The Annals of Probability
1986, Vol. 14, No. 4, 1391–1398

COMPOUND POISSON APPROXIMATIONS FOR SUMS OF RANDOM VARIABLES[1]

BY RICHARD F. SERFOZO

Georgia Institute of Technology

We show that a sum of dependent random variables is approximately compound Poisson when the variables are rarely nonzero and, given they are nonzero, their conditional distributions are nearly identical. We give several upper bounds on the total-variation distance between the distribution of such a sum and a compound Poisson distribution. Included is an example for Markovian occurrences of a rare event. Our bounds are consistent with those that are known for Poisson approximations for sums of uniformly small random variables.

1. Introduction. There have been a number of studies on Poisson approximations for sums of uniformly small random variables. Of paramount interest is the total-variation distance between a sum of random variables and a Poisson variable. The total-variation distance between two probability measures (distributions) F and G on some measurable space is defined by

$$d(F,G) = \sup_B |F(B) - G(B)|, \tag{1.1}$$

where the supremum is over all measurable sets ($2d(F,G)$ is the total variation of the signed measure $F - G$). The total-variation distance between random elements X and Y with the respective distributions F and G is $d(X,Y) = d(F,G)$.

Building on the works of Hodges and Le Cam (1960), Le Cam (1960), Franken (1964), and Freedman (1974), Serfling (1975) proved this result: If $X_1, \dots, X_n$ are nonnegative integer-valued random variables adapted to the increasing σ-fields $\{\mathscr{F}_i\}_{i=0}^n$, then

$$d\left(\sum_{i=1}^n X_i, N\right) \le \sum_{i=1}^n \left[E^2(p_i) + E|p_i - Ep_i| + P(X_i \ge 2)\right], \tag{1.2}$$

where $p_i = P(X_i = 1|\mathfrak{F}_{i-1})$ and N is Poisson with mean $\sum_{i=1}^n Ep_i$. Comparable bounds for other Poisson approximations appear in Barbour and Eagleson (1983), Brown (1983), Chen (1975), Kabanov et al. (1983), Kerstan (1964), Valkeila (1982), and their references. Such bounds are useful for proving limit theorems for random variables and for point processes as well.

In this paper, we present analogues of (1.2) for compound Poisson approximations for sums. We consider sums of random elements that take values in a measurable group S: The group operation, addition, is measurable. If X is a

Received December 1984.
[1]This research was supported in part by AFOSR grant 84-0367.
AMS 1980 *subject classifications*. Primary 60E15, 60F99; secondary 60J10.
Key words and phrases. Compound Poisson distribution, total variation distance, sums of dependent variables, rare Markovian events.

random element of S with the compound Poisson distribution $H(B) = \sum_{n=0}^{\infty} F^{n*}(B)\alpha^n e^{-\alpha}/n!$, then we say X is $\mathscr{CP}(\alpha, F)$. If X has the distribution $EH(\cdot)$, where α or F are random, then we say X is mixed $\mathscr{CP}(\alpha, F)$.

Here is our main result. Let $X_1, \dots, X_n$ be random elements of S adapted to the increasing σ-fields $\{\mathscr{F}_i\}_{i=0}^n$, and define

$$p_i = P(X_i \neq 0 | \mathscr{F}_{i-1}), \qquad F_i(B) = P(X_i \in B | \mathscr{F}_{i-1}, X_i \neq 0).$$

Let F be a distribution on S with $F(\{0\}) = 0$, and define, by (1.1), the random distance $d_i = d(F_i, F)$ (F_i is random but F is not).

