Probability in the Engineering and Informational Sciences, **26**, 2012, 593–611.
doi:10.1017/S0269964812000228

TESTING FOR REVERSIBILITY IN MARKOV CHAIN DATA

TARA L. STEUBER, PETER C. KIESSLER AND ROBERT LUND
Department of Mathematical Sciences
Clemson University
O-106 Martin Hall
Clemson, SC 29634-0975
E-mails: tsteube@clemson.edu; kiesslp@clemson.edu; lund@clemson.edu

This paper introduces two statistics that assess whether (or not) a sequence sampled from a stationary time-homogeneous Markov chain on a finite state space is reversible. The test statistics are based on observed deviations of transition sample counts between each pair of states in the chain. First, the joint asymptotic normality of these sample counts is established. This result is then used to construct two chi-squared-based tests for reversibility. Simulations assess the power and type one error of the proposed tests.

1. INTRODUCTION

Let $\{X_t\}_{t=1}^{L}$ be a realization of a time-homogeneous Markov chain on the finite state space $S = \{1, 2, \ldots, m\}$ with one-step-ahead transition probability matrix $\mathbf{P} = (p_{i,j})_{i,j \in S}$ with entries $p_{i,j} = P[X_{n+1} = j | X_n = i]$. Without a priori information, it is natural to estimate $p_{i,j}$ by

$$\hat{p}_{i,j}^{(N)} = \frac{N_{i,j}}{N_i} 1_{[N_i > 0]},$$

where $N_{i,j} = N_{i,j}(L)$ is the number of times the chain transitions from state i to state j in one step and $N_i = N_i(L)$ is the number of times the chain visits state i in a data realization of length L. The authors of [1,3–5] explore properties of $\hat{p}_{i,j}^{(N)}$ in depth. Some of their results are reviewed/stated in the next section. We will interchangeably use notations such as N_i and $N_i(L)$ below, the latter being preferred when emphasis on the sample size is paramount.

© Cambridge University Press 2012 0269-9648/12 $25.00

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

Suppose that the chain is aperiodic and irreducible. Due to the finite state space, the chain is positive recurrent and admits a unique stationary distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$ (this is also a limiting distribution). The chain is called reversible if it satisfies the so-called detailed balance equations

$$\pi_i p_{i,j} = \pi_j p_{j,i} \quad \text{for all} \quad i,j \in S. \tag{1}$$

Metropolis-Hastings Markov chain Monte Carlo (MCMC) chains, birth and death chains, and random walks on graphs are known to be reversible. In many settings, reversibility can be rationalized through physical reasoning without computation of $\boldsymbol{\pi}$. More examples of reversible Markov chains can be found in [6,11–15].

Knowing whether or not a chain is reversible is advantageous. First, if a chain is known to be reversible, more accurate estimators of the chain's one-step-ahead transition probabilities can be constructed than $\hat{p}_{i,j}^{(N)}$. This issue is discussed in detail in [2]. Second, a reversible chain contains fewer parameters than a non-reversible chain. Elaborating, there are $m(m-1)$ free parameters in the one-step-ahead transition matrix of a chain whose state space has cardinality m (one free parameter is lost in each row since all transition matrix row sums are unity). However, if a chain is known to be reversible, there are only $(m^2 - m + 2)/2$ free parameters due to the restrictions in (1). Hence, a reversible chain on m states has $(m^2 - m - 2)/2$ fewer free parameters than a general chain. Third, the results here can be used in diagnostics. For example, if a supposed Metropolis-Hastings generated chain flunks our reversibility test(s), then its generation algorithm would be called into question as these chains are known a priori to be reversible. Finally, reversibility can be indicative of data properties. For example, Richman and Sharp [12] argues that the successive bands comprising the most desirable types of granite should be reversible; Edwards [8] studies the evolution of a family's social status by examining father-son pairs for reversibility.

Testing for Markov chain reversibility has been previously considered by other authors. For example, Diaconis and Rolles [6] uses a Bayesian approach but does not develop a level-α test; [8], perhaps the work closest to ours, considers the problem and develops asymptotic results from many independent short series runs with a fixed initial distribution. Our work here differs in that we take one trajectory of the chain and assume that the series is sampled from its stationary state. Hence, some of our estimators differ from those in [8], most notably those of the limiting distribution of the chain. While our test statistic T_1 below is similar to the count deviation statistics proposed in [12,14], we are careful to account for all correlations between the different deviations that comprise our statistics and derive a correct degrees of freedom in our chi-squared test. Also, Richman and Sharp [12] considers a test statistic similar to T_1 below; however, no simplifying limit laws are derived there.

The rest of this paper proceeds as follows. Section 2 introduces two test statistics that assess chain reversibility. Section 3 establishes their asymptotic distributions under a null hypothesis of reversibility. Section 4 presents a simulation study showing the efficiency of the statistics in identifying reversibility. Section 5 presents some concluding remarks and an Appendix details two technical calculations.

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

2. REVERSIBILITY TEST STATISTICS

Suppose that state i is visited k times in the first $L-1$ time units; that is, $N_i(L-1)=k$ (we work with time $L-1$ instead of time L because we do not have an observed transition from time L to $L+1$). Then $\mathbf{N}_i=(N_{i,1},\ldots,N_{i,m})'$ is a multivariate multinomial random variable with k trials and success probability vector $\mathbf{p}_i=(p_{i,1},\ldots,p_{i,m})'$. Using this and the central limit theorem for renewal sequences, [1,3,4], and [5] argue that for each $i,j\in S$,

$$\sqrt{L}(\hat{p}_{i,j}^{(N)}(L)-p_{i,j})\xrightarrow{\mathcal{D}} N\left(0,\frac{p_{i,j}(1-p_{i,j})}{\pi_i}\right)$$

as $L\to\infty$. For joint inferences, they also show that

$$\sqrt{L}\begin{pmatrix}\hat{p}_{i,j}^{(N)}(L)-p_{i,j}\\ \hat{p}_{i',j'}^{(N)}(L)-p_{i',j'}\end{pmatrix}\xrightarrow{\mathcal{D}} N\left(\begin{pmatrix}0\\0\end{pmatrix},\boldsymbol{\Sigma}_P\right), \tag{2}$$

where

$$\boldsymbol{\Sigma}_P=\begin{pmatrix}p_{i,j}(1-p_{i,j})/\pi_i & \delta_{i,i'}p_{i,j}(\delta_{j,j'}-p_{i,j'})/\pi_i\\ \delta_{i,i'}p_{i,j}(\delta_{j,j'}-p_{i,j'})/\pi_i & p_{i',j'}(1-p_{i',j'})/\pi_{i'}\end{pmatrix}$$

and $\delta_{i,j}=\mathbf{1}_{[i=j]}$ is the Kronecker delta indicator. For completeness, the convergence in (2) is proven in Appendix A using elementary principles. One motivation for proving (2) is that most of the above cited authors only state the result, and the reader may be surprised that $\hat{p}_{i,j}^{(N)}$ and $\hat{p}_{i',j'}^{(N)}$ are asymptotically uncorrelated when $i\neq i'$. In fact, since the chain cannot be in state i' when it is in state i, one might erroneously rationalize some type of negative dependence.

