David Edwards

Large Sample Tests for Stationarity and Reversibility in Finite Markov Chains

Preprint
February
1979

1

Institute of Mathematical Statistics
University of Copenhagen

David Edwards

LARGE SAMPLE TESTS FOR STATIONARITY AND REVERSIBILITY IN FINITE MARKOV CHAINS

Preprint 1979 No. 1

INSTITUTE OF MATHEMATICAL STATISTICS

UNIVERSITY OF COPENHAGEN

February 1979

Abstract

Large sample tests for stationarity and reversibility, based on repeated observations of a discrete time, time-homogeneous Markov chain with finite state space, are proposed. The asymptotic distributions of the maximum likelihood estimator of the parameters of the chain under stationarity and reversibility are obtained, and explicit maximum likelihood estimates under reversibility are given for the case when the observed samples are of length two. An application to social mobility tables is given.

Key words: Markov chain, stationarity, reversibility, Wald test, likelihood ratio test.

1. Introduction and summary

When repeated sample sequences are observed from a discrete time, finite Markov chain, the initial states can be regarded as fixed (that is to say, we condition on them) or stochastic, generated by the stationary distribution of the chain. Estimation and inference in the latter case are more difficult since explicit maximum likelihood estimators do not exist and mumerical methods must be used. But if stationarity holds and the samples are short, considerable loss of information may result if the initial states are considered fixed. Besides, stationarity is often an interesting hypothesis in itself. For both reasons it is desirable to be able to test stationarity.

Similar considerations apply to the property of reversibility, which may be implied by time-symmetry or correspond to an interesting hypothesis in the theoretical context (see § 6). In the special case of samples of length 2, explicit M.L. estimators under reversibility are given.

Here Wald tests, which do not require the M.L. estimates and likelihood ratio tests, which do, are derived. Likelihood ratio tests follow directly from the asymptotic theory of exponential families and Wald tests can easily be derived once the asymptotic distribution of a certain parameter has been obtained. The Wald test statistics are closely related to the asymptotic distributions of the M.L.E. under stationarity and reversibility, which are also derived. Comparison is made with another approach to testing stationarity suggested by Guilbaud (1977). An application of the Wald tests is made to social mobility tables, taken from Bartholomew (1973).

2. Notation and concepts

Suppose we observe m sample sequences $(X_{i,1},\dots,X_{i,n_i})$, $i=1,\dots m$, from an s-state Markov chain with transition probability matrix $P=(p_{ij})$. Let $u_j = \#\{i \mid X_{i1}=j\}$ be the number of sequences with initial state $j, j=1,\dots s$, $q=(q_1,\dots,q_s)$ be the initial distribution, and $N=(n_{ij})$ be the transition count, i.e. defined by $n_{ij} = \#\{(s,t) \mid X_{s,t}=i, X_{s,t+1}=j \ (s=1,\dots m, t=1,\dots n_s-1)\}$, the number of observed transitions $i \sim j$. Then we can write the log-likelihood as

$$\log L = \sum_j u_j \log q_j + \sum_{i,j} n_{ij} \log p_{ij}. \qquad (1)$$

Consider the parameter $Q=(q_{ij})$ defined by $q_{ij}=q_i p_{ij}$. Since $q_{i\cdot}=q_i p_{i\cdot}=q_i$ and $p_{ij}=q_{ij}/q_i=q_{ij}/q_{i\cdot}, \forall_{i,j}$, we have defined a function that is bijective from $\Theta=\{(P,q) \mid p_{ij} \geqq 0, \ \forall i,j;\ pi\cdot=1, \forall i;\ q_i>0\ \forall i;\ q\cdot=1\}$ to $\tilde{\Theta}=\{Q \mid q_{ij} \geqq 0, \ \forall i,j;\ qi\cdot>0\ \forall i; q_{\cdot\cdot}=1\}$. We assume throughout that $(P,q)\in\Theta$, from which it follows that $P(u_j>0,\forall j) \to 1$ as $m\to\infty$.

Writing (1) with the new parametrisation, we obtain

$$\begin{aligned}\log L &= \sum_j u_j \log q_{j\cdot} + \sum_{ij} n_{ij} (\log q_{ij} - \log q_{i\cdot}) \\ &= \sum_j (u_j - n_{j\cdot}) \log q_{j\cdot} + \sum_{ij} n_{ij} \log q_{ij} \ . \qquad (2)\end{aligned}$$

Let H denote the hypothesis that the initial distribution is arbitrary, and H_o the hypothesis that the initial distribution is stationary. Then since $q_j = \sum_i q_i p_{ij}\ \forall j \Leftrightarrow q_{j\cdot}=q_{\cdot j}, \forall j$, we see that this latter set of constraints specifies H_o.

