Adv. Appl. Prob. **21**, 74–90 (1989)
Printed in N. Ireland

SOME APPLICATIONS OF THE STEIN–CHEN METHOD FOR PROVING POISSON CONVERGENCE

A. D. BARBOUR, **Universität Zürich*
LARS HOLST, ***Uppsala University*

Abstract

Let W be a sum of Bernoulli random variables and U_λ a Poisson random variable having the same mean $\lambda = EW$. Using the Stein–Chen method and suitable couplings, general upper bounds for the variational distance between W and U_λ are given. These bounds are applied to problems of occupancy, using sampling with and without replacement and Pólya sampling, of capture–recapture, of spacings and of matching and ménage.

VARIATIONAL DISTANCE; COUPLING; OCCUPANCY; CAPTURE–RECAPTURE; SPACINGS; MATCHING

1. Introduction

Let $I_1, \cdots, I_r$ be Bernoulli random variables with

$$EI_k = P(I_k = 1) = 1 - P(I_k = 0) = \pi_k, \qquad \text{for } k = 1, \cdots, r.$$

Set

$$W := I_1 + \cdots + I_r,$$
$$\lambda := EW = \pi_1 + \cdots + \pi_r.$$

From now on let U_λ denote a Poisson random variable with mean $\lambda := EW$. Consider the total variation distance between $\mathscr{L}(W)$ and $\mathscr{L}(U_\lambda)$, that is

$$d(W, U_\lambda) := \sup_A |P(W \in A) - P(U_\lambda \in A)| = \tfrac{1}{2} \sum_{j=0}^{\infty} |P(W = j) - P(U_\lambda = j)|.$$

The purpose of this paper is to illustrate how a method, going back to Stein (1970) and Chen (1975a) for proving Poisson convergence can be used very simply, together with suitable couplings, to give upper bounds for $d(W, U_\lambda)$ in a variety of applications. The basic theorems are derived in Section 2. They are then applied in the subsequent sections to problems of occupancy, using sampling with or without replacement and Pólya sampling, of capture–recapture, of spacings and of matching and ménage. Another article with a similar theme, but quite different applications,

Received 10 November 1987.
* Postal address: Institut für Angewandte Mathematik, Universität Zürich, Rämistrasse 74, CH-8001 Zürich, Switzerland.
** Postal address: Uppsala University, Department of Mathematics, Thunbergsv. 3, S-752 38 Uppsala, Sweden.

has been independently prepared by Arratia et al. (1989). It is to be hoped that, as a result of these reviews, the Stein–Chen method will achieve the popularity it deserves.

2. Some basic general results

For each k, let W_k and V_k be random variables defined on the same probability space in such a way that

$$\mathscr{L}(W_k) = \mathscr{L}(W),\ \mathscr{L}(1 + V_k) = \mathscr{L}(W \mid I_k = 1).$$

If $P(I_k = 1) = 0$, then define $1 + V_k = 0$, say.

Theorem 2.1. Under the hypotheses set out above,

$$d(W, U_\lambda) \leqq (1 \wedge \lambda^{-1}) \sum_{k=1}^{r} \pi_k E\,|W_k - V_k|.$$

Proof. For any bounded $h : \mathbb{Z}^+ \to \mathbb{R}$,

$$\begin{aligned} &E(\lambda h(W+1) - Wh(W)) \\ &\quad = \sum_{k=1}^{r} \{\pi_k E(h(W+1)) - E(I_k h(W))\} \\ &\quad = \sum_{k=1}^{r} \pi_k \{E(h(W+1) - E(h(W) \mid I_k = 1)\} \\ &\quad = \sum_{k=1}^{r} \pi_k E(h(W_k + 1) - h(V_k + 1)). \end{aligned}$$

For any integers $i, j \geqq 0$ we have

$$|h(i) - h(j)| \leqq |i - j|\,\Delta h$$

where

$$\Delta h = \sup_{j \geqq 0} |h(j+1) - h(j)|.$$

Hence

$$\begin{aligned} |E(\lambda h(W+1) - Wh(W))| &\leqq \sum_{k=1}^{r} \pi_k E\,|h(W_k + 1) - h(V_k + 1)| \\ &\leqq \sum_{k=1}^{r} \pi_k E\,|W_k - V_k| \,.\, \Delta h. \end{aligned}$$

Let g be defined recursively by

$$g(0) = 0,$$
$$\lambda g(j+1) - jg(j) = I(j \in A) - P(U_\lambda \in A),$$

where $A \subset \mathbb{Z}^+$; recall that U_λ is Poisson with mean λ. In Lemma 4 of Barbour and Eagleson (1983) it is proved that

$$\sup_j |g(j)| \leqq 1 \wedge 1 \cdot 4 \,.\, \lambda^{-\frac{1}{2}},$$

$$\Delta g = \sup_{j \geqq 0} |g(j+1) - g(j)| \leqq \lambda^{-1}(1 - e^{-\lambda}) \leqq 1 \wedge \lambda^{-1}.$$

Thus

$$|P(W \in A) - P(U_\lambda \in A)| = |E(\lambda g(W+1) - Wg(W))|$$
$$\leqq (1 \wedge \lambda^{-1}) \sum_{k=1}^{r} \pi_k E\,|W_k - V_k|,$$

proving the assertion.

