SIAM REVIEW
Vol. 20, No. 3, July 1978

0036-1445/78/2003-0567 $01.00/0

SOME ELEMENTARY RESULTS ON POISSON APPROXIMATION IN A SEQUENCE OF BERNOULLI TRIALS*

R. J. SERFLING†

Abstract. In a finite series of independent success-failure trials, the total number of successes has a binomial probability distribution. It is a classical result that this probability distribution is subject to approximation by a Poisson distribution if the number of trials is sufficiently large and the probability of success in a trial is sufficiently small. Recent interest has focused upon the cases of trials which are possibly dependent, or which have possibly differing success probabilities, or which have more than two possible outcomes. Further, in connection with appropriate metrics for measuring the disparity between two probability distributions, the problem of bounds on the error of approximation has received attention. The present article provides a unified treatment of some of the more elementary of these recent developments. The aim is to offer a springboard for heightened utilization of the Poisson approximation in probability analysis, both theoretical and applied, a source for up-to-date teaching on the subject, and an introduction to a field of research that continues to be interesting and fruitful. The paper also briefly discusses some recent varieties of application of the Poisson approximation, in connection with reliability theory, stochastic processes, urn models, covering problems, and chain letter schemes.

1. Introduction. In a series of n independent trials, each having outcome "success" or "failure," with "success" having occurrence probability p in any given trial, the total number of successes has the binomial (n, p) probability distribution. This is a well-known basic fact of probability theory. Further, as the number of trials increases with p fixed, central limit theory provides approximations to the binomial probabilities based on the normal probability distribution. This approximation was found by de Moivre [25] among others and was refined and popularized by Laplace [22]. If, on the other hand, p is allowed to decrease with increasing n in such fashion that np has a positive finite limit, an approximation based on a Poisson probability distribution becomes valid. Obviously, such an approximation pertains to the case of "success" being "rare". After discovery and publication by Poisson [27], this form of approximation remained overshadowed by the normal approximation until Bortkiewicz [1] recognized and popularized its unique and broad domain of application. Since then a great variety of applications have been found and the elementary proof of convergence of binomial to Poisson probabilities has become a standard item in courses on probability.

That is not the end of the story. In recent decades, attention has turned to the question of Poisson approximation in connection with trials which are *possibly dependent*, or which have *possibly differing success probabilities*, or which have *more than two possible outcomes*. Also, the problem of *bounds on the error of approximation* has been attacked. The present article treats some of the more elementary developments in these directions. The results offer a springboard for heightened utilization of the Poisson approximation in probability analysis, both theoretical and applied, a source for up-to-date teaching on the subject, and an introduction to a field of research that continues to be interesting and fruitful.

In § 2 we introduce a basic framework and consider two distance functions apropos to measuring disparity between two given probability distributions. Key properties of these metrics are examined in § 3.

* Received by the editors October 14, 1976, and in revised form February 23, 1977.

† Department of Statistics and Statistical Consulting Center, Florida State University, Tallahassee, Florida 32306. This research was supported by the U.S. Army, Navy and Air Force under Office of Naval Research Contract N00014-76-C-0608.

Relative to each metric, we present in §§ 4 and 5, respectively, bounds on the error of Poisson approximation to the distribution of the number of successes in a series of independent trials with possibly differing success probabilities. The results are applicable for any choice of the number n of trials and yield as a corollary the classical limit theorem.

Extension to the case of dependent trials is provided in § 6, and to the case of more than two possible outcomes in § 7. A general theorem encompassing these various extensions is presented in § 8.

Section 9 provides a brief discussion of applications of the Poisson approximation, with emphasis upon recent uses in stochastic processes, reliability, urn models and other areas.

Finally, a few complementary remarks are made in § 10.

For general background on probability concepts such as distribution, expectation, independence, etc., used without definitions in the present paper, appropriate sources are Parzen [26], Gnedenko [16], Chung [7] or Loève [24].

2. Basic framework and relevant distance functions. Assume that all random variables under discussion are defined on a common probability space $(\Omega, \mathcal{A}, P)$. As usual, Ω denotes a set of elementary outcomes, $\mathcal{A}$ a σ-algebra of subsets of Ω, and P a probability measure on $\mathcal{A}$.

We shall take special interest in *Bernoulli* variables, i.e., random variables taking values only 0 or 1. For a sequence $X_1, \cdots, X_n$ of such variables, we shall focus attention on the random variable $S_n = \sum_{i=1}^n X_i$, which represents the number of X_i's taking the value 1 (the number of "successes"). The probability distribution of S_n can be quite cumbersome, especially if the success probabilities differ, but for the case of sufficiently small success probabilities it is subject to approximation by a suitably chosen Poisson distribution. For real $\lambda > 0$, let $Y(\lambda)$ denote a random variable having the *Poisson* distribution with mean parameter λ. Thus

$$P[Y(\lambda) \leqq m] = \sum_{j=0}^{m} \frac{e^{-\lambda}\lambda^j}{j!} \qquad m = 0, 1, \cdots.$$

For the purpose of characterizing the efficacy of such an approximation, we consider two choices of *distance function* for measuring disparity between two probability distributions on the nonnegative integers $N_+ = \{0, 1, 2, \cdots\}$. Given two such distributions and letting X and Y denote random variables having the given distributions, we define

$$d(X, Y) = \sup_{A \subset N_+} |P(X \in A) - P(Y \in A)| \tag{2.1}$$

and

$$d_0(X, Y) = \sup_{m \geqq 0} |P(X \leqq m) - P(Y \leqq m)|. \tag{2.2}$$

Clearly, $d_0 \leqq d$. Although d and d_0 depend upon X and Y only through their probability distributions, the notation $d(X, Y)$ and $d_0(X, Y)$ is adopted for convenience and flexibility in certain manipulations. The functions d and d_0 are, in fact, metrics in the space of all probability distributions on N_+.

