}(1-e^{-t})+e^{-2t}]dt \\
& = 3/2.
\end{aligned}$$

The second last equality uses the observation that τ_a and τ_b are independent standard exponential random variables. □

Theorem 3 *Let f_A be defined as in (1). Then*

$$\sup_{n,n-e^{(i)}+e^{(j)}\in\Omega} |f_A(n-e^{(i)}+e^{(j)})-f_A(n)| \le C(p_1,\ldots,p_M)/\sqrt{N},$$

where $C(p_1,\ldots,p_M)=\sup_{i\le j}[C(i,j)\wedge C(j,i)]$ with

$$\begin{aligned}
C(i,j) = & \frac{2}{3}\{(\frac{3}{p_j}+\frac{1}{ep_j})^{1/2}+(\frac{3}{p_i}+\frac{1}{ep_i})^{1/2} \\
& +[\frac{8}{p_i}+\frac{12}{p_j}+\frac{4}{ep_j(1-p_j)}]^{1/2}+\frac{3-p_j}{1-p_j}(2ep_j)^{-1/2}+(2ep_i)^{-1/2}\}.
\end{aligned}$$

PROOF. Let $n, n-e^{(i)}+e^{(j)}\in\Omega$ with $i\neq j$. Also let n', $Z^{(n')}$, $a(t)$, $b(t)$, τ_a and τ_b be defined as in the proof of Theorem 2. We write

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

where $W(t)=(W_1(t),\ldots,W_M(t))$ and $W_m(t)$ denotes the number of balls in urn m [neglecting balls a and b] at time t which has not been picked even once. It is easily seen that

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

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

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

CASE I. Suppose $t < \tau_a \wedge \tau_b$, then

$$\begin{aligned}
& P(Z^{(n')}(t) + b(t) \in A) - P(Z^{(n')}(t) + a(t) \in A) \\
= \;& P(Z^{(n')}(t) + e^{(j)} \in A) - P(Z^{(n')}(t) + e^{(i)} \in A) \\
= \;& \sum_k P(W(t) = k) \sum_{l:l+k\in A} \{P[Y(t) = l - e^{(j)} | W(t) = k] \\
& -P[Y(t) = l - e^{(i)} | W(t) = k]\}. \qquad (3)
\end{aligned}$$

We observe that

$$\begin{aligned}
& |\sum_{l:l+k\in A} \{P[Y(t) = l - e^{(j)} | W(t) = k] \\
& -P[Y(t) = l - e^{(i)} | W(t) = k]\}| \\
\leq \;& \sum_{l:l_j\neq 0} P[Y(t) = l - e^{(j)} | W(t) = k] |1 - \frac{l_i p_j}{l_j p_i}| \\
& +P[Y_j(t) = 0 | W(t) = k] \\
\leq \;& \{E[1 - \frac{L_i p_j}{(L_j + 1) p_i}]^2\}^{1/2} + (1 - p_j)^{N-K-1}, \qquad (4)
\end{aligned}$$

where $L = (L_1, \ldots, L_M) \sim \mathrm{MULT}(N - K - 1, p_1, \ldots, p_M)$. The last inequality uses the Cauchy-Schwarz inequality. To bound the right hand side of (4), we first write

$$\begin{aligned}
& E[1 - \frac{L_i p_j}{(L_j + 1) p_i}]^2 \\
= \;& E\{1 - (\frac{p_j}{p_i}) \frac{2L_i}{L_j + 1} + (\frac{p_j}{p_i})^2 [\frac{L_i(L_i - 1)}{(L_j + 1)(L_j + 2)} + \frac{L_i}{(L_j + 1)(L_j + 2)} \\
& + \frac{L_i(L_i - 1)}{(L_j + 1)^2 (L_j + 2)} + \frac{L_i}{(L_j + 1)^2 (L_j + 2)}]\}. \qquad (5)
\end{aligned}$$

Some straightforward algebra reveal that

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

$$\begin{aligned}
(\frac{p_j}{p_i})^2 E\frac{L_i(L_i-1)}{(L_j+1)(L_j+2)} &\le 1-P(L_j=0),\\
(\frac{p_j}{p_i})^2 E\frac{L_i}{(L_j+1)(L_j+2)} &\le \frac{1}{p_i(N-K)},\\
(\frac{p_j}{p_i})^2 E\frac{L_i(L_i-1)}{(L_j+1)^2(L_j+2)} &\le \frac{3}{p_j(N-K)},\\
(\frac{p_j}{p_i})^2 E\frac{L_i}{(L_j+1)^2(L_j+2)} &\le \frac{1}{p_i(N-K)}.
\end{aligned}$$

