An Improved Poisson Limit Theorem for Sums of Dissociated Random Variables

A. D. Barbour; G. K. Eagleson

Journal of Applied Probability, Vol. 24, No. 3. (Sep., 1987), pp. 586-599.

J. Appl. Prob. **24**, 586–599 (1987)
Printed in Israel

AN IMPROVED POISSON LIMIT THEOREM FOR SUMS OF DISSOCIATED RANDOM VARIABLES

A. D. BARBOUR,* *Universität Zürich*
G. K. EAGLESON,** *CSIRO Division of Mathematics and Statistics*

Abstract

A Poisson limit theorem for sums of dissociated 0–1 random variables is refined by deriving the first terms in an asymptotic expansion. The most natural refinement does not remove all the first-order error in a number of applications to tests of clustering, and a further approximation is derived which gives excellent results in practice. The proofs are based on the technique of Stein and Chen.

CORRELATIONS; INTERPOINT DISTANCES; POISSON APPROXIMATIONS; RATES OF CONVERGENCE

1. Introduction

In a recent paper, Barbour and Eagleson (1985) established a general Poisson approximation theorem (together with an error estimate) for sums of dissociated 0–1 random variables, using a method of proving distributional limit theorems originated by Stein (1970) and adapted to the Poisson context by Chen (1975); the results improve upon the earlier conclusions of Brown and Silverman (1979). However, Stein's technique is also applicable to the derivation of higher-order asymptotic estimates to accompany limit theorems, and it is sensible to ask whether it is possible to improve upon the Poisson approximation by including further corrections of this type. In this paper, two such refined approximations are established. The first of these, whilst the most natural in the context of Stein's method, turns out to be insufficiently sharp when analysing close interpoint distances and a second approximation is derived in order to overcome this problem. The corrections appear, when compared to simulations of interpoint distances on a hypersphere (as in the analysis of correlation coefficients), to be practically as well as theoretically useful.

Received 28 August 1984; revision received 7 August 1986.
* Postal address: Institut für Angewandte Mathematik, Universität Zürich, CH-8001, Zürich, Switzerland.
** Postal address: CSIRO Division of Mathematics and Statistics, P.O. Box 218, Lindfield, NSW 2070, Australia.

The notation and the results are set out in the next section, some examples are discussed in Section 3 and the proofs are presented in Section 4.

2. Statement of results

Let N be a collection of subsets of size m, drawn from the natural numbers, and suppose that, for each $J \in N$, the random variable X_J can take only the values 0 or 1. The random variables $\{X_J;\ J \in N\}$ are said to be *dissociated* (McGinley and Sibson (1975)) if collections of them which have no index in common are independent. That is, $\{X_J; J \in \mathscr{J}\}$ is independent of $\{X_K; K \in \mathscr{K}\}$ whenever $\mathscr{J} \cap \mathscr{K} = \varnothing$. As an example of dissociated random variables, consider a sequence $\{Y_1, Y_2, \cdots\}$ of independent but not necessarily identically distributed random variables and a symmetric 0–1 function of m variables, ϕ. When $J = \{i_1, i_2, \cdots, i_m\}$, define

$$X_J = \phi(Y_{i_1}, \cdots, Y_{i_m}).$$

The $\{X_J\}$ are clearly dissociated.

Let $N_J = \{K \in N: K \cap J \neq \varnothing\}$ and $N_J^- = N_J \setminus \{J\}$, so that $\{X_K;\ K \in N_J\}$ are those random variables which could depend on X_J. Set

$$p_J = EX_J, \quad \beta_{JK} = EX_JX_K, \quad \gamma_{JKL} = EX_JX_KX_L.$$

The statistics which will be considered are of the form

$$T = \sum_{J \in N} X_J$$

and we shall denote

$$ET = \sum_{J \in N} p_J := \lambda.$$

Barbour and Eagleson (1985) showed that, if P_λ denotes a Poisson distribution with parameter λ, then

$$\sup_{A \subset Z^+} |P(T \in A) - P_\lambda(A)| \tag{2.1}$$
$$\leqq \min(1, \lambda^{-1}) \sum_{J \in N} \left\{ p_J^2 + \sum_{K \in N_J^-} (p_J p_K + \beta_{JK}) \right\}.$$

To state our first theorem, let

$$C_J := p_J^2 + \sum_{K \in N_J^-} (p_J p_K + \beta_{JK})$$

and define

$$a_\lambda := \min(1, 1.4\lambda^{-1/2}) \tag{2.2}$$

and

(2.3) $$b_\lambda := \min(1, \lambda^{-1}).$$

Let

$$\delta_1 := \delta + 3b_\lambda \sum_{j\in N} \sum_{K\neq L\in N_j^-} \gamma_{JKL},$$

where

$$\delta = 2\left\{b_\lambda \sum_{j\in N} C_J\right\}^2 + 4a_\lambda b_\lambda \sum_{J\in N} C_J\left(\sum_{K\in N_J} p_K\right)$$

