On the rate of Poisson convergence — Source link

Andrew Barbour, Peter Hall

Institutions: University of Cambridge, Australian National University

Published on: 01 May 1984

Year: 1984

On the rate of Poisson convergence

Barbour, A D; Hall, P

Math. Proc. Camb. Phil. Soc. (1984), **95**, 473
Printed in Great Britain

On the rate of Poisson convergence

BY A. D. BARBOUR†
Gonville and Caius College, Cambridge CB2 1TA

AND PETER HALL
Department of Statistics, The Faculties, Australian National University

(*Received* 25 *May* 1983; *revised* 29 *November* 1983)

1. *Introduction*

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables, and let $p_i = P[X_i = 1]$, $\lambda = \sum_{i=1}^n p_i$ and $W = \sum_{i=1}^n X_i$. Successively improved estimates of the total variation distance between the distribution $\mathscr{L}(W)$ of W and a Poisson distribution P_λ with mean λ have been obtained by Prohorov [5], Le Cam [4], Kerstan [3], Vervaat [8], Chen [2], Serfling [7] and Romanowska [6]. Prohorov, Vervaat and Romanowska discussed only the case of identically distributed X_i's, whereas Chen and Serfling were primarily interested in more general, dependent sequences. Under the present hypotheses, the following inequalities, here expressed in terms of the total variation distance

$$d(\mu, \nu) \equiv \sup_{A \subset Z} |\mu(A) - \nu(A)|,$$

were established respectively by Le Cam, Kerstan and Chen:

$$\left.\begin{aligned} d(\mathscr{L}(W), P_\lambda) &\leqslant \sum_{i=1}^n p_i^2; \\ d(\mathscr{L}(W), P_\lambda) &\leqslant 1\cdot05\lambda^{-1} \sum_{i=1}^n p_i^2, \quad \text{if} \max_i p_i \leqslant \tfrac{1}{4}; \\ d(\mathscr{L}(W), P_\lambda) &\leqslant 5\lambda^{-1} \sum_{i=1}^n p_i^2. \end{aligned}\right\} \qquad (1\cdot1)$$

(Kerstan's published estimate of $2d \leqslant 1\cdot2\lambda^{-1}\sum_{i=1}^n p_i^2$([3], p. 174, equation (1)) is a misprint for $2d \leqslant 2\cdot1\lambda^{-1}\sum_{i=1}^n p_i^2$, the constant $2\cdot1$ appearing twice on p. 175 of his paper.) Here, we use Chen's [2] elegant adaptation of Stein's method to improve the estimates given in (1·1), and we complement these estimates with a reverse inequality expressed in similar terms. Second order estimates, and the case of more general non-negative integer valued X_i's, are also discussed.

In the latter case, it is natural to expect the distribution of W to be almost Poisson only if the contribution to W from X_i's taking values other than 1 is in some sense small. As observed by Serfling, the X_i's can be reduced to 0–1 random variables by replacing all values greater than 1 by 0, at a cost in total variation distance of no more than $\sum_{j=1}^n P[X_j \geqslant 2]$. Thus close approximation to the Poisson is possible when the chance of $\max_i X_i$ exceeding 1 is small. It is shown here that good Poisson approxi-

† Present address: Institut für Angewandte Mathematik der Universität Zürich, CH-8001 Zürich.

mation can also be achieved in another natural situation, in which the expected contribution to W from those X_i's greater than 1 is negligible compared to

$$\left\{\sum_{j=1}^{n} P[X_j = 1]\right\}^{\frac{1}{2}},$$

the standard deviation of the approximating Poisson distribution; upper and lower distance estimates are derived to quantify this approximation.

2. *Upper and lower bounds*

Let x be any real valued function on the non-negative integers. Then, as in Chen [**2**],

$$\begin{aligned} E\{\lambda x(W+1) - Wx(W)\} &= \sum_{j=1}^{n} \{p_j Ex(W+1) - E(X_j x(W))\} \\ &= \sum_{j=1}^{n} p_j E\{x(W+1) - x(W_j+1)\} \\ &= \sum_{j=1}^{n} p_j^2 E\{x(W_j+2) - x(W_j+1)\}, \end{aligned} \tag{2·1}$$

where $W_j = W - X_j$. The following theorem is derived from (2·1) by choosing a suitable function x.

