The Annals of Probability
1975, Vol. 3, No. 3, 534-545

POISSON APPROXIMATION FOR DEPENDENT TRIALS

BY LOUIS H. Y. CHEN

University of Singapore

Let $X_1, \cdots, X_n$ be an arbitrary sequence of dependent Bernoulli random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$. This paper establishes a general method of obtaining and bounding the error in approximating the distribution of $\sum_{i=1}^n X_i$ by the Poisson distribution. A few approximation theorems are proved under the mixing condition of Ibragimov (1959), (1962). One of them yields, as a special case and with some improvement, an approximation theorem of Le Cam (1960) for the Poisson binomial distribution. The possibility of an asymptotic expansion is also discussed and a refinement in the independent case obtained. The method is similar to that of Charles Stein (1970) in his paper on the normal approximation for dependent random variables.

0. Introduction. Let $X_1, X_2, \cdots, X_n$ be independent Bernoulli random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$. There has been considerable theoretical interest in how well the Poisson distribution approximates the distribution of $\sum_{i=1}^n X_i$. Prohorov (1953) showed that, in the case where all $p_i = \lambda/n$,

$$\sum_{k=0}^{\infty} |P(\sum_{i=1}^n X_i = k) - e^{-\lambda}\lambda^k/k!| \leqq (\lambda/n)[2(2\pi e)^{-\frac{1}{2}} + O(\min(1, \lambda^{-\frac{1}{2}}))] .$$

The bound was later improved considerably by Kerstan (1964) to $1.2\lambda/n$ and Vervaat (1969) to $(2^{\frac{1}{2}}\lambda/n)(1 - \lambda/n)^{-\frac{1}{2}}$. Hodges and Le Cam (1960) used an elementary argument to show that, in the general case with $\lambda = \sum_{i=1}^n p_i$, the difference between $P(\sum_{i=1}^n X_i \leqq x)$ and $\sum_{0 \leqq k \leqq x} e^{-\lambda}\lambda^k/k!$ is at most $3(\max_{1 \leqq i \leqq n} p_i)^{\frac{1}{3}}$. At the same time, Le Cam (1960) proved that for every real-valued function h defined on the nonnegative integers such that $|h| \leqq 1$,

$$|Eh(\sum_{i=1}^n X_i) - \mathscr{P}_\lambda h| \leqq 2 \sum_{i=1}^n p_i^2 \tag{0.1}$$

$$|Eh(\sum_{i=1}^n X_i) - \mathscr{P}_\lambda h| \leqq 9 \max_{1 \leqq i \leqq n} p_i \tag{0.2}$$

and

$$|Eh(\sum_{i=1}^n X_i) - \mathscr{P}_\lambda h| \leqq 16\lambda^{-1} \sum_{i=1}^n p_i^2 \quad (\max_{1 \leqq i \leqq n} p_i \leqq \tfrac{1}{4}) \tag{0.3}$$

where

$$\mathscr{P}_\lambda h = \sum_{k=0}^{\infty} e^{-\lambda}\lambda^k h(k)/k!$$

and

$$\lambda = \sum_{i=1}^n p_i .$$

This discrepancy between the distribution of $\sum_{i=1}^n X_i$ and the Poisson distribution considered by Le Cam is equivalent to that considered by Prohorov, Kerstan and Vervaat as mentioned above, in view of the fact that the supremum of

Received May 21, 1973; revised August 5, 1974.
AMS 1970 *subject classifications.* Primary 60F05, 62E20; Secondary 60G99.
Key words and phrases. Poisson approximation, rates of convergence, dependent trials.

$|Eh(\sum_{i=1}^n X_i) - \mathscr{P}_\lambda h|$ over all $|h| \leqq 1$ is attained by setting $h(k) = 1$ or -1 according as whether $P(\sum_{i=1}^n X_i = k) \geqq$ or $< e^{-\lambda}\lambda^k/k!$.

The results of Le Cam were later generalized by Kerstan (1964) to include independent nonnegative integer-valued random variables. One of his results when specialized to Bernoulli random variables yields Le Cam's (0.3), but with the absolute constant 16 replaced by 5.4.

In this paper, we consider the approximation problem in a different direction of generalization by moving from independence to dependence. To this end, we employ a perturbation technique similar to that of Stein (1970). An identity expressing the error in the approximation in terms of h is derived for an arbitrary finite sequence of dependent Bernoulli random variables. This is then used to obtain explicit bounds in special cases of dependence in Section 4, where the results in the m-dependent case yield Le Cam's as special cases and with some improvement. Finally, a refinement of Le Cam's results similar to one of Kerstan's in the independent case is also obtained.

1. Notation. Let $(\Omega, \mathscr{B}, P)$ be a probability space and let $\mathscr{F}$ be a sub-σ-algebra of $\mathscr{B}$. We shall denote the conditional expectation of a random variable X given $\mathscr{F}$ by $E^{\mathscr{F}}X$. In the same way the conditional expectation of X given a random element Y or $Y = y$ will be written as $E^Y X$ or $E^{Y=y}X$. We shall also denote the σ-algebra generated by a set of random variables $X_1, X_2, \cdots$ by $\mathscr{B}(X_1, X_2, \cdots)$ or $\mathscr{B}(X_i : i = 1, 2, \cdots)$. As usual, we shall denote the real line by $\mathbb{R}$ but denote the σ-algebra of Borel subsets of a subset H of $\mathbb{R}^k$ for $k \geqq 1$ by $\mathscr{B}(H)$. The sup norm of a bounded real-valued function f will be denoted by $||f||$. Finally, we adopt the convention that the sum $\sum_a^b$ is empty if $b < a$.

