The Annals of Probability
1992, Vol. 20, No. 4, 1843–1866

COMPOUND POISSON APPROXIMATION FOR NONNEGATIVE RANDOM VARIABLES VIA STEIN'S METHOD

By A. D. Barbour,[1] Louis H. Y. Chen[1] and Wei-Liem Loh[2]

Universität Zürich, National University of Singapore and Purdue University

Dedicated to Charles M. Stein on his seventieth birthday

The aim of this paper is to extend Stein's method to a compound Poisson distribution setting. The compound Poisson distributions of concern here are those of the form POIS(ν), where ν is a finite positive measure on $(0, \infty)$. A number of results related to these distributions are established. These in turn are used in a number of examples to give bounds for the error in the compound Poisson approximation to the distribution of a sum of random variables.

1. Introduction. In 1970, Stein 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. The method was extended from the normal distribution to the Poisson distribution by Chen (1974, 1975a). Since then Stein's method has been an area of intensive research in combinatorics, probability and statistics; see, for example, Arratia, Goldstein and Gordon (1989, 1990), Baldi and Rinott (1989), Barbour (1987), Barbour and Eagleson (1985), Barbour and Hall (1984), Barbour and Holst (1989), Barbour, Holst and Janson (1988), Bolthausen (1984), Bolthausen and Götze (1989), Chen (1987), Götze (1991), Green (1989), Schneller (1989), Stein (1990) and the references cited therein. An excellent account can be found in Stein (1986).

The aim of this paper is to extend Stein's method to a compound Poisson setting. A motivation for doing so is succinctly stated by Aldous (1989). The most interesting potential applications require extensions of the known results on Poisson approximations to the compound Poisson setting: Developing such extensions is an important research topic. In particular, one of the questions that we are interested in is: In situations in which the Poisson approximation is inadequate, when do we have approximately a compound Poisson distribution?

Received September 1990; revised May 1991.
[1]Research supported in part by Schweiz. Nationalsfonds Grant 21-25579.88.
[2]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, compound Poisson distribution, total variation distance, rate of convergence.

The following definition of a compound Poisson distribution is taken from Aldous (1989).

DEFINITION. Let ν be a positive measure on $(0, \infty)$ satisfying

$$\int_0^\infty (1 \wedge x)\nu(dx) < \infty,$$

where $(1 \wedge x) = \min(1, x)$. We say W has a compound Poisson distribution POIS(ν) if

$$E \exp(-\theta W) = \exp\left(-\int_0^\infty (1 - e^{-\theta x})\nu(dx)\right),$$

for all $\theta > 0$.

Thus the usual Poisson distribution with mean λ is POIS($\lambda\delta_1$), where δ_i is the degenerate probability measure at i. In this paper we are interested in the class of compound Poisson distributions of the form POIS(ν), where ν is some finite positive measure on $(0, \infty)$. Writing $\nu = \lambda\mu$, where $\lambda \geq 0$ and μ is a probability measure on $(0, \infty)$, we observe that POIS(ν) is the law of the random variable $\Sigma_{i=1}^N X_i$, where the X_i's are i.i.d. random variables with distribution μ and N is independent of the X_i's with the usual Poisson distribution with mean λ.

There are many studies on the rates of convergence to compound Poisson distributions. Examples of such studies include Le Cam (1960), Chen (1975b), Brown (1983), Serfozo (1986), Michel (1988) and Wang (1991). Arratia, Goldstein and Gordon (1990) have recently introduced an alternative approach to compound Poisson approximation for sums of indicators (also via Stein's method) which involves a "declumping" process. In contrast, we approach the compound Poisson approximation problem using Stein's method directly by considering a compound Poisson identity. In this way we avoid having to "declump." An advantage of this approach is that it applies not only to sums of indicators. Unfortunately, the compound Poisson identity is difficult to solve in general, and even if a solution is obtained it is difficult to obtain an effective bound on it. However, we believe that this approach has the potential of producing the best results when effective bounds are obtained.

The rest of this paper is organized as follows. Section 2 gives some results related to the compound Poisson distribution. In particular, a compound Poisson identity is obtained. This is specialized to discrete compound Poisson distributions in Section 3. In Section 4 these results are used in a number of examples to give bounds for the error in the compound Poisson approximation to the distribution of a sum of random variables. Example 1 gives a lower bound on the total variation distance between the law of a sum of independent discrete random variables and an appropriate compound Poisson distribution. Example 2 deals with random variables under local dependence (of which finitely dependent and m-dependent random variables appear as special cases) and Example 3 treats a problem on equiprobable allocations which involves long-term dependence.

2. A compound Poisson identity. In this section we consider a slight generalization of the compound Poisson distribution, namely the compound Poisson measure. We think that in doing so, the proof of Theorem 1 becomes more transparent.

DEFINITION. Let μ be a finite signed measure on $(0, \infty)$. We define the measure $|\mu|$ on the Borel subsets of $(0, \infty)$ by

$$|\mu|(E) = \sup \sum_{i=1}^{\infty} |\mu(E_i)|,$$

the supremum being taken over all partitions $\{E_i\}$ of E. Furthermore, we write $|\mu|[(0, \infty)] = \gamma_\mu$, the total mass of μ. The compound Poisson measure $S_{\lambda,\mu}$ on $[0, \infty)$, generated by μ and parameter $\lambda \geq 0$, is defined by

$$S_{\lambda,\mu}(\{0\}) = e^{-\lambda},$$

$$S_{\lambda,\mu}(E) = \sum_{i=0}^{\infty} \mu^{i*}(E)\lambda^i e^{-\lambda}/i!, \tag{1}$$

for all Borel subsets E of $(0, \infty)$, where μ^{i*} denotes the i-fold convolution of μ with itself. [Thus, if μ were a probability measure, then $S_{\lambda,\mu} = \text{POIS}(\lambda\mu)$.] We now state the main theorem of this section.

THEOREM 1. *Let g be a bounded function on $[0, \infty)$. The integral equation*

$$wf(w) - \lambda \int tf(w+t)\, d\mu(t) = g(w) \tag{2}$$

has a solution f defined on $[0, \infty)$ such that $\sup_{w>0}|wf(w)| < \infty$ if and only if $\int g\, dS_{\lambda,\mu} = 0$. The solution is unique except at $w = 0$ and for $w > 0$, it is given by

$$f(w) = \sum_{k=0}^{\infty} \lambda^k \int \cdots \int \frac{t_1 \cdots t_k g(w + t_1 + \cdots + t_k)}{w(w+t_1)\cdots(w+t_1+\cdots+t_k)} \prod_{i=1}^{k} d\mu(t_i).$$

Furthermore,

$$\sup_{w>0} |wf(w)| \leq \exp(\lambda\gamma_\mu) \sup_{w\geq 0} |g(w)|.$$

PROOF. Let $\mathscr{Y}$ be the Banach space of all bounded functions defined on $[0, \infty)$ and supplied with the sup norm $\|\cdot\|_{\mathscr{Y}}$ and let $\mathscr{G}$ be the quotient space of $\mathscr{Y}$ with respect to the closed subspace $\mathscr{M} = \{g \in \mathscr{Y}: g = 0 \text{ on } (0, \infty)\}$. Denote the norm of $\mathscr{G}$ by $\|\cdot\|_{\mathscr{G}}$. Also define the following normed linear space on equivalence classes of functions on $[0, \infty)$:

$$\mathscr{X} = \left\{ \tilde{f}: \sup_{w>0} |wf(w)| < \infty \right\},$$

where the norm $\|\cdot\|_{\mathscr{X}}$ is given by

$$\|\tilde{f}\|_{\mathscr{X}} = \sup_{w>0} |wf(w)|.$$

Let ν be a finite signed measure on the Borel subsets of $[0,\infty)$ and let $\tilde{f} \in \mathscr{X}$ or $\mathscr{Z}$. We define

$$\int \tilde{f} d\nu = \int_{(0,\infty)} f d\nu.$$

Since ν defines a linear functional on $\mathscr{Y}$ through integration, where ambiguity does not arise we shall use the same notation to denote the linear functional. Thus $\nu f = \int f d\nu$ for every $f \in \mathscr{Y}$ and $\ker S_{\lambda,\mu} = \{g \in \mathscr{Y}: \int g\, dS_{\lambda,\mu} = 0\}$.