(2.4)

$$+ b_\lambda\left\{\sum_{J\in N}\left[\sum_{K\neq L\in N_J} p_J\beta_{KL} + 2\left(\sum_{K\in N_J^-} p_K\right)\sum_{L\in N_J^-}\beta_{JL}\right]\right\},$$

and let

$$\kappa_2 := -\sum_J\left(p_J^2 + \sum_{K\in N_J^-}(p_Jp_K - \beta_{JK})\right).$$

Denote the nth Charlier polynomial by

$$C_n(\lambda; t) := \sum_{r=0}^{n}\binom{n}{r}(-1)^{n-r}\lambda^{-r}\prod_{m=0}^{r-1}(t-m)$$

and define a measure, $P_\lambda^{(1)}$, on Z^+ by assigning mass

$$\frac{e^{-\lambda}\lambda^j}{j!}\{1 + \tfrac{1}{2}\kappa_2 C_2(\lambda; j)\}$$

to the point j.

Theorem 1. With the above notation

$$\sup_{A\subset Z^+} |P(T\in A) - P_\lambda^{(1)}(A)| \leqq \delta_1.$$

Remarks 1. If the $\{X_J;\ J\in N\}$ are completely independent, the above result simplifies as $\beta_{JK} = p_Jp_K$ and $\gamma_{JKL} = p_Jp_Kp_L$. Let P denote a random variable with distribution P_λ. In the special case of $m = 1$, Theorem 1 reduces to

$$\sup_{A\subset Z^+}\left|P(T\in A) - P_\lambda(A) + \tfrac{1}{2}\lambda^{-2}\left(\sum_{j=1}^{n} p_j^2\right)E\{I(P\in A)(P^2 - (2\lambda+1)P + \lambda^2)\}\right|$$

$$\leqq 8\left\{b_\lambda \sum_{j=1}^{n} p_j^2\right\}^2 + 4a_\lambda b_\lambda \sum_{j=1}^{n} p_j^3,$$

which is of the same form as the result obtained by Barbour and Hall (1984) for this case (see their Theorem 3).

2. Theorem 1, applied to the binomial distribution, $\mathrm{B}(n, \lambda n^{-1})$, gives

$$P_\lambda(A) - \tfrac{1}{2}n^{-1}E\{I(P\in A)(P^2 - (2\lambda+1)P + \lambda^2)\}$$
$$= P(\mathrm{B}(n, \lambda n^{-1})\in A) + O(n^{-2}).$$

Thus, Theorem 1 shows that

$$P[T\in A]=P[\mathrm{B}(n,\lambda n^{-1})\in A]+O(n^{-2},\delta_1)$$

where $\lambda=\Sigma_J p_J$ as usual, and $n=-\lambda^2/\kappa_2$, provided that $\kappa_2<0$. For $\kappa_2>0$, there is an analogous representation using the negative binomial,

$$P[T\in A]=P[\mathrm{NB}(k,q)\in A]+O(n^{-2},\delta_1)$$

where $k=\lambda^2/\kappa_2$ and $k(1-q)/q=\lambda$.

3. For the special sets $A=\{j:j\geqq s\}$, the approximation in Theorem 1 can be given explicitly as

$$P_\lambda^{(1)}=P_\lambda(A)-\tfrac{1}{2}\kappa_2 e^{-\lambda}\lambda^{s-2}\{\lambda-s+1\}/(s-1)!$$

A typical example in which Theorem 1 would be useful is in testing for tendencies towards periodicity in a collection of points distributed in the plane, where one would expect an unusually large number of interpoint distances that were, within a certain tolerance, integer multiples of a given unit of distance. The distribution of the number of such interpoint distances, on the null hypothesis of a random distribution of points, could be estimated using Theorem 1, provided that the tolerance was chosen so that the expected number thereof was of only moderate size. A similar problem, this time on the line, arises in epidemiology when assessing evidence for a latent period of infection from data on the time of presentation of cases of a disease, and the number of pairs of cases whose times of presentation differ by a certain interval is of interest (Mantel and Bailar (1970)). However, in the more usual problem, in which evidence for clustering in a distribution of points is adduced from abnormally large numbers of close pairs, a more subtle analysis is required. This is typically because, if points Y_1 and Y_3 are close to Y_2, then Y_1 is also close to Y_3. So when $J=\{i,j\}$, $K=\{i,k\}$ and $L=\{j,k\}$ the quantity γ_{JKL} can be of the same order of magnitude as β_{JK}, in contrast to the first two examples, where it is of smaller order. So when considering close pairs, these terms must be included in the correction to obtain a sensible approximation. For simplicity, we do this only for the case $m=2$. When $J=\{i,j\}$, $K=\{i,k\}$ denote $\{j,k\}$ by (JK) and let

$$\delta_2:=\delta+8b_\lambda\sum_{J\in N}\sum_{K\in N_J^-}\sum_{\substack{L\in N_J^-\\ L\neq K,(JK)}}\gamma_{JKL}$$
$$+2b_\lambda\sum_{J\in N}\sum_{K\in N_J^-}\beta_{JK}\left\{2a_\lambda\sum_{L\in N_J}p_L+b_\lambda\sum_{L\in N}p_L\right\}$$
$$\kappa_3:=\sum_{J\in N}\sum_{K\in N_J^-}\gamma_{JK(JK)}.$$