Introducing a test function like g and from this obtaining bounds on the variation distance goes back to Stein (1970) and Chen (1975a). The coupling idea used here also appears in Stein (1986), pp. 92–93.

Using symmetry we get the following result.

Corollary 2.2. If the I's are exchangeable, then

$$d(W, U_\lambda) \leqq (\lambda \wedge 1) E\,|W_1 - V_1|.$$

For the bound of $d(W, U_\lambda)$ to be useful, the expectations $E\,|W_k - V_k|$, $k = 1, \cdots, r$, should be small. To accomplish this in specific applications, the coupling should be chosen so that most of the probability mass is on the 'diagonal' $\{W_k = V_k\}$. Note that only the marginal distributions $\mathcal{L}(W_k) = \mathcal{L}(W)$ and $\mathcal{L}(1 + V_k) = \mathcal{L}(W \mid I_k = 1)$ are given, so that any joint distribution having these marginals may be used to obtain a bound.

Next assume that for each k there are Bernoulli random variables $I_{1k}, \cdots, I_{rk}$, $J_{1k}, \cdots, J_{rk}$ defined on the same probability space such that

$$\mathcal{L}(I_{1k}, \cdots, I_{rk}) = \mathcal{L}(I_1, \cdots, I_r),$$
$$\mathcal{L}(J_{1k}, \cdots, J_{rk}) = \mathcal{L}(I_1, \cdots, I_r \mid I_k = 1),$$
$$W_k = I_{1k} + \cdots + I_{rk},$$
$$1 + V_k = J_{1k} + \cdots + J_{rk},$$
$$J_{kk} = 1.$$

Theorem 2.3. If $J_{ik} \leqq I_{ik}$ for each $i \neq k = 1, \cdots, r$, then

$$d(W, U_\lambda) \leqq (1 \wedge \lambda) \,.\, (1 - \operatorname{Var}(W)/\lambda).$$

Proof. As $J_{ik} \leqq I_{ik}$ we have

$$\begin{aligned}\sum_k \pi_k E\,|W_k - V_k| &= \sum_k \pi_k E\left|\sum_i I_{ik} - \sum_{i\neq k} J_{ik}\right| \\ &= \sum_k \pi_k E\Big(I_{kk} + \sum_{i\neq k} (I_{ik} - J_{ik})\Big) \\ &= \sum_k \pi_k\Big(\pi_k + \sum_{i\neq k} E(I_{ik} - J_{ik})\Big) \\ &= \sum_k \pi_k^2 + \sum_{i\neq k}\sum (\pi_k\pi_i - \pi_k E J_{ik}) \\ &= \sum_k \pi_k^2 + \sum_{i\neq k}\sum (\pi_k\pi_i - E I_i I_k) \\ &= \lambda - \sum_k \operatorname{Var}(I_k) - \sum_{i\neq k}\sum \operatorname{Cov}(I_i, I_k) \\ &= \lambda - \operatorname{Var}(W).\end{aligned}$$

Thus the assertion follows from Theorem 2.1.

Note that $J_{ik} \leqq I_{ik}$ implies that

$$\operatorname{Cov}(I_i, I_k) = E(I_i I_k) - \pi_i\pi_k = \pi_i\pi_k(P(J_{ik} = 1 \mid I_{ik} = 1) - 1) \leqq 0.$$

Hence a necessary condition for Theorem 2.3 is that the I's are not positively correlated.

As a simple application of Theorem 2.3 consider independent I's. Then we can take $I_{ik} = I_i$ for all i, k, $J_{ik} = I_i$ for all $i \neq k$, and $\operatorname{Var}(W) = \sum_k \pi_k(1 - \pi_k)$.

Theorem 2.4. For independent I's we have

$$d(W, U_\lambda) \leqq (1 \wedge \lambda)(1 - \operatorname{Var}(W)/\lambda) = (1 \wedge \lambda^{-1}) \sum_{k=1}^{r} \pi_k^2.$$

This theorem is due to Barbour and Hall (1984), where it is also proved that $d(W, U_\lambda) \geqq \frac{1}{32}(1 \wedge \lambda^{-1}) \sum_k \pi_k^2$.

In fact their method of obtaining a lower bound can also be applied under the conditions of Theorem 2.3, giving for any $\theta \geqq e$, an estimate

$$d(W, U_\lambda) \geqq \frac{\varepsilon_\lambda(1 + 3\varepsilon_\lambda/\theta - (3 + \eta_\lambda/\varepsilon_\lambda)/\theta)}{2(2e^{-3/2} + \theta e^{-1})},$$

where

$$\varepsilon_\lambda := \lambda^{-1}(\lambda - \operatorname{Var} W), \qquad \eta_\lambda := \lambda^{-2}(\lambda - \kappa_4(W)),$$

and $\kappa_4(W)$ denotes the fourth cumulant of the distribution of W. Thus $(\lambda \wedge 1)\varepsilon_\lambda$ is the correct order of magnitude for $d(W, U_\lambda)$, for any sequence of processes satisfying the conditions of Theorem 2.3, for which also $\eta_\lambda/\varepsilon_\lambda = O(1)$, i.e. $\kappa_4(W) = \lambda + O(\lambda(\lambda - \operatorname{Var} W))$.