THEOREM 1. *Under the hypotheses set out in the Introduction,*

$$d(\mathscr{L}(W), P_\lambda) \leqslant \lambda^{-1}(1-e^{-\lambda}) \sum_{j=1}^{n} p_j^2. \tag{2·2}$$

Proof. For any $A \subseteq \mathbb{Z}$, define $x = x_{\lambda, A}$ by

$$x(0) = 0; \quad x(m+1) = \lambda^{-m-1} e^{\lambda} m! \, [P_\lambda(A \cap U_m) - P_\lambda(A) P_\lambda(U_m)], \quad m \geqslant 0, \tag{2·3}$$

where $U_m = \{0, 1, \ldots, m\}$. For this x,

$$\lambda x(m+1) - mx(m) = I[m \in A] - P_\lambda(A),$$

and so, from (2·1),

$$|P[W \in A] - P_\lambda(A)| \leqslant \sum_{j=1}^{n} p_j^2 E|x(W_j+2) - x(W_j+1)|. \tag{2·4}$$

It is shown in the appendix to Barbour and Eagleson [1], that, uniformly in A,

$$\|x\| \equiv \sup_{m \geqslant 0} |x(m)| \leqslant 1 \wedge (1{\cdot}4\lambda^{-\frac{1}{2}}), \tag{2·5}$$

and

$$\Delta x \equiv \sup_{m \geqslant 0} |x(m+1) - x(m)| \leqslant \lambda^{-1}(1-e^{-\lambda}). \tag{2·6}$$

The theorem follows immediately from (2·4) and (2·6). ∎

Remark. Theorem 1 improves upon each of the estimates given in (1·1). For $0 \leqslant \lambda \leqslant 1$, set $p_1 = \lambda$; $p_j = 0$, $2 \leqslant j \leqslant n$. Then

$$d(W, P_\lambda) = \lambda(1-e^{-\lambda}) = \lambda^{-1}(1-e^{-\lambda}) \sum_{j=1}^{n} p_j^2,$$

so that the inequality (2·2) is sharp in this case. For integral $\lambda \geqslant 1$, set $p_j = 1$, $1 \leqslant j \leqslant \lambda$; $p_j = 0$, $\lambda < j \leqslant n$. Then

$$d(\mathscr{L}(W), P_\lambda) = 1 - \lambda^\lambda e^{-\lambda}/\lambda! \approx 1 - 1/\sqrt{2\pi\lambda},$$

as compared to the right hand side of (2·2) which takes the value $1-e^{-\lambda}$, both of which tend to one as $\lambda \to \infty$. Thus (2·2) also comes close to being sharp for large λ. There is, however, no reason to suppose that (2·2) could not be improved under added restrictions on the p_i's; for instance, Romanowska's [6] inequality for all p_i's equal to p is sharper than (2·2) when $\frac{1}{2}(1-p)^{-\frac{1}{2}} < 1-e^{-\lambda}$.

As a complement to Theorem 1, we prove the following result: note that

$$(1 \wedge \lambda^{-1}) \geqslant \lambda^{-1}(1-e^{-\lambda}).$$

THEOREM 2. *Under the hypotheses set out in the Introduction,*

$$d(\mathscr{L}(W), P_\lambda) \geqslant \tfrac{1}{32}(1 \wedge \lambda^{-1}) \sum_{j=1}^{n} p_j^2. \tag{2·7}$$

Proof. Take x defined by

$$x(m) = (m-\lambda)\, e^{-(m-\lambda)^2/\theta\lambda}, \quad m \geqslant 0, \tag{2·8}$$

in (2·1), where the constant θ will be chosen later. Since, for P a Poisson variate with mean λ, $E\{\lambda x(P+1) - Px(P)\} = 0$, Equation (2·1) yields the equation

$$E\{[\lambda x(W+1) - Wx(W)] - [\lambda x(P+1) - Px(P)]\} = \sum_j p_j^2 E\{x(W_j+2) - x(W_j+1)\},$$

from which it follows that

$$2d(\mathscr{L}(W), P_\lambda) \sup_j |\lambda x(j+1) - jx(j)| \geqslant \sum_j p_j^2 E\{x(W_j+2) - x(W_j+1)\}. \tag{2·9}$$

