Biometrika (1983), **70**, 1, *pp.* 139–44
Printed in Great Britain

Serial dependence of observations leading to contingency tables, and corrections to chi-squared statistics

BY SIMON TAVARÉ
Department of Statistics, Colorado State University, Fort Collins, Colorado, U.S.A.

AND PATRICIA M. E. ALTHAM
Statistical Laboratory, University of Cambridge

SUMMARY

The asymptotic behaviour of the Pearson goodness-of-fit test and of the test of independence in 2×2 tables is examined when the data are generated by Markov dependent sequences. The results allow simple quantification of the effects of such serial dependence on some standard test statistics.

Some key words: Chi-squared tests; Contingency tables; Dependence; Goodness of fit; Markov chains; Reversibility.

1. INTRODUCTION

In this paper, we discuss the asymptotic behaviour of some standard test statistics in the presence of Markov dependence in a sequence of observations. In particular, we examine the standard Pearson χ^2 goodness-of-fit test, and the 2×2 contingency table test for independence.

Investigations into the effects of Markov dependence seem to have been initiated by Bartlett (1951), who showed that such tests need no longer have the 'usual' asymptotic distribution; see also Patankar (1954). Many other statistical procedures for Markovian data are discussed by Billingsley (1961a) and Basawa & Prakasa Rao (1980). Holt, Scott & Ewings (1980) and Rao & Scott (1981) provide several analyses of categorical data from complex sample surveys. See also Cohen (1976) and Altham (1976, 1979).

The main purpose of this paper is to assess the effects of serially dependent data on standard tests. The choice of Markov chain models arises because they are often a good first approximation to the structure of serially dependent data and because, at least in part, explicit results are available.

2. PRELIMINARIES

Let $\mathscr{X} = \{X_k, k \geqslant 0\}$ be an r-state stationary positive recurrent Markov chain, with stationary distribution $\alpha^{\mathrm{T}} = (\alpha_1, \ldots, \alpha_r)$, $\alpha_i > 0$. From a realization of length n from $\mathscr{X}$, let n_i be the number of times state i is visited, and define

$$m_i = \sqrt{n}(n_i/n - \alpha_i) \quad (1 \leqslant i \leqslant r). \tag{2·1}$$

Set $m^{\mathrm{T}} = (m_1, \ldots, m_r)$.

In this paper the problem of interest is to use the observations $\{n_i\}$ to test hypotheses about α. Of course it would be preferable to have observations on all the transitions

$(X_k = i) \to (X_{k+1} = j)$, say, but we shall suppose in the first instance that only the 'marginal' totals $\{n_i\}$ are available.

The central limit theorem for Markov chains (Billingsley, 1961b) states that

$$m \to N(0, \Omega) \tag{2·2}$$

in distribution as $n \to \infty$. The covariance matrix Ω is defined by

$$\Omega = DZ + Z^{\mathrm{T}} D - D - \alpha\alpha^{\mathrm{T}}, \tag{2·3}$$

where $D = \operatorname{diag}\{\alpha\}$ and $Z^{-1} = I - P + 1\alpha^{\mathrm{T}}$, and P is the transition matrix of $\mathscr{X}$. Then $\alpha^{\mathrm{T}} P = \alpha^{\mathrm{T}}$, $P1 = 1$, and if l, r are left and right eigenvectors of P corresponding to a nonunit eigenvalue λ, $\alpha^{\mathrm{T}} r = l^{\mathrm{T}} 1 = 0$.

Clearly, $Z^{-1} 1 = 1$ and $\alpha^{\mathrm{T}} Z = \alpha^{\mathrm{T}}$. Further

$$Z^{-1} r = (I - P + 1\alpha^{\mathrm{T}}) r = (1 - \lambda) r,$$

so that $Pr = \lambda r$ implies that

$$Zr = (1 - \lambda)^{-1} r, \tag{2·4}$$

and similarly $l^{\mathrm{T}} P = \lambda l^{\mathrm{T}}$ implies that $l^{\mathrm{T}} Z = (1 - \lambda)^{-1} l^{\mathrm{T}}$.