The metric d is relevant to computations involving *probabilities of arbitrary events*. For example, consider a communication system of n relays, each relay receiving an input signal 0 or 1 and transmitting an output signal 0 or 1. Let X_i equal 1 if the ith relay malfunctions, altering the received signal, and 0 otherwise. Then $S_n = \sum_1^n X_i$

represents the total number of reverses encountered by an initial signal passing through the system. Of special interest is the event that a transmitted signal is received correctly at the final receiver, i.e., the event $\{S_n \in A\}$, where $A=\{0, 2, 4, \cdots, 2[\frac{1}{2}n]\}$ and $[\cdot]$ denotes greatest integer part.

The metric d_0 is relevant to computations involving *cumulative distribution functions*. For example, consider a monitoring system of n sensors, and suppose that the system is of k*-out-of-*n type. That is, an alarm becomes sounded if and only if at least k of the n sensors are simultaneously in the activated state. Let X_i equal 1 if the ith sensor is activated and 0 otherwise, and again put $S_n=\sum_1^n X_i$. Then the event that an alarm becomes sounded is given by $\{S_n \geqq k\}$, illustrating the relevance of the d_0 metric in regard to approximations to the alarm probability.

Since $d_0(X, Y) \leqq d(X, Y)$, bounds on d constitute bounds on d_0 also. However, in situations where d_0 is the more relevant metric, one would take advantage of tighter bounds available for d_0.

3. Properties of the metrics d and d_0. First we note two alternative expressions for d:

$$d(X, Y)=\frac{1}{2}\sum_{m=0}^{\infty}|P(X=m)-P(Y=m)| \tag{3.1}$$

and

$$d(X, Y)=\frac{1}{2}\sup_{|h|\leqq 1}|E[h(X)]-E[h(Y)]|, \tag{3.2}$$

where $E[\cdot]$ denotes expectation and the supremum is taken over all measurable functions h such that $|h| \leqq 1$. The equivalence of (2.1), (3.1) and (3.2) is easily verified. Whereas (3.1) exhibits $d(X, Y)$ as the well-known *variation norm*, (3.2) exhibits $2d(X, Y)$ as an *operator norm* on a space of signed measures (see Le Cam [23] for details).

The verification that d and d_0 are metrics is straightforward and entails, in particular, the *triangle inequalities*: for arbitrary N_+-valued random variables X, Y and Z,

$$d^*(X, Y) \leqq d^*(X, Z)+d^*(Y, Z), \tag{3.3}$$

where $d^*=d$ or d_0.

Simple upper bounds for d and d_0 are given by

$$d(X, Y) \leqq P(X \neq Y) \tag{3.4}$$

and

$$d_0(X, Y) \leqq \max\{P(X<Y), P(X>Y)\}. \tag{3.5}$$

Proof. For any $A \subset N_+$, we have

$$P(X \in A) \leqq P(Y \in A)+P(X \neq Y).$$

This leads to (3.4) and a similar argument leads to (3.5). $\square$

The following corollary of (3.4) has been noted by Freedman [13]:

$$d(Y(\mu_1), Y(\mu_2)) \leqq |\mu_1-\mu_2|. \tag{3.6}$$

Proof. Let $\mu_1 < \mu_2$ and take independent random variables $X = Y(\mu_1)$ and $Z = Y(\mu_2 - \mu_1)$. Then

$$d(Y(\mu_1), Y(\mu_2)) = d(X, X+Z) \leqq P(Z \neq 0)$$
$$= 1 - e^{\mu_1 - \mu_2} \leqq \mu_2 - \mu_1.$$

We have used here the basic inequality $\exp(t) \geqq 1 + t$ and the fact that the sum of two independent Poisson variates is itself a Poisson variate. □

The following *contractive* property holds. The distributions of X and Y are changed to closer ones if an independent random variable Z is added to both X and Y. That is, for X and Y arbitrary and Z independent of (X, Y),

$$d^*(X+Z, Y+Z) \leqq d^*(X, Y), \tag{3.7}$$

where $d^* = d$ or d_0. Here X, Y and Z are N_+-valued.

Proof. For any $A \subset N_+$, we have (denoting by $A/\{m\}$ the difference $A - \{m\}$)

$$\begin{aligned} P(X+Z \in A) &= \sum_{m=0}^{\infty} P(X \in A/\{m\}, Z = m) \\ &= \sum_{m=0}^{\infty} P(X \in A/\{m\}) P(Z = m) \\ &\leqq \sum_{m=0}^{\infty} P(Z = m)[P(Y \in A/\{m\}) + d(X, Y)] \\ &= P(Y+Z \in A) + d(X, Y). \end{aligned}$$

This leads to (3.7) for $d^* = d$. A similar argument yields the result for $d^* = d_0$. □

Next we establish important *subadditivity* properties for d and d_0. For $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ all *mutually independent* N_+-valued random variables,

$$d^*\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) \leqq \sum_{i=1}^n d^*(X_i, Y_i), \tag{3.8}$$

where $d^* = d$ or d_0.