Our first task is to bound the supremum on the left in (2·9). Since

$$(d/dw)\,(we^{-w^2/\theta\lambda}) = (1-2w^2/\theta\lambda)\, e^{-w^2/\theta\lambda},$$

which takes only values in the interval $[-2e^{-\frac{3}{2}}, 1]$, then

$$-2e^{-\frac{3}{2}} \leqslant x(w+1) - x(w) \leqslant 1, \tag{2·10}$$

and so

$$\begin{aligned} |\lambda x(j+1) - jx(j)| &= |\lambda\{x(j+1) - x(j)\} - (j-\lambda)^2 \exp\{-(j-\lambda)^2/\theta\lambda\}| \\ &\leqslant \lambda \max(1, 2e^{-\frac{3}{2}} + \theta e^{-1}). \end{aligned} \tag{2·11}$$

Next we treat the series on the right hand side of (2·9). Now,

$$1 - e^{-w^2/\theta\lambda}(1-2w^2/\theta\lambda) \leqslant 3w^2/\theta\lambda,$$

whence, writing $U_j = W_j - \lambda$,

$$1 - \{x(W_j+2) - x(W_j+1)\} \leqslant \int_{U_j+1}^{U_j+2} (3w^2/\theta\lambda)\, dw = (\theta\lambda)^{-1}(3U_j^2 + 9U_j + 7). \tag{2·12}$$

Therefore

$$1 - E\{x(W_j+2) - x(W_j+1)\} \leqslant (\theta\lambda)^{-1}(3\lambda+7),$$

since $E(U_i^2) = \sum_{j \neq i} p_j(1-p_j) + p_i^2$ and $E(U_i) = -p_i$. Consequently,

$$\sum_j p_j^2 E\{x(W_j+2) - x(W_j+1)\} \geqslant \sum_j p_j^2\{1 - (\theta\lambda)^{-1}(3\lambda+7)\}. \tag{2·13}$$

Combining (2·9), (2·11) and (2·13), we see that if $\theta \geqslant e$, $d(\mathscr{L}(W), P_\lambda) \geqslant k \sum_{j=1}^n p_j^2$, where

$$k = \{1 - (\theta\lambda)^{-1}(3\lambda+7)\}/2\lambda(2e^{-\frac{3}{2}} + \theta e^{-1})\}. \tag{2·14}$$

If $\lambda \geqslant 1$ we take $\theta = 21$, which gives

$$\lambda k \geqslant (1-10/\theta)/2(2e^{-\frac{3}{2}} + \theta e^{-1}) \geqslant 1/32,$$

while if $\lambda < 1$ we take $\theta = 21/\lambda$, obtaining

$$k \geqslant (1-10/\theta\lambda)/2(2e^{-\frac{3}{2}}+\theta\lambda e^{-1}) \geqslant 1/32.$$

Theorem 2 follows. ∎

3. *Second-order estimates*

For many choices of $(p_i)_{i=1}^n$, the bounds given in (2·2) and (2·7) can usefully be replaced by an estimate of $d(\mathscr{L}(W), P_\lambda)$ together with a bound on the error of the estimate. This approach was considered by Prohorov [5], Kerstan [3] and Chen [2]: our argument is similar to Chen's, though the error estimate is improved. The method is to take, for any $A \subseteq \mathbb{Z}^+$, the function x defined in (2·3), and then to approximate the right hand side of (2·1) by $(\sum_{j=1}^n p_j^2) E\{x(P+2)-x(P+1)\}$, where P, here and subsequently, denotes a Poisson variate with mean λ. Let $\Delta(A)$ denote the error in this approximation:

$$\Delta(A) \equiv P[W\in A]-P_\lambda(A)-\left(\sum_{j=1}^n p_j^2\right) E\{x(P+2)-x(P+1)\}$$

$$= P[W\in A]-P_\lambda(A)+\tfrac{1}{2}\lambda^{-2}\left(\sum_{j=1}^n p_j^2\right) E\{I[P\in A](P^2-(2\lambda+1)P+\lambda^2)\}. \quad (3·1)$$

