*J. R. Statist. Soc.* B (1983),
**45**, *No.* 1, *pp.* 100–106

# Serial Dependence in Contingency Tables

By SIMON TAVARÉ

*Colorado State University, USA*

[Received November 1981. Revised May 1982]

SUMMARY

A simple and useful representation is given for the asymptotic behaviour of the Pearson $\chi^2$ test of independence in two-way contingency tables. This is used to analyse the statistic when the data are generated by Markov chains. A robustness result, showing when the Markov dependence has no effect on the usual limiting $\chi^2$ distribution, is also given.

*Keywords*: CONTINGENCY TABLES; INDEPENDENCE; MARKOV CHAINS

## 1. INTRODUCTION

Considerable attention has recently been given to the asymptotic behaviour of goodness-of-fit tests and tests of independence in two-way tables for data that arise from complex sample surveys (Holt *et al.*, 1979; Rao and Scott, 1981). A large body of results is also known for the analysis of serially dependent data, as might for instance be generated by types of Markov dependence. See, for example, Bartlett (1951), Patankar (1954), Anderson and Goodman (1957), Billingsley (1961a, b) and Basawa and Prakasa Rao (1980).

In this paper, we focus exclusively on tests of independence for two-way $r \times c$ contingency tables. We assume that we have observations $n_{ij}$ on the joint occurrence of two random sequences $X = \{X_k, k \geqslant 0\}$ and $Y = \{Y_k, k \geqslant 0\}$. That is, let

$$n_{ij} = \text{number of times } X_k = i, \quad Y_k = j, \quad 0 \leqslant k \leqslant n-1,$$

where $X$ has $r$ states, $Y$ has $c$ states. The usual Pearson statistic for testing that $X$ and $Y$ are independent is

$$C_n = \sum_{i,j} \frac{n(n_{ij} - n_{i+}n_{+j}/n)^2}{n_{i+}n_{+j}}, \tag{1.1}$$

where + denotes summation over that index. It is a familiar result that when $X$ and $Y$ are independent trials processes (an assumption equivalent to the usual multinomial model), then $C_n \xrightarrow{\mathscr{D}} \chi_t^2$, as $n \to \infty$, where $t = (r-1)(c-1)$, $\mathscr{D}$ denoting convergence in distribution.

Recently, Holt *et al.*, (1980) and Rao and Scott (1981) have analysed statistics based on (1.1) when the multinomial model is not appropriate. The applications they had in mind involved data generated from complex sample survey designs. We will describe briefly their results, which are based on the following assumption.

### *Assumption*

There is a set of underlying cell probabilities $\alpha_{ij} > 0$, $\alpha_{++} = 1$ such that the asymptotic distribution of the vector of normed proportions

$$\sqrt{n}\left(\frac{n_{11}}{n} - \alpha_{11}, \frac{n_{12}}{n} - \alpha_{12}, \ldots, \frac{n_{rc}}{n} - \alpha_{rc}\right)$$

*Present address*: Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA

 0035–9246/83/45100 $2.00

is $rc$-multivariate $N(0, \boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$, which is singular, reflects the experimental design that generated the counts $n_{ij}$. Tests of the null hypothesis of independence of rows and columns are usually based on the quantities

$$h_{ij}(\hat{\boldsymbol{\alpha}}) = \frac{n_{ij}}{n} - \frac{n_{i+}}{n}\frac{n_{+j}}{n}; \quad i = 1, \ldots, r-1; \quad j = 1, \ldots, c-1. \tag{1.2}$$

Let $h(\boldsymbol{\alpha}) = (h_{11}(\boldsymbol{\alpha}), h_{12}(\boldsymbol{\alpha}), \ldots, h_{(r-1)(c-1)}(\boldsymbol{\alpha}))^{\mathrm{T}}$, and let $\mathbf{H}(\boldsymbol{\alpha})$ be the $(r-1)(c-1) \times rc$ matrix of partial derivatives $\mathbf{H}(\boldsymbol{\alpha}) = \partial h(\boldsymbol{\alpha})/\partial(\boldsymbol{\alpha})$. When no estimate of $\boldsymbol{\Sigma}$ is available, it is common to ignore the sample structure, and use the statistic (1.1).