When the chain is reversible, there are actually more efficient estimators of $p_{i,j}$ than $\hat{p}_{i,j}^{(N)}$. In fact, Annis et al. [2] and Greenwood and Wefelmeyer [10] show that the symmetrized estimator

$$\hat{p}_{i,j}^{(R)}=\frac{N_{i,j}+N_{j,i}}{2N_i}1_{[N_i>0]}$$

is asymptotically at least as efficient as $\hat{p}_{i,j}^{(N)}$ and quantify any efficiency gains. For example, $\hat{p}_{i,j}^{(N)}$ and $\hat{p}_{i,j}^{(R)}$ have unit efficiency when the chain is skip free. Skip free means that when the chain is in state i, the next transition must go to either state $i-1$ or $i+1$. We estimate π_i with $\hat{\pi}_i=N_i(L)/L$.

In view of (1), the deviations

$$\hat{\pi}_i\hat{p}_{i,j}^{(N)}-\hat{\pi}_j\hat{p}_{j,i}^{(N)}=\frac{N_{i,j}-N_{j,i}}{L} \tag{3}$$

should be small for each pair (i,j) with $i<j$ when the chain is reversible. Hence, the quadratic form

$$\sum_{i<j}\sum\left(\frac{N_{i,j}-N_{j,i}}{L}\right)^2 \tag{4}$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

should be statistically small under reversibility. The sum in (4) is inconvenient to quantify asymptotically. This is because the quantities $C_{i,j} := (N_{i,j} - N_{j,i})/L$ are correlated for varying pairs (i,j) with $j > i$.

To quantify the correlations between $C_{i,j}$ and $C_{i',j'}$ when $(i,j) \neq (i',j')$, we will derive the asymptotic distribution of the $m(m-1)/2$ dimensional count discrepancy vector

$$\mathbf{C} = \frac{1}{L}\begin{bmatrix} N_{1,2} - N_{2,1} \\ \vdots \\ N_{1,m} - N_{m,1} \\ N_{2,3} - N_{3,2} \\ \vdots \\ N_{2,m} - N_{m,2} \\ \vdots \\ N_{m-1,m} - N_{m,m-1} \end{bmatrix} \tag{5}$$

and construct a test statistic for reversibility from a quadratic form based on $\mathbf{C}$. This statistic is similar to the quadratic forms in [12,14]; however, the degrees of freedom we obtain for the asymptotic chi-squared statistic differs.

For notation, the element of $\mathbf{C}$ corresponding to $L^{-1}(N_{i,j} - N_{j,i})$ is denoted as $C_{i,j}$. While this indexes a vector with bivariate subscripts, the notation is natural given the transition count components it involves. Let $\mathbf{\Sigma_C} = \lim_{L\to\infty} L\mathrm{Var}(\mathbf{C})$ denote the asymptotic information matrix of $\mathbf{C}$. If one can establish asymptotic normality of $\mathbf{C}$, then the quadratic form statistic

$$T_1 = L\mathbf{C}'\hat{\mathbf{\Sigma}}_{\mathbf{C}}^{+}\mathbf{C}. \tag{6}$$

will have an asymptotic χ^2 distribution, where $\hat{\mathbf{\Sigma}}_{\mathbf{C}}$ is an element-by-element estimator of $\mathbf{\Sigma_C}$. The degrees of freedom of T_1 will be the rank of $\mathbf{\Sigma_C}$, which is not necessarily $m(m-1)/2$ (this corrects [14]). Specifically, $\mathbf{\Sigma_C}$ is not always invertible and the notation $\hat{\mathbf{\Sigma}}_{\mathbf{C}}^{+}$ signifies the estimator for the Moore–Penrose pseudoinverse in (6). These aspects are elaborated upon in detail below.

A second test statistic for reversibility is the maximum of the square of terms in (4); specifically,

$$T_2 = \max_{i<j} D_{i,j}^2, \tag{7}$$

where $D_{i,j}$ is the (i,j)th component in $\mathbf{D}$ defined by $\mathbf{D} = \sqrt{L}(\hat{\mathbf{\Sigma}}_{\mathbf{C}}^{+})^{1/2}\mathbf{C}$. Observe that $\mathbf{D}$ is an $m(m-1)/2$ dimensional vector with components indexed akin to those in $\mathbf{C}$. We show below that T_2 converges asymptotically to the max of κ independent random variables, each of which has a chi-squared distribution with one degree of freedom. Here, κ is the rank of $\mathbf{\Sigma_C}$.

Small values of T_1 and T_2 suggest the null hypothesis of reversibility; that is, reversibility is rejected when T_1 and/or T_2 are too big (statistically). The statistic T_1

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

is an overall sum of deviations, while T_2 is a maximal deviation. Set $\Delta_{i,j} = \pi_i p_{i,j} - \pi_j p_{j,i}$. Should the $\Delta_{i,j}$s be significantly non-zero for all pairs (i,j), then T_1 will be very powerful as it aggregates squared deviations (accumulates information) over all (i,j) pairs. However, if in truth, only a few of the $\Delta_{i,j}$s are non-zero, then the T_2 statistic will isolate on the best (i,j) pair where deviations occur; power is not lost by including pairs (i,j) where no (or very small) differences are taking place. It is not possible to say that one statistic is always more powerful than the other; however, the examples in Section 4 will show that T_1 often works best.

To quantify how big the statistics need to be to warrant rejection of reversibility, we now derive the asymptotic distributions of T_1 and T_2 as $L \to \infty$ under the null hypothesis of reversibility. In the analysis below, we assume that the data sequence $X_1, \ldots, X_L$ is drawn from a time-homogeneous chain that is in its stationary state. As the initial state will not influence limiting behavior, one can apply the results below to any arbitrary initial state when L is large.

3. ASYMPTOTIC DISTRIBUTION OF THE TEST STATISTICS

This section derives the asymptotic distributions of T_1 and T_2 when the chain is reversible. We begin with the following lemma that quantifies the joint normality of the $N_{i,j}(L)/L$s. The result does not readily follow from the joint asymptotic normality of the $\hat{p}_{i,j}$s in (2) because of the randomness in the $\hat{\pi}_i$s. Although joint normality holds for any collection of the $N_{i,j}/L$s, we state the result for two (i,j) pairs only for notational convenience. The proof of the result is presented in Appendix B.

LEMMA 3.1: *Suppose that $\{X_t\}_{t=1}^{L}$ is a sample from a time-homogeneous aperiodic irreducible Markov chain on the finite state space S whose cardinality is $m < \infty$. Then*

$$\sqrt{L}\begin{pmatrix} \dfrac{N_{i,j}(L)}{L} - \pi_i p_{i,j} \\ \dfrac{N_{i',j'}(L)}{L} - \pi_{i'} p_{i',j'} \end{pmatrix} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{\Sigma_N}),$$

where

$$\mathbf{\Sigma_N} = \begin{pmatrix} \gamma_{i,j}^2 & \gamma_{ij,i'j'} \\ \gamma_{ij,i'j'} & \gamma_{i',j'}^2 \end{pmatrix}.$$

Here,

$$\gamma_{i,j}^2 = 2\pi_i p_{i,j} r_{j,i} p_{i,j} + \pi_i p_{i,j}(1 - \pi_i p_{i,j});$$

$$\gamma_{ij,i'j'} = \pi_i p_{i,j} r_{j,i'} p_{i',j'} + \pi_{i'} p_{i',j'} r_{j',i} p_{i,j} - \pi_i p_{i,j} \pi_{i'} p_{i',j'};$$

$$\gamma_{i',j'} = 2\pi_{i'} p_{i',j'} r_{j',i'} + 2\pi_{i'} p_{i',j'}(1 - \pi_{i'} p_{i',j'});$$

and $r_{k,\ell}$ is the (k,ℓ)th entry in the matrix $\mathbf{R} = (\mathbf{I}_{m\times m} - \mathbf{P} + \mathbf{\Pi}_{m\times m})^{-1}$ with $\mathbf{\Pi}_{m\times m}$ being an $m \times m$ dimensional matrix, with each row containing the limiting distribution $\boldsymbol{\pi}$, and $\mathbf{I}_{m\times m}$ is the m-dimensional identity matrix.