Proof. By (3.3) and then (3.7), we have

$$\begin{aligned} d^*(X_1+X_2, Y_1+Y_2) &\leqq d^*(X_1+X_2, X_2+Y_1) + d^*(Y_1+Y_2, X_2+Y_1) \\ &\leqq d^*(X_1, Y_1) + d^*(X_2, Y_2). \end{aligned}$$

The result (3.8) then follows by induction. □

Property (3.8), via somewhat different proofs, has been given for d by Çinlar [3] and for d_0 by Daley [9].

Finally, we present weaker forms of (3.8) which pertain, more generally, to possibly dependent X_i's and Y_i's. For *arbitrary* N_+-valued $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$,

$$d\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) \leqq \sum_{i=1}^n P(X_i \neq Y_i) \tag{3.9}$$

and

$$d_0\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) \leqq \sum_{i=1}^n \max\{P(X_i < Y_i), P(X_i > Y_i)\}. \tag{3.10}$$

Proof. It is easily seen that

$$P\left(\sum_{i=1}^{n} X_i \neq \sum_{i=1}^{n} Y_i\right) \leqq \sum_{i=1}^{n} P(X_i \neq Y_i). \tag{3.11}$$

Also, using

$$P(X_1+X_2 < Y_1+Y_2) \leqq P(X_1 < Y_1) + P(X_2 < Y_2),$$

we find

$$\max\left\{P\left(\sum_{i=1}^{n} X_i < \sum_{i=1}^{n} Y_i\right), P\left(\sum_{i=1}^{n} X_i > \sum_{i=1}^{n} Y_i\right)\right\} \leqq \sum_{i=1}^{n} \max\{P(X_i < Y_i), P(X_i > Y_i)\}. \tag{3.12}$$

Applying (3.4) with (3.11) and (3.5) with (3.13), we obtain (3.9) and (3.10), respectively. □

4. Bounds on the metric d for Poisson approximation to a sum of independent Bernoulli variables. First consider approximation of a *single* Bernoulli variable X by a Poisson variate Y. If $E[X]=E[Y]$ is desired, take $Y=Y(p)$. If $P(X=0)=P(Y=0)$ is desired, take $Y=Y(\lambda)$, where $\lambda=-\ln(1-p)$. For these choices of Y, we have

LEMMA 4.1. *Let X be a Bernoulli variable with success probability p, $0<p<1$. Put $\lambda=-\ln(1-p)$. Then*

$$d(X, Y(p)) = p(1-e^{-p}) \leqq p^2 \tag{4.1}$$

and

$$d(X, Y(\lambda)) = 1-e^{-\lambda} - \lambda e^{-\lambda} \leqq \tfrac{1}{2}\lambda^2. \tag{4.2}$$

Proof. We utilize the identity (3.1). Take the case $Y=Y(p)$. Note that

$$|P(X=0)-P(Y=0)| = |1-p-e^{-p}| = e^{-p}-1+p,$$

$$|P(X=1)-P(Y=1)| = |p-p\,e^{-p}| = p-p\,e^{-p},$$

and, for $m \geqq 2$,

$$|P(X=m)-P(Y=m)| = P(Y=m).$$

Thus

$$d(X,Y) = \tfrac{1}{2}[e^{-p}-1+p+p-p\,e^{-p}+P(Y \geqq 2)],$$

i.e.,

$$d(X, Y(p)) = p(1-e^{-p}),$$

which yields (4.1). A similar development leads to

$$d(X, Y(\lambda)) = p-\lambda e^{-\lambda} = 1-e^{-\lambda}-\lambda e^{-\lambda},$$

which yields the inequality in (4.2) by routine calculation. □

Remark. For $p \leqq .6$, we have $\frac{1}{2}\lambda^2 \leqq \frac{1}{2}p^2+p^3+p^4$. Thus, for p small, the upper bound $\frac{1}{2}\lambda^2$ provided by (4.2) is considerably sharper than the bound p^2 provided by (4.1). However, the approximating Poisson variable in (4.2) entails a less convenient parameter than that in (4.1).

Making use of the subadditivity property (3.8) for d, in conjunction with the preceding lemma, we conclude

THEOREM 4.1. *Let $X_1, \cdots, X_n$ be independent Bernoulli variables with respective success probabilities $p_1, \cdots, p_n$. Assume $0<p_i<1$ and put $\lambda_i=-\ln(1-p_i)$, $1\leqq i\leqq n$. Then*

$$d\left(\sum_{i=1}^{n} X_i, Y\left(\sum_{i=1}^{n} p_i\right)\right)\leqq \sum_{i=1}^{n} p_i^2 \tag{4.3}$$

and

$$d\left(\sum_{i=1}^{n} X_i, Y\left(\sum_{i=1}^{n} \lambda_i\right)\right)\leqq \tfrac{1}{2}\sum_{i=1}^{n} \lambda_i^2. \tag{4.4}$$

An alternative simple proof of (4.3), using indicator random variables, is given in [29]. The inequality (4.3) is a special case of a general result of Le Cam [23] proved, using operator-theoretic methods, for random elements in an Abelian group. Le Cam also establishes by such methods various competitors to (4.3), for example,

$$d\left(\sum_{i=1}^{n} X_i, Y\left(\sum_{i=1}^{n} p_i\right)\right)\leqq 9\max\{p_1, \cdots, p_n\}. \tag{4.5}$$

In § 5 we prove a d_0-metric analogue of (4.4) in which the probability approximation based on $Y(\sum_1^n \lambda_i)$ is actually a *lower bound*, with error bounded as in (4.4). This results from a *stochastic order* relationship between X and $Y(\lambda)$, where X is Bernoulli with success probability p and $\lambda=-\ln(1-p)$.