Using the notation

$$\tilde{\boldsymbol{\alpha}}_r^{\mathrm{T}} = (\alpha_{1+}, \ldots, \alpha_{r-1,+}), \quad \tilde{\boldsymbol{\alpha}}_c^{\mathrm{T}} = (\alpha_{+1}, \ldots, \alpha_{+,c-1}) \tag{1.3}$$

$$\tilde{\mathbf{D}}_r = \operatorname{diag}\{\alpha_{1+}, \ldots, \alpha_{r-1,+}\}, \quad \tilde{\mathbf{D}}_c = \{\alpha_{+1}, \ldots, \alpha_{+,c-1}\},$$

Holt *et al.*, prove that under the independence assumption

$$C_n \xrightarrow{\mathscr{D}} \sum_{i=1}^{(r-1)(c-1)} \rho_i Z_i^2, \quad n \to \infty, \tag{1.4}$$

where $Z_i$ are *i.i.d.* standard Normal variables, and the $\rho_i$ are the eigenvalues of the matrix $\mathbf{E}$, where

$$\mathbf{E} = \mathbf{R}^{-1}\,\mathbf{H}\,\boldsymbol{\Sigma}\,\mathbf{H}^{\mathrm{T}}, \quad \mathbf{R} = (\tilde{\mathbf{D}}_r - \tilde{\alpha}_r\tilde{\alpha}_r^{\mathrm{T}}) \otimes (\tilde{\mathbf{D}}_c - \tilde{\alpha}_c\tilde{\alpha}_c^{\mathrm{T}}). \tag{1.5}$$

Here, $\otimes$ is the usual Kronecker product: $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$.

Tavaré and Altham (1983) were also interested in application of tests based on (1.1), but where the observations $\{n_{ij}\}$ were generated by 2-state Markov chains. If $X$ is a Markov chain whose transition matrix has non-unit eigenvalue $\lambda$, $Y$ a Markov chain whose transition matrix has non-unit eigenvalue $\mu$, then they showed that if $X$ and $Y$ were independent,

$$C_n \xrightarrow{\mathscr{D}} \frac{1+\mu\lambda}{1-\mu\lambda}\chi_1^2, \quad n \to \infty, \tag{1.6}$$

so that the Markov dependence leads to either inflated or deflated values of the $\chi^2$ statistic. It is often possible to estimate consistently both $\mu$ and $\lambda$ from the data, in which case (1.6) can be used as a test of independence for two two-state Markov chains. The Markov assumption is often used as a first approximation to the structure of serially dependent data, and the study of the asymptotics of $C_n$ for such dependence arose because of its frequent appearance in psychological experiments. Further details of this may be found in Tavaré and Altham, and the review article of Castellan (1979).

The purpose of the present paper is to study the asymptotic behaviour of $C_n$ when $X$ and $Y$ are arbitrary (but positive recurrent) Markov chains. In particular, we will extend the result of (1.6) for reversible Markov chains (which includes (1.6) as a special case). The results allow us to quantify the behaviour of $C_n$ in the presence of serially dependent data. As a consequence of the analysis, another representation for (1.4) and (1.5) is obtained, which seems to be easier to work with in practice.

## 2. PRELIMINARIES

Let $\mathbf{A}$ and $\mathbf{B}$ be matrices. The following properties of Kronecker products are well known; it is assumed that matrices are conformable and invertible as required.

$$\begin{aligned}(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= \mathbf{AC} \otimes \mathbf{BD} \\ (\mathbf{A} \otimes \mathbf{B})^{-1} &= \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \\ (\mathbf{A} \otimes \mathbf{B})^{\mathrm{T}} &= \mathbf{A}^{\mathrm{T}} \otimes \mathbf{B}^{\mathrm{T}}\end{aligned} \tag{2.1}$$

For any matrix $\mathbf{A}$ of size $m \times n$, write

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^{\mathrm{T}} \\ \mathbf{a}_2^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_m^{\mathrm{T}} \end{pmatrix},$$

so that $\mathbf{a}_i^{\mathrm{T}}$ is the $i$th row of $A$. We define the vec operator vec $(\mathbf{A})$ by

$$\operatorname{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{a}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_m \end{pmatrix}, \tag{2.2}$$

an $mn \times 1$ column vector of the rows of $\mathbf{A}$. The results of Neudecker (1969) readily give:

*Proposition* 1. For conformable matrices $A$ and $B$,

$$\operatorname{vec}(\mathbf{AB}) = (\mathbf{A} \otimes \mathbf{I}) \operatorname{vec}(\mathbf{B}) = (\mathbf{I} \otimes \mathbf{B}^{\mathrm{T}}) \operatorname{vec}(\mathbf{A}).$$

We now turn to the statistic $C_n$ defined at (1.1). Write

$$C_n = \sum_{i,j} n\left(\frac{n_{ij}}{n} - \frac{n_{i+}}{n}\frac{n_{+j}}{n}\right)^2 \bigg/ \left(\frac{n_{i+}}{n}\frac{n_{+j}}{n}\right). \tag{2.3}$$

The assumption about the asymptotic behaviour of the random variables $n_{ij}/n$ made in the Introduction can be described as follows. Define

$$m_{ij} = \sqrt{n}\left(\frac{n_{ij}}{n} - \alpha_{ij}\right), \text{ and let } \mathbf{M} = (m_{ij}). \text{ Then } \operatorname{vec}(\mathbf{M}) \xrightarrow{\mathscr{D}} N_{rc}(\mathbf{0}, \boldsymbol{\Sigma}) \text{ as } n \to \infty.$$

If we set

$$c_{ij} = \sqrt{n}\left(\frac{n_{ij}}{n} - \frac{n_{i+}}{n}\frac{n_{+j}}{n}\right), \tag{2.4}$$

then under the above assumption, it is straightforward to show that $c_{ij} - g_{ij} \xrightarrow{P} 0$ as $n \to \infty$, where

$$g_{ij} = m_{ij} - \alpha_{i+} m_{+j} - \alpha_{+j} m_{i+}. \tag{2.5}$$

The denominator in (2.3) converges in probability to $\alpha_{i+}\alpha_{+j}$, and it follows that $C_n$ has the same asymptotic distribution as

$$\operatorname{vec}(\mathbf{G})^{\mathrm{T}} \mathbf{D}^{-1} \operatorname{vec}(\mathbf{G}) \tag{2.6}$$

where $\mathbf{G} = (g_{ij})$, $\mathbf{D} = \operatorname{diag}\{\alpha_{i+}\alpha_{+j}\} = \mathbf{D}_r \otimes \mathbf{D}_c$, where

$$\mathbf{D}_r = \operatorname{diag}\{\alpha_{1+}, \ldots, \alpha_{r+}\}, \quad \mathbf{D}_c = \operatorname{diag}\{\alpha_{+1}, \ldots, \alpha_{+c}\}. \tag{2.7}$$

## 3. ASYMPTOTIC REPRESENTATIONS FOR $C_n$

We will use the following notation in what follows:

$\mathbf{I}_m$ is the $m \times m$ identity matrix,

$$\boldsymbol{\alpha}_r^{\mathrm{T}} = (\alpha_{1+}, \ldots, \alpha_{r+}), \tag{3.1}$$

$\boldsymbol{\alpha}_c^{\mathrm{T}} = (\alpha_{+1}, \ldots, \alpha_{+c}),$

$\mathbf{1}_m$ is the $m \times 1$ vector of 1's.

From (2.5), we have

$$g_{ij} = m_{ij} - ((\boldsymbol{\alpha}_r \mathbf{1}_r^{\mathrm{T}})\mathbf{M})_{ij} - (\mathbf{M}(\mathbf{1}_c \boldsymbol{\alpha}_c^{\mathrm{T}}))_{ij}$$

or

$$\mathbf{G} = \mathbf{M} - \mathbf{A}_r^{\mathrm{T}}\mathbf{M} - \mathbf{M}\mathbf{A}_c,$$

where

$$\mathbf{A}_r = \mathbf{1}_r \boldsymbol{\alpha}_r^{\mathrm{T}}, \quad \mathbf{A}_c = \mathbf{1}_c \boldsymbol{\alpha}_c^{\mathrm{T}}. \tag{3.2}$$