Define a measure, $P_\lambda^{(2)}$, on Z^+ by assigning mass

$$\frac{e^{-\lambda}\lambda^j}{j!}\{1 + \tfrac{1}{2}\kappa_2 C_2(\lambda; j) + \tfrac{1}{6}\kappa_3 C_3(\lambda; j)\}$$

to the point j.

Theorem 2. When $m = 2$, and with the above notation,

$$\sup_{A \subset Z^+} |P(T \in A) - P_\lambda^{(2)}(A)| \leqq \delta_2.$$

Remarks 4. When β_{JK} and $\gamma_{JK(JK)}$ are of the same order of magnitude, the two correction terms are clearly of the same order, unless, for $K \in N_J^-$, X_J and X_K are (almost) uncorrelated, in which case the second correction term is the larger. Examples of the various possibilities are given in Section 3.

5. When $A = \{j: j \geqq s\}$, the approximation is given by

$$P_\lambda^{(2)}(A) = P_\lambda^{(1)}(A) + \tfrac{1}{6}\kappa_3 e^{-\lambda}\lambda^{s-3}\{(\lambda - s + 1)^2 - s + 1\}/(s-1)!$$

3. Examples

A natural example is given by considering the $\binom{n}{2}$ different distances between n points independently distributed over some space. A sensible statistic, useful for testing whether the points are random against a clustering alternative, is the number of 'small' interpoint distances. Here,

$$T = \sum_{1 \leqq i < j \leqq n} I(|\mathbf{Y}_i - \mathbf{Y}_j| \leqq \varepsilon),$$

where the $\mathbf{Y}_i$ are independent and identically distributed under the null hypothesis and ε is some suitably chosen measure of closeness, depending on n. The common distribution of the $\mathbf{Y}_i$ determines the correction factors given in Theorem 2. Because the $\{\mathbf{Y}_i\}$ are identically distributed, the values of p_J, β_{JK} and γ_{JKL} are unchanged by permutations of the indices $1 \leqq i \leqq n$.

If one has some information about the expected size of the clustering effect, this should be used to fix ε and hence to determine λ through

$$\lambda = ET = \binom{n}{2} P(|\mathbf{Y}_1 - \mathbf{Y}_2| \leqq \varepsilon). \tag{3.1}$$

A critical region for T of the form $[s, \infty]$ can then be chosen with associated size

$$P(T \geqq s) = \sum_{j=s}^{\infty} e^{-\lambda}\lambda^j/j! + C_\lambda, \tag{3.2}$$

where C_λ is the correction term from Theorem 2. Alternatively, the critical value s and a target size α can be used to determine a value of λ through the equation

$$\alpha = \sum_{j=s}^{\infty} e^{-\lambda}\lambda^j/j! \tag{3.3}$$

and hence a value of ε: a closer approximation to the true size is then given, by (3.2), as $\alpha + C_\lambda$.

To obtain an explicit expression for C_λ, consider n points distributed independently over an open subset Ω of R^k, according to a distribution with a bounded continuous density function f. The probability that a pair of points is separated by a distance less than ε is

$$p_\varepsilon := \int_{\Omega^2} f(\mathbf{x}) f(\mathbf{y}) I[|\mathbf{x}-\mathbf{y}| < \varepsilon] d\mathbf{x} d\mathbf{y}.$$

When ε, and so p_ε also, is small, changing variables by setting $\mathbf{y} = \mathbf{x} + \varepsilon \mathbf{z}$ gives, by dominated convergence,

$$(3.4)\quad p_\varepsilon = \varepsilon^k \int_\Omega f(\mathbf{x}) d\mathbf{x} \left\{ \int_{\Omega_\varepsilon(\mathbf{x})} f(\mathbf{x} + \varepsilon \mathbf{z}) I[|\mathbf{z}| < 1] d\mathbf{z} \right\} = \varepsilon^k V_k \mu_1 + o(\varepsilon^k),$$

where V_k is the volume of the k-dimensional unit sphere, $\mu_r = \int_\Omega f^{r+1}(\mathbf{x}) d\mathbf{x}$ and $\Omega_\varepsilon(\mathbf{x}) = \{\mathbf{z}: \mathbf{x} + \varepsilon \mathbf{z} \in \Omega\}$. If α and s are fixed, they determine λ through (3.3), and the equation

$$(3.5)\qquad p_\varepsilon = \lambda \Big/ \binom{n}{2}$$

determines ε. It follows that, if n is large, $p_J = p_\varepsilon = O(n^{-2})$ and the corresponding ε is $O(n^{-2/k})$.