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

Notice that for a chain whose state space has cardinality m, $\mathbf{\Sigma_N}$ is an $m^2 \times m^2$ matrix.

We will need to derive forms of $\mathbf{\Sigma_C}$ and $\mathbf{\Sigma_N}$ under a null hypothesis of reversibility. These quantities are denoted by $\mathbf{\Sigma}_\mathbf{C}^{(R)}$ and $\mathbf{\Sigma}_\mathbf{N}^{(R)}$, respectively. To compute these information matrices under reversibility, one simply replaces $\pi_j p_{j,i}$ with $\pi_i p_{i,j}$ when $j > i$; these expressions are left unaltered when $j \leq i$. For example, under reversibility, the limiting information for $N_{1,2}$ is written as $\pi_1 p_{1,2} r_{2,1} p_{1,2} + \pi_1 p_{1,2}(1 - \pi_1 p_{1,2})$; the limiting information for $N_{2,1}$ is written as $2\pi_2 p_{2,1} r_{1,2} p_{2,1} + \pi_2 p_{2,1}(1 - \pi_2 p_{2,1})$ in preference to $2\pi_1 p_{1,2} r_{1,2} p_{2,1} + \pi_1 p_{1,2}(1 - \pi_1 p_{1,2})$. The estimators $\hat{\mathbf{\Sigma}}_\mathbf{N}^{(R)}$ and $\hat{\mathbf{\Sigma}}_\mathbf{C}^{(R)}$ are computed by plugging in $\hat{\pi}_i$ for π_i and $\hat{p}_{i,j}^{(N)}$ for $p_{i,j}$ under the above "reflection scheme".

THEOREM 3.2: *Under the assumptions of Lemma* 3.1, $T_1 = L\mathbf{C}'(\hat{\mathbf{\Sigma}}_\mathbf{C}^{(R)})^+\mathbf{C} \xrightarrow{\mathcal{D}} \chi^2_\kappa$ *as* $L \to \infty$, *where* κ *is the rank of* $\mathbf{\Sigma_C}$. *In the case where* $\mathbf{P}$ *has no off-diagonal zero elements or other parameter restrictions,* $\kappa = (m-1)(m-2)/2$.

PROOF: Throughout this proof, we assume that the null hypothesis of reversibility is in force. Then $\pi_i p_{i,j} = \pi_j p_{j,i}$ for all $(i,j) \in \{1, \ldots, m\}$. Let

$$\mathbf{N} = L^{-1} \begin{bmatrix} N_{1,1} - L\pi_1 p_{1,1} \\ N_{1,2} - L\pi_1 p_{1,2} \\ \vdots \\ N_{m,m} - L\pi_m p_{m,m} \end{bmatrix}.$$

Observe that $\mathbf{C}$ is a linear transformation of $\mathbf{N}$ and write $\mathbf{C} = \mathbf{VN}$, where $\mathbf{V}$ is a $m(m-1)/2 \times m^2$ dimensional matrix whose only non-zero entries are positive or negative one. The form of $\mathbf{V}$ is unweildy but the ones and minus ones appear in the following pattern. $\mathbf{V}$ is written as $m-1$ blocks $\mathbf{b}(1), \ldots, \mathbf{b}(m-1)$, with block i containing $m-i$ rows, for $i = 1, 2, \ldots, m-1$. That is, $\mathbf{V}$ has the form

$$\mathbf{V} = \begin{bmatrix} \mathbf{b}(1) \\ \mathbf{b}(2) \\ \vdots \\ \mathbf{b}(m-1) \end{bmatrix}.$$

Row r of the matrix $\mathbf{V}$ is in block i if $(i-1)(m - i/2) + 1 \leq r \leq im - i(i+1)/2$. For block i, the entries in $\mathbf{b}(i)$ are

$$\mathbf{b}(i)_{k,\ell} = \begin{cases} 1, & \text{if } \ell = (i-1)m + i + k \\ -1, & \text{if } \ell = (i-1)m + i + km \\ 0, & \text{otherwise} \end{cases}.$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

When $m = 3$, for example, we have

$$\mathbf{C} = L^{-1}\begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix} \begin{pmatrix} N_{1,1} - L\pi_1 p_{1,1} \\ N_{1,2} - L\pi_1 p_{1,2} \\ N_{1,3} - L\pi_1 p_{1,3} \\ N_{2,1} - L\pi_2 p_{2,1} \\ N_{2,2} - L\pi_2 p_{2,2} \\ N_{2,3} - L\pi_2 p_{2,3} \\ N_{3,1} - L\pi_3 p_{3,1} \\ N_{3,2} - L\pi_3 p_{3,2} \\ N_{3,3} - L\pi_3 p_{3,3} \end{pmatrix}.$$

Under the reversible null, $\sqrt{L}\mathbf{N} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{N}}^{(R)})$ by Lemma 3.1 and hence $\sqrt{L}\mathbf{C} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{C}}^{(R)})$, where $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)} = \mathbf{V}\boldsymbol{\Sigma}_{\mathbf{N}}^{(R)}\mathbf{V}'$. Theorem 4.7.1 in [9] and Slutzky's theorem give

$$(\sqrt{L}\mathbf{C}^T)(\hat{\boldsymbol{\Sigma}}_{\mathbf{C}}^{(R)})^+(\sqrt{L}\mathbf{C}) = L\mathbf{C}^T(\hat{\boldsymbol{\Sigma}}_{\mathbf{C}}^{(R)})^+\mathbf{C} \xrightarrow{\mathcal{D}} \chi^2_\kappa,$$

where κ is the rank of $\boldsymbol{\Sigma}_C^{(R)}$.

To finish the proof, we need to determine the rank of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$. This calculation caused the authors considerable consternation, but is hueristically argued in [8] to be $\kappa = (m-1)(m-2)/2$ in the case where there are no zero one-step-ahead transition probability entries. Zero entries on the main diagonals of $\mathbf{P}$ do not alter the degrees of freedom as the $N_{i,i}$s do not appear in either T_1 or T_2. We provide the following reasoning for the $(m-1)(m-2)/2$ degrees of freedom.

Since $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is an $m(m-1)/2 \times m(m-1)/2$ dimensional matrix, the rank of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is $m(m-1)/2$ minus the rank of the null space of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$. Perhaps surprisingly, the rank of the null space of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is not zero. To see this, we must show that there exist non-zero vectors $\mathbf{u}$ such that $\boldsymbol{\Sigma}_C^{(R)}\mathbf{u} = \mathbf{0}$ and identify how many linearly independent such $\mathbf{u}$ exist.