2. The basic identity. Consider an arbitrary sequence of dependent Bernoulli random variables $X_1, \cdots, X_n$ with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$. Let

$$W = \textstyle\sum_{i=1}^n X_i$$
$$W^{(i)} = \textstyle\sum_{k \neq i} X_k$$
$$V^{(i)} = \textstyle\sum_{|k-i|>m} X_k$$
$$\lambda = \textstyle\sum_{i=1}^n p_i$$

where m is a nonnegative integer. Then for every real-valued function f defined on the nonnegative integers, we have

$$\begin{aligned} E[Wf(W) - \lambda f(W+1)] &\\ = \textstyle\sum_{i=1}^n E[X_i f(W^{(i)}+1)] - \lambda E[f(W+1)] &\\ = \textstyle\sum_{i=1}^n E\{X_i[f(W^{(i)}+1) - f(V^{(i)}+1)]\} &\\ + \textstyle\sum_{i=1}^n E[(X_i - p_i)f(V^{(i)}+1)] &\\ + \textstyle\sum_{i=1}^n E\{p_i[f(V^{(i)}+1) - f(W+1)]\}\,. & \end{aligned} \tag{2.1}$$

Take X_i to be identically zero when $i \leqq 0$ or $\geqq n+1$ and let

$$Y_{ij} = V^{(i)} + \textstyle\sum_{k=i-m}^{j} X_k$$

and

$$Y'_{ij} = V^{(i)} + \sum_{k=i-m,k\neq i}^{j} X_k .$$

Also define $\Delta f(w) = f(w+1) - f(w)$. Then by writing each of $f(W^{(i)}+1) - f(V^{(i)}+1)$ and $f(V^{(i)}+1) - f(W+1)$ as a sum of differences and using the fact that each X_i takes on 0 and 1, (2.1) can be rewritten as

$$\begin{aligned} E[Wf(W) - \lambda f(W+1)] \\ &= \sum\sum_{0<|i-j|\leq m} E[X_i X_j \Delta f(Y'_{i,j-1}+1)] \\ &\quad + \sum_{i=1}^{n} E[(X_i - p_i) f(V^{(i)}+1)] \\ &\quad - \sum\sum_{|i-j|\leq m} p_i E[X_j \Delta f(Y_{i,j-1}+1)] . \end{aligned} \tag{2.2}$$

In order to make the Poisson approximation more apparent, we now choose f in the basic identity (2.2) such that

$$wf(w) - \lambda f(w+1) = h(w) - \mathscr{P}_\lambda h \tag{2.3}$$

where h is a bounded real-valued function defined on the nonnegative integers and

$$\mathscr{P}_\lambda h = e^{-\lambda} \sum_{k=0}^{\infty} h(k)\lambda^k/k! . \tag{2.4}$$

The solution of (2.3) is unique except at $w = 0$, but we see that the value of f at $w = 0$ does not enter into our consideration at all. For $w \geqq 1$, the solution of (2.3) is bounded (as will be seen in the next section) and is given by

$$\begin{aligned} f(w) &= -(w-1)!\, \lambda^{-w} \sum_{k=0}^{w-1} [h(k) - \mathscr{P}_\lambda h]\lambda^k/k! \\ &= (w-1)!\, \lambda^{-w} \sum_{k=w}^{\infty} [h(k) - \mathscr{P}_\lambda h]\lambda^k/k! \end{aligned} \tag{2.5}$$

where the second inequality follows from

$$e^{-\lambda} \sum_{k=0}^{\infty} [h(k) - \mathscr{P}_\lambda h]\lambda^k/k! = 0 .$$

From now on, we shall denote the solution of (2.3) for $w \geqq 1$ by $S_\lambda h(w)$. Substituting (2.5) into (2.2), we obtain

$$\begin{aligned} Eh(W) &= \mathscr{P}_\lambda h + \sum\sum_{0<|i-j|\leq m} E[X_i X_j \Delta S_\lambda h(Y'_{i,j-1}+1)] \\ &\quad + \sum_{i=1}^{n} E[(X_i - p_i) S_\lambda h(V^{(i)}+1)] \\ &\quad - \sum\sum_{|i-j|\leq m} p_i E[X_j \Delta S_\lambda h(Y_{i,j-1}+1)] . \end{aligned} \tag{2.6}$$

Thus it becomes clear that that question of how well the distribution of W can be approximated by the Poisson distribution with parameter λ can be answered by bounding the error terms on the right-hand side of (2.6).

3. Preliminary results. In this section, we shall prove a few lemmas concerning $S_\lambda h$ which will be useful later for bounding the error terms in (2.6).

LEMMA 3.1. *For $\lambda \geqq w$ and $w \geqq 1$,*

$$(w-1)!\, \lambda^{-w} \sum_{k=0}^{w-1} \lambda^k/k! \leqq 2\lambda^{-\frac{1}{2}} . \tag{3.1}$$

PROOF. The lemma is obviously true for $w = 1$. For $w \geqq 2$, let $\phi(\lambda)$ be an increasing function of λ such that $1 \leqq \phi(\lambda) \leqq \lambda$ and $\phi(\lambda) \to \infty$ as $\lambda \to \infty$. Then

$$\begin{aligned}(w-1)!\,\lambda^{-w} \textstyle\sum_{k=0}^{w-1} \lambda^k/k! \\ &= \textstyle\sum_{k=0}^{w-1} (w-1)\cdots(w-k)/\lambda^{k+1} \\ &\leqq \textstyle\sum_{k=0}^{[\lambda]-1} (\lambda-1)\cdots(\lambda-k)/\lambda^{k+1} \\ &\leqq (\textstyle\sum_{k=0}^{[\phi(\lambda)]-1} + \sum_{k=[\phi(\lambda)]}^{[\lambda]-1})(\lambda-1)\cdots(\lambda-k)/\lambda^{k+1} \\ &\leqq \phi(\lambda)/\lambda + \textstyle\sum_{k=0}^{\infty} (\lambda-\phi(\lambda))^k/\lambda^{k+1} \\ &= \phi(\lambda)/\lambda + 1/\phi(\lambda)\,,\end{aligned}$$

where $[a]$ denotes the largest integer $\leqq a$. Clearly the optimal choice of $\phi(\lambda)$ is $\phi(\lambda) = \lambda^{\frac{1}{2}}$ and this proves the lemma.