As a corollary of (4.3), the *classical* Poisson approximation may be deduced. Let S_n have the distribution binomial (n, p) and suppose that $n\to\infty$, $p\to 0$, $np\to\Delta$, $0<\Delta<\infty$. Then, by (3.3), (3.6) and (4.3),

$$d(S_n, Y(\Delta))\leqq d(S_n, Y(np))+d(Y(np), Y(\Delta))\leqq np^2+|\Delta-np|\to 0.$$

5. Bounds on the metric d_0 for Poisson approximation to a sum of independent Bernoulli variables. Again consider first the approximation of a *single* Bernoulli variable by a Poisson variate. The following improvement of (4.1) for d_0 in place of d was noted by Daley [9].

LEMMA 5.1. *Let X be a Bernoulli variable with success probability p, $0<p<1$. Then*

$$d_0(X, Y(p))\leqq \tfrac{1}{2}p^2. \tag{5.1}$$

Proof. For $Y=Y(p)$,

$$|P(X\leqq 0)-P(Y\leqq 0)|=e^{-p}-1+p,$$
$$|P(X\leqq 1)-P(Y\leqq 1)|=1-e^{-p}-p\,e^{-p},$$

and, for $m\geqq 2$,

$$|P(X\leqq m)-P(Y\leqq m)|\leqq|P(X\leqq 1)-P(Y\leqq 1)|.$$

Thus

$$d_0(X, Y)=\max\{e^{-p}-1+p, 1-e^{-p}-p\,e^{-p}\}=e^{-p}-1+p, \tag{5.2}$$

from which (5.1) follows. □

Now observe that if X is Bernoulli (p), and $\lambda = -\ln(1-p)$, then $Y(\lambda)$ is *stochastically larger* than X, i.e., $P(X \geqq t) \leqq P(Y(\lambda) \geqq t)$, for all t. This is seen by the following constructive approach. Given the random variable $Y(\lambda)$, define X by

$$X = I(Y(\lambda) \geqq 1), \tag{5.3}$$

where $I(B)$ denotes the indicator of the event B. Then $X \leqq Y(\lambda)$ and X is clearly Bernoulli (p). Further,

$$P[X \neq Y(\lambda)] = P[Y(\lambda) \geqq 2] = 1 - e^{-\lambda} - \lambda e^{-\lambda} \leqq \tfrac{1}{2}\lambda^2. \tag{5.4}$$

The idea of introducing stochastically ordered random variables in connection with d_0-metric approximation of a sum of Bernoulli variables is due to Gastwirth [15].

The d_0-metric parallel of Theorem 4.1 is

THEOREM 5.1. *Let $X_1, \cdots, X_n$ be independent Bernoulli variables with respective success probabilities $p_1, \cdots, p_n$. Assume $0 < p_i < 1$ and put $\lambda_i = -\ln(1-p_i)$, $1 \leqq i \leqq n$. Then*

$$d_0\left(\sum_{i=1}^{n} X_i, Y\left(\sum_{i=1}^{n} p_i\right)\right) \leqq \frac{1}{2} \sum_{i=1}^{n} p_i^2 \tag{5.5}$$

and, for each $m \in N_+$,

$$P\left[Y\left(\sum_{i=1}^{n} \lambda_i\right) \leqq m\right] \leqq P\left[\sum_{i=1}^{n} X_i \leqq m\right] \leqq P\left[Y\left(\sum_{i=1}^{n} \lambda_i\right) \leqq m\right] + \frac{1}{2} \sum_{i=1}^{n} \lambda_i^2. \tag{5.6}$$

For p_i small, $1 \leqq i \leqq n$, the lower and upper bounds for $P[\sum_1^n X_i \leqq m]$ implied by (5.5) are farther apart than those given by (5.6). On the other hand, since $p_i \leqq \lambda_i$, $1 \leqq i \leqq n$, the analogue of (5.5) implied by (5.6) is weaker than (5.5).

Relation (5.5), due to Daley [9], strengthens a special case of a more general result of Franken [12], by replacing a factor $2/\pi$ by $1/2$. Relation (5.6) is essentially due to Gastwirth [15].

Proof of Theorem 5.1. Application of Lemma 5.1 in conjunction with relation (3.8) gives (5.5). To obtain (5.6), we introduce independent Poisson variates $Y_i = Y(\lambda_i)$, $1 \leqq i \leqq n$, and define $X_i = I(Y_i \geqq 1)$, $1 \leqq i \leqq n$. Then, as remarked in connection with (5.3), $X_1, \cdots, X_n$ are independent Bernoulli variables with respective success probabilities $p_1, \cdots, p_n$ and satisfy

$$\sum_{i=1}^{n} X_i \leqq \sum_{i=1}^{n} Y_i. \tag{5.7}$$

Furthermore, $\sum_1^n Y_i = Y(\sum_1^n \lambda_i)$. Thus the first inequality in (5.6) clearly holds. Likewise, utilizing (5.4) in conjunction with (5.7), we obtain the second inequality in (5.6). □

6. Extension to the case of dependent Bernoulli summands. The following proposition, proved in [29], concerns the possible change in the value of $P[\sum_1^n X_i \in A]$ if the possibly dependent Bernoulli variables $X_1, \cdots, X_n$ become replaced by arbitrarily chosen *independent* Bernoulli variables. Write

$$\theta_1 = P(X_1 = 1) \tag{6.1a}$$

and, for $2 \leqq i \leqq n$,

$$\theta_i(x_1, \cdots, x_{i-1}) = P(X_i = 1 | X_1 = x_1, \cdots, X_{i-1} = x_{i-1}), \tag{6.1b}$$

for $x_1, \cdots, x_{i-1}$ taking values 0 or 1. For $i \geqq 2$, θ_i denotes the conditional probability of the event $X_i = 1$, given the values of the previous X_j's.