3. GOODNESS-OF-FIT TESTS

Consider testing the simple hypothesis H_0: α specified. The Pearson χ^2 statistic is

$$P_n = \sum_{i=1}^{r} \frac{(n_i - n\alpha_i)^2}{n\alpha_i} = m^{\mathrm{T}} D^{-1} m. \tag{3·1}$$

Under H_0, P_n has an asymptotic distribution that can be represented as $\Sigma_i \rho_i Z_i^2$, where the Z_i are independent $N(0, 1)$ random variables, and ρ_i are the nonzero eigenvalues of $D^{-1}\Omega$, and the sum is over $i = 1, \ldots, r-1$. Note that by definition, Ω is of rank $\leqslant r-1$. If data on transitions between states are available, then the ρ_i can often be estimated, and the resultant test for a specified stationary vector should be compared with the standard one given by, for example, Billingsley (1961a, p. 30). The case $\alpha_i = 1/r$ is equivalent to a doubly stochastic transition matrix.

Some progress can be made in identifying the ρ_i explicitly in the important special case when $\mathscr{X}$ is assumed to be reversible. For then $DP = P^{\mathrm{T}} D$, and it readily follows that $DZ = Z^{\mathrm{T}} D$. From (2·3) we have

$$D^{-1}\Omega = 2Z - I - 1\alpha^{\mathrm{T}}. \tag{3·2}$$

But $D^{-1}\Omega 1 = 0$, $Pr = \lambda r$; hence $D^{-1}\Omega r = (1 + \lambda)(1 - \lambda)^{-1} r$, using (2·4). Hence we have the following.

THEOREM 1. *Under the assumption that $\mathscr{X}$ is a reversible Markov chain with transition matrix P, the asymptotic distribution of P_n under H_0 is that of*

$$P_n' = \sum_{i=1}^{r-1} \frac{1 + \lambda_i}{1 - \lambda_i} Z_i^2,$$

where $\{\lambda_i\}$ are the nonunit eigenvalues of P, and $\{Z_i\}$ are independent $N(0, 1)$.

Remark 1. Two-state Markov chains are necessarily reversible, so that

$$\frac{1-\lambda}{1+\lambda}P_n \to \chi_1^2$$

in distribution under H_0, where λ is the nonunit eigenvalue of the 2×2 transition matrix P.

Remark 2. When P is an independent trials process with $P = 1\alpha^{\mathrm{T}}$, then the nonunit eigenvalues of P are all zero, and the familiar result emerges that $P_n \to \chi_{r-1}^2$ in distribution.

At least for reversible chains, Theorem 1 allows us to quantify precisely the effects of Markov dependence on the null distribution of P_n.

Remark 3. An example of a very simple type of departure from an independent trials process is given by the transition matrix $P = (1-\pi)I + \pi 1\alpha^{\mathrm{T}}$, where $1-\pi$ characterizes the departure from mutually independent trials ($0 < \pi < 1$). Then it is easily seen that the $(r-1)$ nonunit eigenvalues of P are all $(1-\pi)$, and so in this case, from Theorem 1, $\pi P_n'/(2-\pi)$ is distributed as χ_{r-1}^2 under H_0.

4. 2×2 CONTINGENCY TABLES

Let $\mathscr{X} = \{X_k, k \geqslant 0\}$ and $\mathscr{Y} = \{Y_k, k \geqslant 0\}$ be two sequences of binary random variables taking values in $S = \{1, 2\}$, say, and define n_{ij} to be the number of pairs (X_k, Y_k) for which $X_k = i$, $Y_k = j$, for $1 \leqslant k \leqslant n$. A common statistic for testing the hypothesis that the processes $\mathscr{X}$ and $\mathscr{Y}$ are mutually independent is the contingency table statistic

$$C_n = \sum_{i,j} \frac{n(n_{ij} - n_{i+}\, n_{+j}/n)^2}{n_{+j}\, n_{i+}}, \tag{4·1}$$

where $+$ denotes summation over that index. When each of the sequences $\mathscr{X}, \mathscr{Y}$ corresponds to mutually independent trials C_n is known to be approximately χ_1^2 if $\mathscr{X}$ and $\mathscr{Y}$ are independent.

The motivation for studying the asymptotic distribution of C_n when $\mathscr{X}$ and $\mathscr{Y}$ are stationary two-state Markov chains arose in part from a recent paper of Gardner, Hartmann & Mitchell (1982).

They used Monte Carlo techniques to simulate the null distribution of C_n when $\mathscr{X}$ and $\mathscr{Y}$ are generated by identical Markov chains, and found that the values of C_n lead to rejection of the independence hypothesis more often than expected. They also cite a number of cases involving possible misuse of the statistic C_n arising from interactional studies in psychological experiments. The review article of Castellan (1979) discusses such problems that arise in behavioural research.