LEMMA 3.2. *For* $0 < \lambda \leqq w$ *and* $w \geqq 1$,

$$(w-1)!\,\lambda^{-w} \textstyle\sum_{k=w}^{\infty} \lambda^k/k! \leqq 2w^{-\frac{1}{2}}\,. \tag{3.2}$$

PROOF. For $w = 1$, 2 or 3, it is easy to check the validity of (3.2). For $w \geqq 4$,

$$\begin{aligned}(w-1)!\,\lambda^{-w} \textstyle\sum_{k=w}^{\infty} \lambda^k/k! \\ &\leqq \textstyle\sum_{k=0}^{\infty} w^{k-1}/(w+1)\cdots(w+k) \\ &= (\textstyle\sum_{k=0}^{[w^{\frac{1}{2}}]-2} + \sum_{k=[w^{\frac{1}{2}}]-1}^{\infty})\, w^{k-1}/(w+1)\cdots(w+k) \\ &\leqq w^{-1}\{[w^{\frac{1}{2}}] - 1 + 1 + w(w+[w^{\frac{1}{2}}])^{-1} \textstyle\sum_{k=0}^{\infty} w^k/(w+w^{\frac{1}{2}})^k\} \\ &= w^{-\frac{1}{2}}\{1 + [w^{\frac{1}{2}}]w^{-\frac{1}{2}} + (w^{\frac{1}{2}} - [w^{\frac{1}{2}}])/(w+[w^{\frac{1}{2}}])\} \\ &\leqq 2w^{-\frac{1}{2}} \qquad \text{by} \quad w^{\frac{1}{2}} \leqq w + [w^{\frac{1}{2}}]\,.\end{aligned}$$

LEMMA 3.3.

$$\|S_\lambda h\| \leqq 4\|h\| \min(\lambda^{-\frac{1}{2}}, 1)\,. \tag{3.3}$$

PROOF. If $\lambda \geqq w$, the lemma follows from the first equation of (2.5) and Lemma 3.1. On the other hand, if $\lambda \leqq w$, then the lemma follows from the second equation of (2.5) and Lemma 3.2.

LEMMA 3.4.

$$\|\Delta S_\lambda h\| \leqq 6\|h\| \min(\lambda^{-\frac{1}{2}}, 1)\,. \tag{3.4}$$

LEMMA 3.5. *For* $w \geqq 1$,

$$|\Delta S_\lambda h(w)| \leqq \lambda^{-1}\|h\|\{2 + 4|w-\lambda| \min(\lambda^{-\frac{1}{2}}, 1)\}\,. \tag{3.5}$$

PROOF OF LEMMAS 3.4 AND 3.5. By (2.3), it is easy to show that for $w \geqq 1$,

$$\Delta S_\lambda h(w) = -\lambda^{-1}[h(w) - \mathscr{P}_\lambda h - (w-\lambda)S_\lambda h(w)] \tag{3.6}$$

and

$$\Delta S_\lambda h(w) = -w^{-1}[h(w) - \mathscr{P}_\lambda h - (w-\lambda)S_\lambda h(w+1)]\,. \tag{3.7}$$

If $\lambda \geqq w$, (3.4) follows from (3.6) and Lemma 3.3. On the other hand, if $\lambda \leqq w$, then (3.4) follows from (3.7) and Lemma 3.3. Finally, (3.5) follows from (3.6) and Lemma 3.3 and this proves the lemmas.

4. Approximation theorems for mixing sequences of trials. It is clear from (2.6) that, under specific dependence assumptions, explicit bounds can be obtained for the discrepancy between the Poisson distribution and the distribution of a sum of n dependent Bernoulli random variables. The type of dependence for which the error terms in (2.6) render themselves to effective bounding is one where the dependence between the random variables decreases as the distance between them increases. Examples of such dependence are the mixing condition of Rosenblatt (1956) and that of Ibragimov (1959), (1962). Other examples can be found in Philipp (1969). In this paper, we shall only be concerned with the mixing condition of Ibragimov (1959), (1962). From the present exposition, it should be clear that results similar to those obtained in this paper could be obtained under other mixing conditions. Ibragimov's mixing condition may be described as follows: Let $X_1, X_2, \cdots$ be a finite or infinite sequence of random variables and let $\mathscr{M}_{a,b} = \mathscr{B}(X_i : a \leqq i \leqq b)$. There exists a monotone decreasing sequence $\phi(k) \downarrow 0$ such that for every $B \in \mathscr{M}_{j+k,\infty}$,

$$(4.1) \qquad |P(B \mid \mathscr{M}_{1j}) - P(B)| \leqq \phi(k) \quad \text{w.p. } 1.$$

A special case of sequences satisfying this mixing condition is one in which $(X_i, \cdots, X_r)$ and $(X_{r+k}, \cdots, X_j)$ are independent whenever $k > m$. Such a sequence is called an m-dependent sequence and is such that $\phi(k) = 0$ for $k > m$. Thus an independent sequence is also a special case satisfying (4.1) and, in this terminology, is 0-dependent. Other sequences of random variables satisfying (4.1) are discussed in Ibragimov (1962) and Billingsley (1968).

In this section, except for Lemmas 4.2 and 4.3, where $X_1, X_2, \cdots$ refer to an arbitrary sequence of random variables satisfying the mixing condition (4.1), $X_1, \cdots, X_n$ will be a sequence of Bernoulli random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$ which satisfies the mixing condition (4.1). All notations will be the same as in Sections 2 and 3.