THEOREM 1.1. *If Z is mixed $\mathscr{CP}(\sum_{i=1}^n p_i, F)$, then*

$$d\left(\sum_{i=1}^n X_i, Z\right) \leq E\left[\sum_{i=1}^n (d_i + p_i^2)\right]. \tag{1.3}$$

If Z is $\mathscr{CP}(\sum_{i=1}^n \alpha_i, F)$ where $\alpha_i = Ep_i$, then

$$d\left(\sum_{i=1}^n X_i, Z\right) \leq E\left[\sum_{i=1}^n \left(d_i + |p_i - \alpha_i| + \alpha_i^2\right)\right]. \tag{1.4}$$

If Z is $\mathscr{CP}(\alpha, F)$, then

$$d\left(\sum_{i=1}^n X_i, Z\right) \leq E\left[\sum_{i=1}^n (d_i + p_i^2) + \left|\sum_{i=1}^n p_i - \alpha\right|\right]. \tag{1.5}$$

This result says, roughly, that $\sum_{i=1}^n X_i$ is approximately compound Poisson when the X_i's are rarely nonzero (the p_i's are small), and given that the X_i's are nonzero, their conditional distributions $F_1, \dots, F_n$ are nearly identical. Note that (1.5) with $\alpha = \sum_{i=1}^n \alpha_i$ is different from (1.4); in some cases the bound in (1.4) is smaller than that in (1.5) but in other cases the reverse is true.

For the degenerate distribution F on R with unit mass at 1, Theorem 1.1 yields bounds for Poisson or mixed Poisson approximations for sums. In this case, (1.4) is the same as (1.2), and (1.5) is consistent with the inequalities of Brown (1983) and Kabanov et al. (1983), which were established by martingale techniques.

Brown (1983) also obtains compound Poisson approximations for certain discrete variables via Poisson approximations. This approach, however, does not apply to the general case. We prove our results by rather direct arguments based on judicious conditioning and the use of (1.1) as a random distance for random distributions. Our approach also brings to light the key role of the F_i's.

From its proof, one can easily see that Theorem 1.1 is also true when the number of variables n in the sum is a stopping time of $\{\mathscr{F}_i\}$. For instance, Theorem 1.1 applies to sums of the form $\sum_{i=1}^{N(t)} X_i$, where $N(t) = \sum_i I(\tau_i \leq t)$ and $\tau_1 < \tau_2 < \cdots$ are stopping times of the increasing σ-fields $\{\mathscr{F}(t)\}$ and $\mathscr{F}_i = \mathscr{F}(\tau_i)$, respectively. Theorem 1.1 also holds when F and α are random; the Z in (1.4) and (1.5) would then be mixed compound Poisson.

The rest of this paper is organized as follows. Section 2 gives some basics on the total-variation distance, Section 3 consists of the proof of Theorem 1.1, and Section 4 gives an example of Markovian occurrences of an event.

2. Basic inequalities for distances. Let X and Y be random elements of some measurable space. A well-known coupling inequality is

$$d(X,Y) \le P(X \ne Y). \tag{2.1}$$

The X, Y in the probability are the random elements—with an arbitrary dependency—defined on a common probability space. Inequality (2.1) follows because $P(X \in B) \le P(X \ne Y) + P(Y \in B)$.

It is natural for us to analyze $d(X,Y)$ in terms of conditional probabilities. Accordingly, we sometimes refer to X as having a distribution $EF(\cdot)$ where F is a random distribution. Typically, $F(B) = P(X \in B|\mathscr{F})$, or F could be defined as a measurable function of random elements.

LEMMA 2.1. *Suppose X and Y have the respective distributions $EF(\cdot)$ and $EG(\cdot)$, where F and G are random distributions. Then*

$$d(X,Y) \le E[d(F,G)]. \tag{2.2}$$

In case $F(B) = P(X \in B|\mathscr{F})$, and $G(B) = P(Y \in B|\mathscr{G})$, for some σ-fields $\mathscr{F}$ and $\mathscr{G}$, then

$$d(X,Y) \le E[d(F,G)] \le E[P(X \ne Y|\mathscr{F},\mathscr{G})]. \tag{2.3}$$

PROOF. Expression (2.2) follows since

$$d(X,Y) = \sup_B |EF(B) - EG(B)| \le \sup_B E|F(B) - G(B)| = E[d(F,G)].$$

Expression (2.3) follows from (2.2) and a random version of (2.1). □

REMARK. Keep in mind that F, G in the expectation in (2.2) are the random distributions on a common probability space and their dependency is arbitrary. A similar comment applies to the $X, Y, \mathscr{F}, \mathscr{G}$ in the probability in (2.3).