Let H_1 denote the hypothesis that the chain is reversible. A finite Markov chain is said to be reversible if for any sequence of states $X_1,\dots,X_r$, $Pr(X_1,\dots,X_r) = Pr(X_r,\dots,X_1)$, and such chains are characterised by being stationary and satisfying $q_i p_{ij} = q_j p_{ji}$ $\forall i,j$, where q is the stationary distribution. Thus H_1 is specified by $q_{ij} = q_{ji}$ $\forall i,j$, and $H_1 \subseteq H_o \subset H$. Stationarity and reversibility are equivalent for 2-state chains, as can easily be seen from this formulation.

3. Likelihood Ratio Tests

We recall the definitions of a regular canonic exponential family and of a differentiable exponential family (see Johansen (1975), Andersen (1969) or Berk (1972)). A regular canonic exponential family upon a topological space X with measure μ is characterised by the family of densities with respect to μ:

$$\frac{\exp(\theta'\cdot t(x))}{\phi(\theta)} \qquad \theta \in D, \tag{3}$$

where $t: X \to \mathbb{R}^k$ is a mapping such that $\{1, t_1(x),\dots,t_k(x)\}$ are linearly independent $[\mu]$, $\phi(\theta) = \int_X \exp(\theta'\cdot t(x))\mu(dx)$ and $D = \{\theta \mid \phi(\theta) < \infty\}$ is open. A differentiable exponential family is characterised by the family of densities

$$\frac{\exp(\pi(\beta)'\cdot t(x))}{\phi(\pi(\beta))} \qquad \beta \in I, \tag{4}$$

where $\pi: \mathbb{R}^m \to \mathbb{R}^k$ $(m < k)$ is a twice differentiable homeomorphism such that $(\frac{d\pi}{d\beta})$ has full rank, and I is an open set in $\mathbb{R}^m$ such that $\pi(I) \subset D$, and where $\exp(\theta'\cdot t(x))/\phi(\theta)$, $\theta \in D$, is a regular canonic family as defined above.

By writing the likelihood (2) as

$$L = q_{s.}^{m-n..}\; q_{ss}^{n..}\; \exp\Big[\sum_{j=1}^{s-1} (u_j - n_{j.}) \log(q_{j.}/q_{s.})$$

$$+ \sum_{(i,j)\neq(s,s)} n_{ij} \log(q_{ij}/q_{ss})\Big], \qquad (5)$$

and observing that $q_{j.}/q_{s.} = (q_{j.}/q_{ss})/(1 + \sum_{t=1}^{s-1} q_{st}/q_{ss})$, so that $\log(q_{j.}/q_{s.})$ is a function of $\{q_{ij}/q_{ss}\}$, $(i,j)\neq(s,s)$, we see immediately that (5) defines a differentiable exponential family with parameter $\beta = \{q_{ij}/q_{ss}\}$ $(i,j)\neq(s,s)$ whose domain is $I = \mathbb{R}_+^{s^2-1}$. Moreover H_0 and H_1 define differentiable hypotheses (subfamilies).

Let $\hat{Q}$, $\hat{Q}_0$ and $\hat{Q}_1$ be the M.L. estimators of Q under H, H_0 and H_1 respectively. In general only $\hat{Q}$ can be given explicitly. Under H, $\hat{Q}$ is asymptotically normal and efficient in the class of Fisher consistent estimators for Q. Similarly $\hat{Q}_0$ under H_0 and $\hat{Q}_1$ under H_1. Let Q_2 be a simple hypothesis within H_1 and let $\ell(\hat{Q}) = -2 \log L(\hat{Q})$, $\ell(\hat{Q}_0) = -2 \log L(\hat{Q}_0)$, etc. Then we can write the following analysis of deviance tables, when all components of Q are positive and $m \to \infty$.