THEOREM 3. *For any $A \subseteq \mathbb{Z}^+$, under the hypotheses set out in the Introduction,*

$$|\Delta(A)| \leqslant 2\lambda^{-1}(1-e^{-\lambda})(1\wedge 1·4\lambda^{-\frac{1}{2}})\sum_{j=1}^n p_j^3+2\{\lambda^{-1}(1-e^{-\lambda})\sum_{j=1}^n p_j^2\}^2. \quad (3·2)$$

Proof. Taking in (2·1) the function x defined in (2·3), it follows immediately that

$$|\Delta(A)| \leqslant 2\lambda^{-1}(1-e^{-\lambda})\sum_{j=1}^n p_j^2 d(\mathscr{L}(W_j), P_\lambda).$$

However, analogously to (2·1), for any $B \subseteq \mathbb{Z}^+$,

$$P[W_j\in B]-P_\lambda(B) = E\{\lambda x_{\lambda,B}(W_j+1)-W_j x_{\lambda,B}(W_j)\}$$
$$= p_j E x_{\lambda,B}(W_j+1)+\sum_{k\neq j} p_k^2 E\{x_{\lambda,B}(W_{jk}+2)-x_{\lambda,B}(W_{jk}+1)\},$$

where $W_{jk} = W_j - X_k$. Thus it follows from (2·5) and (2·6) that

$$d(\mathscr{L}(W_j), P_\lambda) \leqslant p_j(1\wedge 1·4\lambda^{-\frac{1}{2}})+\lambda^{-1}(1-e^{-\lambda})\sum_{k\neq j} p_k^2, \quad (3·3)$$

establishing (3·2). ∎

Remark. Since, by Schwarz's inequality, $\lambda^{-1}(\sum_{j=1}^n p_j^2)^2 \leqslant \sum_{j=1}^n p_j^3$, equation (3·2) implies that $|\Delta(A)| \leqslant 4(1-e^{-\lambda})(1-\frac{1}{2}e^{-\lambda})\lambda^{-1}\sum_{j=1}^n p_j^3$, which improves on Chen's estimate of $(12+48\sqrt{2})\lambda^{-1}\sum_{j=1}^n p_j^3$.

COROLLARY. *Let* $\delta(\lambda) \equiv -\frac{1}{2}\lambda^{-1}E\{(P^2-(2\lambda+1)P+\lambda^2)^-\}$.
Then

$$\left|d(\mathscr{L}(W), P_\lambda)-\lambda^{-1}\left(\sum_{j=1}^n p_j^2\right)\delta(\lambda)\right| \leqslant 2\lambda^{-1}(1-e^{-\lambda})(1\wedge 1·4\lambda^{-\frac{1}{2}})\sum_{j=1}^n p_j^3 + 2\left\{\lambda^{-1}(1-e^{-\lambda})\sum_{j=1}^n p_j^2\right\}^2.$$

Remark. Kerstan gives instead the upper bound

$$0·65\lambda^{-1}\sum_{j=1}^n p_j^3+1·95\left\{\lambda^{-1}\sum_{j=1}^n p_j^2\right\}^2, \quad \text{if } \max_{1\leqslant j\leqslant n} p_j \leqslant \tfrac{1}{4},$$

which is sometimes better than that of the Corollary, and sometimes worse.

The quantity $\delta(\lambda)$ is not in general very neatly expressible, except for moderately small values of λ: for example,

$$\delta(\lambda) = \begin{cases} \lambda(1-\frac{1}{2}\lambda)\,e^{-\lambda}, & 0 \leqslant \lambda \leqslant 2-\sqrt{2} \\ \lambda\left(\frac{1}{2}+\frac{\lambda}{2}-\frac{\lambda^2}{4}\right)e^{-\lambda}, & 2-\sqrt{2} \leqslant \lambda \leqslant 3-\sqrt{3}. \end{cases}$$

However, $y^2-(2\lambda+1)\,y+\lambda^2 \geqslant -(\lambda+\frac{1}{4})$ for all y, and so

$$\delta(\lambda) \leqslant \frac{1}{2}+\frac{1}{8\lambda},$$

for all λ. Furthermore, as $\lambda \to \infty$, $\delta(\lambda) \sim 0{\cdot}242$.