Example. Let $X_1, X_2, \cdots$ be *Markov-dependent* Bernoulli variables with success probability p and transition probabilities $p(1) = P(X_i = 1 | X_{i-1} = 1)$ and $p(0) = P(X_i = 1 | X_{i-1} = 0)$, for $i \geqq 2$. Then the relation $p = p(0)/[1 - p(1) + p(0)]$ must hold. Also, $\theta_1 = p$ and, for $i \geqq 2$,

$$\theta_i(x_1, \cdots, x_{i-1}) = \begin{cases} p(1) & \text{if } x_{i-1} = 1, \\ p(0) & \text{if } x_{i-1} = 0. \end{cases}$$

LEMMA 6.1. *Let $X_1, \cdots, X_n$ be (possibly dependent) Bernoulli variables and let $\theta_1, \cdots, \theta_n$ be given by* (6.1). *Let $X_1^*, \cdots, X_n^*$ be independent Bernoulli variables with respective success probabilities $p_1, \cdots, p_n$. Then*

$$d\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^*\right) \leqq \sum_{i=1}^{n} E|\theta_i - p_i|. \tag{6.2}$$

Remarks. (i) Of course, the bound in (6.2) applies also to $d_0(\sum X_i, \sum X_i^*)$.

(ii) In (6.2), $E|\theta_i - p_i|$ denotes the quantity $E|\theta_i(X_1, \cdots, X_{i-1}) - p_i|$. For $i = 1$, this is simply $|\theta_1 - p_1|$.

(iii) An important feature of the Lemma is that the values p_i may be selected arbitrarily. For example, a natural and convenient choice for p_i is the *mean* $E[\theta_i]$. However, if it is desired to minimize the quantity $E|\theta_i - p_i|$, the appropriate choice is p_i equal to any *median* of the distribution of $\theta_i(X_1, \cdots, X_{i-1})$.

The preceding lemma, by reduction to independent variables, facilitates the utilization of Theorems 4.1 and 5.1. The price paid is an extra term, contributed by (6.2), in the asserted bounds on the error of Poisson approximation. These implications are presented in § 8.

A somewhat different approach toward the case of dependent variables has been found by Galambos [14], who establishes a lemma for reduction from general dependence to a special type of dependence, *exchangeability*. The chief advantage is that the new random variables $X_1^*, \cdots, X_n^*$ satisfy $d(\sum_1^n X_i, \sum_1^n X_i^*) = 0$; that is, the reduction costs nothing as far as the distribution of $\sum_1^n X_i$ is concerned. However, no analogies of Theorems 4.1 and 5.1 are available for exchangeable variables. Nevertheless there do exist certain limit theorems [21], [28] applicable in similar fashion. The relevant implications are presented in [14].

7. Extension to the case of nonnegative integer-valued summands. As in § 6, we handle a general situation by reduction to a special case. Given *arbitrary* N_+-valued random variables $X_1, \cdots, X_n$, define related *Bernoulli* variables

$$X_i' = I(X_i = 1) \qquad 1 \leqq i \leqq n. \tag{7.1}$$

The relations $X_i' \leqq X_i$ and $P(X_i \neq X_i') = P(X_i \geqq 2)$, $1 \leqq i \leqq n$, and (3.4), readily yield the results

$$d\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i'\right) \leqq \sum_{i=1}^{n} P(X_i \geqq 2) \tag{7.2}$$

and, for each $m \in N_+$,

$$P\left[\sum_{i=1}^{n} X_i' \leqq m\right] - \sum_{i=1}^{n} P(X_i \geqq 2) \leqq P\left[\sum_{i=1}^{n} X_i \leqq m\right] \leqq P\left[\sum_{i=1}^{n} X_i' \leqq m\right]. \tag{7.3}$$

The preceding reduction to Bernoulli variables, followed by the reduction in § 6 to independent Bernoulli variables, facilitates utilization of Theorems 4.1 and 5.1. The relevant implications appear in § 8. The effectiveness of the reduction via (7.1) clearly depends upon the negligibility of $P(X_i \geqq 2)$.

Another approach, used by Franken [12], pertains to d_0 and to *independent* N_+-valued $X_1, \cdots, X_n$, and involves the method of characteristic functions. Description of some of Franken's work is contained in [3].

8. A general theorem. We now present an omnibus theorem for the case of a sum of possibly dependent N_+-valued random variables $X_1, \cdots, X_n$. Put

$$\theta_1 = P(X_1 = 1) \tag{8.1a}$$

and, for $2 \leqq i \leqq n$,

$$\theta_i(x_1, \cdots, x_{i-1}) = P(X_i = 1 | X_1 = x_1, \cdots, X_{i-1} = x_{i-1}), \tag{8.1b}$$

for $x_1, \cdots, x_{i-1} \in N_+$.