Table 1

Hypothesis	Test statistic	degrees of freedom
Test for stationarity	$\ell(\hat{Q}) - \ell(\hat{Q}_0)$	$s-1$
Test for reversibility	$\ell(\hat{Q}_0) - \ell(\hat{Q}_1)$	$\frac{1}{2}s(s-1)$
Test for simple hypothesis	$\ell(\hat{Q}_1) - \ell(Q_2)$	$\frac{1}{2}s(s-1)$
Total	$\ell(\hat{Q}) - \ell(Q_2)$	s^2-1

For example, under H_1 $\ell(\hat{Q}_0) - \ell(\hat{Q}_1)$ is asymptotically distributed with the chi-squared distribution with $\frac{s}{2}(s-1)$ degrees of freedom. Similarly for Q_3, a simple hypothesis within H_0, we can write:

Table 2

Hypothesis	Test statistic	degrees of freedom
Test for stationarity	$\ell(\hat{Q}) - \ell(\hat{Q}_0)$	s-1
Test for simple hypothesis	$\ell(\hat{Q}_0) - \ell(Q_3)$	s(s-1)
Total	$\ell(Q) - \ell(Q_3)$	s^2-1

In families of chains where not all transition probabilities are positive it is easy to derive the degrees of freedom by inspection. Let $N = \#\{(i,j) : p_{ij} > 0\}$. Then H specifies $N-s+(s-1) = N-1$ linearly independent parameters, H_0 specifies N-s, and H_1 specifies (N-s)/2. Thus in table 1 we replace the last column by $(s-1, \frac{1}{2}(N-s), \frac{1}{2}(N-s), N-1)'$.

As an example consider an $r^{\underline{th}}$ order chain on t states governed by the t^{r+1} transition probabilities $p_{i_1 \ldots i_{r+1}} = P_r(x_{r+1} = i_{r+1} | x_j = i_j, j=1,\ldots r)$. Then as usual (see f.x. Billingsley, (1961)) we identify the chain with a $1^{\underline{st}}$ order chain on the t^r r-tuples $\underline{i} = (i_1,\ldots,i_r)$ where the probability of transition $\underline{i} \sim \underline{j}$ for $\underline{j} = (j_1,\ldots,j_r)$ is only nonzero if $i_k = j_{k-1}$ $(k=2,\ldots r)$. So here $N = t^{r+1}$, $s = t^r$ and the last column in table 1 becomes $(t^r - 1, \frac{1}{2}t^r(t-1), \frac{1}{2}t^r(t-1), t^{r+1}-1)'$.

Notice that for chains of higher order than 1 we have to revise the characterisation of reversibility: if $\underline{i} = (i_1,\ldots,i_r)$ and $\underline{j} = (i_2,\ldots,i_{r+1})$ we require, instead of $q_{\underline{i}\underline{j}} = q_{\underline{j}\underline{i}}$, $q_{\underline{i}\underline{j}} = q_{\underline{j}^*\underline{i}^*}$

$\forall \underline{i},\underline{j}$, where $\underline{i}^* = (i_r,\ldots,i_1)$ and $\underline{j}^* = (i_{r+1},\ldots,i_2)$, i.e. $\underline{i}$ and $\underline{j}$ in reverse order. This does not affect the degrees of freedom. Notice also that for higher order 2-state chains stationarity is still equivalent to reversibility in this sense.

4. Maximum Likelihood Estimators

Under H the MLE's are of course given by $q_{ij} = \dfrac{u_i n_{ij}}{m\, n_{i.}}$. In general there do not exist explicit MLE's for Q under H_0 and H_1. An important exception to this is under H_1, reversibility, when all the samples are of length 2. We now derive these estimators.

We have now $u_j = n_{j.}, \forall j$, so we can write from (2)

$$\begin{aligned} \log L &= \sum_{ij} n_{ij} \log q_{ij} \\ &= \sum_{ij} n_{ji} \log q_{ij} \qquad \text{(since } q_{ij} = q_{ji} \text{ under } H_1\text{)} \\ &= \sum_{ij} \tfrac{1}{2}(n_{ij} + n_{ji}) \log q_{ij} \qquad \text{(by addition)}. \qquad (6) \end{aligned}$$

Thus $\{\frac{1}{2}(n_{ij} + n_{ji})\}$, $i \leqq j$ say, is sufficient. Furthermore, by writing (6) as

$$L = q_{ss}^{n..} \exp \{ \sum_{i=1}^{s-1} \tfrac{1}{2}(n_{ii} + n_{ii}) \log q_{ii}/q_{ss}$$

$$+ \sum_{i<j\leqq s-1} \tfrac{1}{2}(n_{ij} + n_{ji}) \log q_{ij}/q_{ss}\}$$

we see that we are now within a regular canonic exponential family, and so the maximum likelihood estimator is given uniquely as the solution of the equations

$$E_Q[\tfrac{1}{2}(n_{ij} + n_{ji})] = \tfrac{1}{2}(n_{ij} + n_{ji}), \quad \text{ie.}$$