We write the cell probabilities in the 2×2 table in the form

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix}$$

and assume that these are the equilibrium probabilities for the corresponding 4-state Markov chain. The 'independence' hypothesis is then H_0: $\alpha_1\alpha_4 = \alpha_2\alpha_3$, which is obviously equivalent to $\log\alpha_1 + \log\alpha_4 - \log\alpha_2 - \log\alpha_3 = 0$. Thus a test of H_0 may be

constructed by finding the distribution of the statistic

$$L_n = \sqrt{n}\{\log(n_1/n) + \log(n_4/n) - \log(n_2/n) - \log(n_3/n)\}. \tag{4·2}$$

Write $\alpha^{\mathrm{T}} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, and define m_i, m as in (2·1) and (2·2). Under H_0, L_n has the same asymptotic distribution as $l^{\mathrm{T}} D^{-1} m$, where $D = \mathrm{diag}\{\alpha\}$ as before, and $l^{\mathrm{T}} = (1, -1, -1, 1)$. Since m is asymptotically $N(0, \Omega)$, $l^{\mathrm{T}} D^{-1} m$ is asymptotically $N(0, l^{\mathrm{T}} D^{-1} \Omega D^{-1} l)$. If consecutive trials are independent, so that $P = 1\alpha^{\mathrm{T}}$ then $\Omega = D - \alpha\alpha^{\mathrm{T}}$, whence $l^{\mathrm{T}} D^{-1} m \sim N(0, l^{\mathrm{T}} D^{-1} l)$, under H_0, corresponding for example, to the result given by Plackett (1981, p. 43), since $l^{\mathrm{T}} D^{-1} l = \alpha_1^{-1} + \alpha_2^{-1} + \alpha_3^{-1} + \alpha_4^{-1}$. Define

$$X_l^2 = \left\{\log\left(\frac{n_1 n_4}{n_3 n_2}\right)\right\}^2 \Big/ \left(\frac{1}{n\alpha_1} + \frac{1}{n\alpha_2} + \frac{1}{n\alpha_3} + \frac{1}{n\alpha_4}\right)$$

as the usual logistic χ^2 statistic for the 2×2 table.

For the Markov dependence model, let $P_{\mathscr{X}}$, $P_{\mathscr{Y}}$ be the transition matrices of the processes $\mathscr{X}$ and $\mathscr{Y}$ respectively, with corresponding nonunit eigenvalues λ, μ. Let $l_{\mathscr{X}}^{\mathrm{T}} = l_{\mathscr{Y}}^{\mathrm{T}} = (1, -1)$, thus $l_{\mathscr{X}}^{\mathrm{T}} P_{\mathscr{X}} = \lambda l_{\mathscr{X}}^{\mathrm{T}}$, $l_{\mathscr{Y}}^{\mathrm{T}} P_{\mathscr{Y}} = \mu l_{\mathscr{Y}}^{\mathrm{T}}$. Under H_0, the observations $\{n_i\}$ are generated by the Markov chain with transition matrix $P = P_{\mathscr{X}} \otimes P_{\mathscr{Y}}$, where $\otimes$ denotes the direct product. Since $l^{\mathrm{T}} = l_{\mathscr{X}}^{\mathrm{T}} \otimes l_{\mathscr{Y}}^{\mathrm{T}}$, it can be seen from (2·4) that

$$l^{\mathrm{T}} Z = (1 - \lambda\mu)^{-1} l^{\mathrm{T}}. \tag{4·3}$$

Now $P_{\mathscr{X}}, P_{\mathscr{Y}}$ each correspond to 2-state Markov processes, thus each is reversible, and hence it is easily checked that their product P is also reversible. Hence

$$\begin{aligned} l^{\mathrm{T}} D^{-1} \Omega D^{-1} l &= l^{\mathrm{T}} D^{-1}(2DZ - D - 1\alpha^{\mathrm{T}}) D^{-1} l \\ &= l^{\mathrm{T}}(2Z - I) D^{-1} l = \frac{1 + \mu\lambda}{1 - \mu\lambda} l^{\mathrm{T}} D^{-1} l, \end{aligned}$$

so that under H_0, in distribution as $n \to \infty$

$$L_n \to N\left\{0, \frac{1 + \lambda\mu}{1 - \lambda\mu}(l^{\mathrm{T}} D^{-1} l)\right\}; \tag{4·4}$$

hence, under H_0, $\{(1 - \lambda\mu)/(1 + \lambda\mu)\} X_l^2$ is approximately χ_1^2, in the case of Markovian dependence between consecutive trials. Note that H_0 implies that $l^{\mathrm{T}} D^{-1} l = (\alpha_1 \alpha_4)^{-1}$.