Now define the linear operators U, M from $\mathscr{X}$ into $\mathscr{Y}$ and $\tilde{U}, \tilde{M}$ from $\mathscr{X}$ into $\mathscr{Z}$ as follows:

$$U\tilde{f}(w) = wf(w),$$

$$M\tilde{f}(w) = \int tf(w+t)\, d\mu(t),$$

$$\tilde{U}\tilde{f} = \widetilde{U\tilde{f}},$$

$$\tilde{M}\tilde{f} = \widetilde{M\tilde{f}}.$$

It is clear that U is an isometry and $\|M\| \le \gamma_\mu$ and hence $U - \lambda M$ is a bounded linear operator from $\mathscr{X}$ into $\mathscr{Y}$. Similarly $\tilde{U}$ is an isometry and $\|\tilde{M}\| \le \gamma_\mu$ and hence $\tilde{U} - \lambda\tilde{M}$ is a bounded linear operator from $\mathscr{X}$ into $\mathscr{Z}$. Furthermore, $\tilde{U}$ is bijective and hence $\tilde{U}^{-1}$ is also an isometry. This shows that $\mathscr{X}$ is a Banach space. Next we need a few lemmas.

Lemma 1. *The image of $U - \lambda M$ is contained in* $\ker S_{\lambda,\mu}$.

Proof. For every $\tilde{f} \in \mathscr{X}$,

$$\begin{aligned}\int U\tilde{f}(w)\, dS_{\lambda,\mu}(w) &= \int wf(w)\, dS_{\lambda,\mu}(w)\\ &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\int\cdots\int(t_1+\cdots+t_k)f(t_1+\cdots+t_k)\prod_{i=1}^{k}d\mu(t_i),\end{aligned}$$

which by symmetry, is equal to

$$\begin{aligned}&e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\int\cdots\int t_k f(t_1+\cdots+t_k)\prod_{i=1}^{k}d\mu(t_i)\\ &= \lambda\int\int tf(w+t)\, d\mu(t)\, dS_{\lambda,\mu}(w)\\ &= \int \lambda M\tilde{f}(w)\, dS_{\lambda,\mu}(w).\end{aligned}$$

This proves Lemma 1. □

The next lemma can be proved by induction.

LEMMA 2. *Let $a_1, \dots, a_k$ be complex numbers. Then, provided no denominator vanishes,*

$$\sum_{\pi} \frac{a_{\pi(1)} \cdots a_{\pi(k)}}{a_{\pi(1)}(a_{\pi(1)} + a_{\pi(2)}) \cdots (a_{\pi(1)} + \cdots + a_{\pi(k)})} = 1,$$

where the summation is taken over all permutations π.

LEMMA 3. *The operator $\tilde{U} - \lambda \tilde{M}$ is bijective. Its inverse is given by*

$$(\tilde{U} - \lambda \tilde{M})^{-1} \tilde{g} = \sum_{k=0}^{\infty} A^k \tilde{U}^{-1} \tilde{g}, \tag{3}$$

for every $\tilde{g} \in \mathscr{G}$ where $A = \lambda \tilde{U}^{-1} \tilde{M}$. Moreover,

$$\left\| (\tilde{U} - \lambda \tilde{M})^{-1} \right\| \le \exp(\lambda \gamma_\mu). \tag{4}$$

PROOF OF LEMMA 3. Since $\tilde{U}^{-1}$ exists, we may write

$$\tilde{U} - \lambda \tilde{M} = \tilde{U}(I - \lambda \tilde{U}^{-1} \tilde{M}),$$

where I is the identity operator on $\mathscr{X}$ and $I - \lambda \tilde{U}^{-1}\tilde{M}$ maps $\mathscr{X}$ into itself. Hence the bijectivity of $\tilde{U} - \lambda \tilde{M}$ is equivalent to the existence of $(I - \lambda \tilde{U}^{-1} \tilde{M})^{-1}$. Let $\tilde{f} \in \mathscr{X}$. By Fubini's theorem and symmetry, $A^k \tilde{f} = \tilde{h}$, where for $w > 0$,

$$wh(w) = \lambda \int t_1 f(w + t_1)\, d\mu(t_1)$$

if $k = 1$, and

$$wh(w) = \lambda^k \int \cdots \int \frac{t_1 \cdots t_k f(w + t_1 + \cdots + t_k)}{(w + t_1) \cdots (w + t_1 + \cdots + t_{k-1})} \prod_{i=1}^{k} d\mu(t_i)$$

if $k \ge 2$. So for $k \ge 1$, we have

$$\|A^k \tilde{f}\|_{\mathscr{X}} \le \lambda^k \|\tilde{f}\|_{\mathscr{X}} \int \cdots \int \frac{t_1 \cdots t_k}{t_1(t_1 + t_2) \cdots (t_1 + \cdots + t_k)} \prod_{i=1}^{k} d|\mu|(t_i),$$

which by Lemma 2 and symmetry, is equal to $(\lambda \gamma_\mu)^k \|\tilde{f}\|_{\mathscr{X}} / k!$. Hence

$$\sum_{k=0}^{\infty} \|A^k \tilde{f}\|_{\mathscr{X}} \le \exp(\lambda \gamma_\mu) \|\tilde{f}\|_{\mathscr{X}}. \tag{5}$$

This shows that $\{\sum_{k=0}^{\infty} A^k \tilde{f}\}$ is a Cauchy sequence in $\mathscr{X}$. By standard arguments using the completeness of $\mathscr{X}$ and the boundedness of A, the inverse of $I - A$ exists and is given by

$$(I - A)^{-1} \tilde{f} = \sum_{k=0}^{\infty} A^k \tilde{f}.$$

The bijectivity of $\tilde{U} - \lambda\tilde{M}$ follows. So does (3). Finally (4) follows from (5) and the fact that $\tilde{U}^{-1}$ is an isometry. This proves Lemma 3. □

For every $\tilde{g} \in \tilde{\mathscr{Y}}$, choose a representative g' of $\tilde{g}, g' \in \mathscr{Y}$ such that

$$g'(0) = -\exp(\lambda)\int \tilde{g}\, dS_{\lambda,\mu}.$$

This defines a linear map ϕ from $\tilde{\mathscr{Y}}$ into $\mathscr{Y}$ and clearly $\phi(\tilde{g}) \in \ker S_{\lambda,\mu}$. The next lemma can easily be proved.

LEMMA 4. *The map* ϕ *is injective and* $\operatorname{im}\phi = \ker S_{\lambda,\mu}$. *Furthermore,* $\|\phi^i\| \le 1$, *where* ϕ^i *is the left inverse of* ϕ.

LEMMA 5. *For every* $\tilde{f} \in \tilde{\mathscr{X}}$,

$$(U - \lambda M)\tilde{f} = \big(\phi\circ(\tilde{U} - \lambda\tilde{M})\big)\tilde{f}.$$

PROOF OF LEMMA 5. It suffices to show that

$$(U - \lambda M)\tilde{f}(0) = \big(\phi\circ(\tilde{U} - \lambda\tilde{M})\big)\tilde{f}(0).$$

Indeed the left-hand side equals $-\lambda\int tf(t)\,d\mu(t)$ and the right-hand side equals

$$\begin{aligned}-e^{\lambda}\int(\tilde{U} - \lambda\tilde{M})\tilde{f}dS_{\lambda,\mu} &= -e^{\lambda}\int_{(0,\infty)}(U - \lambda M)\tilde{f}(w)\,dS_{\lambda,\mu}(w)\\ &= e^{\lambda}\int(U - \lambda M)\tilde{f}dS_{\lambda,\mu} - \lambda\int tf(t)\,d\mu(t)\\ &= -\lambda\int tf(t)\,d\mu(t)\end{aligned}$$

by Lemma 1. Hence the lemma. □

Theorem 1 is proved by combining Lemmas 3, 4 and 5 and noting that

$$\|(U - \lambda M)^i\| \le \left\|(\tilde{U} - \lambda\tilde{M})^{-1}\right\| \|\phi^i\| \le \left\|(\tilde{U} - \lambda\tilde{M})^{-1}\right\|,$$

where $(U - \lambda M)^i$ is the left inverse of $U - \lambda M$. □

The following corollary gives a characterization of the compound Poisson measure.

COROLLARY 1. *Let* S *be a finite signed measure on the Borel subsets* $[0,\infty)$ *such that* $\int dS = \int dS_{\lambda,\mu} = \alpha \ne 0$. *Then* $S = S_{\lambda,\mu}$ *if and only if for every* $\tilde{f} \in \tilde{\mathscr{X}}$,

$$\int wf(w)\,dS(w) = \lambda\int\int tf(w + t)\,d\mu(t)\,dS(w).$$

PROOF. The necessity part follows from Lemma 1. Its sufficiency is proved by choosing f to be the solution of (2) with $g = h - \alpha^{-1}\int h\, dS_{\lambda,\mu}$ and $h \in \mathscr{Y}$. □

The class $\mathscr{X}$ in this corollary may be replaced by the smaller class containing $\tilde{f}$ where f is continuous on $(0,\infty)$ or continuous on $(0,\infty)$ with compact support. The next result gives a bound on the "smoothness" of f.