$\frac{m}{2}(q_{ij} + q_{ji}) = mq_{ij} = \frac{1}{2}(n_{ij} + n_{ji})$, giving the MLE's as

$$\hat{q}_{ij} = \frac{n_{ij} + n_{ji}}{2m}, \quad \text{so that}$$

$$\hat{p}_{ij} = \frac{n_{ij} + n_{ji}}{n_{i.} + n_{.i}} \quad \text{and} \quad \hat{q}_i = \frac{n_{i.} + n_{.i}}{2m}. \tag{7}$$

These formulae (7) have been derived previously for two state chains (H. Dalgas Christiansen (1978)). The author considers only stationarity (for two state chains equivalent to reversibility, as seen above). She then attempts to generalise the 'backwards-forwards' estimator $\frac{n_{i.} + n_{.i}}{2m}$ to chains with more than two states: but this is only justifiable under reversibility, where the MLE's are given above. The estimated chains proposed by Christiansen are not in general reversible.

5. Wald Tests

Let in general a composite hypothesis specify a parameter θ as being subject to k restrictions:

$$T_i(\theta) = 0, \quad i = 1, \ldots, k,$$

and let $\hat{\theta}$ be the unrestricted and $\hat{\theta}^*$ the restricted MLE of θ. Then the Wald test of the hypothesis is

$$W_c = \sum_{ij} \lambda^{ij}(\hat{\theta}) T_i(\hat{\theta}) T_j(\hat{\theta}),$$

where (λ^{ij}) is the reciprocal of (λ_{ij}), the asymptotic covariance matrix of $T(\hat{\theta}) = (T_1(\hat{\theta}), \ldots, T_k(\hat{\theta}))'$. If $\hat{\theta}$ is asymptotically normal then $W_c \sim \chi^2_{(k)}$ under the hypothesis. Notice that W_c uses only the unrestricted MLE, $\hat{\theta}$.

We construct Wald tests for the hypotheses $H_0 \subset H$, $H_1 \subset H$ and $H_1 \subset H_0$. We do not consider here families of chains in which some subset of $\{q_{ij}\}$ is always zero. Here $\theta = Q$ and the hypotheses H_0 and H_1 are specified by the linear functions $S_i'Q \equiv q_{i.} - q_{.i} = 0$, $i = 1, \ldots, s-1$, and $R_{ij}'Q \equiv q_{ij} - q_{ji} = 0$, $1 \leqq i < j \leqq s$, respectively. Here Q is considered as a $s^2 \times 1$ column vector. Let also $S = (S_1, \ldots, S_{s-1})$, $R = (R_{12}, \ldots, R_{s-1,s})$, and B be the $s^2 \times 1$ vector with unit components.

We now derive the distribution of $\hat{Q}$ under H. Let I be the information matrix with respect to Q, i.e. the $s^2 \times s^2$ matrix with $((i,j),(k,\ell))^{\underline{th}}$ element

$$I_{(i,j),(k,\ell)} = E\left(-\frac{\partial^2 \log L}{\partial q_{ij}\partial q_{k\ell}}\right) = \delta_{ik}\left(\frac{E(u_i - n_{i.})}{q_{i.}^2} + \delta_{j\ell}\frac{En_{ij}}{q_{ij}^2}\right), \quad (8)$$

from (2). δ_{ik} is the Kronecker delta. Since Q under H satisfies $B'Q = 0$, the asymptotic covariance matrix $\Sigma_{\hat{Q}}$ of $\hat{Q}$ under H is given by (see Silvey (1970), p.81)

$$\begin{pmatrix} \Sigma_{\hat{Q}} & C \\ C & D \end{pmatrix} = \begin{pmatrix} I & B \\ B' & 0 \end{pmatrix}^{-1} \quad (9)$$

for some C and D. When the appropriate inverses exist, the solution to (9) is given by

$$\Sigma_{\hat{Q}} = I^{-1} - I^{-1}B(B'I^{-1}B)^{-1}(I^{-1}B)', \quad (10)$$

$$C = -I^{-1}B(B'I^{-1}B)^{-1}, \quad (11)$$

and $$D = (B'I^{-1}B)^{-1}, \quad (12)$$

(see Rao (1973), p.33). Consider the sxs matrix $A_i = (a^i_{k\ell})$ defined by

$$a^i_{k\ell} = x_i + \delta_{k\ell}/e_{ik} \qquad k,\ell = 1,\dots,s,$$

for some constants x_i and e_{ik}, $k = 1,\dots,s$. It is easy to show that A_i is invertible with inverse $(a_i^{k\ell})$ given by

$$a_i^{k\ell} = \delta_{k\ell} e_{ik} - x_i e_{ik} e_{i\ell}/(1 + x_i \Sigma_j e_{ij}). \tag{13}$$