It is sometimes convenient to use a more detailed coupling than that of Theorem 2.1, conditioning not only on $I_k = 1$, but also on the value of another random element X_k. Suppose that random variables $W_{k,x}$ and $V_{k,x}$ can be defined on the same probability space in such a way that

$$\mathscr{L}(W_{k,x}) = \mathscr{L}(W),$$
$$\mathscr{L}(1 + V_{k,x}) = \mathscr{L}(W \mid I_k = 1, X_k = x).$$

Then, in similar fashion to Theorem 2.1, we can prove the following result.

Theorem 2.5. Under the above hypotheses,

$$d(W, U_\lambda) \leqq (1 \wedge \lambda^{-1}) \sum_{k=1}^{r} \pi_k E(\theta(X_k) \mid I_k = 1)$$

where

$$\theta(x) := E\,|W_{k,x} - V_{k,x}|.$$

3. The occupancy problem

Now we investigate a multinomial scheme with n independent trials and r possible outcomes. Let p_k denote the probability of the kth outcome. The number of outcomes occurring in none of the trials is denoted by W. We can interpret the outcomes as boxes into which n balls are thrown independently of each other, the probability of a ball falling into the kth box being p_k. The random variable W is equal to the number of empty boxes and can be represented as:

$$W = I_1 + \cdots + I_r,$$

where $I_k = 1$ if the kth box is empty and $I_k = 0$ otherwise. We then have

$$\pi_k = P(I_k = 1) = (1 - p_k)^n,$$
$$\lambda = EW = \sum_{k=1}^{r} \pi_k,$$
$$E(I_i I_k) = P(I_i = I_k = 1) = (1 - p_i - p_k)^n,$$
$$\mathrm{Var}\,(W) = \sum_r \pi_k(1 - \pi_k) + \sum_{i \neq k}\sum (E(I_i I_k) - \pi_i \pi_k).$$

Theorem 3.1. Under the hypothesis of this section,

$$d(W, U_\lambda) \leqq (1 \wedge \lambda)(1 - \mathrm{Var}(W)/\lambda).$$

Proof. Introduce the following coupling. Throw each of the balls which have fallen into the kth box independently into one of the other boxes, in such a way that the probability of falling into box $i \neq k$ is $p_i/(1 - p_k)$. Then let $J_{ik} = 1$ if box i is empty, $J_{ik} = 0$ otherwise, and $I_{ik} = I_i$. Evidently $J_{ik} \leqq I_{ik}$ for $i \neq k$, and for each k

$$\mathscr{L}(J_{1k}, \cdots, J_{rk}) = \mathscr{L}(I_1, \cdots, I_r \mid I_k = 1).$$

Thus the assertion follows from Theorem 2.3.

Consider a double array of multinomial schemes with $n = n_r$, $p_k = p_{kr}$, $W = W_r$, etc. The following result is due to Sevast'yanov (1972).

Theorem 3.2. If, as $r \to \infty$,

(i) $$\max_{1 \leqq k \leqq r} \pi_k \to 0,$$

(ii) $$EW \to \lambda_\infty < \infty,$$

then, for $j = 1, 2, 3, \cdots$

$$E(W(W-1)\cdots(W-j+1)) \to \lambda_\infty^j,$$

and

$$W \overset{d}{\Rightarrow} U_{\lambda_\infty}.$$

Under Sevast'yanov's conditions, in view of his conclusion that $EW(W-1) \to \lambda_\infty^2$, Theorem 3.1 adds a useful rate of convergence, which can be overestimated by $\max_k \pi_k + \max_k np_k\pi_k$. However Theorem 3.1 extends the range of Poisson approximation to cover cases in which $EW \to \infty$ provided only that

$$\mathrm{Var}\,(W)/EW \to 1.$$

Note that for $n \geqq 1$,

$$\mathrm{Var}\,(W) < EW.$$

Normal convergence when $\mathrm{Var}\,(W) \to \infty$ has also beeen investigated, see Kolchin et al. (1978).

In the symmetric case $p_k = 1/r$, $k = 1, \cdots, r$, the problem of deriving an exact expression for $\mathcal{L}(W)$, the classical occupancy problem, was solved by De Moivre, see Holst (1986). Von Mises proved Poisson convergence when $EW \to \lambda_\infty$, see Feller (1968). Vatutin and Mikhailov (1982) studied a more general situation, cf. Section 5 below, containing the classical occupancy problem as a special case. From their formula (12) it follows that W has the same distribution as a sum of independent Bernoulli random variables. Hence by the results due to Barbour and Hall (1984), cf. Theorem 2.4 above and remark in connection with that theorem, we have the following result.

Theorem 3.3. For the classical occupancy problem, with $p_k = 1/r$, $1 \leqq k \leqq r$, $\frac{1}{32} \leqq d(W, U_\lambda)/\{(1 \wedge \lambda)(1 - \mathrm{Var}\,(W)/\lambda)\} \leqq 1$.

If $n = ra_r$, it is easy to establish that

$$(1 - \mathrm{Var}\,(W)/EW) \leqq \frac{EW}{r}(1 + r^2a_r/(r-1)^2) \leqq \exp(-a_r)(1 + r^2a_r/(r-1)^2),$$

giving useful Poisson approximation throughout the range $a_r \to \infty$. In the special case

$$a_r = \log r - \log \lambda_\infty + o(1)$$

we have

$$EW \sim r \exp(-a_r) \to \lambda_\infty.$$

Hence $(1 \wedge \lambda_\infty)\lambda_\infty \,.\, (1+n/r)/r$ gives the right rate of convergence to the Poisson as measured by the variation distance. Theorem 3 of Vatutin and Mikhailov (1982) gives the less precise upper bound of $\lambda_\infty^2(1+n/r)/r$. If we had complete independence, the rate would have been $(1 \wedge \lambda_\infty)\lambda_\infty/r$. The slight correlation influences the rate of convergence through the factor $1+n/r \sim \log r$.