In order to compute the finer approximation $\alpha + C_\lambda$ to the size of test obtained, note that $\Omega_\varepsilon(\mathbf{x}) \to R^k$ as $\varepsilon \to 0$, and observe that

$$\begin{aligned}
\beta_\varepsilon := \beta_{\{1,2\},\{1,3\}} &= \int_{\Omega^3} f(\mathbf{x}) f(\mathbf{y}) f(\mathbf{z}) I[|\mathbf{x}-\mathbf{y}| < \varepsilon] I[|\mathbf{x}-\mathbf{z}| < \varepsilon] d\mathbf{x} d\mathbf{y} d\mathbf{z} \\
&= \varepsilon^{2k} \int_\Omega f(\mathbf{x}) \left\{ \int_{\Omega_\varepsilon(\mathbf{x})} f(\mathbf{x} + \varepsilon \mathbf{z}) I[|\mathbf{z}| < 1] d\mathbf{z} \right\}^2 d\mathbf{x} \\
&= \varepsilon^{2k} V_k^2 \mu_2 + o(\varepsilon^{2k}) \\
&= O(n^{-4}),
\end{aligned}$$

again by dominated convergence, and

$$\begin{aligned}
\gamma_3 &:= \gamma_{\{1,2\},\{1,3\},\{2,3\}} \\
&= \int_{\Omega^3} f(\mathbf{x}) f(\mathbf{y}) f(\mathbf{z}) I[|\mathbf{x}-\mathbf{y}| < \varepsilon] I[|\mathbf{x}-\mathbf{z}| < \varepsilon] I[|\mathbf{y}-\mathbf{z}| < \varepsilon] d\mathbf{x} d\mathbf{y} d\mathbf{z} \\
&= \varepsilon^{2k} \int_\Omega f(\mathbf{x}) d\mathbf{x} \left\{ \int_{\Omega_\varepsilon^2(\mathbf{x})} f(\mathbf{x} + \varepsilon \mathbf{u}) f(\mathbf{x} + \varepsilon \mathbf{w}) I[|\mathbf{u}| < 1] I[|\mathbf{w}| < 1] \right. \\
&\qquad\qquad \left. \times I[|\mathbf{w} - \mathbf{u}| < 1] d\mathbf{u} d\mathbf{w} \right\} \\
&= \varepsilon^{2k} W_k \mu_2 + o(\varepsilon^{2k}) \\
&= O(n^{-4}),
\end{aligned}$$

where

$$W_k := \int_{\{|\boldsymbol{u}|<1,|\boldsymbol{w}|<1\}} I\{|\boldsymbol{u}-\boldsymbol{w}|<1\}d\boldsymbol{u}d\boldsymbol{w}.$$

It is also easily seen that all γ's based on four or more different points are of order n^{-6}. Hence if $N = D_n := \{(i,j);\ 1 \leqq i < j \leqq n\}$,

$$\sum_{J\in D_n} \sum_{K\in N_J^-} (\beta_{JK} - p_J p_K) = \binom{n}{2} 2(n-2)(\beta_\varepsilon - p_\varepsilon^2)$$
$$= 4\lambda^2 n^{-1}(\mu_2 - \mu_1^2)/\mu_1^2 + o(n^{-1}),$$

and $\mu_2 > \mu_1^2$ unless f is constant on Ω, which must then be a bounded subset of R^k. Thus the first correction term is of magnitude n^{-1}, unless the n points are *uniformly* distributed over Ω, when only edge effects make any non-zero contribution. For the second correction term,

$$\sum_{J\in D_n} \sum_{K\in N_J^-} E(X_J X_K X_{(JK)}) = \binom{n}{2} 2(n-2)\gamma_\varepsilon$$

is also of order n^{-1}.

Hence, from Theorem 2, we have the approximation

$$P[T \geqq s] = \alpha + \frac{1}{2}\binom{n}{2}\{p_\varepsilon^2 - 2(n-2)(\beta_\varepsilon - p_\varepsilon^2)\}\frac{e^{-\lambda}\lambda^{s-2}(\lambda - s + 1)}{(s-1)!}$$
(3.6)
$$+ \binom{n}{3}\gamma_\varepsilon \frac{e^{-\lambda}\lambda^{s-3}\{(\lambda - s + 1)^2 - s + 1\}}{(s-1)!} + O(n^{-2}).$$

The same approximation clearly holds also for open subsets Ω of uniformly smooth manifolds, such as spheres and toruses.