The Corollary enables some further evaluation of the relative precision of the bounds in Theorems 1 and 2 to be made, in the following sense. Suppose that $\{X_{in}, 1 \leqslant i \leqslant n < \infty\}$ is a double array of Bernoulli random variables, independent within rows, and set $p_{jn} = P[X_{jn} = 1]$, $\lambda_n = \sum_{j=1}^n p_{jn}$ and $W_n = \sum_{j=1}^n X_{jn}$. Suppose also that as $n \to \infty$, $d(\mathscr{L}(W_n), P_{\lambda_n}) \to 0$, or, equivalently, that $\lambda_n^{-1}(1-e^{-\lambda_n})\sum_{j=1}^n p_{jn}^2 \to 0$. Then the error estimate given in the Corollary is of asymptotically smaller order as $n \to \infty$ than $\lambda_n^{-1}\delta(\lambda_n)\sum_{j=1}^n p_{jn}^2$, so that

$$d(\mathscr{L}(W_n), P_{\lambda_n}) \sim \lambda_n^{-1}\delta(\lambda_n)\sum_{j=1}^n p_{jn}^2,$$

as $n \to \infty$. Thus, for example, if also $\lambda_n \to \infty$,

$$u_n \equiv \lambda_n d(\mathscr{L}(W_n), P_{\lambda_n})/\{(1-e^{-\lambda_n})\sum_{j=1}^n p_{jn}^2\} \to 0{\cdot}242,$$

whereas Theorems 1 and 2 guarantee that $\frac{1}{32} \leqslant u_n \leqslant 1$. Actually, the choice of $\theta = 6{\cdot}55$ in (2·14) gives $u_n \geqslant \frac{1}{11}$, provided that λ is sufficiently large, which is not too great a deviation from 0·242. Similar comparisons when $\lambda_n \leqslant 2-\sqrt{2}$ yield

$$u_n \sim \lambda_n(1-e^{-\lambda_n})^{-1}(1-\tfrac{1}{2}\lambda_n)\,e^{-\lambda_n},$$

and so $u_n \to 1$ if $\lambda_n \to 0$. Again, for small enough λ, the choice of $\theta(= 14{\cdot}6)$ in (2·14) improves the bound given by Theorem 2, but this time to $u_n \geqslant \frac{1}{23}$.

4. *Non-negative integer variates*

Let $(Y_j)_{j=1}^n$ be independent non-negative integer valued random variables, and let $p_j = P[Y_j = 1]$, $q_j = P[Y_j \geqslant 2]$, $\lambda = \sum_{j=1}^n p_j$ and $V = \sum_{j=1}^n Y_j$. Define the zero-one random variables $(X_j)_{j=1}^n$ by

$$X_j = \begin{cases} Y_j \text{ if } Y_j = 0 \text{ or } 1; \\ 0 \text{ otherwise,} \end{cases}$$

and set $W = \sum_{j=1}^n X_j$. Note that the X_j's satisfy the conditions outlined in the Introduction. Then, as observed by Serfling [7],

$$d(\mathscr{L}(V), \mathscr{L}(W)) \leqslant \sum_{j=1}^n q_j,$$

and so (2·2) and (2·7) yield the inequalities

$$d(\mathscr{L}(V), P_\lambda) \leqslant \lambda^{-1}(1-e^{-\lambda})\sum_{j=1}^n p_j^2 + \sum_{j=1}^n q_j, \tag{4·1}$$

and

$$d(\mathscr{L}(V), P_\lambda) + \sum_{j=1}^n q_j \geqslant \tfrac{1}{32}(1 \wedge \lambda^{-1})\sum_{j=1}^n p_j^2. \tag{4·2}$$

If the Y_j's have finite second moments, the Stein–Chen method can be applied to get alternatives to (4·1) and (4·2). Let $\nu_j = EY_j$, $\kappa_j = E(Y_j(Y_j-1))$, $\mu_j = E\{Y_j I[Y_j \geq 2]\}$ and $\nu = \sum_{j=1}^n \nu_j$.