To do this, let $N_{i,\cdot} = \sum_{k=1}^m N_{i,k}$ and $N_{\cdot,j} = \sum_{k=1}^m N_{k,j}$. Then $N_{i,\cdot}$ is the total number of transitions out of state i in the data record and $N_{\cdot,i}$ is the total number of transitions into state i in the data record. Physical reasoning shows that $|N_{i,\cdot} - N_{\cdot,i}| \leq 1$ for each $i \in \{1, 2, \ldots, m\}$. As we show below, the lack of full rank for $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ arises from these restrictions (and not reversibility).

For $i \in \{1, 2, \ldots, m\}$, define the $m(m-1)/2$ dimensional vector $\mathbf{u}^{(i)}$ componentwise via

$$u_{h,\ell}^{(i)} = \begin{cases} 1, & \text{if } h = i \text{ and } \ell \in \{i+1, i+1, \ldots, m\}, \\ -1, & \text{if } \ell = i \text{ and } h \in \{i-1, i-2, \ldots, 1\} \\ 0, & \text{otherwise} \end{cases}.$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

Here, we have indexed $\mathbf{u}^{(i)}$ with two components (h, ℓ) drawn over the set $\{1, 2, \ldots, m\}$ with $h < \ell$. For example, when $m = 3$,

$$\mathbf{u}^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}^{(2)} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{u}^{(3)} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}.$$

We now show that $\mathbf{u}^{(i)}$ belongs to the null space of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}$ for $i = 1, 2, \ldots, m$. To see this, observe that for $i = 1, 2, \ldots, m$,

$$N_{i,\cdot} - N_{\cdot,i} = \sum_{j>i} (N_{i,j} - N_{j,i}) - \sum_{j<i} (N_{j,i} - N_{i,j}). \tag{8}$$

The element of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}$ corresponding to $C_{h,\ell}$ and $C_{h',\ell'}$ is

$$\lim_{L\to\infty} L^{-1}\mathrm{E}[(N_{h,\ell} - N_{\ell,h})(N_{h',\ell'} - N_{\ell',h'})] = \lim_{L\to\infty} LE[C_{h,\ell}C_{h',\ell'}].$$

Applying (8) shows that the (h, ℓ)th element of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}\mathbf{u}^{(i)}$ is

$$\lim_{L\to\infty} L^{-1} \left(\sum_{j>i} \mathrm{E}[(N_{h,\ell} - N_{\ell,h})(N_{i,j} - N_{j,i})] - \sum_{j<i} \mathrm{E}[(N_{h,\ell} - N_{\ell,h})(N_{j,i} - N_{i,j})] \right)$$

for $i = 1, 2, \ldots, m$. To show that this is zero, use Eq. (8) to get that the (h, ℓ)th element of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}\mathbf{u}^{(i)}$ is

$$\begin{aligned} &\lim_{L\to\infty} L^{-1}\mathrm{E}\left[(N_{h,\ell} - N_{\ell,h}) \left(\sum_{j>i} (N_{i,j} - N_{j,i}) - \sum_{j<i} (N_{j,i} - N_{i,j}) \right) \right] \\ &\quad = \lim_{L\to\infty} L^{-1}\mathrm{E}[(N_{h,\ell} - N_{\ell,h})(N_{i\cdot} - N_{\cdot i})]. \end{aligned}$$

Since $|N_{i,\cdot} - N_{\cdot,i}| \le 1$ and $\mathrm{Var}\left(\frac{N_{h,\ell}-N_{\ell,h}}{\sqrt{L}}\right) < \infty$ (by Lemma 3.1), the Cauchy–Schwarz inequality gives

$$\begin{aligned} \lim_{L\to\infty} L^{-1}\mathrm{E}[(N_{h,\ell} - N_{\ell,h})(N_{i\cdot} - N_{\cdot i})] &= \lim_{L\to\infty} \mathrm{E}\left[\frac{(N_{h,\ell} - N_{\ell,h})}{\sqrt{L}} \frac{(N_{i\cdot} - N_{\cdot i})}{\sqrt{L}} \right] \\ &\le \lim_{L\to\infty} \left(\mathrm{Var}\left(\frac{N_{h,\ell} - N_{\ell,h}}{\sqrt{L}} \right) \right)^{1/2} \frac{1}{\sqrt{L}} \\ &= 0. \end{aligned}$$

Hence, the (h, ℓ)th element of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}\mathbf{u}^{(i)}$ is zero and $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}$ are in the null space of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}$.

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

It is easy to see that there are $m-1$ linearly independent vectors amongst $\mathbf{u}^{(1)},\ldots,\mathbf{u}^{(m)}$. This can be seen by putting all $\mathbf{u}^{(i)}$ as rows in a matrix and row reducing. A rank of one is lost from m because $\sum_{i=1}^{m}\mathbf{u}^{(i)}=\mathbf{0}$, which is a consequence of $\sum_{i=1}^{m}(N_{i,\cdot}-N_{\cdot,i})=0$.

It follows that the rank of the null space of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is at least $m-1$. When there are no other parametric constraints on the entries of $\mathbf{P}$ (such as zeros), there are no other elements in the null space of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ besides linear combinations of those illuminated above (the next section presents examples with zero transition matrix entries). In this case, we obtain $\kappa=m(m-1)/2-(m-1)=(m-1)(m-2)/2$, which completes our proof. ■

THEOREM 3.3: *As* $L\to\infty$, $T_2\xrightarrow{\mathcal{D}}\max_{1\le\ell\le\kappa}\chi_\ell^2$, *where* $\{\chi_\ell^2\}_{\ell=1}^{\kappa}$ *are independent and identically distributed chi-squared random variables, each having one degree of freedom, and* κ *is the rank of* $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$. *In the case where* $\mathbf{P}$ *has no off-diagonal zero elements or other parameter restrictions,* $\kappa=(m-1)(m-2)/2$.

PROOF: This result follows from the joint asymptotic normality of the $C_{i,j}$s and the fact that the $D_{i,j}$s are asymptotically uncorrelated. ■

4. SIMULATION EXAMPLES

This section presents a simulation study assessing the performance of T_1 and T_2 as test statistics for reversibility. We consider a variety of transition matrices, some reversible, some non-reversible, and some non-reversible but close to reversible. In each case, 10,000 independent chains were generated from each transition matrix with the sample lengths $L=250$, 1,000, and 2,500.

For reversible chains, one expects about 5% of the test statistics to exceed the 95% critical value. For non-reversible chains, one hopes for good power—that most of the test statistics exceed the 95% critical value.

We present six total examples. The first three examples have $m=3$ with no zeros or other parameter restrictions in $\mathbf{P}$. In this case, the rank of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is 1 and the 95% critical value for both T_1 and T_2 is 3.841. Example 4 takes $m=5$ with no zeros or other parameter restrictions in $\mathbf{P}$. Here, the rank of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is 6, the 95% critical value of T_1 is 12.592, and the 95% critical value of T_2 is 17.219. Example 5 considers a one-step-ahead transition matrix with several zero entries when $m=4$. Here, the rank of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is $\kappa=2$, which gives 95% critical values for T_1 and T_2 as 5.991 and 7.352, respectively. Example 6 has a larger state space with $m=16$ and zeros for all diagonal entries of $\mathbf{P}$. The rank of $\boldsymbol{\Sigma}_{\mathbf{C}}^{(R)}$ is 105, which gives 95% critical values of 129.9127 for T_1 and 12.1594 for T_2.