In the particular case of distances between points independently and uniformly distributed on a circle, β_ε and γ_ε can be explicitly calculated as p_ε^2 and $\frac{3}{4}p_\varepsilon^2$ respectively. When considering the distribution of the smallest spacing between any pair of n such points, (3.3) and (3.5) indicate that spacings ε given by

$$\varepsilon = 2\pi\lambda / n(n-1),$$

for fixed λ, are of the appropriate order of magnitude: for such an ε, $p_\varepsilon = \varepsilon/\pi = \lambda/\binom{n}{2}$, and the probability that the smallest spacing is less than ε is then, to a first approximation, $1 - e^{-\lambda} + O(n^{-1})$ from (2.1). Using (3.6), this estimate is improved to

$$1 - e^{-\lambda} + \tfrac{1}{2}\lambda^2 e^{-\lambda}/(n-1) + O(n^{-2}),$$

which agrees with the expansion for the same probability, obtained by explicit calculation. A similar check can be made using the distribution of the second smallest spacing, which is also expressible in terms of the distribution of T.

A rather more interesting example is that of points uniformly distributed on a unit sphere in R^k. Such points have the same distribution as $(k+1)$-dimensional vectors of independent, identically distributed normal random variables, centred at their sample mean and normalized by their sample standard deviation. In this case the 'distance' between two points can be taken to be the cosine of the angle they subtend, so that 'close' points are vectors which are highy correlated. These distances are also pairwise independent (Geisser and Mantel (1962), Brown and Eagleson (1984)) and γ_ε can be obtained from the joint probability density of the three correlation coefficients, r_{12}, r_{13}, r_{23}, calculated between $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$ and $\boldsymbol{Y}_3$, namely:

$$C(k+1)(1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23})^{(k-4)/2}$$
$$\times I(1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23}>0).$$

Once γ_3 is calculated (by numerical integration), Equation (3.6) can be used.

Although (3.6) provides a theoretically better estimate of the critical probability $P[T \geqq s]$ than the target value α computed according to a Poisson approximation, in that the error remaining is of order n^{-2} rather than of order n^{-1}, it is reasonable to ask whether it has practical importance: that is, whether, for n so small that α is not a good approximation to $P[T \geqq s]$, the estimate in (3.6) is any better. Figure 1 shows the results of 23000 simulations of the correlation example with $k=9$, in comparison with the target size α derived from the Poisson approximation ($\alpha = 5\%$) and with the prediction given by (3.6), for $n=5$, 10 and 15. It is clear that the prediction given by (3.6) is extremely reliable, even for small n and relatively large s. Consideration of the error estimates in Theorem 2 indicates that, if a fixed percentage error in the probability to be approximated is tolerable, s can be allowed to grow linearly with n. Figure 1 suggests that a relative error of 20% or less is maintained in this example for $s \leqq 2n$. It is tempting to suppose that a normal approximation might be appropriate for larger s, but this is not in fact the case, since the statistic T has, in the terminology of U-statistics, a degenerate kernel.

4. Proofs

The notation is that of Section 2. In addition, let $x = x_{\lambda,A}$ be the test function used in the proof of (2.1), namely $x\colon Z^+ \to R$ is given by

$$x(0)=0$$
$$x(m+1)=\lambda^{-m-1}e^{\lambda}m!\,[P_\lambda(A\cap U_m)-P_\lambda(A)P_\lambda(U_m)], \qquad m \geqq 0,$$

where $U_m = \{0, 1, \cdots, m\}$. It is shown in the appendix to Barbour and Eagleson (1983) that

$$\sup_{j\geqq 0} |x_{\lambda,A}(j)| \leqq \min(1, 1.4\lambda^{-1/2}) = a_\lambda,$$