Distances involving functions of random elements, such as sums or maxima, can generally be represented as $D = d(h(X), h(Y))$, where $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_n)$, and h is a measurable function from the range space of X and Y to some other measurable space. Here are some bounds on this distance.

LEMMA 2.2. (i) $D \le d(X,Y)$.

(ii) $D \le \sum_{i=1}^n P(X_i \ne Y_i)$.

(iii) *If $X_1, \dots, X_n$ are independent and $Y_1, \dots, Y_n$ are independent, then $D \le \sum_{i=1}^n d(X_i, Y_i)$.*

(iv) *If $X_1, \dots, X_n$ are adapted to the increasing σ-fields $\{\mathscr{F}_i\}_{i=0}^n$ and $Y_1, \dots, Y_n$ are adapted to the increasing σ-fields $\{\mathscr{G}_i\}_{i=0}^n$, and $F_i(B) = P(X_i \in B|\mathscr{F}_{i-1})$, $G_i(B) = P(Y_i \in B|\mathscr{G}_{i-1})$, then*

$$D \le E\left[\sum_{i=1}^n d(F_i, G_i)\right] \le E\left[\sum_{i=1}^n P(X_i \ne Y_i|\mathscr{F}_{i-1}, \mathscr{G}_{i-1})\right]. \tag{2.4}$$

PROOF. Statement (i) is true since

$$D = \sup_B |P(h(X) \in B) - P(h(Y) \in B)|$$
$$= \sup_B |P(X \in h^{-1}(B)) - P(Y \in h^{-1}(B))| \le d(X, Y).$$

Statement (ii) is true since by (i) and (2.1) we have

$$D \le P(X \ne Y) = P\left(\bigcup_{i=1}^{n} \{X_i \ne Y_i\}\right) \le \sum_{i=1}^{n} P(X_i \ne Y_i).$$

Now consider (iii) when $n = 2$. From (i), the triangle inequality for d, and the independence, we have

$$D \le d((X_1, X_2), (Y_1, Y_2)) \le d((X_1, X_2), (Y_1, X_2)) + d((Y_1, X_2), (Y_1, Y_2))$$
$$\le d(X_1, Y_1) + d(X_2, Y_2).$$

Using this inequality and induction yields (iii) for general n.

Under the hypotheses of (iv), it follows by successive conditioning that $P(X \in B_1 \times \cdots \times B_n) = E[F_1(B_1) \cdots F_n(B_n)]$, and a similar statement holds for Y. Then using (i), (2.2), and (iii) we have

$$D \le d(X, Y) \le E[d(F_1 \cdots F_n, G_1 \cdots G_n)] \le E \sum_{i=1}^{n} d(F_i, G_i).$$

The second inequality in (2.4) follows from (2.3). □

The next two results deal with compound Poisson distributions.

LEMMA 2.3. *If X is $\mathscr{CP}(\alpha, F)$ and Y is $\mathscr{CP}(\beta, G)$, with $F(\{0\}) = 0$ and $G(\{0\}) = 0$, then $d(X, Y) \le |\alpha - \beta| + (\alpha \wedge \beta) d(F, G)$.*

PROOF. First consider the case in which $\alpha \le \beta$. Clearly Y is equal in distribution to $Y_1 + Y_2$, where Y_1, Y_2 are independent $\mathscr{CP}(\beta - \alpha, G)$ and $\mathscr{CP}(\alpha, G)$, respectively. Note that the distributions of X and Y_2 can be written as $EF^{N^*}(\cdot)$ and $EG^{N^*}(\cdot)$, respectively, where N is a Poisson random variable with mean α. Then applying the triangle inequality, (2.2), (2.1), and Lemma 2.2(iii) in the form $d(F^{n^*}, G^{n^*}) \le nd(F, G)$ we have

$$d(X, Y) \le d(X, Y_2) + d(Y_2, Y_1 + Y_2) \le Ed(F^{N^*}, G^{N^*}) + P(Y_1 \ne 0)$$
$$\le ENd(F, G) + 1 - e^{-(\beta - \alpha)} \le \alpha d(F, G) + \beta - \alpha.$$

This proves the assertion when $\alpha \le \beta$, and a similar proof applies when $\alpha > \beta$. □

LEMMA 2.4. *Suppose X is a random element of S and let*

$$p = P(X \ne 0) \quad \text{and} \quad F(B) = P(X \in B | X \ne 0). \tag{2.5}$$

If Z is $\mathscr{CP}(p, F)$, then $d(X, Z) \le p^2$.