THEOREM 2. *Let $\lambda \geq 0$ and let μ be a probability measure on $(0,\infty)$. Let $E \subseteq [0,\infty)$, $I_E(\cdot)$ be the indicator function of E and f be defined as in Theorem* 1 *with $g(w) = I_E(w) - P(W \in E)$, for all $w \in [0,\infty)$ and $W \sim POIS(\lambda\mu)$. Then*

$$\sup_{E} \sup_{v \geq w > 0} |w[f(v) - f(w)]| \leq \exp(\lambda).$$

The proof of Theorem 2 is similar to that of Theorem 4 below and hence is omitted.

3. The lattice case. In this section we specialize the class of compound Poisson distributions to those of the form $\mathrm{POIS}(\sum_{i=1}^{\infty}\lambda_i\delta_i)$, where $\sum_{i=1}^{\infty}\lambda_i < \infty$. Due to the special structure of these distributions, somewhat sharper bounds than that in Theorem 1 can be obtained. The next result is the lattice case analogue of Theorem 1.

THEOREM 3. *Let $g\colon \{0,1,2,\dots\} \to R$ be a bounded function, $\lambda_i \geq 0$ whenever $i \geq 1$ and $\sum_{i=1}^{\infty}\lambda_i < \infty$. Then there exists a bounded solution $f\colon \{1,2,3,\dots\} \to R$ of*

$$wf(w) - \sum_{i=1}^{\infty} i\lambda_i f(w+i) = g(w), \qquad \forall\, w \geq 0,$$

if and only if $Eg(W) = 0$, with W having the $POIS(\sum_{i=1}^{\infty}\lambda_i\delta_i)$ distribution. The solution is unique except at $w = 0$ and for $w \geq 1$, it is given by

$$f(w) = \sum_{m=w}^{\infty} a_{m,w}\, g(m), \tag{6}$$

where

$$a_{w,w} = 1/w,$$
$$a_{w+i,w} = \sum_{j=1}^{i} \frac{j\lambda_j}{w+i} a_{w+i-j,w}, \qquad \forall\, i \geq 1. \tag{7}$$

We shall now proceed to bound f. To do so, we need the next few lemmas.

LEMMA 6. *For $w \geq 1$ and $i \geq 1$, we have*

$$wa_{w+i,w} = \sum_{j=1}^{i} j\lambda_j a_{w+i,w+j}. \tag{8}$$

PROOF. Clearly (8) holds for $i = 1$. Now assume that (8) is true for $i \leq k$. Then it follows from (7) that

$$\begin{aligned}\sum_{j=1}^{k+1} j\lambda_j a_{w+k+1,w+j} &= \sum_{l=1}^{k+1} \sum_{j=1}^{k+1-l} a_{w+k+1-l,w+j} j\lambda_j l\lambda_l/(w+k+1) \\ &= \sum_{l=1}^{k+1} wl\lambda_l a_{w+k+1-l,w}/(w+k+1) \\ &= wa_{w+k+1,w}.\end{aligned}$$

The lemma now follows by induction. □

LEMMA 7. *Suppose that W has the compound Poisson distribution $POIS(\sum_{i=1}^{\infty}\lambda_i\delta_i)$. Then*

$$P(W=w) = \exp\left(-\sum_{j=1}^{\infty}\lambda_j\right)\sum_{i=1}^{w} i\lambda_i a_{w,i}, \qquad \forall\, w \geq 1.$$

The proof of this lemma is similar to that of Lemma 6 and is omitted.

LEMMA 8. *Let W be a random variable having $POIS(\sum_{i=1}^{\infty}\lambda_i\delta_i)$ distribution. Then for $1 \leq i \leq w$, we have*

$$ia_{w,i} \leq P(W=w-i)\exp\left(\sum_{j=1}^{\infty}\lambda_j\right).$$

PROOF. First observe that the lemma holds when $w = i$. Next, we assume that $w > i$. By Lemmas 6 and 7, we have

$$P(W=w-i)\exp\left(\sum_{j=1}^{\infty}\lambda_j\right) - ia_{w,i} = \sum_{j=1}^{w-i} j\lambda_j(a_{w-i,j} - a_{w,i+j}) \geq 0,$$

since it is easily seen that $a_{w-i,j} \geq a_{w,i+j}$ whenever $w - i \geq j > 0$. □

LEMMA 9. *Let W be a random variable having $POIS(\sum_{i=1}^{\infty}\lambda_i\delta_i)$ distribution. If $1 \leq i \leq w$, then*

$$\beta_i a_{w,i} \leq P(W=w)\exp\left(\sum_{j=1}^{\infty}\lambda_j\right),$$

where for all $j \geq 1$, *we define*

$$\beta_1 = \lambda_1,$$

$$\beta_{j+1} = (j+1)\lambda_{j+1} + \lambda_1(\beta_j/j) + \sum_{k=1}^{j-1} (j+1-k)\lambda_{j+1-k}(\beta_k/k).$$

PROOF. It follows from Lemma 7 that to prove the above lemma, it suffices only to show that for $1 \leq i \leq w$,

$$\sum_{j=1}^{w} j\lambda_j a_{w,j} = \beta_i a_{w,i} + \sum_{j=i+1}^{w} \left[j\lambda_j a_{w,j} + \sum_{k=1}^{i-1} (\beta_k/k)(j-k)\lambda_{j-k} a_{w,j} \right]. \tag{9}$$

The proof of (9) easily follows from induction and Lemma 6. □

The following result gives a bound on f.

PROPOSITION 1. *Let* f *be defined by* (6). *Then we have for* $i \geq 1$,

$$\sup_{w \geq i} |f(w)| \leq \left[i^{-1} \wedge \min_{1 \leq k \leq i} (\beta_k^{-1}) \right] \exp\left(\sum_{j=1}^{\infty} \lambda_j \right) \sup_{w \geq i} |g(w)|,$$

where thr β_k*'s are defined as in Lemma* 9.

PROOF. We observe that if $w \geq i$, then

$$|f(w)| \leq \sum_{m=w}^{\infty} a_{m,w} |g(w)| \leq \sum_{m=i}^{\infty} a_{m,i} \sup_{w \geq i} |g(w)|.$$

It follows from Lemma 9 that

$$\begin{aligned} \sup_{w \geq i} |f(w)| &\leq \min_{1 \leq k \leq i} \left\{ \exp\left(\sum_{j=1}^{\infty} \lambda_j \right) \sup_{w \geq i} |g(w)| \sum_{m=k}^{\infty} \beta_k^{-1} P(W=m) \right\} \\ &\leq \min_{1 \leq k \leq i} \beta_k^{-1} \exp\left(\sum_{j=1}^{\infty} \lambda_j \right) \sup_{w \geq i} |g(w)|. \end{aligned} \tag{10}$$

On the other hand, it follows from Lemma 8 that

$$\begin{aligned} \sup_{w \geq i} |f(w)| &\leq \exp\left(\sum_{j=1}^{\infty} \lambda_j \right) \sup_{w \geq i} |g(w)| \sum_{m=i}^{\infty} P(W=m-i)/i \\ &= i^{-1} \exp\left(\sum_{j=1}^{\infty} \lambda_j \right) \sup_{w \geq i} |g(w)|. \end{aligned} \tag{11}$$

Proposition 1 follows from (10) and (11). □

The next two theorems, which are needed in the sequel, give bounds on the "smoothness" of f.

THEOREM 4. *Let f be defined by* (6) *and $A \subseteq \{0, 1, 2, \dots\}$. Suppose further that $g(v) = I_A(v) - P(W \in A)$ whenever $v \in \{0, 1, 2, \dots\}$, with W having the POIS$(\sum_{i=1}^{\infty} \lambda_i \delta_i)$ distribution. Then for $i \geq 1$ and $j \geq 0$, we have*

$$\sup_A \sup_{w \geq i} |f(w+j) - f(w)| \leq \left[i^{-1} \wedge \min_{1 \leq k \leq i} (\beta_k^{-1})\right] \exp\left(\sum_{l=1}^{\infty} \lambda_l\right),$$

where the β_k's are defined as in Lemma 9.