Our first example uses the transition matrix

$$\mathbf{P}=\begin{bmatrix}1/5 & 2/5 & 2/5\\ 2/5 & 1/5 & 2/5\\ 2/5 & 2/5 & 1/5\end{bmatrix}.$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

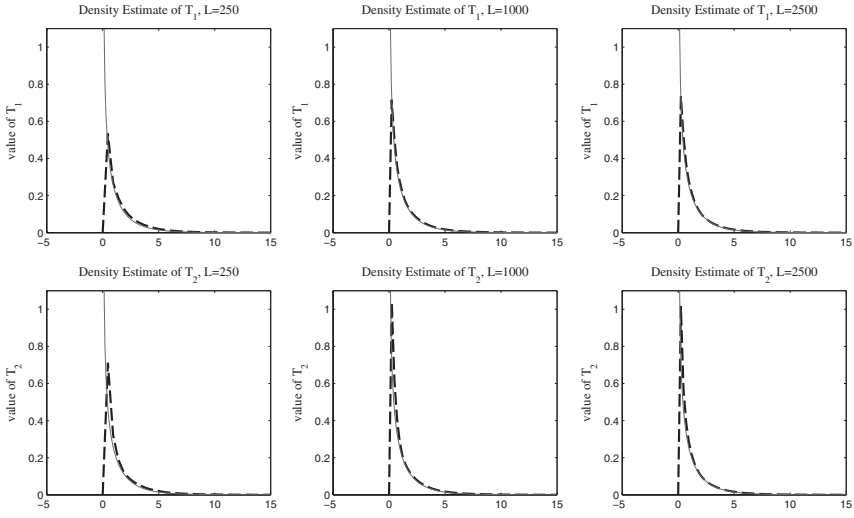


FIGURE 1. Sample kernel density estimates against the $\chi^2(1)$ density.

This chain is reversible, irreducible, doubly stochastic, and aperiodic with stationary distribution $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$. Figure 1 plots a kernel density estimate of the generated values of T_1 and T_2 against a chi-squared probability density function with one degree of freedom. Here, the Epanechnikov kernel function

$$K(x) = \frac{3(1-x^2)}{4}, \quad -1 \le x \le 1,$$

was used with a smoothing bandwidth of 0.25 in all plots. Because the densities in question are supported on $[0, \infty)$, we have wrapped any probability mass that is assigned to $(-\infty, 0)$ by the kernels back to the positive reals. As $L \to \infty$, it is seen that the test statistics match the χ^2 distribution with one degree of freedom reported in Theorems 3.2 and 3.3, the fit improving with increasing L (the fit is so good in some cases that differences are hard to discern from the graphics).

Our second example considers

$$\mathbf{P} = \begin{bmatrix} 1/6 & 1/3 & 1/2 \\ 5/8 & 3/16 & 3/16 \\ 5/24 & 23/48 & 5/16 \end{bmatrix}.$$

This chain is not reversible, but is irreducible and aperiodic. In this case, the generated values of T_1 and T_2 are consistently above the 95% rejection threshold, even for the smaller values of L. Table 1 shows empirical powers of rejection for various L. These powers are reasonable, becoming perfect when $L = 1000$. In our first two examples, the test statistics are making good conclusions.

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

TABLE 1. Empirical powers for Example 2

L	250	1,000	2,500
T_1	99.99%	100.00%	100.00%
T_2	99.99%	100.00%	100.00%

Our third example seeks to fool the methods by considering

$$\mathbf{P} = \begin{bmatrix} 8/15 & 1/3 & 2/15 \\ 9/33 & 20/33 & 4/33 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

This chain is irreducible and aperiodic, but is not reversible. The chain's stationary distribution is $\boldsymbol{\pi} = (55/144, 11/24, 23/144)$. While technically non-reversible, the equations in (1) are close to being satisfied. In particular, Table 2 shows values for both sides of the three reversible balance equations; two of these three equations "nearly hold".

Table 3 shows that the generated test statistics are close to zero and consistently below the 95% critical value. As expected, T_1 and T_2 have difficulty identifying non-reversibility with such small sample sizes.

To investigate this case more deeply, we ran additional simulations with larger sample sizes, particularly $L = 10{,}000$ and $L = 100{,}000$. The reversibility rejection powers reported in Table 4 show that the methods do distinguish reversibility from non-reversibility asymptotically, but that it takes a large sample size L to do this effectively.

Our fourth example considers a larger state space—one where $m = 5$. Here, states one through five correspond to weather conditions of sunny, rainy, foggy, cloudy, and

TABLE 2. Detailed balance equations for Example 3

$\pi_1 p_{1,2} = \dfrac{55}{432} \approx 0.1273148148$	$\pi_2 p_{2,1} = \dfrac{1}{8} \approx 0.125$
$\pi_1 p_{1,3} = \dfrac{1}{72} \approx 0.152777777$	$\pi_3 p_{3,1} = \dfrac{23}{432} \approx 0.0532407407$
$\pi_2 p_{2,3} = \dfrac{1}{18} \approx 0.055555555$	$\pi_3 p_{3,2} = \dfrac{23}{432} \approx 0.0532407407$

TABLE 3. Empirical powers for Example 3

L	250	1,000	2,500
T_1	7.88%	7.19%	9.45%
T_2	7.51%	7.14%	9.45%

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

Table 4. Empirical powers for Example 3 with larger L

L	10,000	100,000
T_1	19.67%	94.08%
T_2	19.65%	94.07%

Table 5. Empirical powers for Example 4

L	250	1,000	2,500
T_1	96.86%	100.00%	100.00%
T_2	61.14%	99.95%	100.00%

partly cloudy, respectively. We take the transition probabilities as

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/16 & 1/16 & 1/8 & 1/4 \\ 1/4 & 3/10 & 1/20 & 1/4 & 3/20 \\ 1/16 & 3/8 & 5/16 & 3/16 & 1/16 \\ 1/10 & 1/4 & 1/10 & 1/4 & 3/10 \\ 3/20 & 3/10 & 1/10 & 1/5 & 1/4 \end{bmatrix}.$$

This chain is not reversible, but is irreducible and aperiodic. Table 5 gives empirical powers of rejection at the 95% percentile. The numbers appear reasonable and the powers increase with increasing L. Here, T_2 performs much worse than T_1.