LEMMA 4.1. *Let $(\Omega, \mathscr{F}, P)$ be a probability space, Y and Z be random elements taking values in the measurable spaces $(R, \mathscr{A})$ and $(S, \mathscr{B})$ respectively such that for every $B \in \mathscr{B}$,*

$$(4.2) \qquad |P(Z \in B \mid Y) - P(Z \in B)| \leqq \alpha \quad \text{w.p. } 1.$$

If $(S, \mathscr{B})$ is a Borel space, then there exists a regular conditional probability $\hat{P}_y$ for Z given $Y = y$ and a set $M \in \mathscr{A}$ with $P(Y \in M) = 1$ such that for every $y \in M$,

$$(4.3) \qquad \sup_{B \in \mathscr{B}} |\hat{P}_y(B) - \hat{P}(B)| \leqq \alpha$$

where $\hat{P}$ is the distribution of Z.

PROOF. Since $(S, \mathscr{B})$ is a Borel space, there exists a regular conditional probability $\hat{P}_y$ for Z given $Y = y$. It remains to show that there exists $M \in \mathscr{A}$ with $P(Y \in M) = 1$ such that (4.3) holds for every $y \in M$. Let $\psi : S \leftrightarrow H \in \mathscr{B}(\mathbb{R})$ be a one-to-one and onto mapping such that ψ and ψ^{-1} are measurable $\mathscr{B}$ and $\mathscr{B}(H)$

respectively. (This is possible, because $(S, \mathscr{B})$ is a Borel space.) Also let

$$\mathscr{C}' = \{H \cap A : A \text{ is a finite disjoint union of left open and right closed intervals in } \mathbb{R} \text{ with rational end points if the end points are finite}\}$$

and

$$\mathscr{C} = \{H \cap A : A \text{ is a finite disjoint union of left open and right closed intervals in } \mathbb{R}\}.$$

Since $\mathscr{C}'$ is countable, we may denote it by $\{C_1, C_2, \cdots\}$. By (4.2) and the properties of regular conditional probabilities, there exists, for every i, $M_i \in \mathscr{A}$ with $P(Y \in M_i) = 1$ such that for every $y \in M_i$,

$$|\hat{P}_y(\psi^{-1}(C_i)) - \hat{P}(\psi^{-1}(C_i))| \leqq \alpha .$$

Let $M = \bigcap_{i=1}^{\infty} M_i$. Then, clearly, $M \in \mathscr{A}$, $P(Y \in M) = 1$ and for every $y \in M$, (4.3) holds for $B = \psi^{-1}(C_i)$, $i = 1, 2, \cdots$, and hence holds for $B = \psi^{-1}(C)$, $C \in \mathscr{C}$, by the denseness of the rationals and the continuity of probability measures. As $\mathscr{C}$ is an algebra of subsets of H and $\mathscr{B}(H)$ the σ-algebra generated by $\mathscr{C}$, the proof of the lemma is completed by applying the monotone class theorem, using the continuity of probability measures and observing that

$$\mathscr{B} = \{B : B = \psi^{-1}(C), C \in \mathscr{B}(H)\} .$$

LEMMA 4.2. *Let $Y = (Y_1, \cdots, Y_r)$ and $Z = (Z_1, \cdots, Z_s)$ be r- and s-dimensional random vectors measurable $\mathscr{M}_{1k}$ and $\mathscr{M}_{k+m,\infty}$ respectively. Then for every bounded and Borel measurable function $f: \mathbb{R}^{r+s} \to \mathbb{R}$, we have*

$$|E^Y f(Y, Z) - E^Y f(Y, Z')| \leqq 2||f||\phi(m) \quad \text{w.p. } 1 \tag{4.4}$$

where Z' has the same distribution as Z and is independent of the sequence $X_1, X_2, \cdots$.

PROOF. Since $(\mathbb{R}^s, \mathscr{B}(\mathbb{R}^s))$ is a Borel space, it follows from Lemma 4.1 that a regular conditional probability $\hat{P}_y$ which satisfies (4.3) with $\alpha = \phi(m)$ exists for Z given $Y = y$. Then for every $y \in M$, the total variation of the signed measure $\hat{P}_y - \hat{P}$ is less than or equal to $2\phi(m)$. Hence for every $y \in M$,

$$|\int f(y, z)\, d\hat{P}_y(z) - \int f(y, z)\, d\hat{P}(z)| \leqq 2||f||\phi(m) .$$

This together with the properties of regular conditional probabilities implies (4.4). Hence the lemma.

LEMMA 4.3. *Let Y, T, Z be random variables measurable $\mathscr{M}_{ab}$, $\mathscr{M}_{cd}$, $\mathscr{M}_{ef}$ respectively where $c - b \geqq m$ and $e - d \geqq m$, and let $g: \mathbb{R} \to \mathbb{R}$ and $f: \mathbb{R}^2 \to \mathbb{R}$ be bounded and Borel measurable. Then*

$$\begin{aligned} &|Ef(Y, Z)g(T) - Ef(Y, Z)Eg(T)| \\ &\qquad \leqq 4\phi(m)||f||E|g(T)| + 2\phi(m)||g||E|f(Y, Z')| \end{aligned} \tag{4.5}$$

where Z' has the same distribution as Z and is independent of $X_1, X_2, \cdots$.