4. Occupancy for drawing without replacement and for Pólya sampling

The multinomial scheme investigated in the previous section could also be formulated using an urn containing N balls of r different colours in proportions $p_1, \cdots, p_r$ from which n balls are drawn with replacement. The random variable W is the number of colours not occurring among the n drawn balls.

Consider instead sampling without replacement. Set $I_k = 1$ if the sample contains no ball of colour k and $I_k = 0$ otherwise. Then

$$W = I_1 + \cdots + I_r,$$

$$\pi_k = P(I_k = 1) = \binom{N(1-p_k)}{n} \Big/ \binom{N}{n},$$

$$\lambda = EW = \sum_{k=1}^{r} \pi_k,$$

$$E(I_i I_k) = P(I_i = I_k = 1) = \binom{N(1-p_i-p_k)}{n} \Big/ \binom{N}{n},$$

$$\mathrm{Var}\,(W) = \sum_k \pi_k(1-\pi_k) + \sum\sum_{i \neq k} (E(I_i I_k) - \pi_i \pi_k).$$

Theorem 4.1. For sampling without replacement,

$$d(W, U_\lambda) \leqq (1 \wedge \lambda)(1 - \mathrm{Var}\,(W)/\lambda).$$

Proof. First draw n balls without replacement from the urn. Set $I_{ik} = I_i$. Then withdraw all balls of colour k from the urn. After that replace all balls of colour k in the sample by balls drawn without replacement from the remaining balls in the urn. Set $J_{ik} = 1$ if the sample now has no ball of colour i and $J_{ik} = 0$ otherwise. Evidently $J_{ik} \leqq I_{ik}$ for $k \neq i$ and

$$\mathscr{L}(J_{1k}, \cdots, J_{rk}) = \mathscr{L}(I_1, \cdots, I_r \mid I_k = 1).$$

The assertion follows from Theorem 2.3.

Note that if we replace 'without replacement' with 'with replacement' in the proof we get the proof of Theorem 3.1.

We have Pólya sampling from the urn if each ball drawn is replaced together with one new ball of the same colour. For this case we see that

$$\pi_k = \binom{-N(1-p_k)}{n} \Big/ \binom{-N}{n}$$

$$E(I_iI_k) = \binom{-N(1-p_i-p_k)}{n} \Big/ \binom{-N}{n}.$$

Changing 'drawing without replacement' to 'Pólya sampling' in the previous proof gives the following result.

Theorem 4.2. For Pólya sampling,

$$d(W, U_\lambda) \leqq (1 \wedge \lambda)(1 - \operatorname{Var}(W)/\lambda).$$

By considering triangular arrays Poisson convergence can be formulated. As in Section 3, we see that as long as $\operatorname{Var}(W)/E(W) \sim 1$ it is reasonable to approximate $\mathscr{L}(W)$ by $\mathscr{L}(U_\lambda)$.

In Holst et al. (1988) some weaker estimates of the variation distance are obtained for some special cases of the above situation, using different methods. Very few results on the rate of Poisson convergence in cases like those above seem to be available elsewhere in the literature.

5. A capture–recapture problem

An urn contains r black balls. A simple random sample of size s is drawn without replacement. After painting each ball in the sample white, the balls are replaced into the urn. This is repeated n independent times with sample sizes $s_1, \cdots, s_n$ respectively. Imagine that the balls are numbered $1, \cdots, r$ and set $I_k = 1$ if ball k is never drawn, $I_k = 0$ otherwise. The number of black balls still in the urn after sampling is

$$W = I_1 + \cdots + I_r.$$

This random variable occurs also in connection with capture–recapture methods for estimating sizes of populations, see Holst (1980) and the references therein. It is easily seen that for each k

$$\pi_1 = \pi_k = P(I_k = 1) = \prod_{j=1}^{n} (1 - s_j/r),$$

and for $i \neq k$

$$\pi_{12} = \pi_{ik} = P(I_i = I_k = 1) = \prod_{j=1}^{n} \left(1 - \frac{s_j}{r}\right)\left(1 - \frac{s_j}{r-1}\right).$$

Hence

$$\lambda = EW = r\pi_1, \qquad \operatorname{Var}(W) = r\pi_1(1-\pi_1) + r(r-1)(\pi_{12} - \pi_1^2).$$

Theorem 5.1. Under the hypothesis of this section,

$$d(W, U_\lambda) \leqq (1 \wedge \lambda)(1 - \mathrm{Var}\,(W)/\lambda).$$

Proof. For each k introduce the following coupling. In each sample where ball k is drawn replace this ball by a randomly chosen ball not in the sample. Let $J_{ik} = 1$ if ball i was never drawn, $J_{ik} = 0$ otherwise, and set $I_{ik} = I_i$. It is readily seen that

$$\mathscr{L}(J_{1k}, \cdots, J_{rk}) = \mathscr{L}(I_1, \cdots, I_r \mid I_k = 1),$$

and $J_{ik} \leqq I_{ik}$. The assertion follows from Theorem 2.3.