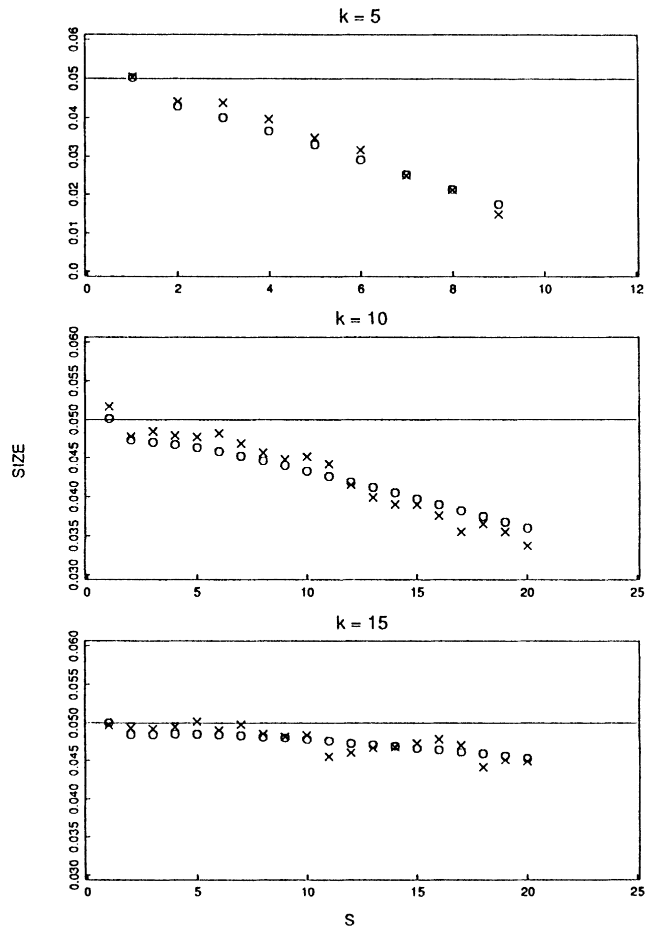

Figure 1. Comparison of the approximations (3.3) and (3.6) to $P[T \geqq s]$ with simulation, in the case of distances between points distributed independently and uniformly over the unit sphere in $\mathbf{R}^9$, with α fixed at 0.05. The approximation (3.6) is denoted by O, and the estimates from 23000 simulations by ×

and

$$\sup_{j \geqq 0} |x_{\lambda,A}(j+1) - x_{\lambda,A}(j)| \leqq \min(1, \lambda^{-1}) = b_\lambda,$$

uniformly in $A \subset Z^+$. To prove Theorem 1, we shall show that

$$\sup_{A \subset Z^+} |P(T \in A) - P_\lambda(A) + \kappa_2 E\{x_{\lambda,A}(P+2) - x_{\lambda,A}(P+1)\}| \leqq \delta_1.$$

As it follows from the definition of $x_{\lambda,A}$ that

$$E\{x(P+2) - x(P+1)\} = -\tfrac{1}{2}E\{I(P \in A)(1 - 2\lambda^{-1}P + \lambda^{-2}P(P-1))\},$$

this will be sufficient to prove the theorem.

Let $Z_J = \Sigma_{K \in N_J} X_K$ and $T_J = T - Z_J$, so that T_J is independent of X_J. Then as in Barbour and Eagleson (1984),

$$P[T \in A] - P_\lambda(A) = \sum_{J \in N} p_J E\{x(T+1) - x(T_J+1)\} \tag{4.1}$$
$$- \sum_{J \in N} E\{X_J[x(T) - x(T_J+1)]\},$$

where x denotes $x_{\lambda,A}$. Consider the first sum. It is immediate that

$$|x(T+1) - x(T_J+1) - Z_J\{x(T_J+2) - x(T_J+1)\}| \leqq 2b_\lambda(Z_J - 1)^+,$$

so that the first sum may be written as

$$\sum_{J \in N} p_J E\left\{\left(X_J + \sum_{K \in N_J^-} X_K\right)(x(T_J+2) - x(T_J+1))\right\} + e_1, \tag{4.2}$$

where

$$|e_1| \leqq 2b_\lambda \sum_{J \in N} p_J E(Z_J - 1)^+. \tag{4.3}$$

Let

$$Z_{JK} = \sum_{L \in N_K \setminus N_J} X_L \quad \text{and} \quad T_{JK} = T_J - Z_{JK},$$

so that T_{JK} is independent of both X_J and X_K: then, directly,

$$|X_K(x(T_J+2) - x(T_J+1)) - X_K(x(T_{JK}+2) - x(T_{JK}+1))| \leqq 2b_\lambda X_K Z_{JK},$$

so that

$$E\{X_K(x(T_J+2) - x(T_J+1))\} = E\{X_K(x(T_{JK}+2) - x(T_{JK}+1))\} + e_{2JK}, \tag{4.4}$$

where

$$|e_{2JK}| \leqq 2b_\lambda E(X_K Z_{JK}). \tag{4.5}$$

Finally, whatever the integer-valued random variable W and function f,

$$|Ef(W) - Ef(P)| \leqq 2 \sup_j |f(j)| d(W, P), \tag{4.6}$$

where $d(W, P)$ denotes the total variation distance between the distributions of W and P,

$$d(W, P) := \sup_{A \subset Z^+} |P(W \in A) - P(P \in A)|.$$

Combining (4.2), (4.4) and (4.6) and using the independence of X_J and T_J and of X_J, X_K and T_{JK}, it follows that the first sum in (4.1) may be written as

$$\sum_{J \in N} p_J \left\{ p_J E(x(P+2) - x(P+1)) + \sum_{K \in N_J^-} p_K E(x(P+2) - x(P+1)) \right\} \tag{4.7}$$
$$+ e_1 + \sum_{J \in N} \left\{ e_{3J} + \sum_{K \in N_J^-} (p_J e_{2JK} + e_{4JK}) \right\},$$

where

$$|e_{3J}| \leqq 2b_\lambda p_J^2 d(T_J, P) \tag{4.8}$$

and

$$|e_{4JK}| \leqq 2b_\lambda p_J p_K d(T_{JK}, P). \tag{4.9}$$

Now both T_J and T_{JK} are dissociated statistics, though with means λ_J and λ_{JK} possibly smaller than λ, and a straightforward extension of the argument leading to (4.1) gives

$$d(T_J, P) \leqq a_\lambda(\lambda - \lambda_J) + b_\lambda \sum_{L\in N} C_L$$

and

$$d(T_{JK}, P) < a_\lambda(\lambda - \lambda_{JK}) + b_\lambda \sum_{L\in N} C_L. \tag{4.10}$$