THEOREM 4. *If* $(Y_j)_{j=1}^n$ *satisfy the above hypotheses, and if* $\nu < \infty$,

$$d(\mathscr{L}(V), P_\nu) \leq \nu^{-1}(1-e^{-\nu}) \sum_{i=1}^n \nu_i p_i + \sum_{i=1}^n \nu_i[\{2(1 \wedge 1{\cdot}4\nu^{-\frac{1}{2}})\, q_i\} \wedge \{\nu^{-1}(1-e^{-\nu})\, \mu_i\}] + [\{2(1 \wedge 1{\cdot}4\nu^{-\frac{1}{2}}) \sum_{i=1}^n \mu_i\} \wedge \{\nu^{-1}(1-e^{-\nu}) \sum_{i=1}^n \kappa_i\}] \tag{4·3}$$

$$\leq \nu^{-1}(1-e^{-\nu}) \left\{\sum_{i=1}^n \nu_i^2 + \sum_{i=1}^n \kappa_i\right\}, \tag{4·4}$$

and

$$d(\mathscr{L}(V), P_\lambda) \leq \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n p_i^2 + \sum_{i=1}^n p_i[\{2(1 \wedge 1{\cdot}4\lambda^{-\frac{1}{2}})\, q_i\} \wedge \{\lambda^{-1}(1-e^{-\lambda})\, \mu_i\}] + (1 \wedge 1{\cdot}4\lambda^{-\frac{1}{2}}) \sum_{i=1}^n \mu_i + [\{2(1 \wedge 1{\cdot}4\lambda^{-\frac{1}{2}}) \sum_{i=1}^n \mu_i\} \wedge \{\lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^n \kappa_i\}] \tag{4·5}$$

$$\leq \lambda^{-1}(1-e^{-\lambda}) \left(\sum_{i=1}^n \nu_i p_i + \sum_{i=1}^n \kappa_i\right) + (1 \wedge 1{\cdot}4\lambda^{-\frac{1}{2}}) \sum_{i=1}^n \mu_i. \tag{4·6}$$

Remark. Estimate (4·1) is clearly better than (4·5) if $\nu = \infty$. On the other hand, for large λ, (4·5) can improve upon (4·1), in circumstances where $\sum_{j=1}^n q_j$ is not small but $\lambda^{-\frac{1}{2}}\sum_{j=1}^n \mu_j$ is: the latter condition is natural, in that it is simply requiring that the expected contribution to V from Y's not taking the values 0 or 1 should be small when compared to the spread $\lambda^{\frac{1}{2}}$ of P_λ. If the Y_j's take only the values 0, 1 and 2, estimate (4·4) reduces to $\nu^{-1}(1-e^{-\nu})\{\sum_{i=1}^n \nu_i^2 + 2\sum_{i=1}^n q_i\}$, and estimate (4·6) to

$$\lambda^{-1}(1-e^{-\lambda}) \left\{\sum_{i=1}^n p_i^2 + 2\sum_{i=1}^n q_i(1+p_i)\right\} + 2(1 \wedge 1{\cdot}4\lambda^{-\frac{1}{2}}) \sum_{i=1}^n q_i,$$

enabling comparison with (4·1) to be easily made.

Estimate (4·4) is typically smaller than (4·6), because $\nu^{-1}(1-e^{-\nu}) \leq \lambda^{-1}(1-e^{-\lambda})$ and, usually, $\nu^{-1}\sum_{i=1}^n \mu_i \nu_i \leq (1 \wedge 1{\cdot}4\lambda^{-\frac{1}{2}})\sum_{i=1}^n \mu_i$. However, in view of (4·1), this does not necessarily imply that it is better to use P_ν than P_λ to approximate the distribution of V.