PROOF. It suffices, by (2.1), to construct X, Z on a common probability space such that $P(X \neq Z) \leq p^2$. To this end, let N, U and $Y_1, \dots, Y_n$ be independent random elements on a common probability space such that N is a Poisson random variable with mean p, each Y_i has the distribution F, and $P(U = 0) = (1 - p)e^p = 1 - P(U = 1)$. Define

$$X = Y_1(1 - I(U = 0, N = 0)) \quad \text{and} \quad Z = \sum_{i=1}^{N} Y_i.$$

An easy check shows that X satisfies (2.5), and Z is clearly $\mathscr{CP}(p, F)$. Furthermore,

$$\begin{aligned} P(X \neq Z) &= P(X \neq Z, N = 0) + P(X \neq Z, N \geq 2) \\ &= P(U = 1)P(N = 0) + P(N \geq 2) = p(1 - e^{-p}) \leq p^2. \end{aligned}$$

This completes the proof. □

We end this section by comparing two random elements that have certain conditional distributions that are equal.

LEMMA 2.5. *Let X and Y be random elements. If there is a measurable set A such that $P(X \in B|X \in A) = P(Y \in B|Y \in A)$ for each measurable B, then*

$$d(X, Y) \leq |P(X \in A) - P(Y \in A)|. \tag{2.6}$$

PROOF. Let U, V, and W be independent random elements on a probability space. Assume that U is uniform on $(0, 1)$ and that V and W take values in A and A^c, respectively, and their distributions are $P(V \in B) = P(X \in B|X \in A)$ and $P(W \in B) = P(X \in B|X \in A^c)$. Let p and q denote the respective probabilities in (2.6), and define $X = VI(U \leq p) + WI(U > p)$ and $Y = VI(U \leq q) + WI(U > q)$. Clearly X and Y satisfy the hypotheses and, moreover, $P(X \neq Y) = P(p \wedge q < U \leq p \vee q) = |p - q|$. Thus the assertion follows by applying (2.1). □

3. Proof of Theorem 1.1. In addition to the notation of Theorem 1.1, we let $G_p(\cdot) = pF(\cdot) + (1 - p)\,\delta_0(\cdot)$, where δ_0 is the Dirac measure with unit mass at 0, and we let Y be a random element with distribution $E(G_{p_1} * \cdots * G_{p_n}(\cdot))$ (recall that p_i is random).

To prove (1.3), consider the inequality

$$d\left(\sum_{i=1}^{n} X_i, Z\right) \leq d\left(\sum_{i=1}^{n} X_i, Y\right) + d(Y, Z). \tag{3.1}$$

By the use of successive conditioning, it is clear that

$$P\left(\sum_{i=1}^{n} X_i \in B\right) = E[F_1' * \cdots * F_n'(B)], \quad \text{where } F_i'(B) = P(X_i \in B|\mathscr{F}_{i-1}).$$

Note that $F_i'(\cdot) = p_i F_i(\cdot) + (1 - p_i)\,\delta_0(\cdot)$, and so $d(F_i', G_{p_i}) = d(F_i, F) = d_i$.

Then applying (2.2) and Lemma 2.2(iii), we have

$$d\left(\sum_{i=1}^{n} X_i, Y\right) \le E\left[d\left(F_1' * \cdots * F_n', G_{p_1} * \cdots * G_{p_n}\right)\right] \le E\left(\sum_{i=1}^{n} d_i\right). \tag{3.2}$$

Similarly, using $P(Z \in B) = E[H_{p_1} * \cdots * H_{p_n}(B)]$, where the distribution H_p is $\mathscr{CP}(p, F)$, and applying Lemmas 2.1, 2.2(iii), and 2.4, we have

$$\begin{aligned} d(Y, Z) &\le E\left[d\left(G_{p_1} * \cdots * G_{p_n}, H_{p_1} * \cdots * H_{p_n}\right)\right] \\ &\le E\left[\sum_{i=1}^{n} d\left(G_{p_i}, H_{p_i}\right)\right] \le E\left(\sum_{i=1}^{n} p_i^2\right). \end{aligned} \tag{3.3}$$

Then combining (3.1)–(3.3) yields the assertion (1.3).