This allows the estimate

$$e_3 = \left|\sum_{J\in N} e_{3J}\right| \leqq 2b_\lambda \sum_{J\in N} p_J^2 \left\{a_\lambda \sum_{K\in N_J} p_K + b_\lambda \sum_{L\in N} C_L\right\} \tag{4.11}$$

and

$$e_4 = \left|\sum_{J\in N}\sum_{K\in N_J^-} e_{4JK}\right| \leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-} p_J p_K \left\{2a_\lambda \sum_{L\in N_J} p_L + b_\lambda \sum_{L\in N} C_L\right\}. \tag{4.12}$$

For e_1 and e_{2JK}, note that

$$(Z_J - 1)^+ \leqq \frac{1}{2} \sum_{\substack{K,L\in N_J \\ K\neq L}} X_K X_L$$

and

$$X_K Z_{JK} \leqq \sum_{L\in N_K^-} X_K X_L:$$

then it follows from (4.3) and (4.5) that

$$|e_1| \leqq b_\lambda \sum_{J\in N} p_J \sum_{\substack{K,L\in N_J \\ K\neq L}} \beta_{KL} \tag{4.13}$$

and

$$e_2 = \left|\sum_{J\in N}\sum_{K\in N_J^-} p_J e_{2JK}\right| \leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-}\sum_{L\in N_K^-} p_J \beta_{KL}. \tag{4.14}$$

Expression (4.7), together with the estimates (4.11)–(4.14), complete the analysis of the first sum in (4.1).

For the second, proceed in a similar fashion. Since

$$\begin{aligned} &|X_J\{x(T) - x(T_J + 1)\} - X_J(Z_J - X_J)\{x(T_J + 2) - x(T_J + 1)\}| \\ &\leqq 2b_\lambda X_J(Z_J - X_J - 1)^+, \end{aligned} \tag{4.15}$$

and for $K \in N_J^-$,

$$
\begin{aligned}
(4.16)\qquad & |X_J X_K\{x(T_J+2)-x(T_J+1)\} - X_J X_K\{x(T_{JK}+2)-x(T_{JK}+1)\}| \\
& \leqq 2b_\lambda X_J X_K Z_{JK},
\end{aligned}
$$

the second sum can be written as

$$(4.17)\qquad \sum_{J\in N}\sum_{K\in N_J^-} E(X_J X_K)E\{x(T_{JK}+2)-x(T_{JK}+1)\} + e_5 + e_6,$$

with

$$(4.18)\qquad |e_5| \leqq 2b_\lambda \sum_{J\in N} E\{X_J(Z_J - X_J - 1)^+\} \leqq b_\lambda \sum_{J\in N}\sum_{K\in N_J^-}\sum_{\substack{L\in N_J^-\\ L\neq K}} \gamma_{JKL}$$

and

$$
\begin{aligned}
|e_6| &\leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-} E(X_J X_K Z_{JK}) \\
(4.19)\qquad &\leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-}\sum_{\substack{L\in N_K^-\\ L\neq J}} \gamma_{JKL}.
\end{aligned}
$$

Expression (4.17) can be reduced further, using (4.6) and (4.10), to

$$(4.20)\qquad \sum_{J\in N}\sum_{K\in N_J^-} E(X_J X_K)E\{x(P+2)-x(P+1)\} + e_5 + e_6 + e_7,$$

where

$$(4.21)\qquad |e_7| \leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-} \beta_{JK}\left\{2a_\lambda \sum_{L\in N_J} p_L + b_\lambda \sum_{L\in N} C_L\right\}.$$

Subtracting (4.20) from (4.7), and using (4.1), it follows that the quantity to be estimated in Theorem 1 is no larger than

$$\sum_{j=1}^{7} |e_j|.$$

Theorem 1 now follows from the estimates (4.11)–(4.14), (4.18), (4.19), and (4.21).