From Proposition 1, it follows that

$$\begin{aligned} \operatorname{vec}(\mathbf{G}) &= \operatorname{vec}(\mathbf{M}) - \operatorname{vec}(\mathbf{A}_r^{\mathrm{T}}\mathbf{M}) - \operatorname{vec}(\mathbf{M}\mathbf{A}_c) \\ &= \operatorname{vec}(\mathbf{M}) - (\mathbf{A}_r^{\mathrm{T}} \otimes \mathbf{I}_c)\operatorname{vec}(\mathbf{M}) - (\mathbf{I}_r \otimes \mathbf{A}_c^{\mathrm{T}})\operatorname{vec}(\mathbf{M}) \\ &= (\mathbf{I}_{rc} - (\mathbf{A}_r^{\mathrm{T}} \otimes \mathbf{I}_c) - (\mathbf{I}_r \otimes \mathbf{A}_c^{\mathrm{T}}))\operatorname{vec}(\mathbf{M}) \\ &= \mathbf{B}^{\mathrm{T}} \operatorname{vec}(\mathbf{M}), \quad \text{say,} \end{aligned}$$

where

$$\begin{aligned} \mathbf{B} &= \mathbf{I}_{rc} - (\mathbf{A}_r \otimes \mathbf{I}_c) - (\mathbf{I}_r \otimes \mathbf{A}_c) \\ &= (\mathbf{I}_r - \mathbf{A}_r) \otimes (\mathbf{I}_c - \mathbf{A}_c) - \mathbf{A}_r \otimes \mathbf{A}_c. \end{aligned} \tag{3.3}$$

Hence $\operatorname{vec}(\mathbf{G}) \xrightarrow{\mathscr{D}} N(\mathbf{0}, \mathbf{B}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{B})$ as $n \to \infty$, and a standard result on the distribution of quadratic forms of Normal random variables, together with (2.6), shows that the asymptotic distribution of $C_n$ can be written as

$$C_n \sim \sum_{i=1}^{rc} \rho_i Z_i^2,$$

where $Z_i$ are *i.i.d.* standard Normal variables, and $\rho_i$ are the eigenvalues of the matrix $\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{B}$. We highlight this discussion as

*Proposition* 2. The statistic $C_n$ is asymptotically distributed as $\Sigma_{i=1}^{rc} \rho_i Z_i^2$, where $Z_i$ are *i.i.d.* standard Normal, and $\rho_i$ are the eigenvalues of $\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}} \boldsymbol{\Sigma} \mathbf{B}$.

*Remark.* This is, in effect, contained in (1.4) and (1.5), although, as will be seen in the next section, this representation seems easier to use than (1.5).

## 4. APPLICATIONS

The following proposition is useful. The proof is straightforward using (2.1) and (3.3). Set $\boldsymbol{\alpha} = \boldsymbol{\alpha}_r \otimes \boldsymbol{\alpha}_c$. Then

*Proposition* 3

(a) $\mathbf{B}$ is $\mathbf{D}$-symmetric, i.e. $\mathbf{D}\mathbf{B} = \mathbf{B}^{\mathrm{T}}\mathbf{D}$ and $\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}} = \mathbf{B}\mathbf{D}^{-1}$

(b) $\mathbf{B}\mathbf{1} = -\mathbf{1}$, $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{B} = -\boldsymbol{\alpha}^{\mathrm{T}}$, where $\mathbf{1} = \mathbf{1}_{rc}$

(c) $\mathbf{B}^2 = (\mathbf{I}_r - \mathbf{A}_r) \otimes (\mathbf{I}_c - \mathbf{A}_c) + \mathbf{1}\,\boldsymbol{\alpha}^{\mathrm{T}}$.

*Multinomial trials*

If $X$ and $Y$ are generated by independent trials processes, then it is well-known that under the null hypothesis that $X$ and $Y$ are independent

$$\boldsymbol{\Sigma} = \mathbf{D} - \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}} \tag{4.1}$$