Hence it follows from (5) that

$$E[1-\frac{L_ip_j}{(L_j+1)p_i}]^2 \le P(L_j=0)+\frac{2}{p_i(N-K)}+\frac{3}{p_j(N-K)} \le \frac{C_1}{N-K},$$

where

$$C_1 = \frac{2}{p_i}+\frac{3}{p_j}+\frac{1}{ep_j(1-p_j)}.$$

Now from (3) and (4), we get

$$\begin{aligned}
&|P(Z^{(n')}(t)+b(t)\in A)-P(Z^{(n')}(t)+a(t)\in A)|\\
\le\ & \sum_k P(W(t)=k)\{[E(1-\frac{L_ip_j}{(L_j+1)p_i})^2]^{1/2}+(1-p_j)^{N-K-1}\}\\
\le\ & \{C_1^{1/2}+[2ep_j(1-p_j)^2]^{-1/2}\}(E\frac{1}{N-\sum_m W_m(t)})^{1/2}\\
\le\ & C_2[(1-e^{-t})N]^{-1/2}, \qquad (6)
\end{aligned}$$

with $C_2 = C_1^{1/2}+[2ep_j(1-p_j)^2]^{-1/2}$. The second last inequality follows from Jensen's inequality and the last inequality uses the observation that $\sum_m W_m(t) \sim \mathrm{B}(N-1, e^{-t})$.

CASE II. Suppose $\tau_a < t < \tau_b$, then

$$\begin{aligned}
&P(Z^{(n')}(t)+b(t)\in A)-P(Z^{(n')}(t)+a(t)\in A)\\
=\ & \sum_k P(W(t)=k)\sum_{l:l+k\in A}\{P[Y(t)=l-e^{(j)}|W(t)=k]\\
& -P[Y(t)+a(t)=l|W(t)=k]\}. \qquad (7)
\end{aligned}$$

We observe that

$$|\sum_{l:l+k\in A}\{P[Y(t)=l-e^{(j)}|W(t)=k]$$

$$
\begin{aligned}
&-P[Y(t)+a(t)=l|W(t)=k]\}| \\
&\leq \; P[Y_j(t)+a_j(t)=0|W(t)=k] \\
&\quad + \sum_{l:l_j\neq 0} P[Y(t)=l-e^{(j)}|W(t)=k]|1-\frac{(N-K)p_j}{l_j}| \\
&\leq \; \{E[1-\frac{(N-K)p_j}{L_j+1}]^2\}^{1/2}+(1-p_j)^{N-K}, \qquad (8)
\end{aligned}
$$

where $a(t)=(a_1(t),\ldots,a_M(t))$ and $L=(L_1,\ldots,L_M)\sim \mathrm{MULT}(N-K-1,p_1,\ldots,p_M)$. Here we have used the fact that given $W(t)=k$, $Y(t)+a(t)\sim \mathrm{MULT}(N-K,p_1,\ldots,p_M)$. As in Case I, it can be shown that

$$
E[1-\frac{(N-K)p_j}{L_j+1}]^2 \leq \frac{C_3}{N-K},
$$

where $C_3=(3/p_j)+(ep_j)^{-1}$. It follows from (7) and (8) that

$$
\begin{aligned}
&|P(Z^{(n')}(t)+b(t)\in A)-P(Z^{(n')}(t)+a(t)\in A)| \\
&\leq \; \sum_k P(W(t)=k)\{(\frac{C_3}{N-K})^{1/2}+[\frac{1}{2ep_j(N-K)}]^{1/2}\} \\
&\leq \; C_4(E\frac{1}{N-\sum_m W_m(t)})^{1/2} \\
&\leq \; C_4[(1-e^{-t})N]^{-1/2}, \qquad (9)
\end{aligned}
$$

where $C_4=C_3^{1/2}+(2ep_j)^{-1/2}$.

CASE III. Suppose $\tau_b<t<\tau_a$, then by symmetry we have

$$
\begin{aligned}
&|P(Z^{(n')}(t)+b(t)\in A)-P(Z^{(n')}(t)+a(t)\in A)| \\
&\leq \; C_5[(1-e^{-t})N]^{-1/2}, \qquad (10)
\end{aligned}
$$

where

$$
C_5=(\frac{3}{p_i}+\frac{1}{ep_i})^{1/2}+(2ep_i)^{-1/2}.
$$

Finally, it follows from (2), (6), (9) and (10) that

$$
\begin{aligned}
&|f_A(n-e^{(i)}+e^{j)})-f_A(n)| \\
&\leq \; \int_0^\infty [(1-e^{-t})N]^{-1/2}[C_2e^{-2t}+(C_4+C_5)e^{-t}(1-e^{-t})]dt \\
&= \; C(i,j)/\sqrt{N},
\end{aligned}
$$

with $C(i,j) = (2/3)(2C_2 + C_4 + C_5)$. Hence from the symmetry of i and j in the above argument, we observe that

$$|f_A(n - e^{(i)} + e^{j)}) - f_A(n)| \leq [C(i,j) \wedge C(j,i)]/\sqrt{N}.$$

This completes the proof. □

Corollary 1 *Let f_A be defined as in (1) and $p_1 = \cdots = p_M = 1/M$. Then*

$$sup_{n,n-e^{(i)}+e^{(j)}\in\Omega}|f_A(n - e^{(i)} + e^{(j)}) - f_A(n)| \leq 7.4\sqrt{M/N}.$$

PROOF. The proof is immediate from Theorem 3. □

3 Applications

3.1 On the base M expansion of a random integer

Let k and M be natural numbers, with $M \geq 2$, and X a random variable uniformly distributed over the set $\{0, \ldots, k-1\}$. Define N to satisfy $M^{N-1} < k \leq M^N$. 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 $a_i, X_i \in \{0, \ldots, M-1\}$. Also define for $i = 1, \ldots, M$,

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

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

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

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 $\mathcal{L}(S)$.