The devices of §§ 6 and 7 permit reduction to the simple case of arbitrarily chosen *independent Bernoulli* variables $X_1^*, \cdots, X_n^*$ with respective success probabilities $p_1, \cdots, p_n$. We assume $0 < p_i < 1$ and put $\lambda_i = -\ln(1-p_i)$, $\gamma_i = E|\theta_i - p_i|$ and $\delta_i = P(X_i \geqq 2)$, $1 \leqq i \leqq n$. We define

$$\alpha = \sum_{i=1}^n p_i^2, \quad \beta = \sum_{i=1}^n \lambda_i^2, \quad \gamma = \sum_{i=1}^n \gamma_i, \quad \delta = \sum_{i=1}^n \delta_i. \tag{8.2}$$

THEOREM 8.1. *Let $X_1, \cdots, X_n$, $p_1, \cdots, p_n$, $\lambda_1, \cdots, \lambda_n$ and $\alpha, \beta, \gamma, \delta$ be as given above. Then*

$$d\left(\sum_{i=1}^n X_i, Y\left(\sum_{i=1}^n p_i\right)\right) \leqq \alpha + \gamma + \delta, \tag{8.3}$$

$$d\left(\sum_{i=1}^n X_i, Y\left(\sum_{i=1}^n \lambda_i\right)\right) \leqq \tfrac{1}{2}\beta + \gamma + \delta, \tag{8.4}$$

$$d_0\left(\sum_{i=1}^n X_i, Y\left(\sum_{i=1}^n p_i\right)\right) \leqq \tfrac{1}{2}\alpha + \gamma + \delta, \tag{8.5}$$

and, for each $m \in N_+$,

$$P\left[Y\left(\sum_{i=1}^n \lambda_i\right) \leqq m\right] - \gamma - \delta \leqq P\left[\sum_{i=1}^n X_i \leqq m\right] \leqq P\left[Y\left(\sum_{i=1}^n \lambda_i\right) \leqq m\right] + \tfrac{1}{2}\beta + \gamma. \tag{8.6}$$

Proof. We show (8.3). Similar arguments lead to (8.4), (8.5) and (8.6). By the reduction device of § 7, we may replace $X_1, \cdots, X_n$ by Bernoulli variables $X_1', \cdots, X_n'$ satisfying $d(\sum_1^n X_i, \sum_1^n X_i') \leqq \delta$. In turn, by the device in § 6, the variables $X_1', \cdots, X_n'$ may be replaced by independent Bernoulli variables $X_1^*, \cdots, X_n^*$ having the given values $p_1, \cdots, p_n$ as respective success probabilities, with (by Lemma 6.1)

$$d\left(\sum_{i=1}^n X_i', \sum_{i=1}^n X_i^*\right) \leqq \sum_{i=1}^n E|\theta_i' - p_i|,$$

where $\theta_1', \cdots, \theta_n'$ are defined relative to $X_1', \cdots, X_n'$ by (6.1). Now $\theta_1' = \theta_1$ and, for $i \geqq 2$, $\theta_i' = E[\theta_i | X_1', \cdots, X_{i-1}']$, whence $E|\theta_i' - p_i| \leqq E|\theta_i - p_i|$. Thus $d(\sum_1^n X_i', \sum_1^n X_i^*) \leqq \gamma$. By Theorem 4.1, $d(\sum_1^n X_i^*, Y(\sum_1^n p_i)) \leqq \alpha$. Combining these results via the triangular inequality (3.3), we have (8.3). □

Remark. Regarding choice of the values $p_1, \cdots, p_n$, the remarks following Lemma 6.1 apply here. For the choice $p_i = E[\theta_i]$, formula (8.3) has been given as (1.3) of [29] and (8.4), slightly weakened, as (1.4) of [29].

Example. Continuing the example of § 6, concerning *Markov-dependent Bernoulli* variables, we note that the random variable $\theta_i(X_1, \cdots, X_{i-1})$ has *mean p*. Taking $p_i \equiv p$ in Theorem 8.1, we have

$$\gamma_i = E|\theta_i - p| = 2p(1-p)|p(1)-p(0)|, \qquad 2 \leqq i \leqq n,$$

so that (8.3) yields

$$d\left(\sum_{i=1}^{n} X_i,\ Y(np)\right) \leqq np^2 + 2(n-1)p(1-p)|p(1)-p(0)|. \tag{8.7}$$

For $p \leqq \frac{1}{2}$, the *median* of $\theta_i(X_1, \cdots, X_{i-1})$ is $p(0)$. Taking $p_i \equiv p(0)$ in Theorem 8.1, we have $\gamma_1 = |p - p(0)|$ and

$$\gamma_i = E|\theta_i - p(0)| = p|p(1)-p(0)|, \qquad 2 \leqq i \leqq n,$$

in which case (8.3) now yields

$$d\left(\sum_{i=1}^{n} X_i,\ Y(np(0))\right) \leqq np^2(0) + |p-p(0)| + (n-1)p|p(1)-p(0)|. \tag{8.8}$$

An associated *limit* theorem may also be deduced, by an argument similar to that at the end of § 4. Let $S_n = \sum_1^n X_i$ and suppose that $n \to \infty$, $p(0) \to 0$, $p(1) \to 0$ and $np \to \Delta$, $0 < \Delta < \infty$. Then, by (3.3), (3.6) and either (8.7) or (8.8), we obtain

$$d(S_n, Y(\Delta)) \to 0, \ n \to \infty.$$

9. Applications. Haight [17, Chap. 7] surveys numerous applications of the Poisson distribution in industry, agriculture, ecology, biology, medicine, telephony, accident theory, commerce, queuing theory, sociology, demography, traffic flow theory, particle counting, military problems and miscellaneous other areas. The purpose of the present discussion is to indicate briefly some recent developments in utilization of the Poisson approximation in probability theory and its applications. In many of these examples, the exact bounds available in results such as Theorem 8.1 play an important role.