Our next example considers a case where the rank of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}$ is not the largest possible. We do this by taking $m = 4$ and imposing several off-diagonal elements of the transition matrix to be zero:

$$\mathbf{P} = \begin{bmatrix} 1/2 & 1/4 & 1/4 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/4 & 1/4 & 1/2 \end{bmatrix}.$$

This chain is reversible, aperiodic, irreducible, and doubly stochastic with $\boldsymbol{\pi} = (1/4, 1/4, 1/4, 1/4)$. Suppose that physical reasoning dictates that $p_{1,4} = p_{4,1} = 0$ but that the other transition probabilities are non-zero. Here, the rank of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}$ is $\kappa = 2$. This is seen by writing out the linear system for the asymptotic null space of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}$ as in the Proof of Theorem 1:

$$\begin{aligned} C_{1,2} + C_{1,3} + C_{1,4} &= 0 \\ -C_{1,2} + C_{2,3} + C_{2,4} &= 0 \\ -C_{1,2} - C_{2,3} + C_{3,4} &= 0 \\ -C_{1,4} - C_{2,4} - C_{3,4} &= 0 \end{aligned}$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

TABLE 6. Empirical type I errors for Example 5

L	250	1,000	2,500
T_1	8.84%	7.41%	6.72%
T_2	5.01%	3.93%	3.56%

However, since we know that $p_{4,1} = 0$ and $p_{1,4} = 0$, $N_{1,4} = N_{4,1} = 0$ and the above linear system reduces to

$$\begin{aligned} C_{1,2} + C_{1,3} &= 0 \\ -C_{1,2} + C_{2,3} + C_{2,4} &= 0 \\ -C_{1,2} - C_{2,3} + C_{3,4} &= 0 \\ -C_{1,4} - C_{2,4} - C_{3,4} &= 0, \end{aligned}$$

which is easily verified to have rank 4. Hence, the rank of $\mathbf{\Sigma}_{\mathbf{C}}^{(R)}$ is $\kappa = 6 - 4 = 2$.

The Type I error probabilities in Table 6 are close to the designed 5% for the T_1 statistic, but smaller than 5% for T_2.

Our last example gives a case with a much larger number of states. Suppose we have a connected graph with $m = 16$ equidistant nodes forming a hexadecagon as depicted in Example 4.32 of [13]. The weights on this graph are $w_{i,j} = \min\{|i - j|, 16 - |i - j|\}$ and represent the distance between (i,j) pairs of nodes. Consider the weights to be traveling costs between the sixteen locations, and let the probability of travel from node i to node j be $p_{i,j} = w_{i,j} / \sum_j w_{i,j} = w_{i,j}/64$. The one-step-ahead transition matrix is irreducible, aperiodic, symmetric, and doubly stochastic and the chain is known to be reversible [13]. The first row of the one-step-ahead transition matrix is

$$\frac{1}{16}[0 \;\; 1 \;\; 2 \;\; 3 \;\; 4 \;\; 5 \;\; 6 \;\; 7 \;\; 8 \;\; 7 \;\; 6 \;\; 5 \;\; 4 \;\; 3 \;\; 2 \;\; 1].$$

Notice that $p_{i,i} = 0$ for all i. Hence, there are zero entries in $\mathbf{P}$. However, since the diagonal counts $N_{i,i}$ do not arise in our tests statistics, the correct degrees of freedom is $\kappa = (m - 1)(m - 2)/2 = 105$.

Because of the large state space, a huge sample size is needed for the limit theory to kick in. The type one errors in Table 7 bear this out.

Overall, it seems that T_1 is superior to T_2, as can be seen by the slightly higher powers in the examples above when the chain is not reversible. Repeating earlier

TABLE 7. Empirical type I errors for Example 6

L	250	1,000	2,500	10,000	100,000	250,000	500,000
T_1	26.06%	25.52%	34.32%	15.11%	5.24%	5.24%	5.04%
T_2	95.81%	51.55%	33.82%	17.07%	7.04%	6.82%	6.14%

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

logic, we expect T_2 to work best (relative to T_1) when only one $\Delta_{i,j}$ is non-zero. When all $\Delta_{i,j}$s are non-zero, T_1 is expected to be superior. Overall, both statistics seem to function reasonably.

5. CONCLUDING REMARKS

This paper considered the problem of detecting reversibility in a Markov chain data sequence. Two statistics were proposed and their asymptotic properties were derived. Both statistics performed reasonably well in simulation studies; however, very large sample sizes are needed for the limit theory to hold when there is a large number of states in the chain. The crux of the mathematical analysis lied with determining the asymptotic rank of an information matrix of a quadratic form.

Acknowledgement

Robert Lund's work was supported by National Science Foundation Grant DMS 0905570. Comments from three referees substantially improved this manuscript.

References

1. Anderson, T. W. & Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics* 28: 89–110.
2. Annis, D. H., Kiessler, P. C., Lund, R. & Steuber, T. L. (2010). Estimation in reversible Markov chains. *The American Statistician* 64: 116–120.
3. Basawa, I. V. & Rao, B. L. S. (1980). *Statistical inference for stochastic processes*. New York: Academic Press, Inc.
4. Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics* 32: 12–40.
5. Derman, C. (1956). Some asymptotic distribution theory for Markov chains with a denumerable number of states. *Biometrika* 43: 285–294.
6. Diaconis, P. & Rolles, S. W. W. (2006). Bayesian analysis for reversible Markov chains. *The Annals of Statistics* 34: 1270–1292.
7. Di Cecco, D. (2011). Conditional exact tests for Markovianity and reversibility in multiple categorical sequences. *TEST* 21: 170–187.
8. Edwards, D. G. (1980). Large sample tests for stationarity and reversibility in finite Markov chains. *Scandinavian Journal of Statistics* 7: 203–206.
9. Graybill, F. A. (1976). *Theory and application of the linear model.* Pacific Grove, CA: Duxbury.
10. Greenwood, P. E. & Wefelmeyer, W. (1999). Reversible Markov chains and optimality of symmetrized empirical estimators. *Bernoulli*, 5, 109–123.
11. Kijima, M. (1997). *Markov processes for stochastic modeling*. London: Chapman & Hall.
12. Richman, D. & Sharp, W. E. (1990). A method for determining the reversibility of a Markov sequence. *Mathematical Geology* 22: 749–761.
13. Ross, S. M. (2007). *Introduction to Probability Models*, 9th ed. Burlington, MA: Academic Press.
14. Sharifdoost, M., Mahmoodi, S. & Pasha, E. (2009). A statistical test for time reversibility of stationary finite state Markov chains. *Applied Mathematical Sciences* 3: 2563–2574.
15. Stroock, D. (2005). *An introduction to Markov processes*. New York: Springer.
16. Ash, R.B. & Doleans-Dade, C.A. (2000). *Probability and measure theory*. New York: Academic Press Inc.

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

APPENDIX A

Proof of Eq. (2)

PROOF: Write the counts up to time $L-1$ (the value of X_L does not affect the asymptotic analysis) as

$$N_i(L-1) = \sum_{n=1}^{L-1} \mathbf{1}_{\{i\}}(X_n), \quad N_{i,j}(L-1) = \sum_{n=1}^{L-1} \mathbf{1}_{\{i,j\}}(X_n, X_{n+1}).$$

Then it is easy to see that $N_{i,j}(L-1) - N_i(L-1)p_{i,j} = \sum_{n=1}^{L-1} \mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j})$ has zero expectation. To handle the covariance term

$$E[\{N_{i,j}(L-1) - N_i(L-1)p_{i,j}\}\{N_{i',j'}(L-1) - N_{i'}(L-1)p_{i',j'}\}],$$

use

$$\begin{aligned}
&\sum_{n=1}^{L-1} \mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j}) \sum_{m=1}^{L-1} \mathbf{1}_{\{i'\}}(X_m)(\mathbf{1}_{\{j'\}}(X_{m+1}) - p_{i',j'}) \\
&\quad = \sum_{n=1}^{L-1} \mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j})\mathbf{1}_{\{i'\}}(X_n)(\mathbf{1}_{\{j'\}}(X_{n+1}) - p_{i',j'}) \\
&\qquad + \sum_{n=1}^{L-2} \sum_{m=n+1}^{L-1} \mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j})\mathbf{1}_{\{i'\}}(X_m)(\mathbf{1}_{\{j'\}}(X_{m+1}) - p_{i',j'}) \\
&\qquad + \sum_{m=1}^{L-2} \sum_{n=m+1}^{L-1} \mathbf{1}_{\{i'\}}(X_m)(\mathbf{1}_{\{j'\}}(X_{m+1}) - p_{i',j'})\mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j}).
\end{aligned}$$