Hence

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{B} &= \mathbf{B}\mathbf{D}^{-1}(\mathbf{D}-\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}})\mathbf{B}\\ &= \mathbf{B}(\mathbf{I}-\mathbf{1}\ \boldsymbol{\alpha}^{\mathrm{T}})\mathbf{B}\\ &= \mathbf{B}^2 - \mathbf{B}\mathbf{1}\ \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{B}\\ &= (\mathbf{I}_r - \mathbf{A}_r)\otimes(\mathbf{I}_c - \mathbf{A}_c)\end{aligned}$$

using Proposition 3. But $\mathbf{I}_r - \mathbf{A}_r$ is idempotent, as is $\mathbf{I}_c - \mathbf{A}_c$. So then is their Kronecker product. But $\mathrm{Tr}\,((\mathbf{I}_r - \mathbf{A}_r)\otimes(\mathbf{I}_c - \mathbf{A}_c)) = \mathrm{Tr}\,(\mathbf{I}_r - \mathbf{A}_r).\ \mathrm{Tr}\,(\mathbf{I}_c - \mathbf{A}_c) = (r-1)(c-1)$. Hence $\mathbf{D}^{-1}\ \mathbf{B}^{\mathrm{T}}\ \boldsymbol{\Sigma}\ \mathbf{B}^{\mathrm{T}}$ has $(r-1)(c-1)$ unit eigenvalues, and $r+c-1$ zero eigenvalues. It follows that under $H_0$, $C_n \sim \chi_t^2$, $t=(r-1)(c-1)$, and we recover the standard test result.

*Markov-dependent trials*

We now assume that $X$ and $Y$ are stationary irreducible aperiodic Markov chains with $r$ and $c$ states respectively. Under the independence assumption, the observations $n_{ij}$ are generated by a Markov chain $W$, say, with transition probability matrix $\mathbf{P} = \mathbf{P}_r \otimes \mathbf{P}_c$, and stationary vector $\boldsymbol{\alpha}^{\mathrm{T}} = \boldsymbol{\alpha}_r^{\mathrm{T}} \otimes \boldsymbol{\alpha}_c^{\mathrm{T}}$, where $\mathbf{P}_r$, $\boldsymbol{\alpha}_r^{\mathrm{T}}$ and $\mathbf{P}_c$, $\boldsymbol{\alpha}_c^{\mathrm{T}}$ are, respectively, the transition matrix and stationary vector of $X$ and $Y$. The central limit theorem for Markov chains (cf. Billingsley, 1961b) shows that

$$\boldsymbol{\Sigma} = \mathbf{D}\mathbf{Z} + \mathbf{Z}^{\mathrm{T}}\mathbf{D} - \mathbf{D} - \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}} \tag{4.2}$$

where $\mathbf{Z}^{-1} = \mathbf{I} - \mathbf{P} + \mathbf{1}\ \boldsymbol{\alpha}^{\mathrm{T}}$.

Then we have

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{B} &= \mathbf{B}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{B}\\ &= \mathbf{B}\mathbf{D}^{-1}(\mathbf{D}\mathbf{Z} + \mathbf{Z}^{\mathrm{T}}\mathbf{D} - \mathbf{D} - \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}})\mathbf{B}\\ &= \mathbf{B}(\mathbf{Z} + \mathbf{D}^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{D} - \mathbf{I} - \mathbf{1}\ \boldsymbol{\alpha}^{\mathrm{T}})\mathbf{B}\end{aligned} \tag{4.3}$$

It remains only to identify the eigenvalues of this matrix. We might expect these eigenvalues to be related to those of $\mathbf{P}$, and hence to those of $\mathbf{P}_r$ and $\mathbf{P}_c$. Clearly, they must be real. There is a large class of Markov chains which have real eigenvalues, namely the reversible chains. $X$ is said to be *reversible* if $\mathbf{P}_r$ is $\mathbf{D}_r$-symmetric, i.e. $\mathbf{D}_r\mathbf{P}_r = \mathbf{P}_r^{\mathrm{T}}\mathbf{D}_r$, and similarly for $Y$. Among reversible chains are independent trials processes, two-state chains, and symmetric doubly stochastic chains (e.g. Iosifescu, 1980, p. 152). If both $X$ and $Y$ are reversible, then so is $W$, since $\mathbf{P} = \mathbf{P}_r \otimes \mathbf{P}_c$, $\mathbf{D} = \mathbf{D}_r \otimes \mathbf{D}_c$. It is then easy to show that $\mathbf{D}\mathbf{Z}^{-1} = (\mathbf{Z}^{\mathrm{T}})^{-1}\mathbf{D}$, so that $\mathbf{D}^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{D} = \mathbf{Z}$. Hence (4.3) reduces to

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{B} &= \mathbf{B}(2\mathbf{Z} - \mathbf{I} - \mathbf{1}\ \boldsymbol{\alpha}^{\mathrm{T}})\mathbf{B}\\ &= 2\mathbf{B}(\mathbf{Z}-\mathbf{I})\,\mathbf{B} + (\mathbf{I}_r - \mathbf{A}_r)\otimes(\mathbf{I}_c - \mathbf{A}_c),\end{aligned} \tag{4.4}$$

using Proposition 3 again.