PROOF. The lemma follows from Lemma 4.2 and the inequality

$$\begin{aligned}|Ef(Y, Z)g(T) - Ef(Y, Z)Eg(T)| \\ &= |Eg(T)[E^{T,Y}f(Y, Z) - E^{T,Y}f(Y, Z')]| \\ &\quad + |Ef(Y, Z')[E^{Y,Z'}g(T) - Eg(T)]| \\ &\quad + |Eg(T)E[E^{Y}f(Y, Z') - E^{Y}f(Y, Z)]| .\end{aligned}$$

LEMMA 4.4.

$$[\mathrm{Var}\,(W)]^{\frac{1}{2}} \leqq \lambda^{\frac{1}{2}}[m + 1 + 2(n\phi(m + 1))^{\frac{1}{2}}] . \tag{4.6}$$

PROOF. By letting $f(w) = w$, (2.2) yields

$$\begin{aligned}\mathrm{Var}\,(W) = \lambda + \textstyle\sum\sum_{0<|i-j|\leqq m} EX_i X_j - \sum\sum_{|i-j|\leqq m} p_i p_j \\ + \textstyle\sum_{i=1}^{n} E[(X_i - p_i)V^{(i)}] .\end{aligned} \tag{4.7}$$

But the first sum on the right-hand side of (4.7) is less than or equal to

$$2m \textstyle\sum_{i=1}^{n} EX_i = 2m\lambda$$

and by Lemma 4.2,

$$\begin{aligned}&|\textstyle\sum_{i=1}^{n} E[(X_i - p_i)V^{(i)}]| \\ &\quad \leqq \textstyle\sum_{i=1}^{n} E[S_i|E^{S_i}X_i - p_i|] + \sum_{i=1}^{n} E[X_i|E^{X_i}T_i - ET_i|] \\ &\quad \leqq 2\phi(m + 1) \textstyle\sum_{i=1}^{n} ES_i + 2n\phi(m + 1) \sum_{i=1}^{n} EX_i \\ &\quad \leqq 4\lambda n\phi(m + 1)\end{aligned} \tag{4.8}$$

where

$$S_i = \textstyle\sum_{k<i-m} X_k \qquad \text{and} \qquad T_i = \sum_{k>i+m} X_k .$$

Thus $\mathrm{Var}\,(W) \leqq \lambda(2m + 1) + 4\lambda n\phi(m + 1)$ and this implies (4.6). Hence the lemma.

LEMMA 4.5.

$$\begin{aligned}\textstyle\sum\sum_{0<|i-j|\leqq m} EX_i X_j \leqq \mathrm{Var}\,(W) - \lambda + (2m + 1) \sum_{i=1}^{n} p_i^2 \\ + 4\lambda n\phi(m + 1) .\end{aligned} \tag{4.9}$$

PROOF. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\textstyle\sum\sum_{|i-j|\leqq m} p_i p_j &\leqq (2m + 1)^{\frac{1}{2}} \textstyle\sum_{i=1}^{n} p_i(\sum_{|i-j|\leqq m} p_j^2)^{\frac{1}{2}} \\ &\leqq (2m + 1)^{\frac{1}{2}}(\textstyle\sum_{i=1}^{n} p_i^2)^{\frac{1}{2}}(\sum\sum_{|i-j|\leqq m} p_j^2)^{\frac{1}{2}} \\ &\leqq (2m + 1) \textstyle\sum_{i=1}^{n} p_i^2 .\end{aligned} \tag{4.10}$$

Then (4.9) follows from (4.7), (4.8) and (4.10), and this proves the lemma.

LEMMA 4.6. *For every bounded function defined on the nonnegative integers, we have*

$$|\textstyle\sum_{i=1}^{n} E[(X_i - p_i)f(V^{(i)} + 1)]| \leqq 6||f||n\phi(m + 1) . \tag{4.11}$$

PROOF. The lemma follows immediately from Lemma 4.3.

Now we prove the approximation theorems.

THEOREM 4.1. *For $m = 0, 1, 2, \cdots$ and $|h| \leqq 1$, we have*

$$|Eh(W) - \mathscr{P}_\lambda h| \leqq 6 \min(\lambda^{-\frac{1}{2}}, 1)[\mathrm{Var}(W) - \lambda + 2(2m+1)\textstyle\sum_{i=1}^n p_i^2 + 4(\lambda+1)n\phi(m+1)]\,. \tag{4.12}$$

PROOF. By Lemmas 3.3, 3.4 and 4.6, (2.6) yields

$$|Eh(W) - \mathscr{P}_\lambda h| \leqq 6 \min(\lambda^{-\frac{1}{2}}, 1)[\textstyle\sum\sum_{0<|i-j|\leqq m} EX_i X_j + \sum\sum_{|i-j|\leqq m} p_i p_j + 4n\phi(m+1)]$$

which by (4.9) and (4.10) implies (4.12) and this proves the theorem.

THEOREM 4.2. *For $m = 0, 1, 2, \cdots$ and $|h| \leqq 1$, we have*

$$\begin{aligned}|Eh(W) - \mathscr{P}_\lambda h| &\leqq 2\lambda^{-1}[8m + 5 + 4(n\phi(m+1))^{\frac{1}{2}}] \\ &\quad \times [\mathrm{Var}(W) - \lambda + 2(2m+1)\textstyle\sum_{i=1}^n p_i^2] \\ &\quad + 32[6m + 3 + (n\phi(m+1))^{\frac{1}{2}}]n\phi(m+1)\,.\end{aligned} \tag{4.13}$$

PROOF. By Lemmas 3.3, 3.5 and 4.6, (2.6) again yields

$$\begin{aligned}|Eh(W) - \mathscr{P}_\lambda h| &\\ \leqq 2\lambda^{-1} &\textstyle\sum\sum_{0<|i-j|\leqq m} E\{X_i X_j[1 + 2\min(\lambda^{-\frac{1}{2}}, 1)|Y'_{i,j-1} + 1 - \lambda|]\} \\ &+ 2\lambda^{-1}\textstyle\sum\sum_{|i-j|\leqq m} p_i E\{X_j[1 + 2\min(\lambda^{-\frac{1}{2}}, 1)|Y_{i,j-1} + 1 - \lambda|]\} \\ &+ 24\min(\lambda^{-\frac{1}{2}}, 1)n\phi(m+1)\,.\end{aligned} \tag{4.14}$$