Proof. Pick any $A \subseteq \mathbb{Z}^+$, set $V_j = V - Y_j$, define x as in (2·3) but with ν for λ, and observe that

$$\begin{aligned} P[V \in A] - P_\nu(A) &= \sum_{j=1}^n \nu_j E\{x(V+1) - x(V_j+1)\} + \sum_{j=1}^n E\{Y_j(x(V_j+1) - x(V))\} \\ &= \sum_{j=1}^n \nu_j p_j E\{x(V_j+2) - x(V_j+1)\} \\ &\quad + \sum_{j=1}^n \nu_j E\{(x(V+1) - x(V_j+1))\, I[Y_j \geq 2]\} \\ &\quad + \sum_{j=1}^n E\{Y_j(x(V_j+1) - x(V))\}. \end{aligned}$$

The three terms are now estimated using (2·5) and (2·6), again with ν for λ, giving (4·3).

The proof of (4·5) is similar, starting from the equation

$$P[V\in A]-P_\lambda(A) = \sum_{j=1}^{n} p_j E\{x(V+1)-x(V_j+1)\}$$
$$+\sum_{j=1}^{n} E\{Y_j(x(V_j+1)-x(V))\}$$
$$-\sum_{j=1}^{n} E\{(Y_j-p_j)\,x(V_j+1)\}. \blacksquare$$

It is also possible to adapt the proof of Theorem 2 so as to get lower bounds for $d(\mathscr{L}(V), P_\nu)$ and $d(\mathscr{L}(V), P_\lambda)$, to contrast with (4·2). The following Theorem establishes such a result, without, however, retaining the elegance of (2·7) or (4·2).

THEOREM 5. *If the hypotheses of Theorem 4 are satisfied, and if, in addition,* $\kappa_j < \infty$, $1 \leqslant j \leqslant n$, *then*

$$d(\mathscr{L}(V),P_\nu)+\tfrac{1}{2}\nu^{-1}\{2e^{-\frac{3}{2}}+21e^{-1}(1\wedge\nu)^{-1}\}^{-1}\times\Bigg\{2e^{-\frac{3}{2}}\sum_{j=1}^{n}\mu_j\nu_j+\sum_{j=1}^{n}\kappa_j+\tfrac{1}{7}(1\wedge\nu^{-1})$$
$$\times\left[\left(\sum_{j=1}^{n}\nu_jp_j\right)\sum_{j=1}^{n}(\kappa_j-\nu_j^2)+2\sum_{j=1}^{n}p_j\nu_j^2(\nu_j-2)\right]\Bigg\}\geqslant\tfrac{1}{32}(1\wedge\nu^{-1})\sum_{j=1}^{n}\nu_jp_j, \quad (4·7)$$

and

$$d(\mathscr{L}(V),P_\lambda)+\tfrac{1}{2}\lambda^{-1}\{2e^{-\frac{3}{2}}+21e^{-1}(1\wedge\lambda)^{-1}\}^{-1}\times\Bigg\{2e^{-\frac{3}{2}}\sum_{j=1}^{n}\mu_jp_j+\sum_{j=1}^{n}\kappa_j$$
$$+(21/2e)^{\frac{1}{2}}(\lambda\vee 1)^{\frac{1}{2}}\sum_{j=1}^{n}\mu_j+\tfrac{1}{7}(1\wedge\lambda^{-1})\left[\left(\sum_{j=1}^{n}p_j^2\right)\sum_{j=1}^{n}(\kappa_j-\nu_j^2)\right.$$
$$\left.+2\sum_{j=1}^{n}p_j^2\nu_j(\nu_j-2)\right]\Bigg\}\geqslant\tfrac{1}{32}(1\wedge\lambda^{-1})\sum_{j=1}^{n}p_j^2. \quad (4·8)$$

Remark. If the Y_j's are in fact 0–1 random variables, the complicated additional term on the left hand side of (4·7) is negative: it only becomes important when the Y_j's are too far from being 0–1 variates. Some such term has to be present, since, if each Y_j is a Poisson variate, $d(\mathscr{L}(V), P_\nu) = 0$, whereas the right hand side of (4·7) is positive.