Consider the first double summation in the expression above. Taking expectations, applying $n < m$, and the Markov property gives

$$\begin{aligned}
&\mathrm{E}[\mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j})\mathbf{1}_{\{i'\}}(X_m)(\mathbf{1}_{\{j'\}}(X_{m+1}) - p_{i',j'})] \\
&\quad = \mathrm{E}[\mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j})\mathrm{E}[\mathbf{1}_{\{i'\}}(X_m)(\mathbf{1}_{\{j'\}}(X_{m+1}) - p_{i',j'})|X_n, X_{n+1}]] \\
&\quad = 0,
\end{aligned}$$

since $\mathrm{E}[\mathbf{1}_{\{i'\}}(X_m)(\mathbf{1}_{\{j'\}}(X_{m+1}) - p_{i',j'})|X_n, X_{n+1}] = 0$. A similar result holds when $n > m$. Hence, the issue lies with evaluating the expectation of the single summation above. When $i \neq i'$ this expectation is zero; however, when $i = i'$,

$$\begin{aligned}
&\sum_{n=1}^{L-1} \mathrm{E}[\mathbf{1}_{\{i\}}(X_n)(\mathbf{1}_{\{j\}}(X_{n+1}) - p_{i,j})(\mathbf{1}_{\{j'\}}(X_{n+1}) - p_{i,j'})] \\
&\quad = \sum_{n=1}^{L-1} \mathrm{E}[\mathbf{1}_{\{i\}}(X_n)\mathrm{E}[\mathbf{1}_{\{j\}}(X_{n+1})\mathbf{1}_{\{j'\}}(X_{n+1}) - p_{i,j}\mathbf{1}_{\{j'\}}(X_{n+1}) \\
&\qquad - p_{i,j'}\mathbf{1}_{\{j\}}(X_{n+1}) + p_{i,j}p_{i,j'}|X_n = i]].
\end{aligned}$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

From here we see that when $j = j'$, the inside conditional expectation equals $p_{i,j} - p_{i,j}^2 = p_{i,j}(1 - p_{i,j})$ and when $j \neq j'$, it equals $-p_{i,j}p_{i,j'}$. Since the chain is assumed to be in its stationarity state, $\mathrm{E}[\mathbf{1}_{\{i\}}(X_n)] = \pi_i$ and

$$\mathrm{E}[(N_{i,j} - N_i p_{i,j})(N_{i',j'} - N_{i'} p_{i',j'})] = \begin{cases} (L-1)\pi_i p_{i,j}(1 - p_{i,j}), & i = i', j = j', \\ -(L-1)\pi_i p_{i,j} p_{i,j'}, & i = i', j \neq j', \\ 0, & i \neq i'. \end{cases}$$

Applying Slutsky's Theorem, using the fact that $N_i/L \to \pi_i$ almost surely, and

$$\sqrt{L}(\hat{p}_{i,j} - p_{i,j}) = \sqrt{L}\left(\frac{L}{N_i}\left(\frac{N_{i,j} - N_i p_{i,j}}{L}\right)\right)$$

verifies the central limit theorem in [1,3–5]; that is,

$$\sqrt{L}\begin{pmatrix} \hat{p}_{i,j}(L) - p_{i,j} \\ \hat{p}_{i',j'}(L) - p_{i',j'} \end{pmatrix} \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \begin{pmatrix} p_{i,j}(1 - p_{i,j})/\pi_i & \delta_{i,i'} p_{i,j}(\delta_{j,j'} - p_{i,j'})/\pi_i \\ \delta_{i,i'} p_{i,j}(\delta_{j,j'} - p_{i,j'})/\pi_i & p_{i',j'}(1 - p_{i',j'})/\pi_{i'} \end{pmatrix}\right).$$

■

APPENDIX B

Proof of Lemma 3.1: The asymptotic normality of the counts is established in [5]; however, the asymptotic information matrix is not derived there. We now pursue this. Express $N_{i,j}$ as the indicator summation

$$N_{i,j} = \sum_{k=1}^{L-1} \mathbf{1}_{\{(i,j)\}}(X_k, X_{k+1})$$

and let $Y_k = (X_k, X_{k+1})$ for simplicity. Then

$$\begin{aligned} \mathrm{E}[N_{i,j}] &= E\left[\sum_{k=1}^{L-1} \mathbf{1}_{\{(i,j)\}}(Y_k)\right] \\ &= \sum_{k=1}^{L-1} P(X_k = i, X_{k+1} = j) \\ &= \sum_{k=1}^{L-1} P(X_k = i, X_{k+1} = j | X_k = i) P(X_k = i) \\ &= \sum_{k=1}^{L-1} \pi_i p_{i,j} \\ &= (L-1)\pi_i p_{i,j}. \end{aligned}$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

A similar tactic allows us to calculate the variance of $N_{i,j}$. Specifically, observe that

$$\begin{aligned}(N_{i,j} - (L-1)\pi_i p_{i,j})^2 &= \sum_{k=1}^{L-1}\sum_{\ell=1}^{L-1}(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i,j)\}}(Y_\ell) - \pi_i p_{i,j}) \\ &= \sum_{k=1}^{L-1}(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})^2 \\ &\quad + 2\sum_{k=1}^{L-2}\sum_{\ell=k+1}^{L-1}(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i,j)\}}(Y_\ell) - \pi_i p_{i,j}).\end{aligned}$$

The expectation of the single summation is simply a sum of variances of indicator functions:

$$\mathrm{E}\left[\sum_{k=1}^{L-1}(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})^2\right] = \sum_{k=1}^{L-1}\mathrm{E}[(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})^2] = (L-1)\pi_i p_{i,j}(1 - \pi_i p_{i,j}).$$

For the second term, we introduce some notation. For a one-step-ahead transition matrix $\mathbf{P}$ with invariant measure $\boldsymbol{\pi}$, we define $\boldsymbol{\Pi}$ as a matrix whose rows are all $\boldsymbol{\pi}$. Define $\mathbf{Q} = \mathbf{P} - \boldsymbol{\Pi}$ and observe that by stationarity $\boldsymbol{\Pi}\mathbf{P} = \boldsymbol{\Pi}$, $\mathbf{P}\boldsymbol{\Pi} = \boldsymbol{\Pi}$, and $\boldsymbol{\Pi}^n = \boldsymbol{\Pi}$. Hence, $\mathbf{Q}^n = \mathbf{P}^n - \boldsymbol{\Pi}$.