Let $V^{(i,j)} = \sum_{|k-i|>m, |k-j|>m} X_k$. Then, for $0 < |i-j| \leqq m$, we have

$$\begin{aligned}&E\{X_i X_j[1 + 2\min(\lambda^{-\frac{1}{2}}, 1)|Y'_{i,j-1} + 1 - \lambda|]\} \\ &\quad\leqq E\{X_i X_j[6m + 3 + 2\min(\lambda^{-\frac{1}{2}}, 1)|V^{(i,j)} - EV^{(i,j)}|]\} \\ &\quad= [EX_i X_j][6m + 3 + 2\min(\lambda^{-\frac{1}{2}}, 1)E|V^{(i,j)} - EV^{(i,j)}|] \\ &\qquad + 2\min(\lambda^{-\frac{1}{2}}, 1)\,\mathrm{Cov}(|V^{(i,j)} - EV^{(i,j)}|, X_i X_j)\,.\end{aligned} \tag{4.15}$$

By Jensen's inequality and Lemma 4.4,

$$E|V^{(i,j)} - EV^{(i,j)}| \leqq [\mathrm{Var}(V^{(i,j)})]^{\frac{1}{2}} \leqq \lambda^{\frac{1}{2}}[m + 1 + 2(n\phi(m+1))^{\frac{1}{2}}] \tag{4.16}$$

where it is observed that the sequence obtained from $X_1, \cdots, X_n$ by omitting some of the X_i's again satisfies the mixing condition (4.1) and that $EV^{(i,j)} \leqq \lambda$. By Lemma 4.3, for $0 < |i-j| \leqq m$, we have

$$\begin{aligned}&\mathrm{Cov}(|V^{(i,j)} - EV^{(i,j)}|, X_i X_j) \\ &\quad\leqq 4n\phi(m+1)EX_i X_j + 2\phi(m+1)E|S_{ij} + T_{ij} - ES_{ij} - ET_{ij}| \\ &\quad\leqq 4np_i\phi(m+1) + 4\lambda\phi(m+1)\end{aligned} \tag{4.17}$$

where

$$S_{ij} = \textstyle\sum_{k<i-m, k<j-m} X_k\,,$$
$$T_{ij} = \textstyle\sum_{k>i+m, k>j+m} X_k'$$

and $(X_1', \cdots, X_n')$ is an independent copy of $(X_1, \cdots, X_n)$. Thus the substitution of (4.16) and (4.17) into (4.15) yields

$$(4.18)\qquad \begin{aligned} &E\{X_i X_j[1 + 2\min(\lambda^{-\frac{1}{2}}, 1)|Y'_{i,j-1} + 1 - \lambda|]\} \\ &\quad \leqq [EX_i X_j][8m + 5 + 4(n\phi(m+1))^{\frac{1}{2}}] \\ &\qquad + 8\min(\lambda^{-\frac{1}{2}}, 1)(np_i + \lambda)\phi(m+1)\,. \end{aligned}$$

In the same way, it can be shown that, for $|i - j| \leqq m$, we have

$$(4.19)\qquad \begin{aligned} &p_i E\{X_j[1 + 2\min(\lambda^{-\frac{1}{2}}, 1)|Y_{i,j-1} + 1 - \lambda|]\} \\ &\quad \leqq p_i p_j[8m + 5 + 4(n\phi(m+1))^{\frac{1}{2}}] \\ &\qquad + 8\min(\lambda^{-\frac{1}{2}}, 1)(np_i + \lambda)\phi(m+1)\,, \end{aligned}$$

By substituting (4.18) and (4.19) into (4.14), by Lemma 4.5 and by (4.10), we obtain (4.13) and this proves the theorem.

Theorems 4.1 and 4.2 immediately yield

THEOREM 4.3. *If $X_1, \cdots, X_n$ are m-dependent, then for $|h| \leqq 1$, we have*

$$(4.20)\qquad \begin{aligned} &|Eh(W) - \mathscr{P}_\lambda h| \\ &\quad \leqq 6\min(\lambda^{-\frac{1}{2}}, 1)[\mathrm{Var}(W) - \lambda + 2(2m+1)\textstyle\sum_{i=1}^n p_i^2] \\ &\quad = 6\min(\lambda^{-\frac{1}{2}}, 1)[\textstyle\sum\sum_{i\neq j}\mathrm{Cov}(X_i, X_j) + (4m+1)\sum_{i=1}^n p_i^2] \end{aligned}$$

and

$$(4.21)\qquad \begin{aligned} &|Eh(W) - \mathscr{P}_\lambda h| \\ &\quad \leqq 2(8m+5)\lambda^{-1}[\mathrm{Var}(W) - \lambda + 2(2m+1)\textstyle\sum_{i=1}^n p_i^2] \\ &\quad = 2(8m+5)\lambda^{-1}[\textstyle\sum\sum_{i\neq j}\mathrm{Cov}(X_i, X_j) + (4m+1)\sum_{i=1}^n p_i^2]\,. \end{aligned}$$

For $m = 0$, we obtain

COROLLARY 4.1 (Le Cam). *If $X_1, \cdots, X_n$ are independent, then for $|h| \leqq 1$, we have*

$$(4.22)\qquad |Eh(W) - \mathscr{P}_\lambda h| \leqq 6\min(\lambda^{-\frac{1}{2}}, 1)\textstyle\sum_{i=1}^n p_i^2$$

and

$$(4.23)\qquad |Eh(W) - \mathscr{P}_\lambda h| \leqq 10\lambda^{-1}\textstyle\sum_{i=1}^n p_i^2$$

where the former is a slight variation of (0.1) *and the latter is an improvement of* (0.3). *The absolute constant in* (4.23) *is not as small as has been obtained by Kerstan* (1964).

However, it should be mentioned that no special attempt has been made in this paper to minimize the absolute constants in the bounds.