Note that for $s_1 = \cdots = s_n = 1$ we get the classical occupancy problem discussed in Section 3.

In Vatutin and Mikhaĭlov (1982) it is proved that, for all $s_1, \cdots, s_n$, W has the same distribution as a sum of independent Bernoulli random variables. Hence, using Theorem 2.4 and the remark following it, we can strengthen the result as follows.

Theorem 5.2.

$$\tfrac{1}{32}(1 \wedge \lambda)(1 - \mathrm{Var}\,(W)/\lambda) \leqq d(W, U_\lambda).$$

Combining the results above we get the following.

Corollary 5.3. Assume, for a sequence of such urn schemes, that

$$E(W) \to \lambda_\infty \in \mathbb{R}^+, \qquad r \to \infty.$$

Then

$$W \overset{\mathrm{d}}{\Rightarrow} U_{\lambda_\infty}$$

if and only if

$$\mathrm{Var}\,(W) \to \lambda_\infty.$$

6. Large spacings

On a circle of circumference 1, r points are taken independently using a uniform distribution on the circumference. Starting with one of the points let $S_1, \cdots, S_r$ denote the spacings, i.e. successive arc-length distances between the points in clockwise order, say. For a given number a, set for each k

$$I_k = I(S_k > a).$$

Then the number of spacings larger than a can be written

$$W = I_1 + \cdots + I_r.$$

We can also think of placing r arcs all of length a at random on the circumference, thus splitting it into gaps, i.e. segments which are not covered by any arc. Clearly W is the number of such gaps. We could also consider $r - 1$ independent observations from a uniform distribution on the unit interval. The successive distances between the order statistics, including the endpoints 0 and 1, are the spacings. In a large number of papers, different aspects of spacings are studied, cf. Holst and Hüsler (1984) and the references therein.

Using symmetry, it is not difficult to derive that

$$P(W=j)=\binom{r}{j}\sum_{v=0}^{r-j}\binom{r-j}{v}(-1)^v(1-(j+v)a)_+^{r-1},$$
$$\pi_1=P(I_k=1)=(1-a)_+^{r-1},$$
$$\pi_{12}=P(I_i=I_k=1)=(1-2a)_+^{r-1},\quad i\neq k,$$
$$\lambda=EW=r\pi_1,\ \mathrm{Var}\,(W)=r\pi_1(1-\pi_1)+r(r-1)(\pi_{12}-\pi_1^2).$$

Theorem 6.1. For the number of large spacings,

$$d(W,U_\lambda)\leqq(1\wedge\lambda)(1-\mathrm{Var}\,(W)/\lambda).$$

In particular, for $a\leqq\frac{1}{2}$,

$$d(W,U_\lambda)\leqq(1\wedge\lambda)\frac{\lambda}{r}\left[\left\{1+\left(\frac{(r-1)a}{1-a}\right)^2\right\}\wedge r\right]$$
$$\leqq(1\wedge\lambda)\exp(-a(r-1))\left[\left\{1+\left(\frac{(r-1)a}{1-a}\right)^2\right\}\wedge r\right].$$

Proof. Consider $r-1$ points taken at random on the unit interval. Conditioned on $S_1>a$, the points come from a uniform distribution on the interval $(a,1)$. Hence

$$\mathscr{L}(S_1-a,S_2,\cdots,S_r\mid S_1>a)=\mathscr{L}((1-a)S_1,(1-a)S_2,\cdots,(1-a)S_r).$$

Set $I_{i1}=I_i=I(S_i>a)$ for all i, $J_{11}=1$, and $J_{i1}=I((1-a)S_i>a)$ for all $i>1$. Then

$$\mathscr{L}(J_{11},\cdots,J_{r1})=\mathscr{L}(I_1,\cdots,I_r\mid I_1=1),$$

and $J_{i1}\leqq I_{i1}$ for $i>1$.

If instead we consider points on the circle we can construct $J_{ik}\leqq I_{ik}=I_i$, $i\neq k$, in analogous fashion. Hence the assertion follows from Theorem 2.3.

Thus, for any sequence (a_r) such that $ra_r\to\infty$, it follows that

$$d(W,U_\lambda)\to 0\quad\text{as}\quad r\to\infty.$$

If for $0<\lambda_\infty<\infty$,

$$\lambda=r(1-a_r)^{r-1}\to\lambda_\infty,\qquad r\to\infty,$$

that is if

$$a_r=\frac{1}{r}(\log r-\log\lambda_\infty+o(1)),$$

then

$$W\overset{\mathrm{d}}{\Rightarrow}U_{\lambda_\infty},$$

with $\lambda_\infty(\log r)^2/r$ as the appropriate rate of convergence. A weaker bound is obtained in Holst et al. (1988) by other methods.

Convergence in distribution of W has been studied before: for example, if $\mathrm{Var}(W) \to \infty$, W is asymptotically normally distributed, see Holst and Hüsler (1984).