Not surprisingly, the same correction factor of $(1 - \lambda\mu)/(1 + \lambda\mu)$ applies also to the traditional χ^2 statistic C_n at (4·1). Write C_n in the form

$$C_n = \frac{n(n_1 n_4/n^2 - n_2 n_3/n^2)^2}{n^{-4}(n_1 + n_2)(n_1 + n_3)(n_2 + n_4)(n_3 + n_4)}.$$

Under H_0, the denominator converges in probability to $\alpha_1 \alpha_4$, and it is easily checked that the numerator has the same asymptotic distribution as $\alpha_1^2 \alpha_4^2 L_n^2$. Putting these together gives Theorem 2.

THEOREM 2. *Let $\mathscr{X}$ and $\mathscr{Y}$ be independent stationary two-state Markov chains with transition matrices $P_{\mathscr{X}}, P_{\mathscr{Y}}$. Let λ and μ be the nonunit eigenvalues of $P_{\mathscr{X}}$ and $P_{\mathscr{Y}}$ respectively, and set $\gamma = (1 - \mu\lambda)/(1 + \mu\lambda)$. Then each of γC_n and γX_l^2 is asymptotically χ_1^2.*

Remark 4. If $\mathscr{X}$ and $\mathscr{Y}$ are independent trials processes, then $\lambda = \mu = 0$, and we recover the usual result. If $P_{\mathscr{X}}$ and $P_{\mathscr{Y}}$ are identical, then $\gamma = (1-\lambda^2)/(1+\lambda^2) < 1$, and we can confirm analytically the observations made by Gardner *et al.* (1982).

Remark 5. This result does provide a robustness property of the test in the case where just one of $\mathscr{X}$ and $\mathscr{Y}$ is an independent trials process. For then $\gamma = 1$, and the statistic is still asymptotically χ_1^2 in the presence of Markov dependence in one component process.

Remark 6. Note that if $\mathscr{X}$ has transition matrix

$$\begin{bmatrix} a_x & 1-a_x \\ 1-b_x & b_x \end{bmatrix}$$

then $\lambda = a_x + b_x - 1$. If data on the transitions for $\mathscr{X}$ are available, then λ can be estimated, and similarly for μ. If such data are not available, it may still be possible to get bounds for λ, μ and hence for γ. The usual test of independence in this context is described by Billingsley (1961a, p. 29).

Remark 7. A referee has pointed out that the correction factor $\gamma^{-1} = (1+\mu\lambda)/(1-\mu\lambda)$ in Theorem 2 is reminiscent of the factor $(1+\rho_1\rho_2)/(1-\rho_1\rho_2)$ given by Bartlett (1935). In this paper Bartlett showed that, if $\{x_r\}, \{y_r\}$ are two long sequences of real observations, which are stationary, with $\rho_1 = \text{corr}(x_r, x_{r+1})$, $\rho_2 = \text{corr}(y_r, y_{r+1})$, then the asymptotic variance of $\hat{\rho}$, the sample correlation coefficient between $\{x_r\}, \{y_r\}$, is $n^{-1}(1+\rho_1\rho_2)(1-\rho_1\rho_2)^{-1}$, if the sequence $\{x_r\}$ is independent of the sequence $\{y_r\}$. Hence, from Bartlett's result, we see that under this null hypothesis $n\hat{\rho}^2(1-\rho_1\rho_2)/(1+\rho_1\rho_2)$ is asymptotically χ_1^2.

It is not surprising that in the special case where $\{x_r\}, \{y_r\}$ are each 2-state Markov chain processes, then the last quantity is identical to γC_n. This is readily checked by evaluating $\hat{\rho}, \rho_1, \rho_2$ in this special case.

In this sense Theorem 2 could have been derived very simply from Bartlett's result, but a more general approach in terms of the distribution of the χ^2 statistic under Markov dependence is still illuminating.

Remark 8. Table 1 shows the results of a simulation to find the values of the statistic C_n from 5000 runs of length n of two independent 2-state Markov chains with identical transition matrices and nonunit eigenvalue λ. This table gives the percentages of observations in which C_n exceeded the nominal 5% and 1% points of the χ_1^2 distribution, and these can be compared with the figures in brackets, which are the expected percentages from Theorem 2.