Finally, Theorems 4.1 and 4.2 also yield

THEOREM 4.4. *If $X_1, \cdots, X_n$ are identically distributed with $\phi(m) = e^{-\alpha m}$ for some $\alpha > 0$, then for $|h| \leqq 1$ and $n \geqq 3$, we have*

$$(4.24)\qquad \begin{aligned} &|Eh(W) - \mathscr{P}_\lambda h| \\ &\quad \leqq C_1(\alpha)\min(\lambda^{-\frac{1}{2}}, 1)[\mathrm{Var}(W) - \lambda + (\lambda+1)^2 n^{-1}\log n] \end{aligned}$$

and

(4.25) $$|Eh(W) - \mathscr{P}_\lambda h| \leqq C_2(\alpha)\lambda^{-1}\log n[\mathrm{Var}(W) - \lambda + \lambda(\lambda+1)n^{-1}\log n]$$

where $C_1(\alpha)$ *and* $C_1(\alpha)$ *depend only on* α.

PROOF. Put $m = [2\alpha^{-1}\log n]$.

5. A refinement. It is possible to obtain expansion of $Eh(W)$ about $\mathscr{P}_\lambda h$ up to any desired order by repeated iteration of the basic identity derived in Section 2. In this section, we shall do so up to the second order for independent Bernoulli random variables and obtain a bound for the error of second order. It should be mentioned that a similar result has been obtained by Kerstan (1964) but by a different method. All notations are the same as in Sections 2 and 3. To facilitate computations, we define two operators L and U_λ by

$$Lf(w) = f(w+1)$$

and

$$U_\lambda h(w) = \Delta S_\lambda h(w+1)\,.$$

LEMMA 5.1. *For* $\lambda \geqq b > 0$ *and* $k = 0, 1, 2, \cdots$,

(5.1) $$|e^{-\lambda}\lambda^k - e^{-b}b^k| \leqq (\lambda - b)(e^{-\lambda}k\lambda^{k-1} + e^{-b}b^k)\,.$$

PROOF. We first obtain

$$|e^{-\lambda}\lambda^k - e^{-b}b^k| \leqq e^{-\lambda}\lambda^k|1 - [1 - \lambda^{-1}(\lambda - b)]^k| + e^{-b}b^k|1 - e^{-\lambda+b}|\,.$$

This together with $1 - kx \leqq (1-x)^k$ for $0 \leqq x \leqq 1$, $k \geqq 1$ and $1 - e^{-x} \leqq x$ yield (5.1).

An immediate consequence of Lemma 5.1 is

LEMMA 5.2. *For* $\lambda \geqq b > 0$,

(5.2) $$|\mathscr{P}_\lambda f - \mathscr{P}_b f| \leqq (\lambda - b)(\mathscr{P}_\lambda|Lf| + \mathscr{P}_b|f|)\,.$$

LEMMA 5.3. *For* $m = 0, 1, 2, \cdots$,

(5.3) $$\mathscr{P}_b|L^m U_\lambda h| \leqq 2\lambda^{-1}||h||[1 + 2|m + 1 + b - \lambda| + 2b^{\frac12}\min(\lambda^{-\frac12}, 1)]\,.$$

PROOF. Let Y be Poisson distributed with parameter b. Then by Lemma 3.5,

$$\begin{aligned}\mathscr{P}_b|L^m U_\lambda h| &= E|U_\lambda h(Y+m)| \\ &\leqq 2\lambda^{-1}||h||[1 + 2\min(\lambda^{-\frac12}, 1)E|Y + m + 1 - \lambda|]\,.\end{aligned}$$

But $E|Y + m + 1 - \lambda| \leqq |m + 1 + b - \lambda| + [E(Y-b)^2]^{\frac12}$. Hence (5.3).

LEMMA 5.4. *For* $w \geqq 1$,

(5.4) $$|S_b U_\lambda h(w)| \leqq 4\lambda^{-1}||h||[(2 + 4|1 + b - \lambda|)\min(b^{-\frac12}, 1) + 3\min(\lambda^{-\frac12}, 1)]\,.$$

PROOF. We have, for $w \geqq 1$,

(5.5) $$\begin{aligned}S_b U_\lambda h(w) &= -(w-1)!\, b^{-w} \textstyle\sum_{k=0}^{w-1} [U_\lambda h(k) - \mathscr{P}_b U_\lambda h]b^k/k! \\ &= (w-1)!\, b^{-w} \textstyle\sum_{k=w}^{\infty} [U_\lambda h(k) - \mathscr{P}_b U_\lambda h]b^k/k!\,.\end{aligned}$$

If $b \leqq w$, (5.4) follows from the second equality in (5.5) and Lemmas 3.2, 3.5 and 5.3. On the other hand, if $1 \leqq w \leqq b$, (5.4) follows from the first equality in (5.5), Lemmas 3.1, 3.5 and 5.3. This proves the lemma.

LEMMA 5.5. *For* $b \leqq \lambda$,

$$\begin{aligned}|U_b U_\lambda h(w)| \leqq 4(\lambda b)^{-1}\|h\|[1 + |1 + b - \lambda| \\ + (b^{\frac{1}{2}} + |w + 2 - \lambda| + 3|w + 1 - \lambda|) \min(\lambda^{-\frac{1}{2}}, 1) \\ + (2 + 4|1 + b - \lambda|)|w + 1 - \lambda| \min(b^{-\frac{1}{2}}, 1)] . \end{aligned} \tag{5.6}$$

PROOF. First we have

$$U_b U_\lambda h(w) = -b^{-1}[U_\lambda h(w + 1) - \mathscr{P}_b U_\lambda h - (w + 1 - b)S_b U_\lambda h(w + 1)] .$$

This together with Lemmas 3.5, 5.3 and 5.4 implies (5.6).

THEOREM 5.1. *If* $X_1, \cdots, X_n$ *are independent and* $\bar{p} = \max_{1\leqq i\leqq n} p_i \leqq \lambda/2$, *then for* $|h| \leqq 1$ *and* $n \geqq 2$,

$$|Eh(W) - \mathscr{P}_\lambda h + (\textstyle\sum_{i=1}^n p_i^2)\mathscr{P}_\lambda U_\lambda h| \leqq C\lambda^{-1} \textstyle\sum_{i=1}^n p_i^3 \tag{5.7}$$

where C is an absolute constant not greater than $24 + 96(2)^{\frac{1}{2}}$.