We have $\boldsymbol{\alpha}_r^{\mathrm{T}}\mathbf{P}_r = \boldsymbol{\alpha}_r^{\mathrm{T}}$, $\mathbf{P}_r\mathbf{1}_r = \mathbf{1}_r$, and similarly for $\mathbf{P}_c$. Let $\boldsymbol{\eta}_r^{\mathrm{T}}$ be a left eigenvector of $\mathbf{P}_r$ with eigenvalue $\lambda \neq 1$, and $\boldsymbol{\eta}_c^{\mathrm{T}}$ be a left eigenvector of $\mathbf{P}_c$ with eigenvalue $\mu \neq 1$. Let $\boldsymbol{\xi}_r$, $\boldsymbol{\xi}_c$ be the corresponding right eigenvectors. Then $\boldsymbol{\eta}_r^{\mathrm{T}}\mathbf{1}_r = \boldsymbol{\eta}_c^{\mathrm{T}}\mathbf{1}_c = \boldsymbol{\alpha}_r^{\mathrm{T}}\ \boldsymbol{\xi}_r = \boldsymbol{\alpha}_c^{\mathrm{T}}\ \boldsymbol{\xi}_c = 0$. It was established in Tavaré and Altham (1983) that if $\boldsymbol{\eta}^{\mathrm{T}} = \boldsymbol{\eta}_r^{\mathrm{T}} \otimes \boldsymbol{\eta}_c^{\mathrm{T}}$, then $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{Z} = (1-\lambda\mu)^{-1}\ \boldsymbol{\eta}^{\mathrm{T}}$, while $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Z} = \boldsymbol{\alpha}^{\mathrm{T}}$. For such $\boldsymbol{\eta}$ we have

$$\boldsymbol{\eta}^{\mathrm{T}}\mathbf{B} = \boldsymbol{\eta}^{\mathrm{T}}, \qquad \boldsymbol{\eta}^{\mathrm{T}}(\mathbf{I}_r - \mathbf{A}_r)\otimes(\mathbf{I}_c - \mathbf{A}_c) = \boldsymbol{\eta}^{\mathrm{T}}$$

and hence from (4.4),

$$\boldsymbol{\eta}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}}\,\boldsymbol{\Sigma}\,\mathbf{B} = \left(\frac{2}{1-\lambda\mu} - 1\right)\boldsymbol{\eta}^{\mathrm{T}} = \frac{1+\lambda\mu}{1-\lambda\mu}\,\boldsymbol{\eta}^{\mathrm{T}}.$$

Now consider any vector $\boldsymbol{\eta}^{\mathrm{T}}$ of the form $\boldsymbol{\eta}^{\mathrm{T}} = \boldsymbol{\alpha}_r^{\mathrm{T}} \times \boldsymbol{\eta}_c^{\mathrm{T}}$.

Then $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{B} = (\boldsymbol{\alpha}_r^{\mathrm{T}} \otimes \boldsymbol{\eta}_c^{\mathrm{T}})((\mathbf{I}_r \otimes \mathbf{I}_c) - (\mathbf{A}_r \otimes \mathbf{I}_c) - (\mathbf{I}_r \otimes \mathbf{A}_c))$

$$= \boldsymbol{\eta}^{\mathrm{T}} - \boldsymbol{\alpha}_r^{\mathrm{T}}\mathbf{A}_r \otimes \boldsymbol{\eta}_c^{\mathrm{T}} - \boldsymbol{\alpha}_r^{\mathrm{T}} \otimes \boldsymbol{\eta}_c^{\mathrm{T}}\mathbf{A}_c = \boldsymbol{\eta}^{\mathrm{T}} - \boldsymbol{\eta}^{\mathrm{T}} = \mathbf{0}^{\mathrm{T}},$$

while $\boldsymbol{\eta}^{\mathrm{T}}(\mathbf{I}_r - \mathbf{A}_r) \otimes (\mathbf{I}_c - \mathbf{A}_c) = \mathbf{0}^{\mathrm{T}}$.