Let now $x_i = E(u_i - n_{i.})/q_{i.}^2$ and $e_{ij} = q_{ij}^2/En_{ij} = q_i q_{ij}/En_{i.}$, so that from (8) $I_{(i,j)k,\ell)} = \delta_{ik} a^i_{k\ell}$. Thus $I = \text{diag}\ (A_1,\dots,A_s)$, $I^{-1} = \text{diag}\ (A_1^{-1},\dots,A_s^{-1})$ and we can write the $((i,j),(k,\ell))^{\underline{th}}$ element of I^{-1} as

$$I^{(i,j),(k,\ell)} = \delta_{ik} q_{ij} q_{i\ell} \left(\frac{1}{Eu_i} - \frac{1}{En_{i.}}\right) + \delta_{ik}\delta_{j\ell} \frac{q_{i.} q_{ij.}}{En_{i.}} \tag{14}$$

Hence $$(I^{-1}B)_{(i,j)} = q_{ij} q_{i.} \left(\frac{1}{Eu_i} - \frac{1}{En_{i.}}\right) + \frac{q_{i.} q_{ij}}{En_{i.}} = \frac{q_{ij}}{m} \tag{15}$$

and $$B' I^{-1} B = q_{..}/m = 1/m\ , \tag{16}$$

so that $D = m$, $C = -Q$, and we obtain from (10) that the $((i,j),(k,\ell))^{\underline{th}}$ element of $\Sigma_{\hat{Q}}$ is given as

$$\sigma_{(i,j)(k,\ell)} = \delta_{ik} q_{ij} q_{i\ell} \left(\frac{1}{Eu_i} - \frac{1}{En_{i.}}\right) + \delta_{ik}\delta_{k\ell} \frac{q_{i.} q_{ij}}{En_{i.}} - \frac{q_{ij}}{m}\, q_{k\ell}$$

$$= q_{ij} q_{k\ell} \left(\delta_{ik}\left(\frac{1}{Eu_i} - \frac{1}{En_{i.}} + \frac{\delta j\ell}{En_{ij}}\right) - \frac{1}{m}\right) . \tag{17}$$

Let $(\gamma_{ij}) = S'\Sigma_{\hat{Q}}S$ and $(\varphi_{(i,j),(k,\ell)})_{(i<j,k<\ell)} = R'\Sigma_{\hat{Q}}R$ be the asymptotic covariance matrices of $S'\hat{Q}$ and $R'\hat{Q}$ respectively, under H. For the sake of reference we write out the formulae.

$$\gamma_{ij} = -\frac{1}{m}(q_{i.} - q_{.i})(q_{j.} - q_{.j}) - q_{ji}q_{j.}\left(\frac{1}{Eu_j} - \frac{1}{En_{j.}}\right) - q_{ij}q_{i.}\left(\frac{1}{Eu_i} - \frac{1}{En_{i.}}\right)$$

$$+ \delta_{ij}\left(\frac{q_{i.}^2}{Eu_i} + \sum_k \frac{q_{ik}^2}{En_{ki}}\right) - \frac{q_{ki}^2}{En_{ji}} - \frac{q_{ij}^2}{En_{ij}} \quad , \qquad (18)$$

$$\varphi_{(i,j)(k,\ell)} = q_{ij}(\delta_{ik}q_{k\ell} - \delta_{i\ell}q_{ik})\left(\frac{1}{Eu_i} - \frac{1}{En_{i.}}\right)$$

$$+ q_{ji}(\delta_{j\ell}q_{\ell k} - \delta_{jk}q_{k\ell})\left(\frac{1}{Eu_j} - \frac{1}{En_{j.}}\right) - (q_{ij} - q_{ji})\frac{(q_{k\ell} - q_{\ell k})}{m} \quad ,$$

$$((i,j) \neq (k,\ell)), \qquad (19)$$

and

$$\varphi_{(i,j),(i,j)} = q_{ij}^2\left(\frac{1}{Eu_i} - \frac{1}{En_{i.}} + \frac{1}{En_{ij}} - \frac{1}{m}\right)$$

$$+ \frac{2q_{ij}q_{ji}}{m} + q_{ji}^2\left(\frac{1}{Eu_j} - \frac{1}{En_{j.}} + \frac{1}{En_{ji}} - \frac{1}{m}\right). \qquad (20)$$