For $\ell > k$,

$$\begin{aligned}\mathrm{E}[(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i,j)\}}(Y_\ell) - \pi_i p_{i,j})] &= \mathrm{E}[\mathbf{1}_{\{(i,j)\}}(Y_k)\mathbf{1}_{\{(i,j)\}}(Y_\ell)] - (\pi_i p_{i,j})^2 \\ &= \pi_i p_{i,j} p_{j,i}^{(\ell-k-1)} p_{i,j} - (\pi_i p_{i,j})^2 \\ &= \pi_i p_{i,j}^2 (p_{j,i}^{(\ell-k-1)} - \pi_i) \\ &= \pi_i p_{i,j}^2 q_{j,i}^{(\ell-k-1)},\end{aligned}$$

where the convention $q_{i,j}^{(0)} = \delta_{i,j}$ has been used. Summing the above equation over k and ℓ gives

$$\begin{aligned}\sum_{k=1}^{L-2}\sum_{\ell=k+1}^{L-1} q_{j,i}^{(\ell-k-1)} &= \sum_{k=1}^{L-2}\sum_{\ell'=0}^{L-k-1} q_{j,i}^{(\ell')} \\ &= \sum_{k=1}^{L-2}\sum_{\ell'=0}^{L-2}\mathbf{1}_{\{\ell' \le L-k-2\}} q_{j,i}^{(\ell')} \\ &= \sum_{\ell'=0}^{L-2}\sum_{k=1}^{L-2}\mathbf{1}_{\{k \le L-\ell'-2\}} q_{j,i}^{(\ell')} \\ &= \sum_{\ell'=0}^{L-2}(L-\ell'-2) q_{j,i}^{(\ell')}.\end{aligned}$$

Combining the above and using that $E[N_{i,j}] = (L-1)\pi_i p_{i,j}$ gives

$$\mathrm{Var}(N_{i,j}) = 2\pi_i p_{i,j}^2 \sum_{\ell=0}^{L-2}(L-2-\ell) q_{j,i}^{(\ell)} + (L-1)\pi_i p_{i,j}(1 - \pi_i p_{i,j}).$$

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

To compute $\mathrm{Cov}(N_{i,j}, N_{i',j'})$, we assume that either $i \neq i'$ or $j \neq j'$ (otherwise, one is merely recalculating $\mathrm{Var}(N_{i,j})$). Then

$$\begin{aligned}
&[N_{i,j} - (L-1)\pi_i p_{i,j}][N_{i',j'} - (L-1)\pi_{i'} p_{i',j'}] \\
&\quad = \sum_{k=1}^{L-1} (\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j}) \sum_{\ell=1}^{L-1} (\mathbf{1}_{\{(i',j')\}}(Y_\ell) - \pi_{i'} p_{i',j'}) \\
&\quad = \sum_{k=1}^{L-2} \sum_{\ell=k+1}^{L-1} (\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i',j')\}}(Y_\ell) - \pi_{i'} p_{i',j'}) \\
&\qquad + \sum_{\ell=1}^{L-2} \sum_{k=\ell+1}^{L-1} (\mathbf{1}_{\{(i',j')\}}(Y_\ell) - \pi_{i'} p_{i',j'})(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j}) \\
&\qquad + \sum_{k=1}^{L-1} (\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i',j')\}}(Y_k) - \pi_{i'} p_{i',j'}). \qquad (9)
\end{aligned}$$

Arguing as above, the expectation of the lowermost single summation in (9) is $-(L-1)\pi_i p_{i,j}\pi_{i'} p_{i',j'}$. As evaluating the expectation of the first and second double summations in (9) are similar, we only consider the first double summation. For $\ell > k$,

$$\begin{aligned}
\mathrm{E}[(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i',j')\}}(Y_\ell) - \pi_{i'} p_{i',j'})] &= \mathrm{E}[\mathbf{1}_{\{(i,j)\}}(Y_k)\mathbf{1}_{\{(i',j')\}}(Y_\ell)] - \pi_i p_{i,j}\pi_{i'} p_{i',j'} \\
&= \pi_i p_{i,j} p_{j,i'}^{(\ell-k-1)} p_{i',j'} - \pi_i p_{i,j}\pi_{i'} p_{i',j'} \\
&= \pi_i p_{i,j} p_{i',j'} (p_{j,i'}^{(\ell-k-1)} - \pi_{i'}) \\
&= \pi_i p_{i,j} p_{i',j'} q_{j,i'}^{(\ell-k-1)}.
\end{aligned}$$

Hence,

$$\sum_{k=1}^{L-2} \sum_{\ell=k+1}^{L-1} \mathrm{E}[(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i',j')\}}(Y_\ell) - \pi_{i'} p_{i',j'})] = \pi_i p_{i,j} p_{i',j'} \sum_{\ell=0}^{L-2} (L-2-\ell) q_{j,i'}^{(\ell)}.$$

When $\ell < k$, we have

$$\sum_{k=1}^{L-2} \sum_{\ell=k+1}^{L-1} \mathrm{E}[(\mathbf{1}_{\{(i,j)\}}(Y_k) - \pi_i p_{i,j})(\mathbf{1}_{\{(i',j')\}}(Y_\ell) - \pi_{i'} p_{i',j'})] = \pi_{i'} p_{i',j'} p_{i,j} \sum_{\ell=0}^{L-2} (L-2-\ell) q_{j',i}^{(\ell)}.$$

Combining the three expectations above gives

$$\mathrm{Cov}(N_{i,j}, N_{i',j'}) = \sum_{\ell=0}^{L-2} (L-2-\ell)[\pi_i p_{i,j} p_{i',j'} q_{j,i'}^{(\ell)} + \pi_{i'} p_{i',j'} p_{i,j} q_{j',i}^{(\ell)}] - (L-1)\pi_i p_{i,j}\pi_{i'} p_{i',j'}.$$

To get the asymptotic covariance matrix of the sample counts, we first note that $\mathbf{Q}^n \to 0$ almost surely and $\mathbf{R} := \sum_{n=0}^{\infty} \mathbf{Q}^n = (\mathbf{I} - \mathbf{Q})^{-1}$. Kronecker's Lemma (see Theorem 6.1.3 in

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press

[16]) gives

$$\frac{1}{L}\sum_{\ell=0}^{L-2}(L-2-\ell)\mathbf{Q}^{\ell}=\sum_{\ell=0}^{L-2}\mathbf{Q}^{\ell}-\frac{1}{L}\sum_{\ell=0}^{L-2}(\ell+2)\mathbf{Q}^{\ell}\to\sum_{\ell=0}^{\infty}\mathbf{Q}^{\ell}=\mathbf{R}.$$

Using the above expectations and taking limits gives

$$\begin{aligned}\lim_{L\to\infty}L\mathrm{Var}\left(\frac{N_{i,j}}{L}\right)&=\lim_{L\to\infty}\frac{1}{L}\mathrm{Var}(N_{i,j})\\&=2\pi_i p_{i,j}^2 r_{j,i}+\pi_i p_{i,j}(1-\pi_i p_{i,j})\end{aligned}$$

and

$$\begin{aligned}\lim_{L\to\infty}L\mathrm{Cov}\left(\frac{N_{i,j}}{L},\frac{N_{i',j'}}{L}\right)&=\lim_{L\to\infty}\frac{1}{L}\mathrm{Cov}(N_{i,j},N_{i',j'})\\&=\pi_i p_{i,j}p_{i',j'}r_{j,i'}+\pi_{i'}p_{i',j'}p_{i,j}r_{j',i}-\pi_i p_{i,j}\pi_{i'}p_{i',j'}.\end{aligned}$$

■

https://doi.org/10.1017/S0269964812000228 Published online by Cambridge University Press