PROOF. From the proof of Proposition 1, it suffices only to show

$$|f(i+j) - f(i)| \leq \sum_{m=i}^{\infty} a_{m,i}. \tag{12}$$

First we observe that

$$\begin{aligned} f(i+j) - f(i) &= -\sum_{m=i}^{\infty} \left[a_{m,i} g(m) - a_{m+j,i+j} g(m+j)\right] \\ &\geq -\sum_{m=i}^{\infty} \left[a_{m,i} - P(W \in A)(a_{m,i} - a_{m+j,i+j})\right] \\ &\geq -\sum_{m=i}^{\infty} a_{m,i}. \end{aligned}$$

Also,

$$\begin{aligned} f(i+j) - f(i) &\leq \sum_{m=i}^{\infty} \left[a_{m+j,i+j} + P(W \in A)(a_{m,i} - a_{m+j,i+j})\right] \\ &\leq \sum_{m=i}^{\infty} a_{m,i}. \end{aligned}$$

This proves (12). □

REMARK. The bound given by Theorem 4 is sharp in the limit as the λ_i's approach 0. Furthermore, there are also other instances in which the bound is reasonably good. For example, suppose $W \sim$ POIS$(\lambda_2 \delta_2)$. By taking A to be the set of nonnegative even integers, it can easily be seen that $\sup_A \sup_{w \geq 1} |f(w+1) - f(w)| > \lambda_2^{-1}(e^{\lambda_2} - 1)$. However, it is enough in applications to find one function f with nice properties such that, for any random variable V,

$$E\left\{V f(V) - \sum_{i=1}^{\infty} i \lambda_i f(V+i)\right\} = P(V \in A) - P(W \in A),$$

where W has the POIS$(\sum_{i=1}^{\infty} \lambda_i \delta_i)$ distribution. In this degenerate example, a better choice of f is obtained by taking

$$g(w) = \begin{cases} I_A(w) - P(W \in A), & \text{if } w \in 2Z, \\ 0, & \text{if } w \in 2Z + 1, \end{cases}$$

which is just as good, since $P(W \in 2Z) = 1$. For this function g, $f(w) = 0$ for $w \in 2Z + 1$, $|f(w)| \le c\lambda^{-1/2}$ and $\sup_w |f(w + 2) - f(w)| \le 1/(2\lambda_2)$ for some constant c. In general, the bound in Theorem 4 is very crude and is far from optimal. It is hoped that this can be remedied in future work.

Under additional assumptions on the λ_i's, substantially better bounds on the "smoothness" of f can be obtained using the probabilistic perturbation technique of Barbour (1988, 1990). Suppose that $j\lambda_j \searrow 0$ as $j \to \infty$. We write

$$\mu_i = i\lambda_i - (i + 1)\lambda_{i+1}, \qquad i \ge 1,$$

$$f(w) = h(w) - h(w - 1), \qquad w \ge 0.$$

It can be easily seen that

$$wf(w) - \sum_{i=1}^{\infty} i\lambda_i f(w + i) = w[h(w) - h(w - 1)] - \sum_{i=1}^{\infty} \mu_i[h(w + i) - h(w)].$$

We observe that the right-hand side of the above equation is of the form $-\mathscr{A}h$, where $\mathscr{A}$ is the infinitesimal generator of an immigration (in groups)-death process whose equilibrium distribution is $\text{POIS}(\sum_{i=1}^{\infty}\lambda_i\delta_i)$. Let Z be the minimal process with the infinitesimal generator $\mathscr{A}$. For $A \subseteq \{0, 1, 2, \ldots\}$, let h_A: $\{0, 1, 2, \ldots\} \to R$ be given by

$$h_A(w) = \int_0^{\infty} [P_w(Z(t) \in A) - P(W \in A)]\, dt,$$

where $W \sim \text{POIS}(\sum_{i=1}^{\infty}\lambda_i\delta_i)$ and P_w denotes the distribution given $Z(0) = w$. We further observe that

$$|h_A(w)| \le \int_0^{\infty} P_w(\tau > t)\, dt = E(\tau) < \infty,$$

where τ is the first coincidence of $Z(t)$ started at w and another independently started with initial distribution $\text{POIS}(\sum_{i=1}^{\infty}\lambda_i\delta_i)$. The proofs of the next two lemmas are very similar to the proofs of Lemmas 1 and 2 of Barbour (1988) and shall be omitted.

LEMMA 10. *Let* $f(w) = h_A(w) - h_A(w - 1)$, $w \ge 1$. *Then* f *satisfies*

$$wf(w) - \sum_{i=1}^{\infty} i\lambda_i f(w + i) = I_A(w) - P(W \in A), \qquad w \ge 0.$$

LEMMA 11. *Let* $f(w) = h_A(w) - h_A(w-1)$, $w \geq 1$. *Then*

$$\sup_A \sup_{w \geq 1} |f(w+1) - f(w)| \leq 1.$$

THEOREM 5. *Suppose* $j\lambda_j \searrow 0$ *as* $j \to \infty$. *Then*

$$\sup_A \sup_{w \geq 1} |f(w+1) - f(w)| \leq 1 \wedge \frac{1}{\lambda_1 - 2\lambda_2}\left[\frac{1}{4(\lambda_1 - 2\lambda_2)} + \log^+ 2(\lambda_1 - 2\lambda_2)\right],$$

where $f(w) = h_A(w) - h_A(w-1)$, $w \geq 1$.

PROOF. Without loss of generality, it follows from Lemma 11 that it suffices only to show

$$\sup_A \sup_{w \geq 1} |f(w+1) - f(w)| \leq \frac{1}{\lambda_1 - 2\lambda_2}\left[\frac{1}{4(\lambda_1 - 2\lambda_2)} + \log^+ 2(\lambda_1 - 2\lambda_2)\right],$$

whenever $2(\lambda_1 - 2\lambda_2) > 1$. From the definition of f, it is easy to see that for $w \geq 0$,

$$f(w+2) - f(w+1) = \int_0^\infty \{P_{w+2}[Z(t) \in A] - 2P_{w+1}[Z(t) \in A] + P_w[Z(t) \in A]\}\,dt.$$

Define the following four coupled immigration (in groups)-death processes: $Z^{(0)}$ is distributed as Z started at w and

$$Z^{(1)}(t) = Z^{(0)}(t) + I_{\{\tau_1 > t\}},$$
$$Z^{(2)}(t) = Z^{(0)}(t) + I_{\{\tau_2 > t\}},$$
$$Z^{(3)}(t) = Z^{(1)}(t) + I_{\{\tau_2 > t\}},$$

where τ_1 and τ_2 are independent standard exponential random variables, independent of $Z^{(0)}$. It follows that

$$f(w+2) - f(w+1) = \int_0^\infty E\{I_A[Z^{(3)}(t)] - I_A[Z^{(2)}(t)] - I_A[Z^{(1)}(t)] + I_A[Z^{(0)}(t)]\}\,dt.$$

We observe that the above integrand is 0 whenever $t \geq (\tau_1 \wedge \tau_2)$. Hence

$$\begin{aligned} f(w+2) - f(w+1) = \int_0^\infty e^{-2t}\{&P[Z^{(0)}(t) \in A - 2] \\ &-2P[Z^{(0)}(t) \in A - 1] \\ &+P[Z^{(0)}(t) \in A]\}\,dt, \end{aligned} \tag{13}$$

where $A - i = \{k\colon k + i \in A\}$. We observe that $Z^{(0)}(t) = W(t) + \sum_{i=1}^\infty Y_i(t)$, where $W(t)$ denotes those of the original w individuals still alive at time t and $Y_i(t)$ denotes those alive at t who immigrated in groups of size i after time 0. Furthermore, we observe that $W(t), Y_1(t), Y_2(t), \dots$ are independent random variables and $Y_1(t) \sim \text{POIS}[(1-e^{-t})\mu_1\delta_1]$. Thus it follows that

$$P\big[Z^{(0)}(t) \in A-2\big] - 2P\big[Z^{(0)}(t) \in A-1\big] + P\big[Z^{(0)}(t) \in A\big]$$

$$(14) \qquad = \sum_k P\left[W(t) + \sum_{i=2}^{\infty} Y_i(t) = k\right] \sum_{l:\, l+k \in A} p_l(t),$$

where $p_l(t) = P[Y_1(t) = l-2] - 2P[Y_1(t) = l-1] + P[Y_1(t) = l]$. As shown in Barbour (1988), we have

$$(15) \qquad \left| \sum_{l:\, l+k \in A} p_l(t) \right| \le \big[(1-e^{-t})\mu_1\big]^{-1}.$$

Hence we conclude from (13), (14) and (15) that

$$|f(w+2) - f(w+1)| \le \int_0^\infty e^{-2t}\Big\{\big[(1-e^{-t})\mu_1\big]^{-1} \wedge 2\Big\}\, dt$$
$$= \frac{1}{\mu_1}\left(\frac{1}{4\mu_1} + \log^+ 2\mu_1\right).$$

This proves the theorem. □

REMARK. We observe from the proof of the above theorem that it should be possible, at the cost of more technical complications, to get better bounds by looking at the transition probabilities of the whole μ-process, not just the μ_1 part.