PROOF. In order that the use of symbols may be simplified, we introduce two random indices I, J independent of the X_i's such that (I, J) is uniformly distributed on $\{(i, j): i, j = 1, 2, \cdots, n; i \neq j\}$. By putting $m = 0$ and using independence, (2.6) yields

$$Eh(W) = \mathscr{P}_\lambda h - nEp_I^2 E^I U_\lambda h(W^*) \tag{5.8}$$

where

$$W^* = \textstyle\sum_{i\neq I} X_i .$$

Now applying (5.8) again to $E^I U_\lambda h(W^*)$ but using the random index J, we obtain

$$\begin{aligned} Eh(W) &= \mathscr{P}_\lambda h - Ep_I^2[\mathscr{P}_{\lambda^*} U_\lambda h - (n - 1)E^I p_J^2 U_{\lambda^*} U_\lambda h(W^{**})] \\ &= \mathscr{P}_\lambda h - (\textstyle\sum_{i=1}^n p_i^2)\mathscr{P}_\lambda U_\lambda h - nEp_I^2(\mathscr{P}_{\lambda^*} U_\lambda h - \mathscr{P}_\lambda U_\lambda h) \\ &\quad + n(n - 1)Ep_I^2 p_J^2 E^{I,J} U_{\lambda^*} U_\lambda h(W^{**}) \end{aligned}$$

where

$$\lambda^* = \textstyle\sum_{i\neq I} p_i$$

and

$$W^{**} = \textstyle\sum_{i\neq I,J} X_i .$$

By Lemmas 5.2 and 5.3,

$$\begin{aligned} |\mathscr{P}_{\lambda^*} U_\lambda h - \mathscr{P}_\lambda U_\lambda h| &\leqq p_I[\mathscr{P}_\lambda |LU_\lambda h| + \mathscr{P}_{\lambda^*}|U_\lambda h|] \\ &\leqq 24 p_I \lambda^{-1} , \end{aligned}$$

and by Lemma 5.5,

$$\begin{aligned} |E^{I,J} U_{\lambda^*} U_\lambda h(W^{**})| &\leqq 8(\lambda\lambda^*)^{-1}[7 + 5(\lambda^*)^{-\frac{1}{2}} E^{I,J}|W^{**} - \lambda^{**}|] \\ &\leqq 8(\lambda\lambda^*)^{-1}[7 + 5(\lambda^*)^{-\frac{1}{2}}(\mathrm{Var}^{I,J}(W^{**})^{\frac{1}{2}}] \\ &\leqq 96(\lambda\lambda^*)^{-1} , \end{aligned}$$

where

$$\lambda^{**} = \sum_{i \neq I,J} p_i \,.$$

Thus

$$(5.9) \qquad |Eh(W) - \mathscr{P}_\lambda h + (\sum_{i=1}^n p_i^2)\mathscr{P}_\lambda U_\lambda h| \\ \leqq 24\lambda^{-1} nEp_I^3 + 96\lambda^{-1} n(n-1)E(\lambda^*)^{-1} p_I^2 p_J^2 \,.$$

But by Jensen's inequality,

$$\begin{aligned} \lambda^{-1}n(n-1)E(\lambda^*)^{-1}p_I^2p_J^2 &= \lambda^{-1} \sum_{i=1}^n p_i^2 \sum_{j\neq i} p_j^2/\lambda^{(i)} \\ &\leqq \lambda^{-1} \sum_{i=1}^n p_i^2 (\sum_{j\neq i} p_j^3/\lambda^{(i)})^{\frac{1}{2}} \\ &\leqq (\lambda - \bar{p})^{-\frac{1}{2}} (\sum_{j=1}^n p_j^3)^{\frac{1}{2}} \sum_{i=1}^n p_i^2/\lambda \\ &\leqq [\lambda(\lambda - \bar{p})]^{-\frac{1}{2}} \sum_{i=1}^n p_i^3 \\ &\leqq 2^{\frac{1}{2}} \lambda^{-1} \sum_{i=1}^n p_i^3 \,, \end{aligned}$$

where $\lambda^{(i)} = \sum_{j\neq i} p_j$. Hence the theorem.

Acknowledgment. This paper is based on a part of the author's Ph. D. dissertation written at Stanford University in 1971 under the supervision of Professor Charles Stein. The author wishes to thank Professor Stein for his guidance and suggestions. He also wishes to thank the referee for suggesting the use of regular conditional probabilities which has led to Lemma 4.1 and the improvement of Lemmas 4.2 and 4.3.

REFERENCES

[1] BILLINGSLEY, P. (1968) *Convergence of Probability Measures*. Wiley, New York.

[2] HODGES, J. L. and LE CAM, L. (1960). The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.* **31** 737-740.

[3] IBRAGIMOV, I. A. (1959). Some limit theorems for strict-sense stationary stochastic processes. *Dokl. Akad. Nauk SSSR* **125** 711-714. (In Russian).

[4] IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theor. Probability Appl.* **7** 349-382.

[5] KERSTAN, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** 173-179.

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

[7] PHILIPP, W. (1969). The central limit problem for mixing sequences of random variables. *Z. Wahrscheinlichkeitstheorie und Verw.Gebiete* **12** 155-171.

[8] PROHOROV, J. V. (1953). Asymptotic behavior of the binomial distribution. *Uspehi Mat. Nauk* **8** 3: 135-142.

[9] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43-47.

[10] STEIN, C. (1970). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2**. Univ. of California Press.

[11] VERVAAT, W. (1969). Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution. *Statistica Neerlandica* **23** 79-86.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SINGAPORE
SINGAPORE 10
REPUBLIC OF SINGAPORE