(i) *Reliability of k-out-of-n systems*. An example of such a system is the monitoring system discussed in § 2 to illustrate relevance of the d_0 metric. The exploitation of Theorem 8.1, relations (8.5) and (8.6), in this context is carried out in [31].

(ii) *Occurrences of rare but persistent meteorological elements*. Because of the serial dependence of observations such as the daily rainfall at a site over a sequence of days, the classical Poisson approximation with its assumption of independence of trials has been used with restraint in probability analysis in the field of meteorology. However, the recent extensions for dependence now permit an expanded role for the Poisson approximation. See [30].

(iii) *Order statistics; extreme value theory*. Let $\xi_1, \cdots, \xi_n$ be random variables on a probability space and denote their ordered values by $\xi_1^* \leqq \xi_2^* \leqq \cdots \leqq \xi_n^*$. Note that

$$P(\xi_j^* < x) = P\left[\sum_{i=1}^{n} I(\xi_i \geqq x) \leqq n - j\right].$$

Thus, for large values of x, the distribution of ξ_j^* is essentially characterized by that of a sum of Bernoulli variables with small "success" probabilities. The development of Poisson approximation theory in this regard is found in [14].

(iv) *Superpositions of stochastic point processes.* A random subset of the real line, such as the times of failure of a component in a machine, is called a "stochastic point process". In particular, a "Poisson process' is a stochastic point process for which the numbers of points in fixed disjoint "time" intervals are *independent Poisson* variates. The superposition of a large number of "sparse" point processes is approximately a Poisson process. See [3].

(v) *Crossings of a high level by a general stochastic process.* Crossings of a sufficiently high level by a given stochastic process in a short period of time are "rare", but often important, events. The total number of crossings in a long period of time is approximately a Poisson random variable. See Cramér and Leadbetter [8] for extensive treatment and Brillinger [2] for recent discussion in the point process context and application in neurophysiology.

(vi) *Urn models.* Let balls be thrown uniformly and independently into n urns. Denote by $N_m(n, k)$ the number of throws after which k urns will first contain at least m balls. For the case $k = n$, Erdös and Renyi [10] proved

$$\lim_{n\to\infty} P[N_m(n, n) \leqq b_n(x)] = \exp\left(-\frac{e^{-x}}{(m-1)!}\right), \tag{9.1}$$

where

$$b_n(x) = [n\ (\ln n + (m-1) \ln \ln n + x)],$$

and $[\cdot]$ denotes greatest integer part. More recently, making application of a Poisson approximation theorem of Hodges and Le Cam [18], Kaplan [19] generalized (9.1) to

$$\lim_{n\to\infty} \max_{0\leqq k\leqq n} \left| P[N_m(n, k) \leqq b_n(x)] - P\left[Y\!\left(\frac{e^{-x}}{(m-1)!}\right) \leqq n-k \right] \right| = 0, \tag{9.2}$$

where $Y(\cdot)$ again denotes a Poisson variate as defined in § 2.

Another urn problem has also been attacked by a Poisson approximation approach, by Sevast'yanov [32] and further by Kaplan [20].

(vii) *Covering problems.* Let arcs of length α, $0 < \alpha < 1$, be thrown independently and uniformly on a circle with unit circumference. Denote by $N_m(\alpha)$ the number of arcs needed to cover the circle at least m times and let

$$c_\alpha(x) = \frac{1}{\alpha}\left(\ln \frac{1}{\alpha} + m \ln \ln \frac{1}{\alpha} + x\right).$$

Flatto [11] proved

$$\lim_{\alpha\to 0} P[N_m(\alpha) \leqq c_\alpha(x)] = \exp\left(-\frac{e^{-x}}{(m-1)!}\right). \tag{9.3}$$

Kaplan [20] obtains a simpler proof via the Poisson approximation result of Freedman [13].

(viii) *"Pyramid" or "chain letter" schemes.* In pyramid schemes, a participant's initial investment is recouped primarily by recruiting new participants. Gastwirth [15] develops novel methodology for Poisson approximation, as given in (5.6), and makes application to approximate key probabilities in pyramid schemes.

(ix) *Markov-dependent Bernoulli trials.* See [29], [30], [31].

10. Complements. (i) *Other metrics.* Metrics other than d and d_0 have received attention. See Franken [12], Vervaat [35], Simons and Johnson [33] and Chen [4] for a variety of results.

(ii) *Other approaches.* In the present paper we have used only elementary methods. For the basis of a general approach using *characteristic functions*, see Tsaregradskii [34]. As noted in § 7, this approach has been utilized in [12] and expounded in [3]. For the basis of a general approach using *operator-theoretic* methods, see Le Cam [23].

(iii) *Other extensions for dependent summands.* An approach due to Galambos [14] has been described in § 6. Still other approaches are given by Freedman [13] and Chen [5].

(iv) *Extensions to random elements of arbitrary Abelian groups.* See Le Cam [23] and Chen [6].

(v) *Functions of Bernoulli sequences.* It may happen that the trials of primary interest are generated by a more fundamental series of trials. For example, a sequence $\{X_i\}$ of Bernoulli variables generates a further sequence $\{Z_i\}$ through the relations $Z_i = X_iX_{i-1}$, $i = 2, 3, \cdots$. The sequence $\{Z_i\}$ records occurrences of successes *twice in succession* in the initial sequence. If the X_i's are replaced by X_i^*'s conveniently selected for purposes of approximation, then $\{Z_i\}$ becomes replaced by $\{Z_i^*\}$ but

$$d\left(\sum_{i=1}^n Z_i, \sum_{i=1}^n Z_i^*\right) \leqq \sum_{i=1}^n P(X_i \neq X_i^*).$$

Some application of this device has been made in [30].