7. Permutations, matching and ménage

In this section, we illustrate some applications where Theorem 2.5 is used. Let $(c_{ij})_{i,j=1}^r$ be an $r \times r$ matrix of zeros and ones, and let $I_i = c_{i\sigma(i)} = c(i, \sigma(i))$, where σ is a random permutation, i.e. a permutation drawn from the uniform distribution on all permutations of $\{1, 2, \cdots, r\}$. Let

$$W := \sum_{i=1}^r I_i,$$

$$\pi_i := \frac{1}{r}\sum_{j=1}^r c_{ij} = EI_i = P(I_i = 1),$$

$$\rho_j := \frac{1}{r}\sum_{i=1}^r c_{ij},$$

$$\lambda := EW = \frac{1}{r}\sum\sum_{i,j=1}^r c_{ij}.$$

Theorem 7.1.

$$d(W, U_\lambda) \leqq (1 \wedge \lambda^{-1})\tfrac{3}{2}\Big(\sum_i \pi_i^2 + \sum_j \rho_j^2\Big).$$

Remark 1. Since $\pi_i = P(I_i = 1)$, the first part of the estimate is comparable to the estimates for sums of independent indicators. The second part can be understood in exactly the same way, since W can equally well be written as $\sum_j I_j'$ where

$$I_j' = c(\sigma^{-1}(j), j).$$

Remark 2. Chen (1975b), Corollary 3.2, gives a similar estimate, but with the constant $45.25/2 = 22.625$ in place of $\frac{3}{2}$. The proof below can also be adapted to prove (2.2) of his more general Theorem 2.1, again with 3 in place of 45.25 in the constant multiplier.

Proof. We use Theorem 2.5, coupling W with random variables

$$V_{kj} = \sum_{i \neq k} c(i, \sigma'(i)),$$

where $\sigma'(k) = j$, $\sigma'(\sigma^{-1}(j)) = \sigma(k)$, and $\sigma'(i) = \sigma(i)$ otherwise, and where j is any index such that $c_{kj} = 1$; it is easy to see that

$$\mathscr{L}(1 + V_{kj}) = \mathscr{L}(W \mid I_k = 1, \sigma(k) = j).$$

Now

$$\begin{aligned}\theta(j) &= E\,|W - V_{kj}|\\ &\leqq E\,|c(k, \sigma(k)) + c(\sigma^{-1}(j), j) - c(\sigma^{-1}(j), \sigma(k))|\\ &= \frac{1}{r}c_{kj} + \frac{1}{r(r-1)}\sum_{l\neq j}\sum_{m\neq k}|c_{kl} + c_{mj} - c_{ml}|\\ &\leqq \frac{1}{r}c_{kj} + \frac{1}{r}(c_{k.} - c_{kj}) + \frac{1}{r}(c_{.j} - c_{kj}) + \frac{1}{r(r-1)}(c_{..} - c_{.j} - c_{k.} + c_{kj})\\ &= \{\lambda + (r-2)(\pi_k + \rho_j - r^{-1}c_{kj})\}/(r-1),\end{aligned}$$

where $c_{k.} := \sum_{j=1}^{r} c_{kj}$ and so on. Since also

$$\pi_k E(\theta(\sigma(k)) \mid I_k = 1) = \frac{1}{r}\sum_{j=1}^{r} c_{kj}\theta(j),$$

it follows from Theorem 2.5 that

$$d(W, U_\lambda) \leqq (1 \wedge \lambda^{-1})\left\{\lambda^2 + (r-2)\left(\sum_{k=1}^{r}\pi_k^2 + \sum_{j=1}^{r}\rho_j^2\right)\right\}\Big/(r-1),$$

and the result follows because, using Schwarz's inequality,

$$2\lambda^2 \leqq r\left(\sum_k \pi_k^2 + \sum_j \rho_j^2\right).$$

We apply the theorem first to a matching problem. Let A and B be two decks of cards, let A have a_1 cards of type $1, \cdots, a_n$ cards of type n and let $b_1, \cdots, b_n$ be the corresponding numbers for B. Further, let r be the size of A equal to that of B:

$$r = \sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j.$$

Now the two decks are matched at random, i.e. each card from A is put into correspondence with one card from B in such a way that each of the $r!$ ways of doing this has probability $1/r!$ The r pairs so obtained are examined for matches where a pair is said to give a match if the two cards in the pair are of the same type. The distribution of W, the total number of matches, has been investigated by previous authors, see Lanke (1973) and the references therein. Lanke proved that under general conditions W is asymptotically Poisson distributed as $r \to \infty$.

We give a bound on the variation distance using Theorem 7.1. The matrix (c_{ij}) corresponding to this problem can be constructed as follows. Define $A_0 = B_0 = 0$ and

$$A_s = \sum_{j=1}^{s} a_j, \qquad B_s = \sum_{j=1}^{s} b_j, \qquad 1 \leqq s \leqq n,$$

$$\alpha(i) = s \quad \text{if} \quad A_{s-1} < i \leqq A_s,$$

$$\beta(j) = s \quad \text{if} \quad B_{s-1} < j \leqq B_s,$$

and set

$$c_{ij} = 1(0) \quad \text{if} \quad \alpha(i) = \beta(j) \quad \text{(otherwise)}.$$

Then

$$\lambda = EW = \sum_{j=1}^{n} a_j b_j / r,$$

$$\sum_{k=1}^{r} \pi_k^2 = \sum_{j=1}^{n} a_j^2 b_j / r^2,$$

$$\sum_{j=1}^{r} \rho_j^2 = \sum_{j=1}^{n} a_j b_j^2 / r^2,$$

and we have the following result.