Table 1. *The percentages of simulations in which C_n exceeded the* 5% *and* 1% *points of the χ_1^2 distributions, and the corresponding expected percentages in brackets, for two independent identical* 2*-state Markov chains each with nonunit eigenvalue λ, with* 5000 *runs each of length n in each case*

	$\lambda = -0{\cdot}5$	$\lambda = -0{\cdot}25$	$\lambda = 0$	$\lambda = 0{\cdot}25$	$\lambda = 0{\cdot}5$		$\lambda = -0{\cdot}5$	$\lambda = -0{\cdot}25$	$\lambda = 0$	$\lambda = 0{\cdot}25$	$\lambda = 0{\cdot}5$
5%	(12·86)	(6·58)	(5·00)	(6·58)	(12·86)	1%	(4·56)	(1·56)	(1·00)	(1·56)	(4·56)
$n = 75$	14·30	7·30	6·14	6·52	12·56	$n = 75$	4·82	1·88	1·30	1·56	4·94
$n = 100$	12·74	6·86	4·72	7·22	12·74	$n = 100$	4·64	1·32	1·16	1·74	4·48
$n = 150$	13·08	6·08	5·28	6·52	12·66	$n = 150$	4·46	1·60	1·16	1·38	4·54
$n = 200$	13·92	6·78	5·10	7·32	12·64	$n = 200$	4·44	1·70	0·92	1·70	4·36

5. Discussion

The results derived here show how misleading the application of standard tests can be in the presence of serially correlated observations. For reversible chains, for example 2-state chains, explicit results are available for the true asymptotic distribution of the test statistics. If the data on transitions in the original sequences $\mathscr{X}$ and $\mathscr{Y}$ are available, the relevant eigenvalues can be estimated consistently. If not, then great care must be taken in the subsequent analysis.

The results given for testing independence of two 2-state processes $\mathscr{X}$ and $\mathscr{Y}$ say, can be readily extended to testing independence of the r-state process $\mathscr{X}$ and the c-state process $\mathscr{Y}$ say, provided that each of $\mathscr{X}$ and $\mathscr{Y}$ is reversible. The relevant eigenvalues of the matrix corresponding to $D^{-1}\Omega$ in (3·2) will then be

$$\left\{\frac{1+\lambda_i\mu_j}{1-\lambda_i\mu_j};\ i=1,\ldots,r-1;\ j=1,\ldots,c-1\right\},$$

where $\lambda_1,\ldots,\lambda_{r-1}$ are the nonunity eigenvalues of the transition matrix of $\mathscr{X}$, and $\mu_1,\ldots,\mu_{c-1}$ the corresponding quantities for $\mathscr{Y}$. Details of this will be given elsewhere.

We are grateful to Professor L. J. Snell for helpful comments, and to Mr W. Gardner for letting us see his results prior to publication.

References

ALTHAM, P. M. E. (1976). Discrete variable analysis for individuals grouped into families. *Biometrika* **63**, 263–9.

ALTHAM, P. M. E. (1979). Detecting relationships between categorical data observed over time: A problem of deflating a χ^2 statistic. *Appl. Statist.* **28**, 115–25.

BARTLETT, M. S. (1935). Some aspects of the time-correlation problem in regard to tests of significance. *J. R. Statist. Soc.* **98**, 536–43.

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

BASAWA, I. V. & 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.

BILLINGSLEY, P. (1961b). Statistical methods in Markov chains. *Ann. Math. Statist.* **32**, 12–40.

CASTELLAN, N. J. (1979). The analysis of behavior sequences. In *The Analysis of Social Interaction,* Ed. R. B. Cairns, pp. 81–116. Hillsdale, New Jersey: Lawrence Earlbaum Assoc.

COHEN, J. E. (1976). The distribution of the χ^2 statistic under clustered sampling from contingency tables. *J. Am. Statist. Assoc.* **71**, 665–70.

GARDNER, W. P., HARTMANN, D. P. & MITCHELL, C. (1982). The effects of serial dependence on the use of χ^2 for analysing data in dyadic interactions. *Behavioral Assessment* **4**, 75–82.

HOLT, D., SCOTT, A. J. & EWINGS, P. D. (1980). Chi-squared tests with survey data. *J. R. Statist. Soc.* A **143**, 303–20.

PATANKAR, V. N. (1954). The goodness of fit of frequency distributions obtained from stochastic processes. *Biometrika* **41**, 450–62.

PLACKETT, R. L. (1981). *The Analysis of Categorical Data,* 2nd edition. London: Griffin.

RAO, J. N. K. & 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. Am. Statist. Assoc.* **76**, 221–30.

[*Received March* 1981. *Revised May* 1982]