Proof. Take x as defined in (2·8), but with ν for λ, and deduce, from the proofs of Theorems 2 and 4, that, for $\theta \geqslant e$,

$$2d(\mathscr{L}(V),P_\nu)\,\nu(2e^{-\frac{3}{2}}+\theta e^{-1})\geqslant\sum_{j=1}^{n}\nu_jp_jE\{x(V_j+2)-x(V_j+1)\}$$
$$+\sum_{j=1}^{n}\nu_jE\{(x(V+1)-x(V_j+1))\,I[Y_j\geqslant 2]\}+\sum_{j=1}^{n}E\{Y_j(x(V_j+1)-x(V))\}.$$

Thus, from the argument leading to (2·10), it follows that

$$2d(\mathscr{L}(V),P_\nu)\,\nu(2e^{-\frac{3}{2}}+\theta e^{-1})+2e^{-\frac{3}{2}}\sum_{j=1}^{n}\nu_j\mu_j+\sum_{j=1}^{n}\kappa_j$$
$$\geqslant\sum_{j=1}^{n}\nu_jp_jE\{x(V_j+2)-x(V_j+1)\}$$
$$\geqslant\sum_{j=1}^{n}\nu_jp_j\{1-(\theta\nu)^{-1}(3E(V_j-\nu)^2+9E(V_j-\nu)+7)\}, \quad (4·9)$$

where the last line is a consequence of (2·12). Evaluating the moments of $V_j - \nu$ enables the right hand side of (4·9) to be estimated as no smaller than

$$\sum_{j=1}^{n} \nu_j p_j\{1-(\theta\nu)^{-1}(3\nu+7)\} - 3(\theta\nu)^{-1}\sum_{j=1}^{n}\nu_j p_j\left\{\sum_{j=1}^{n}(\kappa_j-\nu_j^2)+2\nu_j(\nu_j-2)\right\},$$

and the proof of (4·7) is concluded in the same way as the proof of Theorem 2, taking $\theta = 21(1\wedge\nu)^{-1}$. The upper bound (4·8) is proved in a similar way. ∎

Some simplification of (4·7) and (4·8) can often be achieved. The next Corollary, which follows directly from (4·7), illustrates the possibilities. The inequality obtained reduces to (2·7) when the Y_j's are 0–1 random variables.

COROLLARY. *If, in addition to the conditions of Theorem* 5, $\nu_j \leqslant 2$ *for all* j,

$$d(\mathscr{L}(V),P_\nu)+\frac{3}{100}\left\{\sum_{j=1}^{n}\mu_j\nu_j+\frac{9}{4}\left(1+\frac{1}{7}\sum_{j=1}^{n}p_j\nu_j\right)\sum_{j=1}^{n}\kappa_j\right\}\geqslant\frac{1}{32}\sum_{j=1}^{n}\nu_j p_j \quad \text{if} \quad \nu<1 \tag{4·10}$$

and

$$d(\mathscr{L}(V),P_\nu)+\frac{11\nu^{-1}}{400}\left\{\sum_{j=1}^{n}\mu_j\nu_j+\frac{9}{4}\left(1+\frac{\nu^{-1}}{7}\sum_{j=1}^{n}p_j\nu_j\right)\sum_{j=1}^{n}\kappa_j\right\}\geqslant\frac{\nu^{-1}}{32}\sum_{j=1}^{n}\nu_j p_j \quad \text{if} \quad \nu\geqslant 1. \tag{4·11}$$

The first author gratefully acknowledges the help afforded by a Visiting Fellowship at the Australian National University, during which this work was begun.

REFERENCES

[1] A. D. BARBOUR and G. K. EAGLESON. Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.* **15** (1983), 585–600.

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

[3] J. KERSTAN. Verallgemeinerung eines Satzes von Procharov und Le Cam. *Z. Wahrsch. Verw. Gebiete* **2** (1964), 173–179.

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

[5] J. V. PROHOROV. Asymptotic behaviour of the binomial distribution. *Uspekhi Mat. Nauk* **8** (1953), 135–142.

[6] M. ROMANOWSKA. A note on the upper bound for the distance in total variation between the binomial and the Poisson distribution. *Statist. Neerlandica* **31** (1977), 127–130.

[7] R. J. SERFLING. A general Poisson approximation theorem. *Ann. Probab.* **3** (1975), 726–731.

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