When the samples are all of length 2, formulae (17) - (20) become

$$\sigma_{(i,j),(k,\ell)} = \frac{q_{ij}}{m}(\delta_{ij}\delta_{k\ell} - q_{k\ell}), \qquad (21)$$

(the well-known multinomial expression),

$$\gamma_{ij} = \frac{1}{m}[\delta_{ij}(q_{i.} + q_{.i}) - (q_{i.} - q_{.i})(q_{j.} - q_{.j}) - q_{ij} - q_{ji}] \qquad (22)$$

$$\varphi_{(i,j)(k,\ell)} = -\frac{1}{m}(q_{ij} - q_{ji})(q_{k\ell} - q_{\ell k}), \quad (\ell,j) \neq (k,\ell) \qquad (23)$$

and

$$\varphi_{(i,j)(i,j)} = \frac{1}{m}[q_{ij}(1 - q_{ij}) + q_{ji}(1 - q_{ji}) + 2q_{ij}q_{ji}] \qquad (24)$$

Let $\hat{\Sigma}_{\hat{Q}}$ be the estimate of $\Sigma_{\hat{Q}}$ formed by substituting the observed values of u_i, n_{ij} etc. for their expected values, in (17). Then the Wald test for $H_0 \subset H$ is given by

$$W_0 = (S'\hat{Q})'(S'\hat{\Sigma}_{\hat{Q}}S)^{-1}(S'\hat{Q}), \qquad (25)$$

which has asymptotically a $\chi^2_{(s-1)}$ distribution under H_0. Similarly the Wald test for $H_1 \subset H$ is given by

$$W_1 = (R'\hat{Q})^1 (R'\Sigma_{\hat{Q}} R)^{-1} (R'\hat{Q}) \tag{26}$$

which is asymptotically $\chi^2_{\frac{1}{2}s(s-1)}$ under H_1.

To derive the Wald test for $H_1 \subset H_0$ we need the asymptotic covariance matrix $\Sigma_{\hat{Q}_0}$ of $\hat{Q}_0$. As in (9) this is given by

$$\begin{pmatrix} \Sigma_{\hat{Q}_0} & C_0 \\ C_0' & D \end{pmatrix} = \begin{pmatrix} I & B & S \\ B' & 0 & 0 \\ S & 0 & 0 \end{pmatrix}^{-1} = \left(\begin{array}{cc|c} I_0 & & S \\ & & 0 \\ \hline S' & 0 & 0 \end{array}\right)^{-1} , \tag{27}$$

where $I_0 = \begin{pmatrix} I & B \\ B' & 0 \end{pmatrix}$. After a little manipulation using (10) we obtain

$$\Sigma_{\hat{Q}_0} = \Sigma_{\hat{Q}} - (\Sigma_{\hat{Q}} S)(S'\Sigma_{\hat{Q}} S)^{-1} (\Sigma_{\hat{Q}} S)' . \tag{28}$$

Similarly the asymptotic covariance matrix $\Sigma_{\hat{Q}_1}$ of $\hat{Q}_1$ under H_1 is given by

$$\Sigma_{\hat{Q}_1} = \Sigma_{\hat{Q}} - (\Sigma_{\hat{Q}} R)(R'\Sigma_{\hat{Q}} R)^{-1} (\Sigma_{\hat{Q}} R)' . \tag{29}$$

We can now construct the Wald test for $H_1 \subset H_0$. Since $q_{i.} - q_{.i} = 0 \; \forall i$, $\{q_{ij} - q_{ji}\}$, $1 \leqq i < j \leqq s$ are not linearly independent under H_0. It is easily seen though that $\{q_{ij} - q_{ji}\}$, $1 \leqq i < j \leqq s-1$ are linearly independent under H_0 and that $q_{ij} - q_{ji} = 0$, $1 \leqq i < j \leqq s-1$, specify H_1 under H_0. Let therefore $R^* = (R_{12}, R_{13}, \ldots, R_{s-2,s-1})$. Then $R^{*\prime}\Sigma_{\hat{Q}_0} R^*$ is nonsingular and the Wald test for $H_1 \subset H_0$ is given by

$$W_2 = (R^{*'} \hat{Q}_0)(R^{*'} \hat{\Sigma}_{\hat{Q}_0} R^{*})^{-1}(R^{*'} \hat{Q}_0)^{1} \quad (30)$$

which is asymptotically $\chi^2_{\frac{1}{2}(s-2)(s-1)}$ under H_1.