Theorem 4 *For $M \geq 2$, we have*

$$d(\mathcal{L}(U), MULT(N, 1/M, \ldots, 1/M)) \leq 3.7(M-1)M\sqrt{M/N}.$$

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 $\{0, \ldots, M-1\}$ and J be a random variable uniformly distributed over $\{1, \ldots, N\}$ 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 $i = 1, \ldots, M$,

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

It is clear by symmetry that (U, U^*) is an exchangeable pair of random variables. Next we consider 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

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

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

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

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

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

and hence

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

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

$$|P(U \in A) - P(W \in A)| \leq 3.7(M-1)M\sqrt{M/N}.$$

This completes the proof. □

3.2 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, …, α_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 $i = 1, \ldots, M$, let V_i denote the number of individuals of type i found in the sample. Then the random vector $V = (V_1, \ldots, V_M)$ is said to have the multivariate hypergeometric distribution with parameters $N, \alpha_1, \ldots, \alpha_M$ [see, for example, 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 $\mathrm{MULT}(N, p_1, \ldots, p_M)$, where $p_i = \alpha_i/N_0$, $i = 1, \ldots, M$.

Theorem 5 *With the above notation,*

$$\begin{aligned} & d(\mathcal{L}(V), MULT(N, p_1, \ldots, p_M)) \\ \leq\ & \min\{(N-1)N/(2N_0), [(3/2) \wedge C(p_1, \ldots, p_M)/\sqrt{N}] \\ & \times(1 - \sum_{i=1}^{M} \alpha_i^2/N_0^2)N^2/N_0\}, \end{aligned}$$

where $C(p_1, \ldots, p_M)$ is defined as in Theorem 3.

PROOF. Let $W \sim \mathrm{MULT}(N, p_1, \ldots, p_M)$. 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

$$\begin{aligned} d(\mathcal{L}(V), \mathcal{L}(W)) & \leq P(V \neq W) \\ & \leq 1 - \prod_{i=1}^{N-1}(1 - i/N_0) \\ & \leq (N-1)N/(2N_0). \end{aligned} \tag{13}$$

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

As in the previous 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 subsection. 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 $i = 1, \ldots, M$, V_i^* to be the number of individuals of type i found in the revised sample. Write $V^* = (V_1^*, \ldots, V_M^*)$.

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

$$0 = E[f_A(V)I_{\{V=V^*-e^{(i)}+e^{(j)}\}} - f_A(V^*)I_{\{V^*=V-e^{(i)}+e^{(j)}\}}]$$

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

From Theorems 1, 2 and 3, we have

$$\begin{aligned}
& |P(V \in A) - P(W \in A)| \\
\leq & \sup_{n, n - e^{(i)} + e^{(j)} \in \Omega} |f_A(n - e^{(i)} + e^{(j)}) - f_A(n)| E \sum_{i,j: i \neq j} V_i V_j / N_0 \\
\leq & [(3/2) \wedge C(p_1, \ldots, p_M)/\sqrt{N}] \sum_{i,j: i \neq j} (EV_i)(EV_j)/N_0 \\
\leq & [(3/2) \wedge C(p_1, \ldots, p_M)/\sqrt{N}](1 - \sum_{i=1}^{M} \alpha_i^2 / N_0^2) N^2 / N_0. \qquad (14)
\end{aligned}$$

In the second 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 2 *Let* $p_1 = \ldots, p_M = 1/M$. *Then*

$$\begin{aligned}
& d(\mathcal{L}(V), MULT(N, 1/M, \ldots, 1/M)) \\
\leq & \min\{(N-1)N/(2N_0), 7.4(1 - 1/M)\sqrt{M} N^{3/2}/N_0\}.
\end{aligned}$$

PROOF. This is immediate from Corollary 1 and Theorem 5. □

4 Acknowledgments

I would like to thank Professors Charles Stein, Louis Chen and Andrew Barbour for introducing this subject to me.

References

[1] ARRATIA, R., GOLDSTEIN, L. and GORDEN, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method. *Ann. Probab.* **17** 9-25.

[2] ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen-Stein method. To appear in *Statist. Sci.*

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

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

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

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

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

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

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

[10] DELANGE, H. (1975). Sur la fonction sommatoire de la fonction "somme des chiffres." *L'Enseignement mathematique, deuxième serie* **21** 31-47.

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

[12] GÖTZE, F. (1989). On the rate of convergence in the multivariate CLT. Preprint.

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

[14] 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.

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

[16] KARLIN, S. and MCGREGOR, J. (1965). Ehrenfest urn models. *J. Appl. Probab.* **2** 352-376.

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

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

[19] STEIN, C. M. (1970). 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. of California Press.

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

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