REMARK. Theorem 5 is most useful if $\lambda_1 \gg \lambda_2$ (as is often the case in "perturbation" problems).

Finally we end this section with a corresponding bound for f. Although this result is not needed in the sequel, we think it is of some independent interest.

PROPOSITION 2. *Suppose* $j\lambda_j \searrow 0$ *as* $j \to \infty$. *Then*

$$\sup_A \sup_{w \ge 1} |f(w)| \le \begin{cases} 1, & \text{if } \lambda_1 - 2\lambda_2 \le 1, \\ \left(1/\sqrt{\lambda_1 - 2\lambda_2}\right)\left[2 - \left(1/\sqrt{\lambda_1 - 2\lambda_2}\right)\right], & \text{if } \lambda_1 - 2\lambda_2 > 1. \end{cases}$$

where $f(w) = h_A(w) - h_A(w-1)$, $w \ge 1$.

The proof of Proposition 2 is similar to that of Theorem 5 and hence is omitted.

4. Applications. In this section we shall use the results of the previous sections to obtain bounds for the error in the compound Poisson approximation to the distribution of a sum of random variables. Here the total variation distance is used to measure how close the distribution of the random variable of interest is to a compound Poisson distribution.

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

$$d(F,G) = \sup_E |F(E) - G(E)|,$$

where the supremum is taken over all measurable sets of the real line. We observe that $2d(F,G) = \gamma_{F-G}$, the total mass of the signed measure $F - G$ (see Section 2). Also for simplicity, we denote the law of a random variable X by $\mathscr{L}(X)$.

4.1. *Sum of independent discrete random variables.* Upper bounds on the total variation distance between the distribution of a sum of independent random variables and an appropriate compound Poisson distribution have been obtained by Le Cam (1960) and Chen (1975b) using different methods. Their techniques are more direct and general and give more reasonable upper bounds than those considered here. As such, we shall give a complementary lower bound instead.

Let $Y_1, \dots, Y_m$ be independent random variables taking on only the values $0, 1, 2, \dots, n$. Define for $1 \le i \le n$, $1 \le j \le m$,

$$p_{i,j} = P(Y_j = i), \qquad V = \sum_{j=1}^{m} Y_j,$$

$$\lambda_i = \sum_{j=1}^{m} p_{i,j}, \qquad V_j = \sum_{i \ne j} Y_i.$$

The proof of the following theorem is similar to that given by Barbour and Hall (1984) for getting lower bounds in Poisson approximations.

THEOREM 6. *With the above notation, we have*

$$d\left(\mathscr{L}(V), POIS\left(\sum_{i=1}^{n} \lambda_i \delta_i\right)\right) \ge \frac{1}{32n^2}\left(1 \wedge \left(\sum_{i=1}^{n} i\lambda_i\right)^{-1}\right) \sum_{j=1}^{m} \left(\sum_{k=1}^{n} kp_{k,j}\right)^2.$$

PROOF. Let $h: \{0, 1, 2, \dots\} \to R$ be a bounded function and W be a random variable having $\mathrm{POIS}(\sum_{i=1}^{n} \lambda_i \delta_i)$ distribution. We observe that

$$\begin{aligned} &E\left[\sum_{i=1}^{n} i\lambda_i h(V+i) - Vh(V)\right] \\ &\quad = E\left\{\sum_{i=1}^{n}\left[i\lambda_i h(V+i) - \sum_{j=1}^{m} ip_{i,j} E\big(h(V) \mid Y_j = i\big)\right]\right\} \\ &\quad = E\left\{\sum_{j=1}^{m} \sum_{i,k=1}^{n} ip_{i,j} p_{k,j}\big[h(V_j + i + k) - h(V_j + i)\big]\right\}. \end{aligned}$$

It is easy to check that $E[Wh(W) - \sum_{i=1}^n i\lambda_i h(W+i)] = 0$ and hence

$$E\left[\sum_{i=1}^n i\lambda_i h(V+i) - Vh(V) - \sum_{i=1}^n i\lambda_i h(W+i) + Wh(W)\right]$$
$$= E\left\{\sum_{j=1}^m \sum_{i,k=1}^n ip_{i,j}p_{k,j}\left[h(V_j+i+k) - h(V_j+i)\right]\right\}.$$

This implies that

$$(16)\qquad \begin{aligned} &2d(\mathscr{L}(V),\mathscr{L}(W)) \sup_j \left|\sum_{i=1}^n i\lambda_i h(i+j) - jh(j)\right| \\ &\quad \geq \sum_{j=1}^m \sum_{i,k=1}^n ip_{i,j}p_{k,j}E\left[h(V_j+i+k) - h(V_j+i)\right]. \end{aligned}$$

Let $\lambda = \sum_{i=1}^n i\lambda_i$ and

$$h(j) = (j-\lambda)\exp\left(-(j-\lambda)^2/(\theta\lambda)\right), \qquad \forall j \geq 0,$$

where θ is some positive number to be determined later. To get an upper bound for $\sup_j|\sum_{i=1}^n i\lambda_i h(i+j) - jh(j)|$, we observe that

$$(17)\qquad \begin{aligned} &\left|\sum_{i=1}^n i\lambda_i h(i+j) - jh(j)\right| \\ &\quad = \left|\sum_{i=1}^n i\lambda_i[h(i+j) - h(j)] - (j-\lambda)^2\exp\left(-(j-\lambda)^2/(\theta\lambda)\right)\right| \\ &\quad \leq \lambda \max\{n, 2ne^{-3/2} + \theta e^{-1}\}. \end{aligned}$$

To get a lower bound for the right-hand side of (16), we observe that

$$1 - d(we^{-w^2/(\theta\lambda)})/dw \leq 3w^2/(\theta\lambda).$$

Hence writing $U_j = V_j - \lambda$ for all $1 \leq j \leq m$, we have

$$\int_{U_j+i}^{U_j+i+k}\left[1 - d(we^{-w^2/(\theta\lambda)})/dw\right]dw \leq \int_{U_j+i}^{U_j+i+k} 3w^2/(\theta\lambda)\,dw.$$

This implies that

$$\begin{aligned} &k - h(V_j+i+k) + h(V_j+i) \\ &\quad \leq \left[k^3 + 3ik^2 + 3i^2k + 3U_j^2k + 3U_j(k^2+2ik)\right]/(\theta\lambda). \end{aligned}$$

Furthermore, it is easy to see that $EU_j = -\sum_{k=1}^n kp_{k,j}$ and $EU_j^2 \leq n\lambda$. Hence it follows that

$$(18)\qquad E\left[k - h(V_j+i+k) + h(V_j+i)\right] \leq k(7n^2 + 3n\lambda)/(\theta\lambda).$$

It follows from (16), (17) and (18) that if we take $\theta \geq ne$,

$$\begin{aligned} &d(\mathscr{L}(V),\mathscr{L}(W)) \\ &\quad \geq \sum_{j=1}^m \left(\sum_{k=1}^n kp_{k,j}\right)^2\left[1 - \frac{7n^2+3n\lambda}{\theta\lambda}\right]/\left[2\lambda(2ne^{-3/2} + \theta e^{-1})\right]. \end{aligned}$$

As in Barbour and Hall (1984), if $\lambda \ge 1$, we take $\theta = 21n^2$ and if $\lambda < 1$, we take $\theta = 21n^2/\lambda$. In both cases, we have

$$d(\mathscr{L}(V), \mathscr{L}(W)) \ge \frac{1}{32n^2}\left(1 \wedge \left(\sum_{i=1}^{n} i\lambda_i\right)^{-1}\right)\sum_{j=1}^{m}\left(\sum_{k=1}^{n} kp_{k,j}\right)^2.$$

This completes the proof. □

REMARK. In the case where $n = 1$, Theorem 6 reduces to Theorem 2 of Barbour and Hall (1984).

4.2. *Random variables under local dependence.* In this subsection, we shall approximate the distribution of a sum of locally dependent random variables by that of a suitably chosen compound Poisson distribution.