As it can be shown that

$$E\{x_{\lambda,A}(P+3) - 2x_{\lambda,A}(P+2) + x_{\lambda,A}(P+1)\} = \tfrac{1}{3}E\{I(P\in A)C_3(\lambda; P)\},$$

Theorem 2 will follow if we show that

$$
\begin{aligned}
\sup_{A\subset Z^+} & |P(T\in A) - P_\lambda(A) + \kappa_2 E\{x_{\lambda,A}(P+2) - x_{\lambda,A}(P+1)\} \\
& + \tfrac{1}{2}\kappa_3 E\{x_{\lambda,A}(P+3) - 2x_{\lambda,A}(P+2) + x_{\lambda,A}(P+1)\}| \\
& \leqq \delta_2.
\end{aligned}
$$

To do this, observe first that the awkward terms involving $\gamma_{JK(JK)}$ arise only from the estimates of e_5 and e_6, introduced through the errors in the approximations (4.15) and (4.16). In fact, since $Z_{JK} = \Sigma_{L\in N_K\setminus N_J} X_L$, the estimates (4.19) of e_6 can immediately be improved to

$$|e_6| \leqq 2b_\lambda \sum_{J\in N} \sum_{K\in N_J^-} \sum_{\substack{L\in N_J^- \\ L\neq K,(JK)}} \gamma_{JKL}, \tag{4.22}$$

which means that the error introduced in the approximation (4.16) is in fact of the required order. However, (4.15) is essentially too weak, and must be replaced by

$$\Big| X_J\{x(T_J+Z_J)-x(T_J+1)\} - X_J(Z_J-X_J)\{x(T_J+2)-x(T_J+1)\}$$
$$- \tfrac{1}{2}X_J \sum_{K\in N_J^-} X_K X_{(JK)}\{x(T_J+3)-2x(T_J+2)+x(T_J+1)\}\Big|$$
$$\leqq 4b_\lambda X_J \sum_{K\in N_J^-} \sum_{\substack{L\in N_J^- \\ L\neq K,(JK)}} X_K X_L .$$

Thus e_5 is replaced by e_8, with

$$|e_8| \leqq 4b_\lambda \sum_{J\in N} \sum_{K\in N_J^-} \sum_{\substack{L\in N_J^- \\ L\neq K,(JK)}} \gamma_{JKL}, \tag{4.23}$$

avoiding the awkward error terms at the expense of introducing an extra correction term

$$-\frac{1}{2}\sum_{J\in N}\sum_{K\in N_J^-} E[X_J X_K X_{(JK)}\{x(T_J+3)-2x(T_J+2)+x(T_J+1)\}],$$

which, as in the proof of Theorem 1, can be reduced to

$$-\frac{1}{2}\left\{\sum_{J\in N}\sum_{K\in N_J^-} E(X_J X_K X_{(JK)})\right\} E\{x(P+3)-2x(P+2)+x(P+1)\} + e_9 + e_{10},$$

where

$$|e_9| \leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-} E(X_J X_K Z_{JK}), \tag{4.24}$$

again of the form (4.22), and

$$\begin{aligned} |e_{10}| &\leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-} \beta_{JK} d(T_{JK}, P) \\ &\leqq 2b_\lambda \sum_{J\in N}\sum_{K\in N_J^-} \beta_{JK}\left\{2a_\lambda \sum_{L\in N_J} p_L + b_\lambda \sum_{L\in N} C_L\right\}. \end{aligned} \tag{4.25}$$

Theorem 2 follows: the discrepancy is essentially the estimate of

$$\sum_{\substack{j=1\\ j\neq 5}}^{10} e_j$$

provided by (4.11)–(4.14) and (4.21)–(4.25).

References

Barbour, A. D. and Eagleson, G. K. (1983) Poisson approximation for some statistics based on exchangeable trials. *Adv. Appl. Prob.* **15**, 585–600.

Barbour, A. D. and Eagleson, G. K. (1985) Poisson convergence for dissociated statistics, *J.R. Statist. Soc.* B **46**, 397–402.

Barbour, A. D. and Hall, P. (1984) On the rate of Poisson convergence. *Math. Proc. Camb. Phil. Soc.* **95**, 473–480.

Brown, T. C. and Eagleson, G. K. (1984) A useful property of some symmetric statistics. *Amer. Statistician* **38**, 63–65.

Brown, T. C. and Silverman, B. W. (1979) Rates of Poisson convergence for *U*-statistics. *J. Appl. Prob.* **16**, 428–432.

Chen, L. H. Y. (1975) Poisson approximation for dependent trials. *Ann. Prob.* **3**, 534–545.

Geisser, S. and Mantel, N. (1962) Pairwise independence of jointly dependent variables. *Ann. Math. Statist.* **33**, 290–291.

Mantel, N. and Bailar, J. C. (1970) A class of permutational and multinomial tests arising in epidemiological research. *Biometrics* **26**, 687–700.

McGinley, W. G. and Sibson, R. (1975) Dissociated random variables. *Math. Proc. Camb. Phil. Soc.* **77**, 185–188.

Silverman, B. W. and Brown, T. C. (1978) Short distances, flat triangles and Poisson limits. *J. Appl. Prob.* **15**, 815–825.

Stein, C. (1970) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. 6th Berkeley Symp. Math. Statist. Prob.* **2**, 583–602.