We mention briefly another approach to testing for stationarity suggested implicitly by Guilbaud (1977). His paper is concerned with the asymptotic distribution of $\hat{p}$, the eigenvector of $\hat{P}$, under H, i.e. the MLE of the stationary distribution under H. The results are primarily for samples of length 2 but can be extended to samples of greater but equal length. He derives the asymptotic distribution of $\hat{q} - \hat{p}$, i.e. the difference between the initial and stationary distribution estimates, and constructs confidence regions for $q - p$ by means of the Bonferroni inequality. If they do not include zero he concludes that the chain is not stationary. Use of the Bonferroni inequality involves loss of power and so the tests given here are to be preferred. However, the asymptotic distribution of $\hat{q} - \hat{p}$ could be used to construct a test for $q = p$ to which the criticism of loss of power would not apply. It is difficult to evaluate the relative merits of these tests. From a practical point of view the Wald test is easiest to perform.

6. An Application

Bartholomew (1973, chapter 2), discusses extensively data given by Glass and Hall (Glass (1954)), which provide the basis for a social mobility model. The data consists of a sample of 3500 father/son pairs classified by the father's and son's social categories, (7 categories). A Markov chain model is assumed and Bartholomew remarks that the data seem to come from a stationary chain, but that 'a complete answer to the question of sampling error is not available'. Later he also discusses reversibility, which in this context means equal expected exchange between classes, i.e. $mq_{ij} = mq_{ji}$, $\forall i,j$. He calculates (ibid.,p.39), the matrix $(q^*_{ij}) = (\hat{p}_i \hat{p}_{ij})$, where $\hat{p}$ is as before the MLE of the stationary distribution under H, and examines it for symmetry. This has a curious interpretation. It is concievable that observed sample sequences from a stationary chain, due perhaps to some sampling scheme, do not have the stationary distribution as their initial one. In this case one could not test for stationarity, but assuming stationarity one could test for reversibility by examining whether $p_i p_{ij} = p_j p_{ji}$ $\forall i,j$. This would require the asymptotic distribution of (q^*_{ij}).

Wald tests for stationarity and reversibility (within H) were performed on the data, and the estimates $\hat{P}$, $\hat{q}$, $\hat{Q}$, $S'\hat{\Sigma}_{\hat{Q}}S$ and $(S'\hat{\Sigma}_{\hat{Q}}S)^{-1}$ are given in table 3. The test statistics were $W_0 = 38.70$ and $W_1 = 51.24$, which are highly significant ($P < 0.0001$). Guilbaud (1977) also analysed this data and concluded that stationarity must be rejected. Likelihood ratio tests were not performed due to the difficulty of determining the MLE under stationarity.

Table 3

$$\hat{P} = \begin{bmatrix} 0.388 & 0.146 & 0.202 & 0.062 & 0.140 & 0.047 & 0.015 \\ 0.107 & 0.267 & 0.227 & 0.120 & 0.206 & 0.053 & 0.020 \\ 0.035 & 0.101 & 0.188 & 0.191 & 0.357 & 0.067 & 0.061 \\ 0.021 & 0.039 & 0.112 & 0.212 & 0.430 & 0.124 & 0.062 \\ 0.009 & 0.024 & 0.075 & 0.123 & 0.473 & 0.171 & 0.125 \\ 0.000 & 0.013 & 0.041 & 0.088 & 0.391 & 0.312 & 0.155 \\ 0.000 & 0.008 & 0.036 & 0.083 & 0.364 & 0.235 & 0.274 \end{bmatrix}$$

$$\hat{q} = (0.037,\ 0.043,\ 0.098,\ 0.148,\ 0.432,\ 0.131,\ 0.111)'$$

$$\hat{Q} = \begin{bmatrix} 0.014 & 0.005 & 0.007 & 0.002 & 0.005 & 0.002 & 0.001 \\ 0.005 & 0.011 & 0.010 & 0.005 & 0.009 & 0.002 & 0.001 \\ 0.003 & 0.010 & 0.018 & 0.019 & 0.035 & 0.007 & 0.006 \\ 0.003 & 0.006 & 0.017 & 0.031 & 0.064 & 0.018 & 0.009 \\ 0.004 & 0.010 & 0.032 & 0.053 & 0.204 & 0.074 & 0.054 \\ 0.000 & 0.002 & 0.005 & 0.012 & 0.051 & 0.041 & 0.020 \\ 0.000 & 0.001 & 0.004 & 0.009 & 0.040 & 0.026 & 0.030 \end{bmatrix}$$