DEFINITION. Let I be an arbitrary index set. A nonempty family of random variables $\{X_\alpha: \alpha \in I\}$ is said to be locally dependent if for each $\alpha \in I$, there exist $A_\alpha \subseteq B_\alpha \subseteq I$ with $\alpha \in A_\alpha$ such that X_α is independent of $\{X_\beta: \beta \in A_\alpha^c\}$ and $\{X_\beta: \beta \in A_\alpha\}$ is independent of $\{X_\beta: \beta \in B_\alpha^c\}$. Let A and B be nonempty subsets of I. The set $\{X_\alpha: \alpha \in B\}$ is said to be a locally dependent set of $\{X_\alpha: \alpha \in A\}$ if the latter is independent of $\{X_\alpha: \alpha \in B^c\}$.

DEFINITION. A nonempty family of random variables $\{X_\alpha: \alpha \in I\}$ is said to be finitely dependent if for every nonempty finite subset A of I there exists another finite subset $B = B(A)$ (including A) such that $\{X_\alpha: \alpha \in A\}$ is independent of $\{X_\alpha: \alpha \in B^c\}$ and such that $\sup_A \inf_B |B|/|A| < \infty$, where $|\cdot|$ denotes the order of a set. The order of dependence of the family is defined to be the smallest integer not less than $\sup_A \inf_B |B|/|A|$. Let C and D be nonempty finite subsets of I. The set $\{X_\alpha: \alpha \in D\}$ is said to be a finitely dependent set of $\{X_\alpha: \alpha \in C\}$ if the latter is independent of $\{X_\alpha: \alpha \in D^c\}$ and $|D|/|C|$ does not exceed the order of dependence.

We observe that m-dependence is a special case of finite dependence which in turn is a special case of local dependence. We refer the reader to Chen (1978) for examples of finitely dependent random variables.

For each $n \ge 1$, let $\{X_\alpha^{(n)}: \alpha \in I\}$ be a locally dependent family of nonnegative random variables. For each α, let $\{X_\beta^{(n)}: \beta \in A_\alpha^{(n)}\}$ be a locally dependent set of $\{X_\alpha^{(n)}\}$ and $\{X_\beta^{(n)}: \beta \in B_\alpha^{(n)}\}$ a locally dependent set of $\{X_\beta^{(n)}: \beta \in A_\alpha^{(n)}\}$. Also let

$$Y_\alpha^{(n)} = \sum_{\beta \in A_\alpha^{(n)}} X_\beta^{(n)}, \qquad \lambda^{(n)} = \sum_{\alpha \in I} EX_\alpha^{(n)}\left(Y_\alpha^{(n)}\right)^{-1}.$$

Here we adopt the convention that $0/0 = 0$ and we assume that $\lambda^{(n)} \in (0, \infty)$. Define the probability measure $\mu^{(n)}$ on the Borel subsets of $(0, \infty)$ by

$$\mu^{(n)}(E) = \left(\lambda^{(n)}\right)^{-1} \sum_{\alpha \in I} EX_\alpha^{(n)}\left(Y_\alpha^{(n)}\right)^{-1} I_{\{Y_\alpha^{(n)} \in E\}},$$

for every Borel subset E of $(0,\infty)$. Furthermore,

$$p_\alpha^{(n)} = P\big(X_\alpha^{(n)} > 0\big), \qquad \xi_\alpha^{(n)} = P\Bigg(\sum_{\beta\in B_\alpha^{(n)}} X_\beta^{(n)} > 0\Bigg),$$
$$\tau^{(n)} = \sum_{\alpha\in I} p_\alpha^{(n)}, \qquad \tilde p^{(n)} = \max_{\alpha\in I} p_\alpha^{(n)}. \tag{19}$$

THEOREM 7. *With the above notation,*

$$d\Bigg(\mathscr{L}\Bigg(\sum_{\alpha\in I} X_\alpha^{(n)}\Bigg), POIS\big(\lambda^{(n)}\mu^{(n)}\big)\Bigg) \le 2\exp(\lambda^{(n)}) \sum_{\alpha\in I} p_\alpha^{(n)}\xi_\alpha^{(n)}$$
$$\le 2\exp(\lambda^{(n)}) \sum_{\alpha\in I}\sum_{\beta\in B_\alpha^{(n)}} p_\alpha^{(n)} p_\beta^{(n)}.$$

Furthermore, if for every $n \ge 1$, $\{X_\alpha^{(n)}: \alpha\in I\}$ is a finitely dependent family with order of dependence r, we have the following complementary limiting result: If $d(\mathscr{L}(\sum_{\alpha\in I} X_\alpha^{(n)}), POIS(\lambda^{(n)}\mu^{(n)})) \to 0$, $\tilde p^{(n)} \to 0$ and $\lambda^{(n)}$ remains bounded as $n\to\infty$, then $\sum_{\alpha\in I} p_\alpha^{(n)}\xi_{(\alpha)}^{(n)} \to 0$ as $n\to\infty$, where for $\alpha\in I$, $n\ge 1$, $\{X_\beta^{(n)}: \beta\in B_\alpha^{(n)}\}$ is a finitely dependent set of $\{X_\beta^{(n)}: \beta\in A_\alpha^{(n)}\}$.

PROOF. For simplicity we drop the superscript (n) but will pick up the superscript when the need arises. Let $\{X'_\alpha: \alpha\in I\}$ be an independent copy of $\{X_\alpha: \alpha\in I\}$ and let

$$W = \sum_{\alpha\in I} X_\alpha, \qquad V_\alpha = \sum_{\beta\in A_\alpha^c} X_\beta,$$
$$\tilde V_\alpha = \sum_{\beta\in B_\alpha^c} X_\beta, \qquad Z_\alpha = \sum_{\beta\in B_\alpha} X_\beta,$$
$$T_\alpha = \sum_{j\in B_\alpha - A_\alpha} X_\beta, \qquad Y'_\alpha = \sum_{\beta\in A_\alpha} X'_\beta.$$

Let $E \subseteq [0,\infty)$, $h(w) = I_E(w)$ and f be a solution of the equation

$$wf(w) - \lambda\int tf(w+t)\,d\mu(t) = h(w) - \int h\,dS_{\lambda,\mu}.$$

We observe that

$$\begin{aligned} EWf(W) &= \sum_{\alpha\in I} EX_\alpha f(V_\alpha + Y_\alpha) \\ &= \sum_{\alpha\in I} EI_{\{X_\alpha>0\}}\big[X_\alpha f(V_\alpha+Y_\alpha) - X_\alpha f\big(\tilde V_\alpha + Y_\alpha\big)\big] \\ &\quad + \sum_{\alpha\in I} EI_{\{X'_\alpha>0\}}\big[X'_\alpha f\big(\tilde V_\alpha + Y'_\alpha\big) - X'_\alpha f(W+Y'_\alpha)\big] \\ &\quad + \sum_{\alpha\in I} EX'_\alpha f(W+Y'_\alpha) \\ &= R_1 + R_2 + \lambda E\int tf(W+t)\,d\mu(t), \end{aligned}$$

where R_i denotes the ith sum on the right-hand side of the second equality. Now

$$R_1 = \sum_{\alpha \in I} EI_{\{X_\alpha > 0\}} I_{\{T_\alpha > 0\}} \left[X_\alpha f(V_\alpha + Y_\alpha) - X_\alpha f(\tilde{V}_\alpha + Y_\alpha) \right]$$

and so it follows from Theorem 2 that

$$\begin{aligned} |R_1| &\le \exp(\lambda) \sum_{\alpha \in I} EI_{\{X_\alpha > 0\}} I_{\{T_\alpha > 0\}} \\ &\le \exp(\lambda) \sum_{\alpha \in I} p_\alpha \xi_\alpha \\ &\le \exp(\lambda) \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta. \end{aligned}$$

Similarly,

$$R_2 = \sum_{\alpha \in I} EI_{\{X'_\alpha > 0\}} I_{\{Z_\alpha > 0\}} \left[X'_\alpha f(\tilde{V}_\alpha + Y'_\alpha) - X'_\alpha f(W + Y'_\alpha) \right]$$

and by the same argument as above,

$$|R_2| \le \exp(\lambda) \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta.$$

Hence the first part of the theorem is easily proved using Theorem 1.