Corollary 7.2. For the matching problem described above

$$d(W, U_\lambda) \leqq (1 \wedge \lambda^{-1}) \tfrac{3}{2} \sum_{j=1}^{n} a_j b_j (a_j + b_j) / r^2.$$

Now consider a triangular array such that $a_j = a_{jr}$, $b_j = b_{jr}$, $n = n_r$ with $EW = \sum_{j=1}^{n} a_j b_j / r \to \lambda_\infty$, $r \to \infty$.

Corollary 7.3. If $(1/r) \max_j ((a_j + b_j) I(a_j b_j > 0)) \to 0$ as $r \to \infty$, then

$$W \overset{d}{\Rightarrow} U_{\lambda_\infty}.$$

Proof. The conditions clearly imply that the right-hand side in Corollary 7.2 tends to 0 as $r \to \infty$.

It is easy to see that

$$\sum_{j=1}^{n} a_j b_j (a_j + b_j) / r^2 \to 0$$

and

$$\frac{1}{r} \max_j ((a_j + b_j) I(a_j b_j > 0)) \to 0$$

are equivalent if $EW \to \lambda_\infty$, and that they are uniform asymptotic negligibility conditions, see Lanke (1973). Lanke also proves that the conditions are necessary for Poisson convergence.

Corollary 7.4. If $r = nc$ and $a_j = b_j = c$, for all j, then $EW = c$ and

$$d(W, U_\lambda) \leqq 3c/r.$$

Thus Poisson approximation is useful as long as the probability of a given card being matched, c/r, is small.

The ménage problem asks for the number of seatings of r man–woman couples at a circular table, with men and women alternating, so that no one sits next to his or her partner.

Consider the following generalization. Let $(\sigma_1, \cdots \sigma_r)$ be a random permutation of $\{1, \cdots, r\}$ and let $1 \leqq c \leqq r$ be a given integer. Set $I_k = 1$, if $k \in \{\sigma_k, \sigma_{k+1}, \cdots, \sigma_{k+c-1}\}$, $I_k = 0$ otherwise, where $\sigma_{r+j} = \sigma_j$. We say that there is a match at place k if $I_k = 1$. We wish to study the distribution of the total number of matches

$$W = I_1 + \cdots + I_r.$$

As $P(I_k = 1) = c/r$, we have $\lambda = EW = c$. Note that determining $P(W = 0)$, when $c = 1$, is the classical rencontre or matching problem, and that $r!\,P(W = 0)$, when $c = 2$, is the number asked for in the ménage problem.

Corollary 7.5. For the generalized ménage problem,

$$d(W, U_\lambda) \leqq 3c/r.$$

Hence, for fixed c,

$$W \overset{\mathrm{d}}{\Rightarrow} U_c \quad \text{as} \quad r \to \infty.$$

Proof. Define (c_{ij}) by

$$c_{ij} = \begin{cases} 1 & \text{if} \quad i \leqq j \leqq i + c - 1 \quad \text{or} \quad 1 \leqq j \leqq i + c - 1 - r \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mathscr{L}(W) = \mathscr{L}\Big(\sum_i c(i, \sigma(i))\Big),$$

and $\pi_k = \rho_j = c/r$. The result follows from Theorem 7.1.

For the classical matching problem, that is the case $c = 1$, it is easily seen, from the explicit expression for the distribution of W, that the convergence rate to Poisson (1) is much faster than $3/r$ as given by Corollary 7.5.

For $c = 2$, the ménage problem, an explicit form of the distribution of W is also available, but the convergence to the Poisson (2) distribution is not as fast. By the formula (29) in Takács (1981) the convergence rate seems to be $O(r^{-1})$, as given by Corollary 7.5.

For $c > 2$ we are not aware of any results on $\mathscr{L}(W)$ or on its convergence.

The interesting papers Takács (1981) and Bogart and Doyle (1986) contain many results connected with the ménage problem and give many old references both on the matching and the ménage problem.

Results similar in spirit to Theorem 7.1 can also be established for arrays with more than two indices, though the computations become heavier. The following theorem gives an example with statistical applications.

Theorem 7.6. Let a and b be symmetric $r \times r$ matrices with elements 0 or 1, σ a random permutation of $\{1, \cdots, r\}$

$$I_{ik} := a_{ik} b_{\sigma(i)\sigma(k)},$$

$$W := \sum_{i<k}\sum I_{ik},$$

and let

$$\lambda := EW = \binom{r}{2} \bar{a}_{..} \bar{b}_{..},$$

where $\bar{c}_{..}$ denotes $\sum\sum_{i<j} c_{ij} / \binom{r}{2}$. Then

$$d(W, U_\lambda) \leqq (1 \wedge \lambda^{-1})\{\lambda(\bar{a}_{..} + \bar{b}_{..}) + r^2(r-1)(\bar{a}_{..}^2 + a_1)(\bar{b}_{..}^2 + b_1)\}$$

where

$$c_1 := \sum_{i=1}^{r} c_{i.}(c_{i.} - 1)/r^2(r-1), \qquad c_{i.} := \sum_{k \neq i} c_{ik}.$$