Now consider the assertion (1.4). Here Z is $\mathscr{CP}(\sum_{i=1}^{n}\alpha_i, F)$. Let $U_1, \dots, U_n$ be independent random elements with the respective distributions $G_{\alpha_1}, \dots, G_{\alpha_n}$. Then by applications of (3.2), Lemmas 2.1, 2.2(iii), and 2.5 (with $A = S \setminus \{0\}$), we have

$$\begin{aligned} d\left(\sum_{i=1}^{n} X_i, Z\right) &\le d\left(\sum_{i=1}^{n} X_i, Y\right) + d\left(Y, \sum_{i=1}^{n} U_i\right) + d\left(\sum_{i=1}^{n} U_i, Z\right) \\ &\le E\left(\sum_{i=1}^{n} d_i\right) + E\left[d\left(G_{p_1} * \cdots * G_{p_n}, G_{\alpha_1} * \cdots * G_{\alpha_n}\right)\right] \\ &\quad + d\left(G_{\alpha_1} * \cdots * G_{\alpha_n}, H_{\alpha_1} * \cdots * H_{\alpha_n}\right) \\ &\le E\left[\sum_{i=1}^{n} d_i + |p_i - \alpha_i| + \alpha_i^2\right]. \end{aligned}$$

Finally, to prove (1.5), consider the inequality

$$d\left(\sum_{i=1}^{n} X_i, Z\right) \le d\left(\sum_{i=1}^{n} X_i, Z'\right) + d(Z', Z), \tag{3.4}$$

where Z is $\mathscr{CP}(\alpha, F)$ and Z' is mixed $\mathscr{CP}(\sum_{i=1}^{n} p_i, F)$. By Lemmas 2.1 and 2.3 we have $d(Z', Z) \le E|\sum_{i=1}^{n} p_i - \alpha|$. Applying this and (1.3) to (3.4) yields (1.5). □

4. A compound Poisson approximation for Markovian occurrences of an event. Suppose that $Y_0, Y_1, \dots$ is a Markov chain with states 0 and 1 that represent the nonoccurrence and occurrence, respectively, of a certain event $\mathscr{E}$. Let $\varepsilon = P(Y_1 = 1 | Y_0 = 0)$ and $p = P(Y_1 = 1 | Y_0 = 1)$, and assume that ε and p are not zero or one. The stationary distribution of this Markov chain is

$$\pi(0) = (1-p)/(1-p+\varepsilon), \qquad \pi(1) = \varepsilon/(1-p+\varepsilon).$$

Consequently, when ε is small, then the event $\mathscr{E}$ is rare.

Consider the sum $N_n = \sum_{i=1}^{n} Y_i$, which is the number of occurrences of the event $\mathscr{E}$ in time n. We assume, for simplicity, that the Markov chain is stationary. Isham (1980) and Böker and Serfozo (1983) showed that if ε varies with n such that $\varepsilon \to 0$ and $n\varepsilon \to \alpha > 0$ as $n \to \infty$, then N_n converges in distribution to a random variable Z that is $\mathscr{CP}(\alpha, F)$ with $F(\{k\}) = p^{k-1}(1-p)$, $k \geq 1$. A bound on the rate of this convergence is given in the following result. Brown (1983) obtained a variation of this bound by another approach.