For the second part of the theorem, we assume that for each $n \ge 1$, $\{X_\alpha^{(n)}: \alpha \in I\}$ is a finitely dependent family with order of dependence r. It is easy to see that $\sup_{\alpha \in I} |B_\alpha^{(n)}| \le \sup_{\alpha \in I} r|A_\alpha^{(n)}| \le r^2$. Hence it suffices to show that $\tau^{(n)}$ [defined as in (19)] remains bounded as $n \to \infty$ since $\xi_\alpha^{(n)} \le r^2 \tilde{p}^{(n)}$. It follows from Lemma 12 below that $\{X_\alpha^{(n)}: \alpha \in I\}$ can be partitioned into r subsets of independent random variables with index sets $I_1^{(n)}, \dots, I_r^{(n)}$. Let $\tau_i^{(n)} = \sum_{\alpha \in I_i^{(n)}} p_\alpha^{(n)}$. Suppose $\{\tau^{(n)}\}$ is unbounded. Then there exists a subsequence $\{n'\}$ of $\{n\}$ and a sequence $\{k_n\}$ of numbers from $\{1, \dots, r\}$ such that $\tau_{k_n}^{(n')} \to \infty$ as $n \to \infty$. So

$$P\left(\sum_{\alpha \in I} X_\alpha^{(n')} = 0 \right) \le P\left(\sum_{\alpha \in I_{k_n}^{(n')}} X_\alpha^{(n')} = 0 \right) \le \exp\left(-\tau_{k_n}^{(n')} \right),$$

which tends to 0 as $n \to \infty$. This is a contradiction and the proof of the theorem is complete. □

LEMMA 12. *Let $\mathscr{F} = \{X_\alpha: \alpha \in I\}$ be a finitely dependent family of random variables with order of dependence r. Then $\mathscr{F}$ can be partitioned into r subfamilies of independent random variables.*

PROOF. By Zorn's lemma, every random variable in a finitely dependent family generates a maximal subfamily of independent random variables, that is, every random variable in the family is contained in a maximal subfamily of independent random variables. We note that every subfamily of a finitely dependent family is itself finitely dependent. So we partition $\mathscr{F}$ as follows.

Start with a particular $X_1 \in \mathscr{F}$ and let $\mathscr{M}_i \subset \mathscr{F} - \bigcup_{k=1}^{i-1} \mathscr{M}_k$ be a maximal subfamily of independent random variables generated by $X_i \in \mathscr{F} - \bigcup_{k=1}^{i-1} \mathscr{M}_k$, $i \geq 1$. This process cannot continue beyond $i = r$. For if it did, then by virtue of the maximality of each $\mathscr{M}_i$, every locally dependent set of every X_{r+1} belonging to $\mathscr{M}_{r+1}$ would have a nonempty intersection with each $\mathscr{M}_i$, $i = 1, \dots, r$. This contradicts the assumption that the order of dependence is r. Hence the lemma. □

In the case of a sum of locally dependent indicators, the bound in Theorem 7 can be improved. Let $\{X_\alpha : \alpha \in I\}$ be a family of locally dependent Bernoulli random variables with $p_\alpha = P(X_\alpha = 1) = 1 - P(X_\alpha = 0) > 0$. For each $\alpha \in I$, let A_α, B_α be a locally dependent set of $\{X_\alpha\}$, A_α respectively. We define for $i \geq 1$,

$$W = \sum_{\alpha \in I} X_\alpha, \qquad \lambda = EW,$$

$$Y_\alpha = \sum_{\beta \in A_\alpha} X_\beta, \qquad \lambda_i = (1/i) \sum_{\alpha \in I} EX_\alpha I_{\{Y_\alpha = i\}}.$$

We assume that $\lambda \in (0, \infty)$.

THEOREM 8. *With the above notation,*

$$d\left(\mathscr{L}(W), POIS\left(\sum_{i=1}^{\infty} \lambda_i \delta_i\right)\right) \leq 2(1 \wedge \lambda_1^{-1}) \exp\left(\sum_{j=1}^{\infty} \lambda_j\right) \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta. \tag{20}$$

However, if we have the additional condition that $j\lambda_j \searrow 0$ as $j \to \infty$, then the bound can be improved to

$$d\left(\mathscr{L}(W), POIS\left(\sum_{i=1}^{\infty} \lambda_i \delta_i\right)\right) \leq 2\left\{1 \wedge \frac{1}{\lambda_1 - 2\lambda_2}\left[\frac{1}{4(\lambda_1 - 2\lambda_2)} + \log^+ 2(\lambda_1 - 2\lambda_2)\right]\right\} \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta.$$

PROOF. The proof is similar to that of Theorem 7 with the exception that the first (second) part of the theorem uses Theorem 4 (Theorem 5) to bound $\sup_A \sup_{w \geq 1} |f(w+1) - f(w)|$ respectively. □

Arratia, Goldstein and Gordon (1989) have shown that

$$d(\mathscr{L}(W), POIS(\lambda\delta_1)) \le \lambda^{-1}(1 - e^{-\lambda})\Big(\mathrm{Var}(W) - \lambda + 2\sum_{\alpha\in I}\sum_{\beta\in A_\alpha} p_\alpha p_\beta\Big).$$

This, together with Theorem 8, implies that in the case that

$$\lambda^{-1}(1 - e^{-\lambda})[\mathrm{Var}(W) - \lambda]$$

is large, for which the Poisson approximation fails, we still have an approximation—the compound Poisson approximation—provided the right-hand side of (20) is small.

For an illustration of a concrete application of Theorem 8, we refer the reader to Chen (1990), who considered a problem involving head runs. In particular, an error bound is obtained for the compound Poisson approximation of the distribution of the random variable which counts the number of locations among the first n tosses of a coin at which a head run of length at least t begins. The asymptotic distribution of the length of the longest run of heads beginning in the first n tosses of a coin is also considered there.

REMARK. Though it will not be covered in this paper, we wish to remark that Chen (1976) has also used Stein's method to obtain a number of limit theorems and asymptotic expansions involving the compound Poisson approximation of the distribution of a sum of finitely dependent random variables.

4.3. *Equiprobable allocations.* Let there be ν balls and k urns. The balls are placed independently and randomly randomly (uniformly) among the k urns. Let N_i denote the number of urns containing exactly i balls. We are interested in the random variables $U = 2N_0 + N_1$ and $V = N_1 + 2N_2$. Here U can be interpreted as the minimum number of additional balls needed to ensure that each urn has at least 2 balls and V the total number of balls contained in urns with exactly 1 or 2 balls. It is easily seen that for $i = 0, 1, \dots, \nu$,

$$E(N_i) = kC_i^\nu\left(\frac{1}{k}\right)^i\left(1 - \frac{1}{k}\right)^{\nu - i} = \lambda_i,$$

where C_i^ν denotes the number of ways of choosing i objects from ν objects. In this example, we shall approximate the distributions of U, V by POIS($\lambda_1\delta_1 + \lambda_0\delta_2$), POIS($\Sigma_{i=1}^2\lambda_i\delta_i$) respectively. For convenience, we write

$$Y_{ij} = \begin{cases} 1, & \text{if the } j\text{th urn contains exactly } i \text{ balls},\\ 0, & \text{otherwise}. \end{cases}$$

Thus $N_i = \Sigma_{j=1}^k Y_{ij}$. Let $A \subset \{0, 1, 2, \dots\}$ and $g(v) = I_A(v) - P(W \in A)$, where

$W \sim \mathrm{POIS}(\sum_{i=1}^{2}\lambda_i\delta_i)$. By Theorem 3, there exists a bounded function f satisfying

$$\begin{aligned} Eg(V) &= E\left[Vf(V) - \sum_{i=1}^{2} i\lambda_i f(V+i)\right] \\ &= E\sum_{i=1}^{2}\left\{\sum_{j=1}^{k} iP(Y_{ij}=1)E[f(V)|Y_{ij}=1] - i\lambda_i f(V+i)\right\} \\ &= E\sum_{i=1}^{2} i\lambda_i\{E[f(V)|Y_{i1}=1] - f(V+i)\} \\ &= E\sum_{i=1}^{2} i\lambda_i[f(V_i) - f(V+i)], \end{aligned}$$

where V_i has the conditional distribution of V given that $Y_{i1} = 1$. Hence

$$|Eg(V)| \le \sum_{i=1}^{2} i\lambda_i E\left|V + i - V_i\right| \sup_{w\ge 1}\left|f(w+1) - f(w)\right|. \tag{21}$$

To obtain a reasonable bound for $E|V + i - V_i|$, we shall couple V and V_i on the same probability space as follows. Distribute ν balls at random (uniformly) among k urns. This determines V. Let Z_j denote the number of balls contained in the jth urn. If $Z_1 > i$, distribute $Z_1 - i$ balls from urn 1 uniformly among the remaining urns. If $Z_1 = i$, do nothing. If $Z_1 < i$, select $i - Z_1$ balls uniformly among the balls in the remaining urns and put them in urn 1. This determines V_i.