$$m.(S'\hat{\Sigma}_{\hat{Q}}S) = \begin{bmatrix} 0.038 & 0.010 & 0.011 & 0.006 & 0.009 & 0.001 \\ 0.010 & 0.066 & 0.020 & 0.011 & 0.019 & 0.004 \\ 0.011 & 0.020 & 0.155 & 0.035 & 0.067 & 0.012 \\ 0.006 & 0.011 & 0.035 & 0.216 & 0.117 & 0.029 \\ 0.009 & 0.019 & 0.067 & 0.117 & 0.431 & 0.124 \\ 0.001 & 0.004 & 0.012 & 0.029 & 0.124 & 0.218 \end{bmatrix}$$

$$m^{-1}(S'\hat{\Sigma}_{\hat{Q}}S)^{-1} = \begin{bmatrix} 34.314 & 10.696 & 8.092 & 6.477 & 5.667 & 4.973 \\ 10.696 & 22.250 & 7.846 & 6.385 & 5.611 & 4.975 \\ 8.092 & 7.846 & 12.204 & 6.242 & 5.524 & 4.855 \\ 6.477 & 6.385 & 6.242 & 9.756 & 5.469 & 4.934 \\ 5.667 & 5.611 & 5.524 & 5.469 & 6.434 & 4.851 \\ 4.973 & 4.975 & 4.855 & 4.934 & 4.851 & 8.419 \end{bmatrix}$$

References

Andersen, A.H. (1969). Asymptotic results for exponential families, Bull. Inst. Internat. Statist. 2, 241-2.

Bartholomew, David (1973). Stochastic Models for Social Processes, Wiley, New York.

Berk, R.H. (1972). Consistency and asymptotic normality of MLE's for exponential models, Ann. Math. Stat. 43, 1, 193-204.

Billingsley, P. (1961). Statistical methods in Markov chains, Ann. Math. Stat. 32, 12-40.

Dalgas Christensen, H. (1978). Estimation of transition matrices for Markov processes in equilibrium, unpublished manuscript.

Glass, D.V. (Ed)(1954). Social Mobility in Britain, Routledge and Paul, London.

Guilbaud, O. (1977). Estimating the probability eigenvalue and related characteristics of an ergodic transition matrix, Scand. J. Statist., 4, 97-104.

Johansen, Søren (1975). Eksponentielle familier, lecture notes, Institute for Mathematical Statistics, Copenhagen.

Rao, C.R. (1973). Linear Statistical Inference and its Applications, Wiley, New York.

Silvey, S.D. (1970). Statistical Inference, Harmondsworth, Penguin.

PREPRINTS 1978

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Tjur, Tue: Statistical Inference under the Likelihood Principle.

No. 2 Hering, Heinrich: The Non-Degenerate Limit for Supercritical Branching Diffusions.

No. 3 Henningsen, Inge: Estimation in M/G/1-Queues.

No. 4 Braun, Henry: Stochastic Stable Population Theory in Continuous Time.

No. 5 Asmussen, Søren: On some two-sex population models.

No. 6 Andersen, Per Kragh: Filtered Renewal Processes with a Two-Sided Impact Function.

No. 7 Johansen, Søren & Ramsey, Fred L.: A Bang-Bang Representation for 3x3 Embeddable Stochastic Matrix.

No. 8 Braun, Henry: A Simple Method for Testing Goodness of Fit in the Presence of Nuisance Parameters.

No. 9 Lauritzen, S.L. & Speed, T.P. & Vijayan, K. : Decomposable Graphs and Hypergraphs.

No. 10 Hald, Anders: On the Statistical Theory of Sampling Inspection by Attributes.

No. 11 Darroch, J.N. & Lauritzen, S.L. & Speed, T.P.: Markov Fields and Log-linear Interaction Models for Contingency Tables.

No. 12 Speed, T.P.: A Note on Nearest-Neighbour Gribbs and Markov Probabilities.

No. 13 Andersson, Steen A.: Canonical Correlations with Respect to a Complex Structure.

PREPRINTS 1979

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Edwards, David: Large Sample Tests for Stationarity and Reversibility in Finite Markov Chains.