Hence $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}}\,\boldsymbol{\Sigma}\,\mathbf{B} = \mathbf{0}^{\mathrm{T}}$. Similarly, if $\boldsymbol{\eta}^{\mathrm{T}}$ is either of the form $\boldsymbol{\eta}^{\mathrm{T}} = \boldsymbol{\alpha}_r^{\mathrm{T}} \otimes \boldsymbol{\alpha}_c^{\mathrm{T}}$ or $\boldsymbol{\eta}^{\mathrm{T}} = \boldsymbol{\eta}_r^{\mathrm{T}} \otimes \boldsymbol{\alpha}_c^{\mathrm{T}}$, then $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}}\,\boldsymbol{\Sigma}\,\mathbf{B} = \mathbf{0}^{\mathrm{T}}$. We have then established:

*Theorem* 1. Let $X$ and $Y$ be reversible Markov chains. Then under the assumption that $X$ and $Y$ are independent, the statistic $C_n$ is asymptotically distributed as

$$\sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\left(\frac{1+\lambda_i\mu_j}{1-\lambda_i\mu_j}\right)Z_{ij}^2.$$

where $Z_{ij}$ are *i.i.d.N*(0, 1), and $\lambda_i, \mu_j$ are the non-unit eigenvalues of $\mathbf{P}_r$ and $\mathbf{P}_c$ respectively.

*Remarks*

(i) When $r = c = 2$, the chains are necessarily reversible, and then

$$C_n \sim \left(\frac{1+\lambda\mu}{1-\lambda\mu}\right)\chi_1^2$$

as in (1.6).

(ii) When $\mathbf{P}_r = \mathbf{1}_r\,\boldsymbol{\alpha}_r^{\mathrm{T}}$, $\mathbf{P}_c = \mathbf{1}_c\,\boldsymbol{\alpha}_c^{\mathrm{T}}$, then $X$ and $Y$ are (reversible) independent trials processes. The non-unit eigenvalues of $\mathbf{P}_r$ and $\mathbf{P}_c$ are 0, and hence $C_n \sim \chi_t^2$, $t = (r-1)(c-1)$. This recovers the "multinomial trials" case.

(iii) Even when the chains cannot be assumed to be reversible, it is often possible to find consistent estimators of the eigenvalues of $\mathbf{D}^{-1}\mathbf{B}^{\mathrm{T}}\,\boldsymbol{\Sigma}\,\mathbf{B}$ from the data. There is a well-known test of independence for two Markov chains which is described, for example, by Billingsley (1961a).

*Example*. The following example, due to Dr P.M.E. Altham, displays the effects of Markov dependence in a simple way. Suppose that the transition matrix $\mathbf{P}_r$ of $X$ is given by

$$\begin{aligned}(\mathbf{P}_r)_{ij} &= \alpha_r + (1-\alpha_r)\,\theta_i^{(r)}, \quad j = i\\ &= (1-\alpha_r)\,\theta_j^{(r)}, \quad j \neq i\end{aligned} \tag{4.5}$$

where each $\theta_i^{(r)} > 0$, $\Sigma_{i=1}^{r}\,\theta_i^{(r)} = 1$, and $\alpha_r$ is chosen to make $(\mathbf{P}_r)_{ij} \geqslant 0$ for all $i, j$. When $\alpha_r = 0$, $\mathbf{P}_r$ reduces to an independent trials process. It is readily checked that $\mathbf{P}_r$ is reversible, and that $\mathbf{P}_r$ has eigenvalues $\alpha_r(r-1$ times) and 1. Suppose that $\mathbf{P}_c$ has the same structure as (4.5) for some set $\{\theta_i^{(c)}\}$, and $\alpha_c$. From Theorem 1, it follows that

$$C_n \xrightarrow{\mathscr{D}} \frac{1+\alpha_r\alpha_c}{1-\alpha_r\alpha_c}\,\chi_t^2, \tag{4.6}$$

where $t = (r-1)(c-1)$. Thus $(1-\alpha_r\alpha_c)(1+\alpha_r\alpha_c)^{-1}C_n$ has asymptotically the "standard" $\chi^2$ distribution. If $\alpha_r, \alpha_c > 0$ (as would usually be the case) the statistic $C_n$ must be deflated to allow for Markov dependence. This shows that the application of the standard test (treating $C_n$ as though it were a $\chi_t^2$ random variable) is likely to have misleading consequences.