(vi) *Limit theorems.* This paper has emphasized *exact bounds* on errors of approximation. Under typical assumptions apropos to asymptotic analysis, the bounds decrease to zero in the limit, so that a host of limit theorems are contained in the results presented. Examples were noted in § 4 and § 8.

Acknowledgment. The author is very grateful to D. J. Daley and to the referee for many valuable suggestions.

REFERENCES

[1] L. VON BORTKIEWICZ, *Das Gesetz der kleinen Zahlen*, Teubner, Leipzig, 1898.

[2] D. R. BRILLINGER, *Measuring the association of point processes: A case history*, Amer. Math. Monthly, 83 (1976), pp. 16–21.

[3] E. ÇINLAR, *Superposition of point processes*, Stochastic Point Processes, P. A. Lewis, ed., 1972, pp. 549–606.

[4] L. H. Y. CHEN, *On the convergence of Poisson binomial to Poisson distributions*, Ann. Probability, 2 (1974), pp. 178–180.

[5] ———, *Poisson approximation for dependent trials*, Ibid., 3 (1975), pp. 534–545.

[6] ———, *An approximation theorem for convolutions of probability measures*, Ibid., 3 (1975), pp. 992–999.

[7] K.-L. CHUNG, *A Course in Probability Theory*, 2nd ed., Academic Press, New York, 1974.

[8] H. CRAMÉR AND M. R. LEADBETTER, *Stationary and Related Stochastic Processes*, John Wiley, New York, 1967.

[9] D. J. DALEY, Private communication, 1975.

[10] P. ERDÖS AND A. RENYI, *On a classical problem in probability theory*, Magyar Tud. Akad. Kutato Int. Közl., 6 (1961), pp. 215–219.

[11] L. FLATTO, *A limit theorem for random coverings of a circle*, Israel J. Math., 15 (1973), pp. 167–184.

[12] P. FRANKEN, *Approximation der Verteilungen von Summen unabhängiger nichtnegativer ganzzahler Zufallsgrossen durch Poissonsche Verteilungen*, Math. Nachr., 27 (1964), pp. 303–340.

[13] D. FREEDMAN, *The Poisson approximation for dependent events*, Ann. Probability, 2 (1974), pp. 256–269.

[14] J. GALAMBOS, *A general Poisson limit theorem of probability theory*, Duke Math. J., 40 (1973), pp. 581–486.
[15] J. L. GASTWIRTH, *A probability model of a pyramid scheme*, Amer. Statist., 31 (1977), pp. 79–82.
[16] B. V. GNEDENKO, *The Theory of Probability*, 4th ed. Chelsea, New York, 1967.
[17] F. A. HAIGHT, *Handbook of the Poisson Distribution*, John Wiley, New York, 1967.
[18] J. L. HODGES AND L. LE CAM, *The Poisson approximation to the binomial distribution*, Ann. Math. Statist., 31 (1960), pp. 737–740.
[19] N. KAPLAN, *A generalization of a result of Erdös and Renyi and a related problem*, J. Appl. Probability, 14 (1977), to appear.
[20] ———, *Two applications of a Poisson approximation for dependent events*, Ann. Probability, 5 (1977), to appear.
[21] D. G. KENDALL, *On finite and infinite sequences of exchangeable events*, Studia Sci. Math. Hungar., 2 (1967), pp. 319–327.
[22] P. S. LAPLACE, *Théorie Analytique des Probabilités*, Paris, 1812.
[23] L. LE CAM, *An approximation theorem for the Poisson binomial distribution*, Pacific J. Math., 10 (1960), pp. 1181–1197.
[24] M. LOÈVE, *Probability Theory*, 3rd ed., Van Nostrand, NJ, 1963.
[25] A. DE MOIVRE, *The Doctrine of Chances*, 1718.
[26] E. PARZEN, *Modern Probability Theory and its Applications*, John Wiley, New York, 1960.
[27] S. D. POISSON, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, Précédées des Regles Générales du Calcul des Probabilités, 1837.
[28] C. J. RIDLER-ROWE, *On two problems of exchangeable events*, Studia Sci. Math. Hungar., 2 (1967), pp. 415–418.
[29] R. J. SERFLING, *A general Poisson approximation theorem*, Ann. Probability, 3 (1975), pp. 726–731.
[30] ———, *The Poisson distribution for the frequency of rare levels of persistent meteorological elements*, Proc. Fourth Conference on Probability and Statistics in Atmospheric Sciences, 1975, pp. 135–138.
[31] ———, *The role of the Poisson distribution in approximating system reliability of k-out-of-n structures*, Proc. Conference on Theory and Applications of Reliability, 1975, to appear.
[32] B. A. SEVAST'YANOV, *Poisson limit law for a scheme of sums of dependent random variables*, Theor. Probability Appl., 17 (1972), pp. 695–699.
[33] G. SIMONS AND N. L. JOHNSON, *On the convergence of binomial to Poisson distributions*, Ann. Math. Statist., 42 (1971), pp. 1735–1736.
[34] I. P. TSAREGRADSKII, *On uniform approximation of the binomial distribution by infinitely divisible laws*, Theor. Probability Appl., 3 (1958), pp. 434–438.
[35] W. VERVATT, *Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution*, Statistica Neerlandica, 23 (1969), pp. 79–86.