LEMMA 13. *With respect to the above probability space, for* $\nu \ge 1$, $k \ge 2$,

$$E|V + 1 - V_1| \le \left(1 - \frac{1}{k}\right)^{\nu-3}\left[\left(\frac{\nu}{k}\right)^3 + 4\left(\frac{\nu}{k}\right)^2 + 2\left(\frac{\nu}{k}\right) + 3\right].$$

PROOF. We observe that

$$E|V + 1 - V_1| = E(V + 1 - V_1)_+ + E(V_1 - V - 1)_+ \tag{22}$$

and

$$\begin{aligned} E(V + 1 - V_1)_+ &= E[(V + 1 - V_1)_+|Z_1 = 0]P(Z_1 = 0) + P(Z_1 = 1) \\ &\quad + \sum_{j=2}^{\nu} E[(V + 1 - V_1)_+|Z_1 = j]P(Z_1 = j). \end{aligned}$$

Furthermore, it can be seen that

$$\begin{aligned} E[(V + 1 - V_1)_+|Z_1 = 0]P(Z_1 = 0) &= E\left[\frac{N_1 + 2N_2}{\nu}\middle|Z_1 = 0\right]P(Z_1 = 0) \\ &\le \left(1 - \frac{1}{k}\right)^{\nu} \end{aligned}$$

and

$$\sum_{j=2}^{\nu} E[(V+1-V_1)_+|Z_1=j]P(Z_1=j)$$
$$\leq P(Z_1=2)E\left[\frac{N_0+N_1}{k-1}+\frac{4(N_2-1)}{k-1}+\frac{2\Sigma_{i\geq 3}N_i}{k-1}\middle|Z_1=2\right]$$
$$+\sum_{j=3}^{\nu} P(Z_1=j)E\left[2(j-1)\frac{N_1+N_2}{k-1}\middle|Z_1=j\right]$$
$$\leq 3\left(\frac{\nu}{k}\right)^2\left(1-\frac{1}{k}\right)^{\nu-2}+\left(\frac{\nu}{k}\right)^3\left(1-\frac{1}{k}\right)^{\nu-3}.$$

Hence we conclude that

$$E(V+1-V_1)_+\leq\left(1-\frac{1}{k}\right)^{\nu}+\left(\frac{\nu}{k}\right)\left(1-\frac{1}{k}\right)^{\nu-1}+3\left(\frac{\nu}{k}\right)^2\left(1-\frac{1}{k}\right)^{\nu-2}+\left(\frac{\nu}{k}\right)^3\left(1-\frac{1}{k}\right)^{\nu-3}. \tag{23}$$

Similarly it can be shown that

$$E(V_1-V-1)_+\leq 2\left(1-\frac{1}{k}\right)^{\nu}+\left(\frac{\nu}{k}\right)\left(1-\frac{1}{k}\right)^{\nu-1}+\left(\frac{\nu}{k}\right)^2\left(1-\frac{1}{k}\right)^{\nu-2}. \tag{24}$$

The lemma follows directly from (22), (23) and (24). □

LEMMA 14. *With respect to the above probability space, for $\nu \geq 2$, $k \geq 2$,*

$$E|V+2-V_2|\leq\left(1-\frac{1}{k}\right)^{\nu-3}\left[\left(\frac{\nu}{k}\right)^3+4\left(\frac{\nu}{k}\right)^2+4\left(\frac{\nu}{k}\right)+6\right].$$

The proof of this lemma is similar to that of Lemma 13 and hence is omitted.

THEOREM 9. *With the above notation, we have for $\nu \geq 2$, $k \geq 2$,*

$$d\left(\mathscr{L}(V), POIS\left(\sum_{i=1}^{2}\lambda_i\delta_i\right)\right)\leq\left(1-\frac{1}{k}\right)^{\nu-3}\left[\left(\frac{\nu}{k}\right)^3+4\left(\frac{\nu}{k}\right)^2+4\left(\frac{\nu}{k}\right)+6\right]$$
$$\times(1\wedge\lambda_1^{-1})\exp\left(\sum_{j=1}^{2}\lambda_j\right)\sum_{i=1}^{2}i\lambda_i.$$

PROOF. The proof follows immediately from (21), Theorem 4 and Lemmas 13 and 14. □

REMARK. An immediate corollary is that $d(\mathscr{L}(V), \mathrm{POIS}(\Sigma_{i=1}^{2}\lambda_i\delta_i))$ tends to 0 whenever ν and k tend to ∞ in such a way that $\nu/k \to \infty$ and λ_2 remains bounded.

The essential difference in the treatment of U and V comes from the observation that $\lambda_0 \ll \lambda_1 \ll \lambda_2$ for ν/k large. Hence in the case of U, Theorem 5 can be used to bound $\sup_A \sup_{w \geq 1}|f(w+1) - f(w)|$ instead. The corresponding result for U is stated next. The proof is similar to that of Theorem 9 and is omitted.

THEOREM 10. *With the above notation, we have for* $\nu \geq 2(k-1)$, $k \geq 2$,

$$d(\mathscr{L}(U), POIS(\lambda_1\delta_1 + \lambda_0\delta_2))$$
$$\leq \left(1 - \frac{1}{k}\right)^{\nu-2}\left[\left(\frac{\nu}{k}\right)^2 + 2\left(\frac{\nu}{k}\right) + 2\right](2\lambda_0 + \lambda_1)$$
$$\times\left\{1 \wedge \frac{1}{\lambda_1 - 2\lambda_0}\left[\frac{1}{4(\lambda_1 - 2\lambda_0)} + \log^+ 2(\lambda_1 - 2\lambda_0)\right]\right\}.$$

REMARK. A corollary is that $d(\mathscr{L}(U), \mathrm{POIS}(\lambda_1\delta_1 + \lambda_0\delta_2))$ tends to 0 whenever ν and k tend to ∞ in such a way that $\nu/k \to \infty$ and λ_1 remains bounded.

Acknowledgments. We are indebted to Professor Charles M. Stein for suggesting to us Theorem 1 and its relevance to the compound Poisson approximation and to Professor Leonard Y. H. Yap for his comments on its proof. We would also like to thank Professor Peter Ney, an Associate Editor and a referee for their comments and suggestions.

REFERENCES

ALDOUS, D. J. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.

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 (with discussion). *Statist. Sci.* **5** 403–434.

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

BARBOUR, A. D. (1987). Asymptotic expansions in the Poisson limit theorem. *Ann. Probab.* **15** 748–766.

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. and EAGLESON, G. K. (1985). Multiple comparisons and sums of dissociated random variables. *Adv. in Appl. Probab.* **17** 147–162.

BARBOUR, A. D. and HALL, P. G. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.

BARBOUR, A. D. and HOLST, L. (1989). Some applications of the Stein–Chen method for proving Poisson convergence. *Adv. in Appl. Probab.* **21** 74–90.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1988). Poisson approximation with the Stein–Chen method and coupling. Report, Dept. Mathematics, Uppsala Univ.

BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379–386.

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

BROWN, T. C. (1983). Some Poisson approximations using compensators. *Ann. Probab.* **11** 726–744.

CHEN, L. H. Y. (1974). On the convergence of Poisson binomial to Poisson distributions. *Ann. Probab.* **2** 178–180.

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

CHEN, L. H. Y. (1975b). An approximation theorem for convolutions of probability measures. *Ann. Probab.* **3** 992–999.

CHEN, L. H. Y. (1976). Compound Poisson approximation for nonnegative random variables. Unpublished manuscript.

CHEN, L. H. Y. (1978). Two central limit problems for dependent random variables. *Z. Wahrsch. Verw. Gebiete* **43** 223–243.

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

CHEN, L. H. Y. (1990). Comment on "Poisson approximation and the Chen–Stein method" by R. Arratia, L. Goldstein and L. Gordon. *Statist. Sci.* **5** 429–432.

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.

LE CAM, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197.

MICHEL, R. (1988). Improved error bound for the compound Poisson approximation of a nearly homogeneous portfolio. *Astin Bull.* **17** 165–169.

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

SERFOZO, R. F. (1986). Compound Poisson approximations for sums of random variables. *Ann. Probab.* **14** 1391–1398.

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 354, Dept. Statistics, Stanford Univ.

WANG, Y. H. (1991). A compound Poisson convergence theorem. *Ann. Probab.* **19** 452–455.

A. D. BARBOUR
INSTITUT FÜR ANGEWANDTE MATHEMATIK
DER UNIVERSITÄT ZÜRICH
RÄMISTRASSE 74
CH-8001 ZÜRICH
SWITZERLAND

LOUIS H. Y. CHEN
DEPARTMENT OF MATHEMATICS
NATIONAL UNIVERSITY OF SINGAPORE
KENT RIDGE
SINGAPORE 0511
REPUBLIC OF SINGAPORE

WEI-LIEM LOH
DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907