## 5. A ROBUSTNESS RESULT

Examination of Theorem 1 shows that $X$ is a (reversible) Markov chain, and $Y$ is an independent trials process, then $C_n \sim \chi_t^2$, $t = (r-1)(c-1)$, so that *no* correction is required to the asymptotic distribution of $C_n$. The dependence structure in $X$ is in some sense "swamped" by the independence of $Y$. It is intuitively clear that this result should hold when $X$ is an arbitrary (irreducible) Markov chain, and $Y$ an independent trials process. This is the content of the next theorem, the proof of which is straightforward, and is thus omitted.

*Theorem* 2. Let $X$ be an arbitrary irreducible Markov chain, and let $Y$ be an independent trials process. Then, under the assumption that $X$ and $Y$ are independent,

$$C_n \xrightarrow{\mathscr{D}} \chi_t^2, \quad t = (r-1)(c-1), \quad \text{as } n \to \infty.$$

## 6. CONCLUSIONS

In this paper, we have described a simple representation of the Pearson statistic $C_n$ for testing independence in two-way contingency tables. The representation is applied to the case of serially dependent observations when the data are generated by stationary Markov chains. Explicit results are available when the chains are assumed to be reversible (for example, two-state chains, independent trials processes and symmetric doubly stochastic chains). The results show that formal use of such a test statistic can be extremely misleading in the presence of serial dependence in *both* constituent processes.

An interesting robustness result emerges however. When one component process is independent trials, and the other an arbitrary stationary Markov chain the statistic $C_n$ is still distributed as $\chi^2$ with $(r-1)(c-1)$ degrees of freedom; thus serial dependence of the above form in one component process does not effect the asymptotic behaviour of the "usual" test statistic.

## ACKNOWLEDGEMENTS

I would like to thank Dr P.M.E. Altham for several critical comments on an earlier draft of this paper, and for suggesting the example of Section 4. I also thank the referees, whose comments have improved the presentation of the results.

## REFERENCES

Anderson, T.W. and Goodman, L. A. (1957) Statistical inference about Markov chains. *Ann. Math. Stat.*, **28**, 89–109.

Bartlett, M. S. (1951) The frequency goodness of fit test for probability chains. *Proc. Comb. Phil. Soc.*, **47**, 86–95.

Basawa, I. V. and Prakasa Rao, B. L. S. (1980) *Statistical Inference for Stochastic Processes*. New York: Academic Press.

Billingsley, P. (1961a) *Statistical Inference for Markov Processes*. University of Chicago Press.

——— (1961b) Statistical methods in Markov chains. *Ann. Math. Stat.*, **32**, 1–40.

Castellan, N. J. (1979) The analysis of behavior sequences. In *The Analysis of Social Interaction* (R. B. Cairns, ed.) pp. 81–116. Lawrence Earlbaum Associates.

Holt, D., Scott, A. J. and Ewings, P. D. (1980) Chi-squared tests with survey data. *J. R. Statist. Soc.* A, **143**, 303–320.

Iosifescu, M. (1980) *Finite Markov Processes and their Applications*. New York: Wiley.

Neudecker, H. (1969) Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Amer. Statist. Ass.*, **64**, 953–963.

Patankar, V. (1954) The goodness of fit of frequency distributions obtained from stochastic processes, *Biometrika*, **41**, 450–462.

Rao, J. N. K. and Scott, A. J. (1981) The analysis of categorical data from complex sample surveys. Chi squared tests for goodness-of-fit and independence in two-way tables. *J. Amer. Statist. Ass.*, **76**, 221–230.

Tavaré, S. and Altham, P. M. E. (1983) Dependence in goodness of fit tests and contingency tables. *Biometrika*, **70**, in press.