THEOREM 4.1.

$$d(N_n, Z) \leq |n\varepsilon - \alpha| + \varepsilon(1 + p + \varepsilon n(2-p))/(1 - p + \varepsilon). \tag{4.1}$$

PROOF. Define the random variables

$$X_i = \sum_{k=1}^{\infty} k(1 - Y_{i-1})Y_i \cdots Y_{i+k-1}(1 - Y_{i+k}), \qquad i = 1, \dots, n,$$

$$X_1' = \sum_{k=1}^{\infty} kY_1Y_2 \cdots Y_k(1 - Y_{k+1}).$$

When the Markov chain begins a sojourn in state 1 at time i (a success run of the event $\mathscr{E}$), then X_i records the length of that sojourn. Clearly

$$\begin{aligned} p_i &:= P(X_i \geq 1 | Y_0, \dots, Y_{i-1}) \\ &= \sum_{k=1}^{\infty} (1 - Y_{i-1})\varepsilon p^{k-1}(1-p) = \varepsilon(1 - Y_{i-1}), \\ F_i(k) &:= P(X_i \leq k | Y_0, \dots, Y_{i-1}, X_i \geq 1) = F(k). \end{aligned}$$

Let $T_n = \sum_{i=1}^n X_i$ and $T_n' = T_n + X_1'$, and consider

$$d(N_n, Z) \leq d(N_n, T_n') + d(T_n', T_n) + d(T_n, Z). \tag{4.2}$$

Clearly

$$d(N_n, T_n') \leq P(N_n \neq T_n') = P(Y_n = 1, Y_{n+1} = 1) = \pi(1)p, \tag{4.3}$$

$$d(T_n', T_n) \leq P(X_1' \neq 0) = P(Y_1 = 1) = \pi(1), \tag{4.4}$$

and by (1.5)

$$\begin{aligned} d(T_n, Z) &\leq \sum_{i=1}^n Ep_i^2 + E\left|\sum_{i=1}^n p_i - \alpha\right| \\ &= n\varepsilon^2\pi(0) + E\left|\varepsilon\sum_{i=1}^n (1 - Y_{i-1}) - \alpha\right| \\ &\leq n\varepsilon^2\pi(0) + \varepsilon n\pi(1) + |n\varepsilon - \alpha|. \end{aligned} \tag{4.5}$$

Combining (4.2)–(4.5) yields (4.1).

REMARK. Note that the preceding proof applies (1.5) to the auxiliary sum T_n instead of to the original sum N_n. One could also apply (1.4) to T_n, but this would yield (4.1) with $2 - p$ replaced by $(2-p)^2$, which is worse.

REFERENCES

[1] BARBOUR, A. D. and EAGLESON, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. Appl. Probab.* **15** 583–600.

[2] BÖKER, F. and SERFOZO, R. F. (1983). Ordered thinnings of point processes and random measures. *Stochastic Process. Appl.* **15** 113–132.

[3] BROWN, T. C. (1983). Some Poisson approximations using compensators. *Ann. Probab.* **11** 726–744.
[4] CHEN, L. H. Y. (1975). An approximation theorem for convolutions of probability measures. *Ann. Probab.* **3** 992–999.
[5] FRANKEN, P. (1964). Approximation der Verteilungen von Summen unabhängiger nichtnegativer ganzzahler Zufallsgrössen durch Poissonsche Verteilungen. *Math. Nachr.* **27** 303–340.
[6] FREEDMAN, D. (1974). The Poisson approximation for dependent events. *Ann. Probab.* **2** 256–269.
[7] HODGES, J. L. and LE CAM, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.* **31** 737–740.
[8] ISHAM, V. (1980). Dependent thinning of point processes. *J. Appl. Probab.* **17** 987–995.
[9] KABANOV, Y. M., LIPTSER, R. S. and A. N. SHIRYAYEV (1983). Weak and strong convergence of the distributions of counting processes. *Theor. Probab. Appl.* **28** 303–336.
[10] KERSTAN, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam. *Z. Wahrsch. verw. Gebiete* **2** 173–179.
[11] LE CAM, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10** 1181–1197.
[12] SERFLING, R. J. (1975). A general Poisson approximation theorem. *Ann. Probab.* **3** 726–731.
[13] VALKEILA, E. (1982). A general Poisson approximation theorem. *Stochastics* **7** 159–171.

SCHOOL OF INDUSTRIAL
AND SYSTEMS ENGINEERING
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332