Proof. For any pair $i < k$ such that $a_{ik} = 1$, let $X_{ik} := (\sigma(i), \sigma(k))$, and compute $\theta_{ik}(j, m)$ of Theorem 2.5 for any pair (j, m) such that $b_{jk} = 1$, using the coupling $\sigma'(i) = j$, $\sigma'(k) = m$, $\sigma'(\sigma^{-1}(j)) = \sigma(i)$, $\sigma'(\sigma^{-1}(m)) = \sigma(k)$ and $\sigma'(s) = \sigma(s)$ otherwise, to define

$$I'_{st} := a_{st} b_{\sigma'(s)\sigma'(t)}$$

and

$$1 + V_{ik,jm} = \sum_{s<t}\sum I'_{st}.$$

Then

$$\begin{aligned}
W - V_{ik,jm} &= \sum_{s<t}\sum (I_{st} - I'_{st}) + 1 \\
&= b_{\sigma(i)\sigma(k)} + [a(i, \sigma^{-1}(m)) - a(k, \sigma^{-1}(j))][b_{\sigma(i)m} - b_{j\sigma(k)}] \\
&\quad + a(\sigma^{-1}(j), \sigma^{-1}(m))(1 - b_{\sigma(i)\sigma(k)}) \\
&\quad + \sum_{s=1}^{r} \{a_{is}[b_{\sigma(i)\sigma(s)} - b_{j\sigma(s)}] + a(\sigma^{-1}(j), s)[b_{j\sigma(s)} - b_{\sigma(i)\sigma(s)}] \\
&\quad + a_{ks}[b_{\sigma(k)\sigma(s)} - b_{m\sigma(s)}] + a(\sigma^{-1}(m), s)[b_{m\sigma(s)} - b_{\sigma(k)\sigma(s)}]\} \\
&\quad \times I(s \neq i, k, \sigma^{-1}(j), \sigma^{-1}(m)).
\end{aligned}$$

Hence, by direct but laborious computation,

$$\begin{aligned}
\theta_{ik}(j, m) \leqq{}& \bar{a}_{..} + \bar{b}_{..} \\
&+ \frac{1}{r}[(a_{i.} - 1)(b_{j.} - 1) + (a_{k.} - 1)(b_{m.} - 1)] \\
&+ \bar{a}_{..}[(b_{j.} - 1) + (b_{m.} - 1)] + \bar{b}_{..}[(a_{i.} - 1) + (a_{k.} - 1)] \\
&+ 2(r-1)\bar{a}_{..}\bar{b}_{..},
\end{aligned}$$

and since from Theorem 2.5,

$$d(W, U_\lambda) \leqq \frac{1}{2r(r-1)} \sum_{i \neq k}\sum a_{ik} \sum_{j \neq m}\sum b_{jm}\theta_{ik}(j, m),$$

the theorem follows.

As an application of the theorem, consider the data of Knox (1964), in which $1 \leqq i \leqq r = 96$ indexes cases of childhood leukaemia, and $a_{ij} = 1$ if children i and j lived close to each other, $b_{ij} = 1$ if cases i and j presented at times close to one another. In this data $\frac{1}{2}a_{..} = 25$ and $\frac{1}{2}b_{..} = 152$. Values of a_1 and b_1 are not given, but values of 10^{-5} and 10^{-3} respectively are not unreasonable. Theorem 7.6 then gives the upper bound for the variation distance 0.074, as compared to 0.71 from Barbour and Eagleson (1983). They also obtained by Monte Carlo simulation that the true value of the variation distance is around 0.016, see Barbour and Eagleson (1983), p. 594.

References

Arratia, R., Goldstein, L. and Gordon, L. (1989) Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17** (January).

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

Bogart, K. P. and Doyle, P. (1986) Non-sexist solution of the ménage problem. *Amer. Math. Monthly.* **93,** 514–518.

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

Chen, L. H. Y. (1975b) An approximation theorem for sums of certain randomly selected indicators. *Z. Wahrscheinlichkeitsth.* **33,** 69–74.

Feller, W. (1968) *An Introduction to Probability Theory and Its Applications.* Vol. 1, 3rd edn. Wiley, New York.

Holst, L. (1980) On matrix occupancy, commitee, and capture–recapture problems. *Scand. J. Statist.* **7,** 139–146.

Holst, L. (1986) On birthday, collectors', occupancy and other classical urn problems. *Int. Statist. Rev.* **54,** 15–27.

Holst, L. and Hüsler, J. (1984) On the random coverage of the circle. *J. Appl. Prob.* **21,** 558–566.

Holst, L., Kennedy, J. and Quine, M. (1988) Rates of Poisson convergence for some coverage and urn problems using coupling. *J. Appl. Prob.* **25,** 717–724.

Knox, G. (1964) Epidemiology of childhood leukaemia in Northumberland and Durham. *Brit. J. Prev. Soc. Med.* **18,** 17–24.

Kolchin, V. F., Sevast'yanov, B. A. and Chistyakov, V. P. (1978) *Random Allocations.* Winston, Washington DC.

Lanke, J. (1973) Asymptotic results on matching distributions. *J. R. Statist. Soc. B.* **35,** 117–122.

Sevast'yanov, B. A. (1972) Poisson limit law for a scheme of sums of dependent random variables. *Theory Prob. Appl.* **17,** 695–699.

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.

STEIN, C. (1986) *Approximate Computation of Expectations.* Inst. Math. Statist. Lecture Notes—Monograph Series Vol. **7**, Hayward, California.

TAKÁCS, L. (1981) On the 'Problème des Ménages'. *Discrete Math.* **36,** 289–297.

VATUTIN, V. A. AND MIKHAILOV, V. G. (1982) Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory Prob. Appl.* **27,** 734–743.